

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2016 年 7 月 28 日 (28.07.2016)



(10) 国际公布号
WO 2016/115968 A1

- (51) 国际专利分类号:
H04N 19/30 (2014.01)
- (21) 国际申请号: PCT/CN2015/100056
- (22) 国际申请日: 2015 年 12 月 31 日 (31.12.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
2015100252011 2015 年 1 月 20 日 (20.01.2015) CN
- (71) 申请人: 北京工业大学 (BEIJING UNIVERSITY OF TECHNOLOGY) [CN/CN]; 中国北京市北京市朝阳区平乐园 100 号, Beijing 100124 (CN)。
- (72) 发明人: 刘鹏宇 (LIU, Pengyu); 中国北京市北京市朝阳区平乐园 100 号, Beijing 100124 (CN)。 贾克斌 (JIA, Kebin); 中国北京市北京市朝阳区平乐园 100 号, Beijing 100124 (CN)。
- (74) 代理人: 北京思海天达知识产权代理有限公司 (BEIJING SIHAI TIANDA INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市北京市朝阳区平乐园 100 号北京工业大学旧图东配楼 302 室, Beijing 100124 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

(54) Title: VISUAL PERCEPTION FEATURE-FUSED SCALED VIDEO CODING METHOD

(54) 发明名称: 一种融合视觉感知特征的可分层视频编码方法

(57) Abstract: Disclosed is a visual perception feature-fused scaled video coding method, comprising: the setting of a priority of a visual region-of-interest and setting of a video coding resource allocation solution. The former comprises: in consideration of the richness of contents in a video image and a vision selective attention mechanism for human eyes, respectively performing a time domain and a space domain visual feature prominence analysis on the contents of the video, and finishing labelling of the priority of the visual region-of-interest. The latter demonstrates that: in order to improve the video coding real-time performance while ensuring a video coding quality and compression efficiency, according to the priority of the visual region-of-interest, firstly, optimization of coding resources of a macro block in the region-of-interest is satisfied, scaled coding is realized, and the contradiction between the coding complexity and the coding efficiency is effectively relieved; and compared with an H.264/AVC international video coding standard, an average coding speed is improved by about 80% under the premise of ensuring a higher video image quality and a low compression code rate, and the video coding performance are obviously improved.

(57) 摘要: 公开了一种融合视觉感知特征的可分层视频编码方法, 包括视觉感兴趣区域优先级的设定以及视频编码资源分配方案的设定, 前者为: 鉴于视频图像内容的丰富性和人眼视觉选择性注意机制, 将视频内容分别进行时域和空域视觉特征显著度分析, 完成视觉感兴趣区域优先级标注, 后者表现为: 为在保证视频编码质量和压缩效率的同时, 提高视频编码实时性能, 依据视觉感兴趣区域优先级, 首先满足感兴趣区域宏块的编码资源最优化, 实现分层编码, 有效缓解了编码复杂度与编码效率之间的矛盾, 与 H.264/AVC 国际视频编码标准相比, 能够在保持较高视频图像质量和低压缩码率的前提下, 将编码速度平均提高约 80%, 显著改善了视频编码性能。

WO 2016/115968 A1

一种融合视觉感知特征的可分层视频编码方法

技术领域

本发明涉及视频编码方法，尤其涉及一种融合视觉感知特征的可分层视频编码方法。

背景技术

随着多媒体信息处理和通信技术的飞速发展，IPTV、PDA、立体电影、自由视点视频等多样化视频业务相继推出，视频编码在信息化处理以及相关领域展现了广阔的发展前景。然而，在网络带宽和存储空间受限的同时，人们对视频质量的要求却在不断提高、数字视频的性能指标，如分辨率、质量、帧率等不断提升，对现有的视频编码标准提出了新要求。

为了获得具有低复杂度、高质量和高压缩率的视频编码方法，继 2003 年由国际电信联盟 ITU-T 和国际标准化组织 ISO/IEC 联合推出视频压缩标准 H.264/AVC 后，2010 年 1 月，ISO/IEC 和 ITU-T 联合成立了 JCT-VC (Joint Collaborative Team on Video Coding) 小组，并且发布了下一代视频编码技术提案 HEVC(High Efficiency Video Coding)。提案指出，HEVC 依然沿用 H.264/AVC 的混合编码框架，着力研究新的编码技术，旨在解决现有视频编码标准在压缩率与编码复杂度之间的矛盾，使之适应多类型的网络传输，承载更多的信息处理业务。具有“实时性”、“高压缩率”和“高清晰度”的视频编码标准及其应用技术，已成为信号与信息处理领域的研究热点之一。

截至目前，众多学者围绕视频快速编码或者视觉感知分析开展了大量研究工作，但是很少将二者结合在一个编码框架内联合实现对视频编码性能的优化。

在视觉感知特征分析方面，有的研究人员采用颜色、亮度、方向和肤色四种视觉特征进行感兴趣区域计算，但忽略了运动视觉特征；有的研究人员融合了运动、亮度强度、人脸和文字等视觉特征，构建视觉注意模型实现感兴趣提取；也有的研究人员采用运动和纹理信息获取感兴趣区域；或者有人提出在压缩域或者基于小波变换的方法获得感兴趣区域。由于现有的全局运动估计算法复杂度都较大，因此视觉感兴趣区域提取算法复杂度过高。上述基于人类视觉系统 HVS(Human Visual System) 的视频编码技术集中研究了比特资源优化分配的方法，在比特资源受限时保证感兴趣区域的视频图像质量，但欠缺对计算资源分配问题的考虑，并且对进行视觉感知分析时引入的额外计算复杂度，也没有引起足够的关注，其计算效率有待提高。

在快速视频编码方面，有的研究人员通过控制运动估计点数、以损失率失真性能为代价，实现快速编码；有的研究人员通过进行编码参数控制实现快速编码。但上述方法并不区分视频图像中不同区域在视觉意义上的重要程度，对所有编码内容采用相同的快速编码方案，忽略了 HVS 对视频场景感知的差异性。

发明内容

本发明针对上述问题，提出一种融合视觉感知特征的可分层视频编码方法，包括视觉感兴趣区域优先级的设定以及视频编码资源分配方案的设定两部分；

所述视觉感兴趣区域优先级的设定主要为：鉴于视频图像内容的丰富性和人眼视觉选择性注意机制，视频内容通常同时具有时域和空域双重视觉特征，标注视觉特征显著度区域的计算公式可表示为：

$$ROI(x,y) = \begin{cases} 3, & ((T(x,y,MV) = 2 \text{ or } T(x,y,MV) = 1) \parallel S(x,y,Mode) = 1) \text{ or } S(x,y,Mode) = 2 \\ 2, & (T(x,y,MV) = 2 \text{ or } T(x,y,MV) = 1) \parallel S(x,y,Mode) = 0 \\ 1, & T(x,y,MV) = 0 \parallel S(x,y,Mode) = 1 \\ 0, & T(x,y,MV) = 0 \parallel S(x,y,Mode) = 0 \end{cases}$$

式中， $ROI(x,y)$ 代表当前编码宏块视觉感兴趣优先级； $T(x,y,MV)$ 代表当前编码宏块的时域视觉特征显著度； $S(x,y,Mode)$ 代表当前编码宏块的空域视觉特征显著度； (x,y) 表示当前编码宏块的位置坐标；

所述视频编码资源分配方案的设定表现为：为在保证视频编码质量和压缩效率的同时，改善视频编码实时性能，首先满足感兴趣区域宏块的编码最优化，

采用快速帧内预测算法，利用宏块灰度直方图描述宏块平坦程度，依据宏块平坦度自适应地选取可能的帧内预测模式集合；

采用快速帧间预测算法，通过分析各种帧间预测模式出现概率的统计特性，对特定模式进行预判以提前终止不必要的帧间预测模式搜索和率失真代价计算，减少编码耗时；

采用快速运动估计搜索算法，基于编码块运动矢量相关性，依据编码块运动程度，判定搜索层次，实现高效搜索。

所述视觉感兴趣区域优先级的设定中，首先，进行时域视觉显著度区域标注：具体分为两个步骤：步骤1运动矢量噪声检测和步骤2平移运动矢量检测，分别用于削弱由于运动矢量噪声和摄像机运动而产生的平移运动矢量对于时域视觉显著度区域检测准确性的影响，完成前景与背景的分离，得到较为准则的、符合人眼视觉特征的时域视觉显著度区域标注结果；然后，进行空域视觉显著度区域标；最后，依据时域、空域视觉特征显著度区域标注结果，完成视觉特征显著度区域的标注。

在进行时域视觉显著度区域标注时，所述步骤1运动矢量噪声检测的计算公式表示为：

$$T_1(x,y,MV) = \begin{cases} 3, & \text{if } |\overrightarrow{V_{rr}}| = 0 \\ 2, & \text{elseif } \left| \overrightarrow{V_s} \right| \geq \left| \overrightarrow{V_{rr}} \right| \\ T_2(x,y,MV) & \text{esle} \end{cases} \quad (1)$$

(1)式中， (x,y) 为当前编码块位置坐标， $\overrightarrow{V_s}$ 表示当前编码块的运动矢量 MV ， $\overrightarrow{V_{rr}}$ 表示运动参考区域 C_{rr} 内的平均运动矢量，定义为：

$$\bar{V}_{rr} = \frac{\sum_{\vec{v}_{rr} \in C_{rr}} \vec{v}_{rr}}{num_{C_{rr}}}$$

$\vec{v}_{rr}$ 表示运动参考区域 C_{rr} 所包含宏块的运动矢量, $num_{C_{rr}}$ 表示累加次数;

运动参考区域 C_{rr} 的定义如下, 以使参考区域 C_{rr} 的形状、位置、面积能够随当前运动矢量 $\vec{V}_s$ 的变化而实现自适应调整:

将分别位于 C_{rr} 的左上、右上、左下和右下的四个宏块表示为 MB_1, MB_2, MB_3, MB_4 , 其位置坐标定义为:

$$MB_k(x_k, y_k)_{k=1,2,3,4} = \begin{cases} MB_1(x-j_c, y-i_c) \\ MB_2(x-j_c, y+i_c) \\ MB_3(x+j_c, y-i_c) \\ MB_4(x+j_c, y+i_c) \end{cases}$$

其中, $i_c = \left\lceil \frac{\left\lfloor \vec{V}_{sx} \right\rfloor}{w_s} + 1 \right\rceil$, $j_c = \left\lceil \frac{\left\lfloor \vec{V}_{sy} \right\rfloor}{h_s} + 1 \right\rceil$, $\left\lfloor \vec{V}_{sx} \right\rfloor$ 和 $\left\lfloor \vec{V}_{sy} \right\rfloor$ 分别是当前运动矢量 $\vec{V}_s$ 在水平和垂直方向上的

运动幅度, w_s 和 h_s 分别表示当前编码块的宽度和高度, $\lceil \cdot \rceil$ 表示取整计算;

如果 $\left\lfloor \vec{V}_{rr} \right\rfloor = 0$, 说明在运动参考区域 C_{rr} 内不存在运动矢量, 则认为 $\vec{V}_s$ 是由运动噪声引起的, 应予以滤除, $\vec{V}_s$ 被置为 0, 标记 $T_1(x, y, MV) = 3$;

如果 $\left\lfloor \vec{V}_s \right\rfloor \geq \left\lfloor \vec{V}_{rr} \right\rfloor$, 说明当前编码块与邻近宏块相比具有较显著的运动特征, 属于前景动态区域,

标记 $T_1(x, y, MV) = 2$;

否则, 说明当前编码块与邻近宏块具有相似的运动特性, 时域显著性并不明显, 需要进一步进行平移运动矢量检测, 以判定该编码块是属于背景区域, 还是属于前景平移区域, 标记为 $T_2(x, y, MV)$;

所述步骤 2 平移运动矢量检测的计算公式可表示为:

$$T_2(x, y, MV) = \begin{cases} 1, & \text{if } SAD_{(x,y)} \geq \bar{S}_c \\ 0, & \text{else} \end{cases} \quad (2)$$

(2)式中, (x, y) 表示当前编码块的位置坐标; $\bar{S}_c$ 为动态阈值; $SAD_{(x,y)}$ 为当前编码块与前一帧对应位置块之间的绝对差值和 SAD (Sum of Absolute Differences, SAD), 用于表征相邻两帧对应编码块的变化程度, 定义如下:

$$SAD_{(x,y)} = \sum_{i=1}^M \sum_{j=1}^N |s(i, j) - c(i, j)|$$

其中, $s(i, j)$ 为当前编码块的像素值; $c(i, j)$ 为前一帧对应位置块的像素值; M, N 分别为当前编码块的长、宽尺寸;

动态阈值 $\bar{S}_c$ 为前一帧中被确定为背景区域内的所有编码块 SAD 的均值, 定义如下:

$$\bar{S}_c = \frac{\sum_{x,y \in S_c} SAD_{(x,y)}}{Num}$$

其中, S_c 表示前一帧的背景区域; $\sum_{x,y \in S_c} SAD_{(x,y)}$ 表示 S_c 内包含编码块对应 SAD 值的累加和; Num 表示累加次数;

综合上述(1)和(2)两个处理步骤, 进行时域视觉显著度区域标注的计算公式可描述为:

$$T(x, y, MV) = \begin{cases} 3, & \text{if } |\vec{V}_{rr}| = 0 \\ 2, & \text{elseif } |\vec{V}_s| \geq |\vec{V}_{rr}| \\ 1, & \text{elseif } SAD_{(x,y)} \geq \bar{S}_c \\ 0, & \text{else} \end{cases} \quad (3)$$

(3)式中, 各参数的定义与式(1)和式(2)相同;

然后, 进行空域视觉显著度区域标注, 其计算公式可描述为:

$$S(x, y, Mode) = \begin{cases} 2, & Mode_p \in \{Intra\} \\ 1, & Mode_p \in \{Inter8\} \text{ or } Mode_I \in Intra4 \times 4 \\ 0, & Mode_p \in \{Inter16\} \text{ or } Mode_I \in Intra16 \times 16 \end{cases} \quad (4)$$

(4)式中, (x, y) 表示当前编码块的位置坐标; $Mode$ 表示编码块的预测模式; $mode_p$ 表示 P 帧编码中当前编码块的预测模式; $mode_I$ 表示 I 帧编码中当前编码块的预测模式;

如果 $mode_p$ 选择了帧内预测模式, 说明空域视觉特征显著度最高, 属于敏感区, 标记 $S(x, y, Mode) = 2$;

如果 $mode_p$ 选择了子块帧间预测模式集合 $Inter8(8 \times 8, 8 \times 4, 4 \times 8, 4 \times 4)$ 或者 $mode_I$ 选择了 $Intra4 \times 4$ 预测模式, 说明空间细节丰富, 也具有较高空域视觉特征显著度, 属于关注区, 标记 $S(x, y, Mode) = 1$;

如果 $mode_p$ 选择了宏块帧间预测模式集合 $Inter16(Skip, 16 \times 16, 16 \times 8, 8 \times 16)$ 或者 $mode_I$ 选择了 $Intra16 \times 16$ 预测模式, 说明空间变化平缓, 空间视觉特征显著度低, 属于非显著区, 标记 $S(x, y, Mode) = 0$;

最后, 依据时域、空域视觉特征显著度区域标注结果, 标注视觉特征显著度区域。
采用快速帧内预测算法的具体过程:

步骤 1: 计算当前编码宏块亮度分量 Y 的灰度直方图, 记录其最大像素数 $Max\ Value$;

步骤 2: 设定上限阈值 Th_{high} 和下限阈值 Th_{low} , Th_{high} 和 Th_{low} 均为[1, 256]间的整数;

步骤 3: 若 $Max\ Value \geq Th_{high}$, 认为宏块平坦, 舍弃 Intra4×4 预测模式集合, 选择 Intra16×16 预测模式集合, 并将率失真开销最小的模式作为最优帧内预测模式; 同时更新上限阈值:

$$Th_{high-new} = \frac{Th_{high} + MaxValue}{2}; \text{ 否则, 进入步骤 4;}$$

步骤 4: 若 $Max\ Value \leq Th_{low}$, 认为宏块细节丰富, 舍弃 Intra16×16 预测模式集合, 选择 Intra4×4 预测模式集合, 并将率失真开销最小的模式作为最佳帧内预测模式; 同时更新下限阈值:

$$Th_{low-new} = \frac{Th_{low} + MaxValue}{2}; \text{ 否则, 进入步骤 5;}$$

步骤 5: 若 $Th_{low} < Max\ Value < Th_{high}$, 认为宏块平坦度特征不显著, 采用标准帧内预测算法;

本发明中的上限阈值 Th_{high} 和下限阈值 Th_{low} 分别设定为150和50。

采用快速帧间预测算法的具体过程:

步骤 1: Skip 模式的预判

步骤 1.1: 计算 Skip 模式 (mode0) 的率失真代价值 J_{skip} , 如果小于阈值 T 则停止对其他模式的搜索, 选定 Skip 作为最佳预测模式, 跳转至步骤 4; 否则执行步骤 1.2;

其中, $T = (0.7 - Min_cost / 18000) \times Min_cost$, Min_cost 为前一个编码宏块的最优率失真代价值;

步骤 1.2: 计算 Inter16×16 模式 (mode1) 的率失真代价值 $J_{16 \times 16}$, 如果 $J_{16 \times 16} > J_{skip}$, 则仍选定 Skip 作为最佳编码模式, 跳转至步骤 4; 否则执行步骤 2;

步骤 2: 宏块/子块帧间预测模式的预判

步骤 2.1: 计算 Inter16×16 模式和 Inter8×8 模式的率失真代价值 $J_{16 \times 16}$ 和 $J_{8 \times 8}$, 若 $J_{8 \times 8} - J_{16 \times 16} > T_0$, 则选定 Inter16×16 模式作为最佳帧间编码模式, 跳转至步骤 4; 否则, 执行步骤 2.2;

其中, $T_0 = 0.2 \times Min_cost$, 是根据实验数据得出的自适应经验域值, 可以在保证快速判决模式的同时, 尽量减少误判率, Min_cost 为前一个编码宏块的最优率失真代价值;

步骤 2.2: 计算 Inter4×4 模式的率失真代价值 $J_{4 \times 4}$, 如果 $J_{4 \times 4} < \min(J_{16 \times 16}, J_{8 \times 8})$, 则对该宏块进行亚分割, 将子块帧间预测模式 Inter8×8, Inter8×4, Inter4×8 和 Inter4×4 (mode4~ mode7) 作为帧间备选模式集合; 否则, 将宏块帧间预测模式 Inter16×16, Inter16×8, Inter8×16 (mode1~ mode3) 作为帧间预测备选模式集合, 舍弃亚分割预测;

步骤 3: 帧内模式的预判

步骤 3.1: 计算当前编码宏块的平均边缘误差值 ABE (Average Boundary Error) 与边缘误差总和 SBE (Summation Boundary Error), 平均边缘误差值 ABE 反映了宏块的时间相关性;

$$ABE = SBE / 64$$

其中, $SBE = \sum_{i=0}^{15} |Y_{\text{orig}}(x+i, y) - Y_{\text{rec}}(x+i, y-1)| + \sum_{i=0}^{15} |Y_{\text{orig}}(x, y+i) - Y_{\text{rec}}(x-1, y+i)|$

式中: Y_{orig} 为当前编码宏块的像素值; Y_{rec} 为重构宏块的像素值; (x, y) 表示当前编码宏块的位置坐标;

步骤 3.2: 计算当前编码宏块的平均比特率 AR (Average Rate), 平均比特率 AR 反映了宏块的空间相关性;

$$AR = \lambda \text{Rate}/384$$

式中: λ 为拉格朗日乘数因子; Rate 为宏块编码所需的比特数;

步骤 3.3: 比较宏块的平均边缘误差和平均比特率, 若 $ABE < C \cdot AR (C = 0.95)$, 则说明该宏块的空域冗余小于时域冗余, 舍弃对帧内预测模式的遍历计算, 进入步骤 4; 否则, 保留帧内预测模式遍历计算, 进入步骤 4;

步骤 4: 根据率失真准则计算并选择出最优帧间预测模式, 完成帧间预测编码。

采用快速运动估计搜索算法的具体过程:

步骤 1: 描述宏块运动特征

步骤 1.1: 基于率失真准则, 计算当前编码宏块运动估计的率失真代价值 $RD \text{ cost}_{\text{motion}}$:

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

式中, s 为当前宏块像素值; mv 是宏块运动矢量, $pred$ 为预测矢量; ref 为选择的参考帧; $r(ref, mv)$ 为参考宏块的像素值; R 是运动矢量进行差分编码消耗的比特数, 包括运动矢量与其预测值的差值的编码比特数和参考帧的编码比特数; λ_{motion} 为拉格朗日乘子; SAD 为当前块与参考块之间的绝对误差和, 其定义为:

$$SAD(s, c(m)) = \sum_{x=1, y=1}^{M, N} |s[x, y] - c[x - m_x, y - m_y]|$$

式中 M 和 N 分别表示当前编码宏块的宽度和高度; x, y 表示宏块所在位置; s 表示真实值; c 表示预测值; $m = (m_x, m_y)^T$ 表示宏块运动矢量, m_x 和 m_y 分别表示宏块在水平和垂直方向的运动矢量;

步骤 1.2: 基于率失真准则, 计算在 mode 模式下的率失真代价值 $RD \text{ cost}_{\text{mode}}$:

$$J_{\text{mode}}(s, c, \text{mode} | \lambda_{\text{mode}}) = SSD(s, c, \text{mode} | QP) + \lambda_{\text{mode}} \times R(s, c, \text{mode} | QP)$$

式中, mode 为当前宏块的帧间编码模式; s 为原始视频信号; c 为 mode 模式编码后的重建视频信号; λ_{mode} 为拉格朗日乘子; $R(s, c, \text{mode} | QP)$ 是与模式和量化参数有关的包括宏块头信息、运动矢量和所有 DCT 块信息的总的二进制位数; QP 是编码量化步长; $SSD(s, c, \text{mode})$ 为原始信号与重建信号之间的平方差值和, 即:

$$SSD(s, c, \text{mode} | QP) = \sum_{i=1, j=1}^{B, B_2} (s_Y[x, y] - c_Y[x, y, \text{mode} | QP])^2 + \sum_{i=1, j=1}^{B, B_2} (s_U[x, y] - c_U[x, y, \text{mode} | QP])^2 + \sum_{i=1, j=1}^{B, B_2} (s_V[x, y] - c_V[x, y, \text{mode} | QP])^2$$

其中, B_1 和 B_2 分别表示编码块的水平像素数和垂直像素数, 其取值可为 16, 8, 4; $s_r[x, y]$, $c_r[x, y, mode|QP]$ 表示原始视频和重建视频亮度信号的取值; c_U, c_V 和 s_U, s_V 表示相应的色差信号的取值;

步骤 1.3: 从 $RD \text{ cost}_{\text{motion}}$ 和 $RD \text{ cost}_{\text{mode}}$ 中选取最小率失真代价值, 记为 $RD_mincost$;

步骤 2: 判定宏块运动剧烈程度

判定宏块运动剧烈程度的计算公式为:

$$\begin{cases} RD_mincost < (1+\gamma) \times pred_mincost & \text{运动程度较低} \\ (1+\gamma) \times pred_mincost < RD_mincost < (1+\delta) \times pred_mincost & \text{运动程度中等} \\ RD_mincost > (1+\delta) \times pred_mincost & \text{运动程度较高} \end{cases}$$

式中, γ, δ 为判别宏块运动程度的调整因子, 分别定义为:

$$\gamma = \frac{Bsize[blocktype]}{pred_mincost^2} - \alpha_1[blocktype], \quad \delta = \frac{Bsize[blocktype]}{pred_mincost^2} - \alpha_2[blocktype]$$

其中, $Bsize[blocktype]$ 为当前编码宏块尺寸, 有 7 种取值: 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4; $pred_mincost$ 由 UMHexagonS 算法起始搜索点选择的运动矢量预测方式而定:

(1) 如果起始搜索点采用时间预测方式的运动矢量, 则 $pred_mincost$ 选取参考帧预测运动矢量 reference frame MV predictor;

(2) 如果起始搜索点没有采用时间预测方式的运动矢量, 再分为以下两类情况:

(2.1) 若当前运动估计预测宏块选择的是 16×16, 16×8, 8×16 大尺寸帧间预测模式, 则 $pred_mincost$ 选取中值预测运动矢量 median MV predictor;

(2.2) 若当前运动估计预测宏块选择的是 8×8, 8×4, 4×8, 4×4 小尺寸帧间预测模式, 则 $pred_mincost$ 选取上层预测运动矢量 uplayer MV predictor;

根据大量实验测试数据, 数组 $\alpha_1[blocktype]$ 和数组 $\alpha_2[blocktype]$ 分别定义为:

$$\alpha_1[blocktype] = [-0.23, -0.23, -0.23, -0.25, -0.27, -0.27, -0.28];$$

$$\alpha_2[blocktype] = [-2.39, -2.40, -2.40, -2.41, -2.45, -2.45, -2.48];$$

步骤 3: 确定宏块运动估计搜索层次

步骤 3.1: 当宏块运动程度较低时, 在 UMHexagonS 算法的“非均匀 4 层次的六边形格网搜索”步骤中, 只进行内层的第 1 层和第 2 层的非均匀六边形网格搜索;

步骤 3.2: 当宏块运动程度中等时, 在 UMHexagonS 算法的“非均匀 4 层次的六边形格网搜索”步骤中, 进行第 1 层到第 3 层的非均匀六边形网格搜索;

步骤 3.3: 当宏块运动程度较高时, 在 UMHexagonS 算法的“非均匀 4 层次的六边形格网搜索”步骤中, 才进行 4 层的非均匀六边形网格搜索。

本发明采用视频编码层和视觉感知分析层的双层结构实现快速编码。一方面, 视觉感知分析层利用编码层的码流信息, 进行视觉特征显著度分析, 标识视觉感兴趣区域优先级, 大大缩短了视觉感知分析的计算耗时; 另一方面, 编码层复用感知分析层输出的视觉特征显著度分析结果, 优化分配编码计算资源, 实现了可分层视频编码, 提高了编码速度。本发明既保证了视频图像质量和编码效率, 又整体提高了编码计算速度, 在编码速度、主观视频质量及压缩码率三方面达到了平衡。

附图说明

图 1 所示为本发明中率失真性能比较结果图；

图 2 所示为本发明中计算复杂度比较结果图；

具体实施方式

下面结合图表和具体实施方式对本发明作进一步详细地说明。

一种融合视觉感知特征的可分层视频编码方法，包括视觉感兴趣区域优先级的设定以及视频编码资源分配方案的设定两部分；

所述视觉感兴趣区域优先级的设定主要为：鉴于视频图像内容的丰富性和人眼视觉选择性注意机制，视频内容通常同时具有时域和空域双重视觉特征，为了降低对视频内容的时域和空域视觉特征的计算复杂度，提出利用已有视频编码码流信息来进行人眼对时域和空域视觉显著度区域的标注，

首先，进行时域视觉显著度区域标注：具体分为两个步骤：（1）运动矢量噪声检测和（2）平移运动矢量检测，分别用于削弱由于运动矢量噪声和摄像机运动而产生的平移运动矢量对于时域视觉显著度区域检测准确性的影响，完成前景与背景的分离，得到较为准则的、符合人眼视觉特征的时域视觉显著度区域标注结果，

（1）运动矢量噪声检测：

进行运动矢量噪声检测的计算公式可表示为：

$$T_1(x, y, MV) = \begin{cases} 3, & \text{if } |\overline{V_{rr}}| = 0 \\ 2, & \text{elseif } \left| \vec{V_s} \right| \geq \left| \overline{V_{rr}} \right| \\ T_2(x, y, MV) & \text{esle} \end{cases} \quad (1)$$

(1)式中， (x, y) 为当前编码块位置坐标， $\vec{V_s}$ 表示当前编码块的运动矢量 MV ， $\overline{V_{rr}}$ 表示运动参考区域 C_{rr} 内的平均运动矢量，定义为：

$$\overline{V_{rr}} = \frac{\sum_{v_{rr} \in C_{rr}} \vec{v_{rr}}}{num_{C_{rr}}}$$

$\vec{v_{rr}}$ 表示参考区域 C_{rr} 所包含宏块的运动矢量， $num_{C_{rr}}$ 表示累加次数。

为了使得运动参考区域 C_{rr} 的形状、位置、面积能够随当前运动矢量 $\vec{V_s}$ 的变化而实现自适应调整，运动参考区域 C_{rr} 的定义如下，：

将分别位于 C_{rr} 的左上、右上、左下和右下的四个宏块表示为 MB_1, MB_2, MB_3, MB_4 ，其位置坐标定义为：

$$MB_k(x_k, y_k)_{k=1,2,3,4} = \begin{cases} MB_1(x - j_c, y - i_c) \\ MB_2(x - j_c, y + i_c) \\ MB_3(x + j_c, y - i_c) \\ MB_4(x + j_c, y + i_c) \end{cases}$$

$$\text{其中, } i_c = \left\lceil \frac{\left| \vec{V}_{sx} \right|}{w_s} + 1 \right\rceil, \quad j_c = \left\lceil \frac{\left| \vec{V}_{sy} \right|}{h_s} + 1 \right\rceil, \quad \left| \vec{V}_{sx} \right| \text{ 和 } \left| \vec{V}_{sy} \right| \text{ 分别是当前运动矢量 } \vec{V}_s \text{ 在水平和垂直方}$$

向上的运动幅度, w_s 和 h_s 分别表示当前编码块的宽度和高度, $\lceil \cdot \rceil$ 表示取整计算。

如果 $\left| \vec{V}_{rr} \right| = 0$, 说明在运动参考区域 C_{rr} 内不存在运动矢量, 则认为 $\vec{V}_s$ 是由运动噪声引起的, 应予以滤除, $\vec{V}_s$ 被置为 0, 标记 $T_1(x, y, MV) = 3$ 。

如果 $\left| \vec{V}_s \right| \geq \left| \vec{V}_{rr} \right|$, 说明当前编码块与邻近宏块相比具有较显著的运动特征, 属于前景动态区域,

标记 $T_1(x, y, MV) = 2$ 。

否则, 说明当前编码块与邻近宏块具有相似的运动特性, 时域显著性并不明显, 需要进一步进行平移运动矢量检测, 以判定该编码块是属于背景区域, 还是属于由于摄像机的移动而产生的平移运动矢量的前景平移区域, 标记为 $T_2(x, y, MV)$ 。

(2) 平移运动矢量检测:

进行平移运动矢量检测的计算公式可表示为:

$$T_2(x, y, MV) = \begin{cases} 1, & \text{if } SAD_{(x,y)} \geq \bar{S}_c \\ 0, & \text{else} \end{cases} \quad (2)$$

(2)式中, (x, y) 表示当前编码块的位置坐标; $\bar{S}_c$ 为动态阈值; $SAD_{(x,y)}$ 为当前编码块与前一帧对应位置块之间的绝对差值和 SAD (Sum of Absolute Differences, SAD), 用于表征相邻两帧对应编码块的变化程度, 定义如下:

$$SAD_{(x,y)} = \sum_{i=1}^M \sum_{j=1}^N |s(i, j) - c(i, j)|$$

其中, $s(i, j)$ 为当前编码块的像素值; $c(i, j)$ 为前一帧对应位置块的像素值; M、N 分别为当前编码块的长、宽尺寸。

动态阈值 $\bar{S}_c$ 为前一帧中被确定为背景区域内的所有编码块 SAD 的均值, 定义如下:

$$\bar{S}_c = \frac{\sum_{x,y \in S_c} SAD_{(x,y)}}{Num}$$

其中, S_c 表示前一帧的背景区域; $\sum_{x,y \in S_c} SAD_{(x,y)}$ 表示 S_c 内包含编码块对应 SAD 值的累加和; Num 表示累加次数。

综合上述(1)和(2)两个处理步骤, 进行时域视觉显著度区域标注的计算公式可描述为:

$$T(x, y, MV) = \begin{cases} 3, & \text{if } |\overrightarrow{V_{rr}}| = 0 \\ 2, & \text{elseif } |\overrightarrow{V_s}| \geq |\overrightarrow{V_{rr}}| \\ 1, & \text{elseif } SAD_{(x,y)} \geq \overline{S_c} \\ 0, & \text{else} \end{cases} \quad (3)$$

(3)式中, 各参数的定义与式(1)和式(2)相同。

然后, 进行时域视觉显著度区域标注:

$$S(x, y, Mode) = \begin{cases} 2, & Mode_p \in \{Intra\} \\ 1, & Mode_p \in \{Inter8\} \text{ or } Mode_l \in Intra4 \times 4 \\ 0, & Mode_p \in \{Inter16\} \text{ or } Mode_l \in Intra16 \times 16 \end{cases} \quad (4)$$

(4)式中, (x, y) 表示当前编码块的位置坐标; $Mode$ 表示编码块的预测模式; $mode_p$ 表示 P

帧编码中当前编码块的预测模式; $mode_l$ 表示 I 帧编码中当前编码块的预测模式。

如果 $mode_p$ 选择了帧内预测模式, 说明空域视觉特征显著度最高, 属于敏感区, 标记

$$S(x, y, Mode) = 2;$$

如果 $mode_p$ 选择了子块帧间预测模式集合 Inter8(8×8 , 8×4 , 4×8 , 4×4)或者 $mode_l$ 选择了 Intra 4×4 预测模式, 说明空间细节丰富, 也具有较高空域视觉特征显著度, 属于关注区, 标记

$$S(x, y, Mode) = 1;$$

如果 $mode_p$ 选择了宏块帧间预测模式集合 Inter16(Skip, 16×16 , 16×8 , 8×16)或者 $mode_l$ 选择了 Intra 16×16 预测模式, 说明空间变化平缓, 空间视觉特征显著度低, 属于非显著区, 标记

$$S(x, y, Mode) = 0。$$

最后, 依据时域、空域视觉特征显著度区域标注结果, 标注视觉特征显著度区域:

计算公式可表示为:

$$ROI(x, y) = \begin{cases} 3, & ((T(x, y, MV) = 2 \text{ or } T(x, y, MV) = 1) \| S(x, y, Mode) = 1) \text{ or } S(x, y, Mode) = 2 \\ 2, & (T(x, y, MV) = 2 \text{ or } T(x, y, MV) = 1) \| S(x, y, Mode) = 0 \\ 1, & T(x, y, MV) = 0 \| S(x, y, Mode) = 1 \\ 0, & T(x, y, MV) = 0 \| (S(x, y, Mode) = 0) \end{cases}$$

(5)

(5)式中, $ROI(x, y)$ 代表当前编码宏块视觉感兴趣优先级; $T(x, y, MV)$ 代表当前编码宏块的时域视觉特征显著度; $S(x, y, Mode)$ 代表当前编码宏块的空域视觉特征显著度; (x, y) 表示当前编码宏块的位置坐标;

视频编码资源分配方案的设定表现为: 为在保证视频编码质量和压缩效率的同时, 改善视频编码

实时性能，首先满足感兴趣区域宏块的编码最优化，制定的分层编码方案见表 1。

表 1

编码方案		帧内预测模式 ^[1]		帧间预测模式 ^[2]		搜索层数 ^[3]	参考帧数
		Intra16×16	Intra4×4	Inter16	Inter8		
P 帧	$ROI(x,y)=3$	Intra16×16	Intra4×4	—	Inter8	第 2~4 层搜索	5
	$ROI(x,y)=2$	—	—	Inter16	—	第 1~3 层搜索	3
	$ROI(x,y)=1$	—	—	—	Inter8	第 1~2 层搜索	2
	$ROI(x,y)=0$	—	—	Inter16	—	第 1 层搜索	1
I 帧	$ROI(x,y)=1$	—	Intra4×4	—	—	—	—
	$ROI(x,y)=0$	Intra16×16	—	—	—	—	—

表1采用快速帧内预测算法，利用宏块灰度直方图描述宏块平坦程度，依据宏块平坦度自适应地选取可能的帧内预测模式集合。

基本原理为：

宏块灰度直方图描述了该宏块所包含的灰度级信息。从数学上来说，灰度直方图统计了宏块中各个灰度级出现的次数或概率；从图形上来说，灰度直方图是一个二维图形，横坐标表示宏块包含的灰度级，从全黑到全白的灰度级范围为[0,255]，纵坐标表示宏块中的像素点在各个灰度级上出现的次数。

宏块灰度直方图的形状可以反映宏块纹理的丰富程度。在宏块灰度直方图的纵坐标轴上，必然存在一个纵坐标最大的灰度级(即峰值)，那么把属于该灰度级的像素点总数定义为该宏块的最大像素数，记为 *Max Value*。如果 *Max Value* 明显高于宏块灰度直方图中其他灰度级出现的次数，则说明它是该宏块的主要灰度分量，宏块内像素的空间相关性较大，即宏块平坦，适合采用 Intra16×16 预测模式集合；反之，*Max Value* 与宏块灰度直方图中的其他灰度级出现的次数相当，则说明该宏块覆盖了多个灰度级，宏块内像素灰度变化剧烈，即宏块纹理丰富，适合采用 Intra4×4 预测模式集合。

具体过程为：

步骤 1：计算当前编码宏块亮度分量 Y 的灰度直方图，记录其最大像素数 *Max Value*；

步骤 2：设定上限阈值 Th_{high} 和下限阈值 Th_{low} ， Th_{high} 和 Th_{low} 均为[1, 256]间的整数；

步骤 3：若 $Max Value \geq Th_{high}$ ，认为宏块平坦，舍弃 Intra4×4 预测模式集合，选择 Intra16×16 预测模式集合，并将率失真开销最小的模式作为最优帧内预测模式；同时更新上限阈值：

$$Th_{high-new} = \frac{Th_{high} + MaxValue}{2}；\text{否则，进入步骤 4；}$$

步骤 4：若 $Max Value \leq Th_{low}$ ，认为宏块细节丰富，舍弃 Intra16×16 预测模式集合，选择 Intra4×4 预测模式集合，并将率失真开销最小的模式作为最佳帧内预测模式；同时更新下限阈值：

$$Th_{low-new} = \frac{Th_{low} + MaxValue}{2}；\text{否则，进入步骤 5；}$$

步骤 5: 若 $Th_{low} < Max\ Value < Th_{high}$, 认为宏块平坦度特征不显著, 采用标准帧内预测算法。

本发明中的上限阈值 Th_{high} 和下限阈值 Th_{low} 分别设定为150和50。

表1采用快速帧间预测算法, 通过分析各种帧间预测模式出现概率的统计特性, 对特定模式进行预判以提前终止不必要的帧间预测模式搜索和率失真代价计算, 减少编码耗时。

基本原理为:

为了提高编码精度, H.264/AVC 视频编码标准在帧间编码中采用 7 种可变块的预测模式, 每个编码宏块可划分为 Inter16×16, Inter16×8, Inter8×16 和 Inter8×8 模式, 其中 Inter8×8 模式还可以进行亚分割, 划分为 Inter8×8, Inter8×4, Inter4×8 和 Inter4×4 模式。此外, H.264/AVC 帧间预测还支持 Skip 模式以及 Intra16×16 和 Intra4×4 两种帧内预测模式。H.264/AVC 对每一个编码宏块遍历所有可能选择的预测模式, 以获得率失真性能的最优化, 达到最佳预测效果。灵活多样的可选帧间预测模式的引入成为 H.264/AVC 较其它视频编码标准获得更高编码效率的重要因素, 但块划分模式组合的增多也使帧间预测模式判决过程异常复杂, 造成编码计算复杂度急剧增加。

研究发现, 视频图像基本可以分为背景纹理平坦区域、背景纹理细致区域和运动区域 3 大类: 通常背景纹理平坦区域在视频内容中占有很大比重, 对于这类平坦区域和运动平滑区域, 大多采取 Skip 模式 (mode0) 或者宏块级帧间预测模式 Inter16×16, Inter16×8, Inter8×16 (mode1~mode3) 进行预测; 在运动复杂的情况下需要使用更多的编码模式进行预测, 才会使用到帧间亚分割预测模式 Inter8×8, Inter8×4, Inter4×8 和 Inter4×4 (mode4~mode7); 只有在视频图像的边缘部分才会使用到 Intra16×16 和 Intra4×4 帧内预测模式 (I16 MB, I4 MB), 其出现的几率非常低。因此, 可以根据帧间预测模式统计特性进行预先判决, 分集筛选, 以排除掉出现机率很小的编码模式, 提高编码速度。

具体过程为:

步骤 1: Skip 模式的预判

步骤 1.1: 计算 Skip 模式 (mode0) 的率失真代价值 J_{skip} , 如果小于阈值 T 则停止对其他模式的搜索, 选定 Skip 作为最佳预测模式, 跳转至步骤 4; 否则执行步骤 1.2;

其中, $T = (0.7 - \text{Min_cost} / 18000) \times \text{Min_cost}$ 。Min_cost 为前一个编码宏块的最优率失真代价值。

步骤 1.2: 计算 Inter16×16 模式 (mode1) 的率失真代价值 $J_{16 \times 16}$, 如果 $J_{16 \times 16} > J_{skip}$, 则仍选定 Skip 作为最佳编码模式, 跳转至步骤 4; 否则执行步骤 2。

步骤 2: 宏块/子块帧间预测模式的预判

步骤 2.1: 计算 Inter16×16 模式和 Inter8×8 模式的率失真代价值 $J_{16 \times 16}$ 和 $J_{8 \times 8}$, 若 $J_{8 \times 8} - J_{16 \times 16} > T_0$, 则选定 Inter16×16 模式作为最佳帧间编码模式, 跳转至步骤 4; 否则, 执行步骤 2.2;

其中, $T_0 = 0.2 \times \text{Min_cost}$, 是根据实验数据得出的自适应经验域值, 可以在保证快速判决模式的同时, 尽量减少误判率。Min_cost 为前一个编码宏块的最优率失真代价值。

步骤 2.2: 计算 Inter4×4 模式的率失真代价值 $J_{4 \times 4}$, 如果 $J_{4 \times 4} < \min(J_{16 \times 16}, J_{8 \times 8})$, 则对该宏块进行亚分割, 将子块帧间预测模式 Inter8×8, Inter8×4, Inter4×8 和 Inter4×4 (mode4~mode7) 作为帧间

备选模式集合；否则，将宏块帧间预测模式 Inter16×16，Inter16×8，Inter8×16（mode1~ mode3）作为帧间预测备选模式集合，舍弃亚分割预测。

步骤 3：帧内模式的预判

步骤 3.1：计算当前编码宏块的平均边缘误差值 ABE（Average Boundary Error）与边缘误差总和 SBE（Summation Boundary Error），平均边缘误差值 ABE 反映了宏块的时间相关性；

$$ABE = SBE / 64$$

$$\text{其中, } SBE = \sum_{i=0}^{15} |Y_{\text{orig}}(x+i, y) - Y_{\text{rec}}(x+i, y-1)| + \sum_{i=0}^{15} |Y_{\text{orig}}(x, y+i) - Y_{\text{rec}}(x-1, y+i)|$$

式中： Y_{orig} 为当前编码宏块的像素值； Y_{rec} 为重构宏块的像素值； (x, y) 表示当前编码宏块的位置坐标。

步骤 3.2：计算当前编码宏块的平均比特率 AR（Average Rate），平均比特率 AR 反映了宏块的空间相关性；

$$AR = \lambda \text{Rate} / 384$$

式中： λ 为拉格朗日乘数因子；Rate 为宏块编码所需的比特数。

步骤 3.3：比较宏块的平均边缘误差和平均比特率，若 $ABE < C \cdot AR (C = 0.95)$ ，则说明该宏块的空域冗余小于时域冗余，舍弃对帧内预测模式的遍历计算，进入步骤 4；否则，保留帧内预测模式遍历计算，进入步骤 4。

步骤 4：根据率失真准则计算并选择出最优帧间预测模式，完成帧间预测编码。

表1采用快速运动估计搜索算法，基于编码块运动矢量相关性，依据编码块运动程度，判定搜索层次，实现高效搜索。

基本原理为：

H.264/AVC 标准中所采用的 UMHexagonS 算法，是目前效果最好的运动估计算法之一。但通过大量实验统计数据发现，最佳匹配点在 UMHexagonS 算法各个搜索步骤上是非均匀分布的，但 UMHexagonS 算法在“非均匀 4 层次的六边形网格搜索”步骤中，并没有对编码块运动特征与搜索范围之间的关联性进行分析，不论当前编码宏块的运动程度如何，都必须在完成 4 层非均匀六边形搜索（4 层×16 个搜索点/层=64 个搜索点）以后才能进入下一步搜索，计算量相当可观。对于视频序列中占有较大比重的运动平缓区域的宏块而言，过大的搜索半径及分布在外层上的搜索点对提高运动估计准确度的贡献甚微，但却消耗了较多的运动估计时间。反之，对于少数运动程度剧烈的编码块，耗费在内层上搜索点的遍历计算也造成了编码耗时。可见，当前编码宏块的运动剧烈程度与其最佳匹配点所在运动估计搜索层次有着必然的联系。如果能够根据宏块运动程度自适应地选择搜索层数，无疑将大大节约搜索点数，降低运动估计的计算复杂度。而选取何种特征和准则判别宏块运动程度成为优化运动估计算法的关键所在。

由此，本发明将原 UMHexagonS 算法中的 4 层非均匀六边形网格搜索改进为搜索层数随宏块运

动程度自适应变化的非均匀六边形网格搜索。首先描述宏块运动特征；然后将宏块运动程度划分为三个档次：运动程度较低、运动程度中等、运动程度较高；最后根据运动程度选择相应的搜索层次。具体过程为：

步骤 1：描述宏块运动特征

步骤 1.1：基于率失真准则，计算当前编码宏块运动估计的率失真代价值 $RD \text{ cost}_{\text{motion}}$ ：

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

式中， s 为当前宏块像素值； mv 是宏块运动矢量， $pred$ 为预测矢量； ref 为选择的参考帧； $r(ref, mv)$ 为参考宏块的像素值； R 是运动矢量进行差分编码消耗的比特数，包括运动矢量与其预测值的差值的编码比特数和参考帧的编码比特数； λ_{motion} 为拉格朗日乘子； SAD 为当前块与参考块之间的绝对误差和，其定义为：

$$SAD(s, c(m)) = \sum_{x=1, y=1}^{M, N} |s[x, y] - c[x - m_x, y - m_y]|$$

式中 M 和 N 分别表示当前编码宏块的宽度和高度； x, y 表示宏块所在位置； s 表示真实值； c 表示预测值； $m = (m_x, m_y)^T$ 表示宏块运动矢量， m_x 和 m_y 分别表示宏块在水平和垂直方向的运动矢量。

步骤 1.2：基于率失真准则，计算在 mode 模式下的率失真代价值 $RD \text{ cost}_{\text{mode}}$ ：

$$J_{\text{mode}}(s, c, \text{mode} | \lambda_{\text{mode}}) = SSD(s, c, \text{mode} | QP) + \lambda_{\text{mode}} \times R(s, c, \text{mode} | QP)$$

式中， mode 为当前宏块的帧间编码模式； s 为原始视频信号； c 为 mode 模式编码后的重建视频信号； λ_{mode} 为拉格朗日乘子； $R(s, c, \text{mode} | QP)$ 是与模式和量化参数有关的包括宏块头信息、运动矢量和所有 DCT 块信息的总的二进制位数； QP 是编码量化步长； $SSD(s, c, \text{mode})$ 为原始信号与重建信号之间的平方差值和，即：

$$SSD(s, c, \text{mode} | QP) = \sum_{i=1, j=1}^{B_1, B_2} (s_Y[x, y] - c_Y[x, y, \text{mode} | QP])^2 + \sum_{i=1, j=1}^{B_1, B_2} (s_U[x, y] - c_U[x, y, \text{mode} | QP])^2 + \sum_{i=1, j=1}^{B_1, B_2} (s_V[x, y] - c_V[x, y, \text{mode} | QP])^2$$

其中， B_1 和 B_2 分别表示编码块的水平像素数和垂直像素数，其取值可为 16, 8, 4； $s_Y[x, y]$ ， $c_Y[x, y, \text{mode} | QP]$ 表示原始视频和重建视频亮度信号的取值； c_U, c_V 和 s_U, s_V 表示相应的色差信号的取值。

步骤 1.3：从 $RD \text{ cost}_{\text{motion}}$ 和 $RD \text{ cost}_{\text{mode}}$ 中选取最小率失真代价值，记为 $RD_mincost$ 。

步骤 2：判定宏块运动剧烈程度

判定宏块运动剧烈程度的计算公式为：

$$\begin{cases} RD_mincost < (1+\gamma) \times pred_mincost & \text{运动程度较低} \\ (1+\gamma) \times pred_mincost < RD_mincost < (1+\delta) \times pred_mincost & \text{运动程度中等} \\ RD_mincost > (1+\delta) \times pred_mincost & \text{运动程度较高} \end{cases}$$

式中， γ, δ 为判别宏块运动程度的调整因子，分别定义为：

$$\gamma = \frac{Bsize[blocktype]}{pred_mincost^2} - \alpha_1[blocktype], \quad \delta = \frac{Bsize[blocktype]}{pred_mincost^2} - \alpha_2[blocktype]$$

其中, $Bsize[blocktype]$ 为当前编码宏块尺寸, 有 7 种取值: 16×16 , 16×8 , 8×16 , 8×8 , 8×4 , 4×8 , 4×4 ; $pred_mincost$ 由 UMHExagonS 算法起始搜索点选择的运动矢量预测方式而定:

(1) 如果起始搜索点采用时间预测方式的运动矢量, 则 $pred_mincost$ 选取参考帧预测运动矢量 reference frame MV predictor;

(2) 如果起始搜索点没有采用时间预测方式的运动矢量, 再分为以下两类情况:

(2.1) 若当前运动估计预测宏块选择的是 16×16 , 16×8 , 8×16 大尺寸帧间预测模式, 则 $pred_mincost$ 选取中值预测运动矢量 median MV predictor;

(2.2) 若当前运动估计预测宏块选择的是 8×8 , 8×4 , 4×8 , 4×4 小尺寸帧间预测模式, 则 $pred_mincost$ 选取上层预测运动矢量 uplayer MV predictor。

根据大量实验测试数据, 数组 $\alpha_1[blocktype]$ 和数组 $\alpha_2[blocktype]$ 分别定义为:

$$\alpha_1[blocktype] = [-0.23, -0.23, -0.23, -0.25, -0.27, -0.27, -0.28];$$

$$\alpha_2[blocktype] = [-2.39, -2.40, -2.40, -2.41, -2.45, -2.45, -2.48]。$$

步骤 3: 确定宏块运动估计搜索层次

步骤 3.1: 当宏块运动程度较低时, 在 UMHExagonS 算法的 “非均匀 4 层次的六边形格网格搜索” 步骤中, 只进行内层的第 1 层和第 2 层的非均匀六边形网格搜索;

步骤 3.2: 当宏块运动程度中等时, 在 UMHExagonS 算法的 “非均匀 4 层次的六边形格网格搜索” 步骤中, 进行第 1 层到第 3 层的非均匀六边形网格搜索;

步骤 3.3: 当宏块运动程度较高时, 在 UMHExagonS 算法的 “非均匀 4 层次的六边形格网格搜索” 步骤中, 才进行 4 层的非均匀六边形网格搜索。

在 P 帧编码中, 由公式(5):

$ROI(x, y) = 3$ 时, 情况①, 编码宏块属于前景动态区域 ($T(x, y, MV) = 2$) 或者前景平移区域 ($T(x, y, MV) = 1$), 具有时域视觉特征, 并且 $S(x, y, Mode) = 1$, 说明该宏块选择了帧间子块预测模式集合 Inter8, 也具有空域视觉特征, 属于时域视觉特征显著且纹理丰富区; 情况②当 $S(x, y, Mode) = 2$, 说明 P 帧编码宏块采用了帧内预测模式, 属于空域视觉特征敏感区。上述两种情况下人眼关注度最高, 遍历执行快速的帧内预测和帧间子块预测模式集合 Inter8, 运动估计搜索执行第 2~4 层搜索, 允许的参考帧数为 5 个。

$ROI(x, y) = 2$ 时, 编码宏块具有时域视觉特征 ($T(x, y, MV) = 2$ 或 $T(x, y, MV) = 1$), 且 $S(x, y, Mode) = 0$, 说明该宏块选择了帧间宏块预测模式集合 Inter16, 空域视觉特征不显著, 属于时域视觉特征显著且纹理平坦区, 人眼关注度次之, 略过帧内预测, 仅遍历执行快速帧间宏块预测模式集合 Inter16, 运动估计搜索执行第 1~3 层搜索, 参考帧数为 3 个。

$ROI(x, y) = 1$ 时, 编码宏块不具有时域视觉特征 ($T(x, y, MV) = 0$), 属于非动态背景区域, 且 $S(x, y, Mode) = 1$, 说明该宏块选择了帧间子块预测模式集合 Inter8, 具有空域视觉特征, 属于空域视觉特征关注区, 人眼关注度再次之, 略过帧内预测, 仅遍历执行快速帧间子块预测模式集合 Inter8, 运动估计搜索执行第 1~2 层搜索, 参考帧数为 1 个。

$ROI(x, y) = 0$ 时, 说明当前编码宏块不具有时域特征和空域视觉特征, 属于平坦静止背景区域; 人眼关注度最低, 仅遍历执行快速帧间宏块预测模式集合 Inter16, 运动估计搜索执行第 1 层搜索, 参考帧数为 1 个。

在 I 帧编码中, 由公式(5):

$ROI(x, y) = 1$ 时, 编码宏块不具有时域视觉特征($T(x, y, MV) = 0$), 且 $S(x, y, Mode) = 1$,

说明该宏块选择了 Intra4×4 预测模式, 空间细节丰富, 也具有较高空域视觉特征显著度, 属于关注区, 略过 Intra16×16 预测。

$ROI(x, y) = 0$ 时, 编码宏块不具有时域特征和空域视觉特征, 属于平坦静止背景区域; 人眼关注度最低, 仅执行 Intra16×16 预测。

综上所述, 本发明首先依据底层编码信息实现了高效视觉感知特征分析与检测, 再根据视觉感兴趣区域优先级标注结果, 指导编码方案选择, 简化了预测编码的备选模式集和运动估计搜索范围, 减少了参考帧数量, 降低了视频编码层的计算复杂度。

本发明同时公开了仿真测试与统计结果;

表 2 为本发明方法与 H.264/AVC(JM17.0)标准算法的性能比较结果。

表 2

视频序列名称	量化步长 QP	编码时间变化 $\Delta \text{Time}(\%)$	输出码率变化 $\Delta \text{Bit}(\%)$	亮度分量的峰值信噪比变化 $\text{PSNR-Y } \Delta$ (dB)	感兴趣区域亮度分量的峰值信噪比变化 $\text{ROI-PSNR-Y } \Delta$ (dB)
Foreman	28	-71.24	+2.19	-0.30	-0.25
	32	-71.92	+2.15	-0.29	-0.21
	36	-70.81	+2.06	-0.17	-0.15
Hall	28	-83.64	+1.24	-0.21	-0.19
	32	-84.27	+1.17	-0.22	-0.14
	36	-84.98	+1.12	-0.19	-0.13
Salesman	28	-76.43	+2.12	-0.26	-0.21
	32	-75.92	+1.87	-0.20	-0.15
	36	-76.01	+1.83	-0.14	-0.12
Car-phone	28	-73.85	+2.84	-0.32	-0.24
	32	-73.36	+2.57	-0.21	-0.19
	36	-74.96	+1.69	-0.14	-0.12
News	28	-85.47	+2.14	-0.13	-0.11
	32	-85.84	+2.13	-0.12	-0.07
	36	-85.93	+1.86	-0.09	-0.08
Coastguard	28	-61.24	+2.47	-0.28	-0.24
	32	-62.35	+2.13	-0.29	-0.26
	36	-62.62	+1.76	-0.21	-0.18
Paris	28	-84.12	+1.07	-0.29	-0.23
	32	-84.61	+1.15	-0.22	-0.21
	36	-84.97	+1.03	-0.21	-0.17
Silent	28	-80.14	+1.51	-0.24	-0.18
	32	-80.86	+1.42	-0.18	-0.15
	36	-81.32	+1.17	-0.11	-0.09
Mother &	28	-83.76	+1.85	-0.18	-0.14

Daughter	32	-83.89	+1.64	-0.10	-0.10
	36	-84.21	+0.09	-0.08	-0.05
Akiyo	28	-85.64	+1.87	-0.11	-0.09
	32	-85.76	+1.21	-0.08	-0.07
	36	-86.41	+0.08	-0.07	-0.06
平均统计结果					
QP	28	-78.55	+1.93	-0.188dB	-0.153dB
	32	-78.88	+1.74		
	36	-79.22	+1.27		

表2统计了具有不同运动特点的10个典型标准测试序列在本发明提出方法下相对于H.264/AVC (JM17.0)标准算法的编码性能比较结果。

在量化步长QP分别为28, 32, 36时, 本发明方法与H.264/AVC标准算法相比, 编码时间平均节省约80%; 输出码率增加平均控制在2%以内; PSNR-Y平均降低-0.188dB, 其中视觉感兴趣区域的PSNR-Y平均降低-0.153dB, 优先保证了视觉感知特征显著区域的编码质量, 符合人眼对非感兴趣区域的降质不敏感的视觉感知特性。

在输出码率控制方面, 图1中两条率失真性能R-D曲线非常接近, 说明本发明方法较好地继承了H.264/AVC标准算法低码率、高质量的编码优势。

在视频图像重建质量方面, 本发明方法保证了PSNR-Y平均降幅在-0.2dB以内, 远小于人眼对图像质量变化感知的最小灵敏度(-0.5dB), 保持了良好的重建视频图像质量。

在编码速度提高方面, 图2统计数据表明, 本发明方法与H.264/AVC标准算法和现有算法相比, 具有更低的计算复杂度。对于运动平缓、纹理平坦的Akiyo, News等序列, 与H.264/AVC(JM17.0)相比, 编码时间平均节省均在85%以上。

本发明提出的融合视觉感知特征的视频编码方法, 可以在大幅提高编码速度的前提下, 保持良好的视频图像主观质量, 试验结果证明了充分利用编码信息进行低复杂度视觉感知分析的可行性, 视觉感知特征显著度分析结果与HVS的一致性, 验证了基于视觉感知特征制定可分级编码方案的合理性。

权 利 要 求 书

1.一种融合视觉感知特征的可分层视频编码方法，其特征在于，包括视觉感兴趣区域优先级的设定以及视频编码资源分配方案的设定两部分；

所述视觉感兴趣区域优先级的设定主要为：鉴于视频图像内容的丰富性和人眼视觉选择性注意机制，视频内容通常同时具有时域和空域双重视觉特征，标注视觉特征显著度区域的计算公式可表示为：

$$ROI(x,y)=\begin{cases} 3, & ((T(x,y,MV)=2 \text{ or } T(x,y,MV)=1) \parallel (S(x,y,Mode)=1)) \text{ or } S(x,y,Mode)=2 \\ 2, & (T(x,y,MV)=2 \text{ or } T(x,y,MV)=1) \parallel S(x,y,Mode)=0 \\ 1, & T(x,y,MV)=0 \parallel S(x,y,Mode)=1 \\ 0, & T(x,y,MV)=0 \parallel (S(x,y,Mode)=0) \end{cases}$$

式中， $ROI(x,y)$ 代表当前编码宏块视觉感兴趣优先级； $T(x,y,MV)$ 代表当前编码宏块的时域视觉特征显著度； $S(x,y,Mode)$ 代表当前编码宏块的空域视觉特征显著度； (x,y) 表示当前编码宏块的位置坐标；

所述视频编码资源分配方案的设定表现为：为在保证视频编码质量和压缩效率的同时，改善视频编码实时性能，首先满足感兴趣区域宏块的编码最优化，

采用快速帧内预测算法，利用宏块灰度直方图描述宏块平坦程度，依据宏块平坦度自适应地选取可能的帧内预测模式集合；

采用快速帧间预测算法，通过分析各种帧间预测模式出现概率的统计特性，对特定模式进行预判以提前终止不必要的帧间预测模式搜索和率失真代价计算，减少编码耗时；

采用快速运动估计搜索算法，基于编码块运动矢量相关性，依据编码块运动程度，判定搜索层次，实现高效搜索。

2. 根据权利要求1所述的融合视觉感知特征的可分层视频编码方法，其特征在于，所述视觉感兴趣区域优先级的设定中，首先，进行时域视觉显著度区域标注：具体分为两个步骤：步骤1运动矢量噪声检测和步骤2平移运动矢量检测，分别用于削弱由于运动矢量噪声和摄像机运动而产生的平移运动矢量对于时域视觉显著度区域检测准确性的影响，完成前景与背景的分离，得到较为准则的、符合人眼视觉特征的时域视觉显著度区域标注结果；然后，进行空域视觉显著度区域标注；最后，依据时域、空域视觉特征显著度区域标注结果，完成视觉特征显著度区域的标注。

3. 根据权利要求2所述的融合视觉感知特征的可分层视频编码方法，其特征在于，在进行时域视觉显著度区域标注时，所述步骤1运动矢量噪声检测的计算公式表示为：

$$T_1(x, y, MV) = \begin{cases} 3, & \text{if } |\vec{V}_{rr}| = 0 \\ 2, & \text{elseif } |\vec{V}_s| \geq |\vec{V}_{rr}| \\ T_2(x, y, MV) & \text{esle} \end{cases} \quad (1)$$

(1)式中, (x, y) 为当前编码块位置坐标, $\vec{V}_s$ 表示当前编码块的运动矢量 MV , $\vec{V}_{rr}$ 表示运动参考区域 C_{rr} 内的平均运动矢量, 定义为:

$$\vec{V}_{rr} = \frac{\sum_{\vec{v}_{rr} \in C_{rr}} \vec{v}_{rr}}{num_{C_{rr}}}$$

$\vec{v}_{rr}$ 表示运动参考区域 C_{rr} 所包含宏块的运动矢量, $num_{C_{rr}}$ 表示累加次数;

运动参考区域 C_{rr} 的定义如下, 以使参考区域 C_{rr} 的形状、位置、面积能够随当前运动矢量 $\vec{V}_s$ 的变化而实现自适应调整:

将分别位于 C_{rr} 的左上、右上、左下和右下的四个宏块表示为 MB_1, MB_2, MB_3, MB_4 , 其位置坐标定义为:

$$MB_k(x_k, y_k)_{k=1,2,3,4} = \begin{cases} MB_1(x - j_c, y - i_c) \\ MB_2(x - j_c, y + i_c) \\ MB_3(x + j_c, y - i_c) \\ MB_4(x + j_c, y + i_c) \end{cases}$$

$$\text{其中, } i_c = \left\lceil \frac{|\vec{V}_{sx}|}{w_s} + 1 \right\rceil, \quad j_c = \left\lceil \frac{|\vec{V}_{sy}|}{h_s} + 1 \right\rceil, \quad |\vec{V}_{sx}| \text{ 和 } |\vec{V}_{sy}| \text{ 分别是当前运动矢量 } \vec{V}_s \text{ 在水平和}$$

垂直方向上的运动幅度, w_s 和 h_s 分别表示当前编码块的宽度和高度, $\lceil \cdot \rceil$ 表示取整计算;

如果 $|\vec{V}_{rr}| = 0$, 说明在运动参考区域 C_{rr} 内不存在运动矢量, 则认为 $\vec{V}_s$ 是由运动噪声引起的, 应予以滤除, $\vec{V}_s$ 被置为 0, 标记 $T_1(x, y, MV) = 3$;

如果 $|\vec{V}_s| \geq |\vec{V}_{rr}|$, 说明当前编码块与邻近宏块相比具有较显著的运动特征, 属于前景

动态区域, 标记 $T_1(x, y, MV) = 2$;

否则, 说明当前编码块与邻近宏块具有相似的运动特性, 时域显著性并不明显, 需要进一步进行平移运动矢量检测, 以判定该编码块是属于背景区域, 还是属于前景平移

区域, 标记为 $T_2(x, y, MV)$;

所述步骤 2 平移运动矢量检测的计算公式可表示为:

$$T_2(x, y, MV) = \begin{cases} 1, & \text{if } SAD_{(x,y)} \geq \bar{S}_c \\ 0, & \text{else} \end{cases} \quad (2)$$

(2)式中, (x, y) 表示当前编码块的位置坐标; $\bar{S}_c$ 为动态阈值; $SAD_{(x,y)}$ 为当前编码块与前一帧对应位置块之间的绝对差值和 SAD (Sum of Absolute Differences, SAD), 用于表征相邻两帧对应编码块的变化程度, 定义如下:

$$SAD_{(x,y)} = \sum_{i=1}^M \sum_{j=1}^N |s(i, j) - c(i, j)|$$

其中, $s(i, j)$ 为当前编码块的像素值; $c(i, j)$ 为前一帧对应位置块的像素值; M, N 分别为当前编码块的长、宽尺寸;

动态阈值 $\bar{S}_c$ 为前一帧中被确定为背景区域内的所有编码块 SAD 的均值, 定义如下:

$$\bar{S}_c = \frac{\sum_{x,y \in S_c} SAD_{(x,y)}}{Num}$$

其中, S_c 表示前一帧的背景区域; $\sum_{x,y \in S_c} SAD_{(x,y)}$ 表示 S_c 内包含编码块对应 SAD 值的累加和; Num 表示累加次数;

综合上述(1)和(2)两个处理步骤, 进行时域视觉显著度区域标注的计算公式可描述为:

$$T(x, y, MV) = \begin{cases} 3, & \text{if } |\overrightarrow{V_{rr}}| = 0 \\ 2, & \text{elseif } |\overrightarrow{V_s}| \geq |\overrightarrow{V_{rr}}| \\ 1, & \text{elseif } SAD_{(x,y)} \geq \bar{S}_c \\ 0, & \text{else} \end{cases} \quad (3)$$

(3)式中, 各参数的定义与式(1)和式(2)相同;

然后, 进行空域视觉显著度区域标注, 其计算公式可描述为:

$$S(x, y, Mode) = \begin{cases} 2, & Mode_p \in \{Intra\} \\ 1, & Mode_p \in \{Inter8\} \text{ or } Mode_I \in Intra4 \times 4 \\ 0, & Mode_p \in \{Inter16\} \text{ or } Mode_I \in Intra16 \times 16 \end{cases} \quad (4)$$

(4)式中, (x, y) 表示当前编码块的位置坐标; $Mode$ 表示编码块的预测模式; $mode_p$ 表示 P 帧编码中当前编码块的预测模式; $mode_I$ 表示 I 帧编码中当前编码块的预测模式;

如果 $mode_p$ 选择了帧内预测模式, 说明空域视觉特征显著度最高, 属于敏感区, 标记 $S(x, y, Mode) = 2$;

如果 $mode_p$ 选择了子块帧间预测模式集合 Inter8(8×8 , 8×4 , 4×8 , 4×4) 或者 $mode_l$ 选择了 Intra 4×4 预测模式, 说明空间细节丰富, 也具有较高空域视觉特征显著度, 属于关注区, 标记 $S(x, y, Mode) = 1$;

如果 $mode_p$ 选择了宏块帧间预测模式集合 Inter16(Skip, 16×16 , 16×8 , 8×16) 或者 $mode_l$ 选择了 Intra 16×16 预测模式, 说明空间变化平缓, 空间视觉特征显著度低, 属于非显著区, 标记 $S(x, y, Mode) = 0$;

最后, 依据时域、空域视觉特征显著度区域标注结果, 标注视觉特征显著度区域。

4. 根据权利要求1所述的融合视觉感知特征的可分层视频编码方法, 其特征在于, 采用快速帧内预测算法的具体过程:

步骤 1: 计算当前编码宏块亮度分量 Y 的灰度直方图, 记录其最大像素数 $Max\ Value$;

步骤 2: 设定上限阈值 Th_{high} 和下限阈值 Th_{low} , Th_{high} 和 Th_{low} 均为 $[1, 256]$ 间的整数;

步骤 3: 若 $Max\ Value \geq Th_{high}$, 认为宏块平坦, 舍弃 Intra 4×4 预测模式集合, 选择 Intra 16×16 预测模式集合, 并将率失真开销最小的模式作为最优帧内预测模式; 同时更新上限阈值: $Th_{high-new} = \frac{Th_{high} + MaxValue}{2}$; 否则, 进入步骤 4;

步骤 4: 若 $Max\ Value \leq Th_{low}$, 认为宏块细节丰富, 舍弃 Intra 16×16 预测模式集合, 选择 Intra 4×4 预测模式集合, 并将率失真开销最小的模式作为最佳帧内预测模式; 同时更新下限阈值: $Th_{low-new} = \frac{Th_{low} + MaxValue}{2}$; 否则, 进入步骤 5;

步骤 5: 若 $Th_{low} < Max\ Value < Th_{high}$, 认为宏块平坦度特征不显著, 采用标准帧内预测算法;

本发明中的上限阈值 Th_{high} 和下限阈值 Th_{low} 分别设定为 150 和 50。

5. 根据权利要求1所述的融合视觉感知特征的可分层视频编码方法, 其特征在于, 采用快速帧间预测算法的具体过程:

步骤 1: Skip 模式的预判

步骤 1.1: 计算 Skip 模式 (mode0) 的率失真代价值 J_{skip} , 如果小于阈值 T 则停止对其他模式的搜索, 选定 Skip 作为最佳预测模式, 跳转至步骤 4; 否则执行步骤 1.2;

其中, $T = (0.7 - Min_cost / 18000) \times Min_cost$, Min_cost 为前一个编码宏块的最优率失真代价值;

步骤 1.2: 计算 Inter 16×16 模式 (mode1) 的率失真代价值 $J_{16 \times 16}$, 如果 $J_{16 \times 16} > J_{skip}$,

则仍选定 Skip 作为最佳编码模式，跳转至步骤 4；否则执行步骤 2；

步骤 2：宏块/子块帧间预测模式的预判

步骤 2.1：计算 Inter16×16 模式和 Inter8×8 模式的率失真代价值 $J_{16 \times 16}$ 和 $J_{8 \times 8}$ ，若 $J_{8 \times 8} - J_{16 \times 16} > T_0$ ，则选定 Inter16×16 模式作为最佳帧间编码模式，跳转至步骤 4；否则，执行步骤 2.2；

其中， $T_0 = 0.2 \times \text{Min_cost}$ ，是根据实验数据得出的自适应经验域值，可以在保证快速判决模式的同时，尽量减少误判率，Min_cost 为前一个编码宏块的最优率失真代价值；

步骤 2.2：计算 Inter4×4 模式的率失真代价值 $J_{4 \times 4}$ ，如果 $J_{4 \times 4} < \min(J_{16 \times 16}, J_{8 \times 8})$ ，则对该宏块进行亚分割，将帧间亚分割预测模式 Inter8×8，Inter8×4，Inter4×8 和 Inter4×4 (mode4~ mode7) 作为帧间备选模式集合；否则，将宏块级帧间预测模式 Inter16×16，Inter16×8，Inter8×16 (mode1~ mode3) 作为帧间预测备选模式集合，舍弃亚分割预测；

步骤 3：帧内模式的预判

步骤 3.1：计算当前编码宏块的平均边缘误差值 ABE (Average Boundary Error) 与边缘误差总和 SBE (Summation Boundary Error)，平均边缘误差值 ABE 反映了宏块的时间相关性；

$$\text{ABE} = \text{SBE} / 64$$

$$\text{其中，SBE} = \sum_{i=0}^{15} |Y_{\text{orig}}(x+i, y) - Y_{\text{rec}}(x+i, y-1)| + \sum_{i=0}^{15} |Y_{\text{orig}}(x, y+i) - Y_{\text{rec}}(x-1, y+i)|$$

式中： Y_{orig} 为当前编码宏块的像素值； Y_{rec} 为重构宏块的像素值； (x, y) 表示当前编码宏块的位置坐标；

步骤 3.2：计算当前编码宏块的平均比特率 AR (Average Rate)，平均比特率 AR 反映了宏块的空间相关性；

$$\text{AR} = \lambda \text{Rate} / 384$$

式中： λ 为拉格朗日乘数因子；Rate 为宏块编码所需的比特数；

步骤 3.3：比较宏块的平均边缘误差和平均比特率，若 $\text{ABE} < C \cdot \text{AR} (C = 0.95)$ ，则说明该宏块的空域冗余小于时域冗余，舍弃对帧内预测模式的遍历计算，进入步骤 4；否则，保留帧内预测模式遍历计算，进入步骤 4；

步骤 4：根据率失真准则计算并选择出最优帧间预测模式，完成帧间预测编码。

6. 根据权利要求 1 所述的融合视觉感知特征的可分层视频编码方法，其特征在于，采用快速运动估计搜索算法的具体过程：

步骤 1：描述宏块运动特征

步骤 1.1: 基于率失真准则, 计算当前编码宏块运动估计的率失真代价值 $RD \text{ cost}_{\text{motion}}$:

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

式中, s 为当前宏块像素值; mv 是宏块运动矢量, $pred$ 为预测矢量; ref 为选择的参考帧; $r(ref, mv)$ 为参考宏块的像素值; R 是运动矢量进行差分编码消耗的比特数, 包括运动矢量与其预测值的差值的编码比特数和参考帧的编码比特数; λ_{motion} 为拉格朗日乘子; SAD 为当前块与参考块之间的绝对误差和, 其定义为:

$$SAD(s, c(m)) = \sum_{x=1, y=1}^{M, N} |s[x, y] - c[x - m_x, y - m_y]|$$

式中 M 和 N 分别表示当前编码宏块的宽度和高度; x, y 表示宏块所在位置; s 表示真实值; c 表示预测值; $m = (m_x, m_y)^T$ 表示宏块运动矢量, m_x 和 m_y 分别表示宏块在水平和垂直方向的运动矢量;

步骤 1.2: 基于率失真准则, 计算在 $mode$ 模式下的率失真代价值 $RD \text{ cost}_{\text{mode}}$:

$$J_{\text{mode}}(s, c, mode | \lambda_{\text{mode}}) = SSD(s, c, mode | QP) + \lambda_{\text{mode}} \times R(s, c, mode | QP)$$

式中, $mode$ 为当前宏块的帧间编码模式; s 为原始视频信号; c 为 $mode$ 模式编码后的重建视频信号; λ_{mode} 为拉格朗日乘子; $R(s, c, mode | QP)$ 是与模式和量化参数有关的包括宏块头信息、运动矢量和所有 DCT 块信息的总的二进制位数; QP 是编码量化步长;

$SSD(s, c, mode)$ 为原始信号与重建信号之间的平方差值和, 即:

$$SSD(s, c, mode | QP) = \sum_{i=1, j=1}^{B_1, B_2} (s_Y[x, y] - c_Y[x, y, mode | QP])^2 + \sum_{i=1, j=1}^{B_1, B_2} (s_U[x, y] - c_U[x, y, mode | QP])^2 + \sum_{i=1, j=1}^{B_1, B_2} (s_V[x, y] - c_V[x, y, mode | QP])^2$$

其中, B_1 和 B_2 分别表示编码块的水平像素数和垂直像素数, 其取值可为 16, 8, 4; $s_Y[x, y]$, $c_Y[x, y, mode | QP]$ 表示原始视频和重建视频亮度信号的取值; c_U, c_V 和 s_U, s_V 表示相应的色差信号的取值;

步骤 1.3: 从 $RD \text{ cost}_{\text{motion}}$ 和 $RD \text{ cost}_{\text{mode}}$ 中选取最小率失真代价值, 记为 $RD_min \text{ cost}$;

步骤 2: 判定宏块运动剧烈程度

判定宏块运动剧烈程度的计算公式为:

$$\begin{cases} RD_mincost < (1+\gamma) \times pred_mincost & \text{运动程度较低} \\ (1+\gamma) \times pred_mincost < RD_mincost < (1+\delta) \times pred_mincost & \text{运动程度中等} \\ RD_mincost > (1+\delta) \times pred_mincost & \text{运动程度较高} \end{cases}$$

式中, γ, δ 为判别宏块运动程度的调整因子, 分别定义为:

$$\gamma = \frac{Bsize[blocktype]}{pred_mincost^2} - \alpha_1[blocktype], \quad \delta = \frac{Bsize[blocktype]}{pred_mincost^2} - \alpha_2[blocktype]$$

其中, $Bsize[blocktype]$ 为当前编码宏块尺寸, 有 7 种取值: 16×16 , 16×8 , 8×16 , 8×8 , 8×4 , 4×8 , 4×4 ; $pred_mincost$ 由 UMHexagonS 算法起始搜索点选择的运动矢量预测方式而定:

(1) 如果起始搜索点采用时间预测方式的运动矢量, 则 $pred_mincost$ 选取参考帧预测运动矢量 reference frame MV predictor;

(2) 如果起始搜索点没有采用时间预测方式的运动矢量, 再分为以下两类情况:

(2.1) 若当前运动估计预测宏块选择的是 16×16 , 16×8 , 8×16 大尺寸帧间预测模式, 则 $pred_mincost$ 选取中值预测运动矢量 median MV predictor;

(2.2) 若当前运动估计预测宏块选择的是 8×8 , 8×4 , 4×8 , 4×4 小尺寸帧间预测模式, 则 $pred_mincost$ 选取上层预测运动矢量 uplayer MV predictor;

根据大量实验测试数据, 数组 $\alpha_1[blocktype]$ 和数组 $\alpha_2[blocktype]$ 分别定义为:

$$\alpha_1[blocktype] = [-0.23, -0.23, -0.23, -0.25, -0.27, -0.27, -0.28];$$

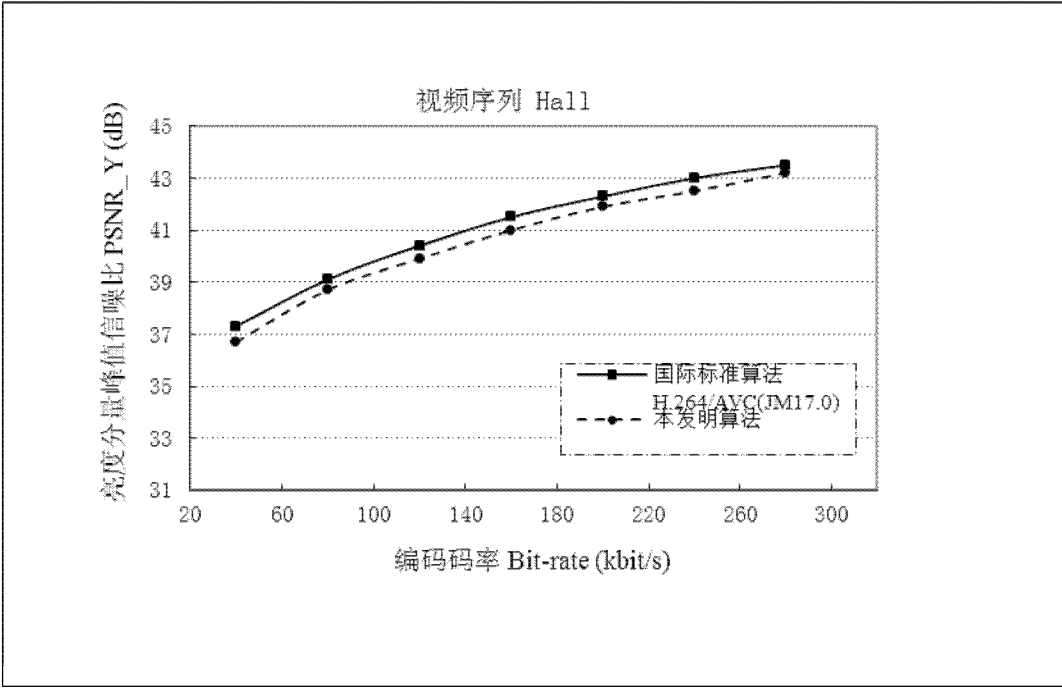
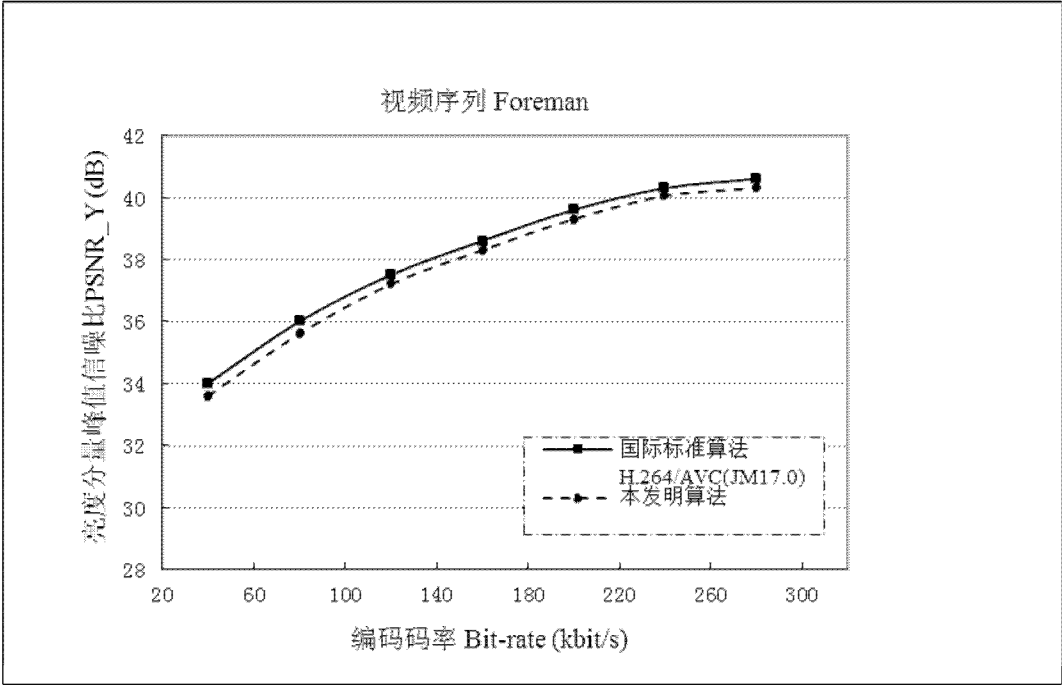
$$\alpha_2[blocktype] = [-2.39, -2.40, -2.40, -2.41, -2.45, -2.45, -2.48];$$

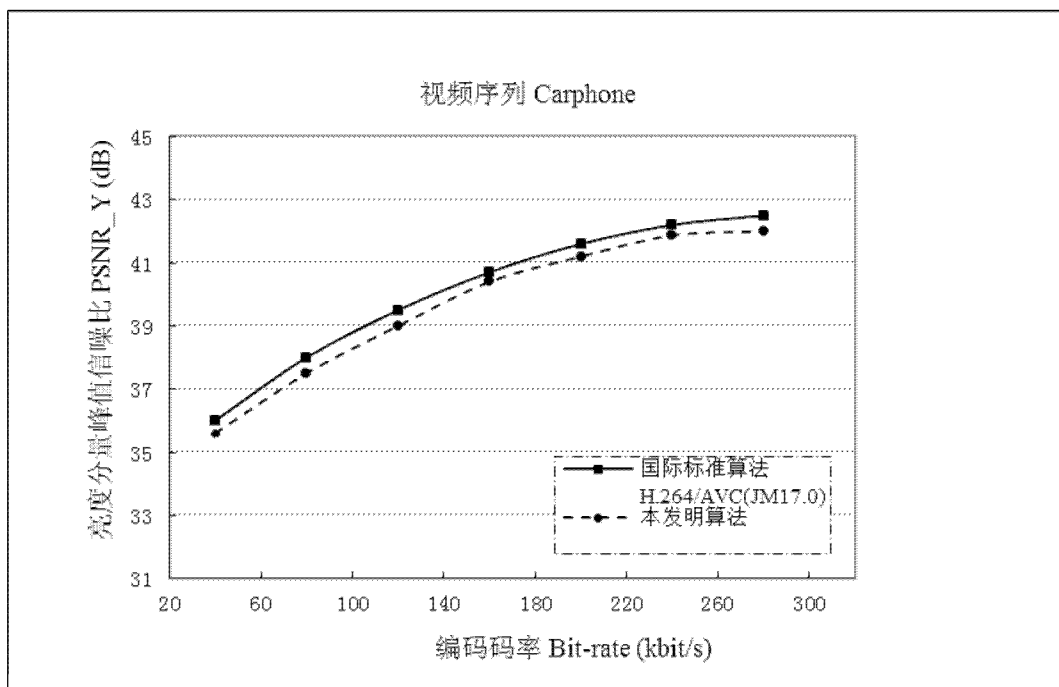
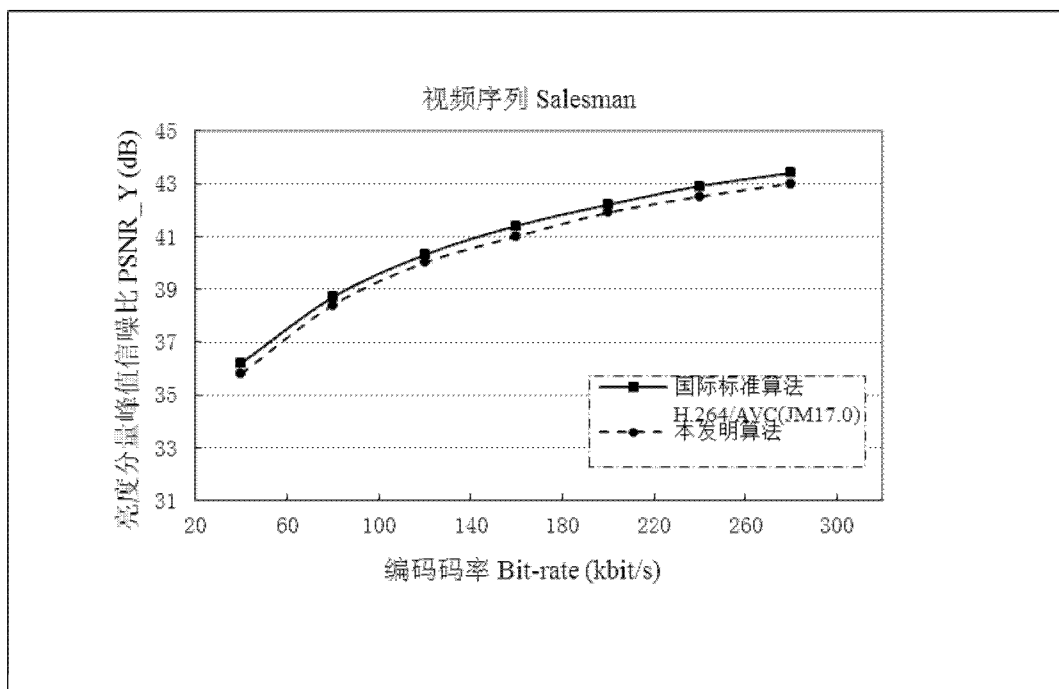
步骤 3: 确定宏块运动估计搜索层次

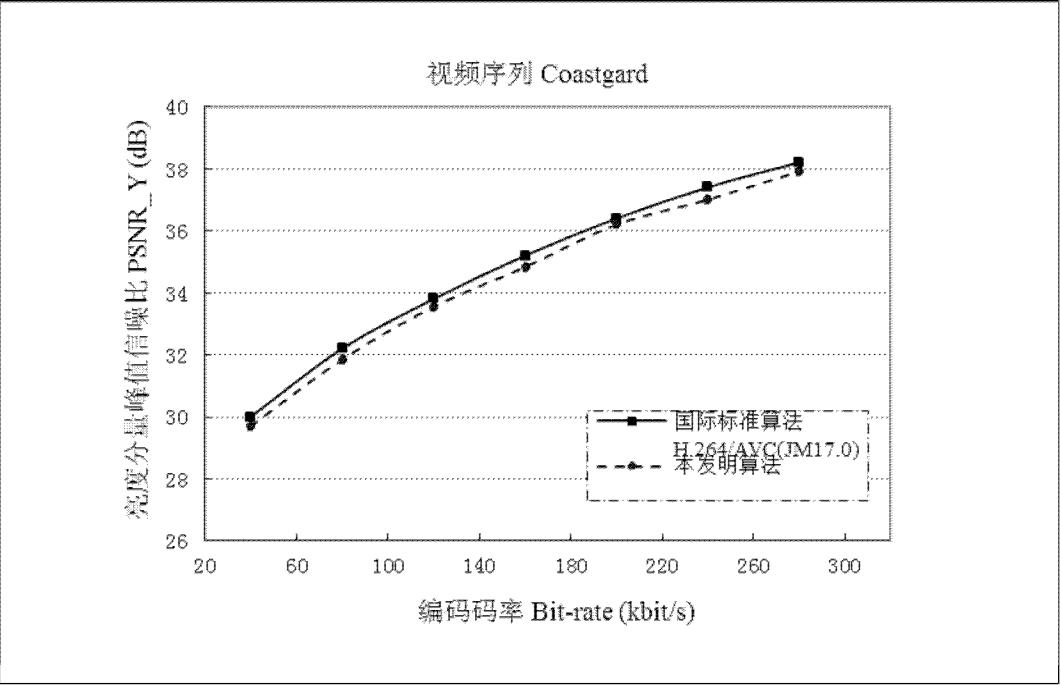
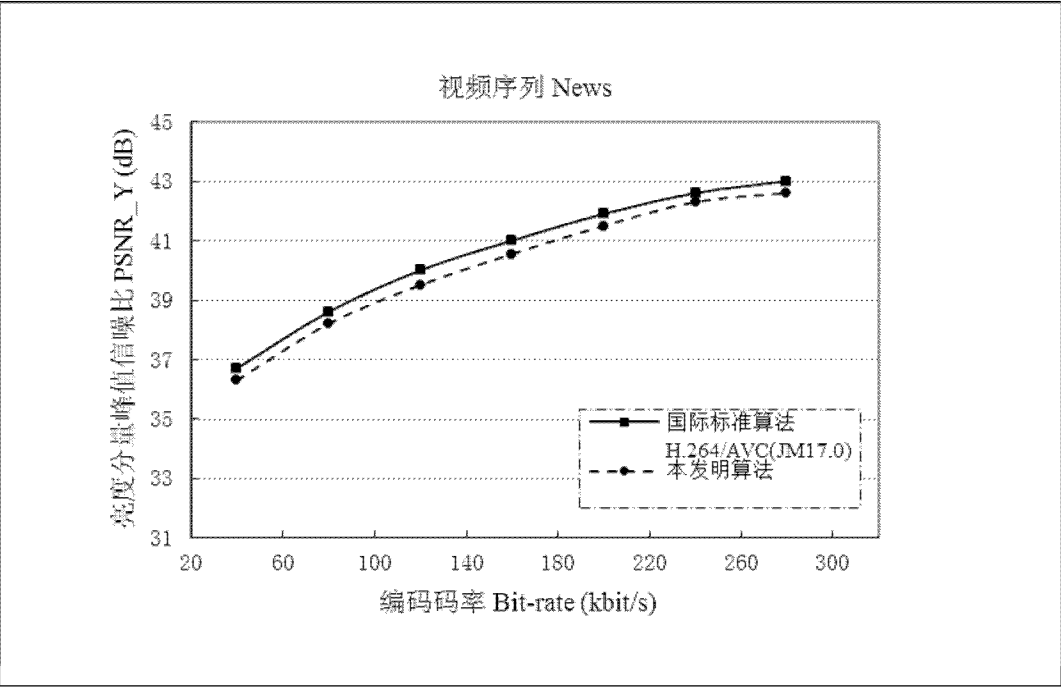
步骤 3.1: 当宏块运动程度较低时, 在 UMHexagonS 算法的“非均匀 4 层次的六边形格网搜索”步骤中, 只进行内层的第 1 层和第 2 层的非均匀六边形网格搜索;

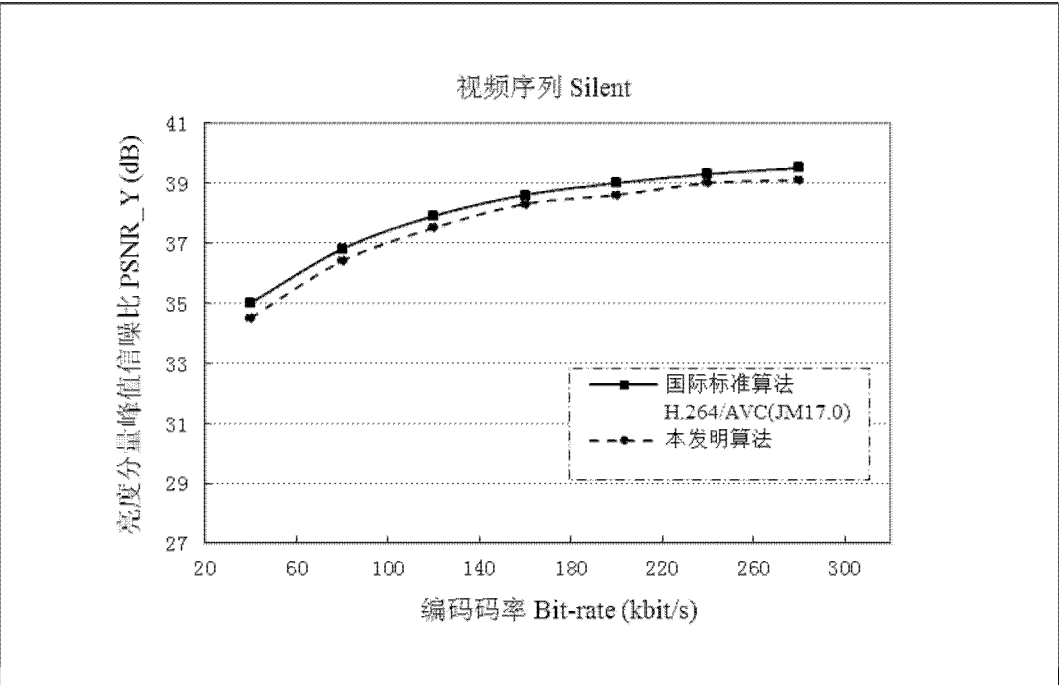
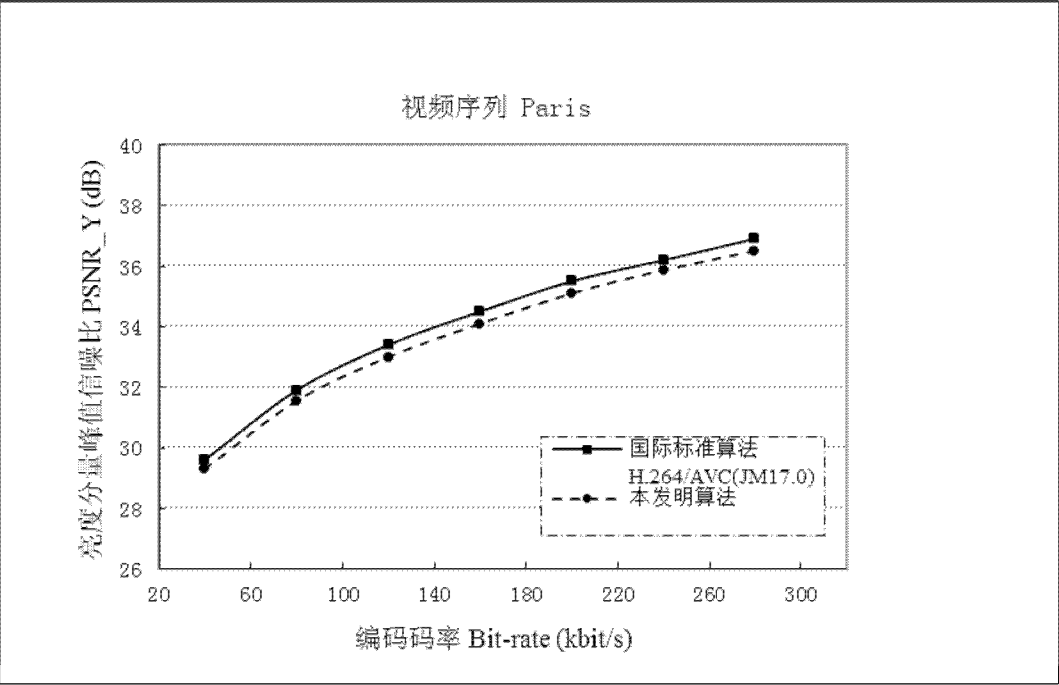
步骤 3.2: 当宏块运动程度中等时, 在 UMHexagonS 算法的“非均匀 4 层次的六边形格网搜索”步骤中, 进行第 1 层到第 3 层的非均匀六边形网格搜索;

步骤 3.3: 当宏块运动程度较高时, 在 UMHexagonS 算法的“非均匀 4 层次的六边形格网搜索”步骤中, 才进行 4 层的非均匀六边形网格搜索。









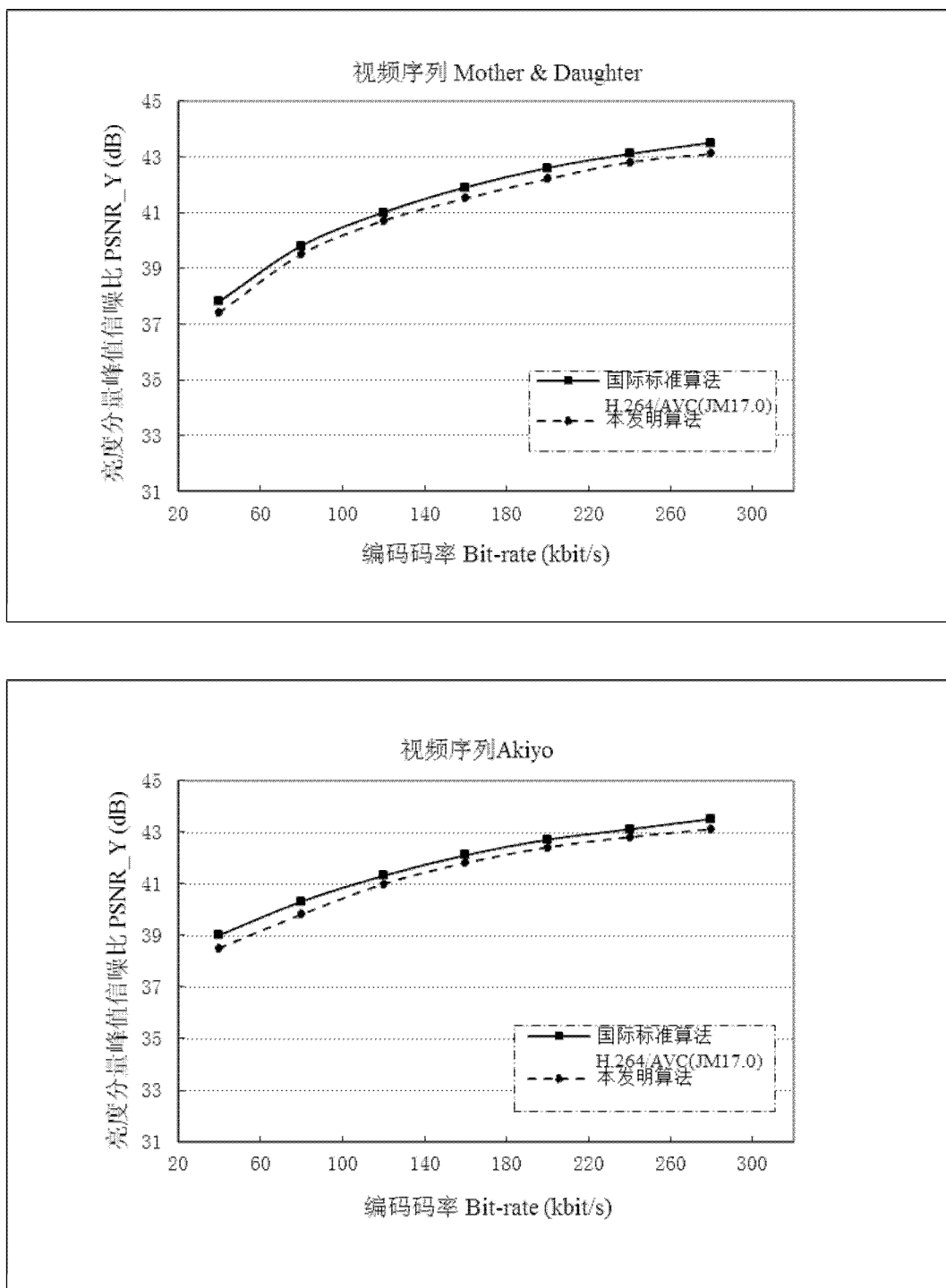


图 1

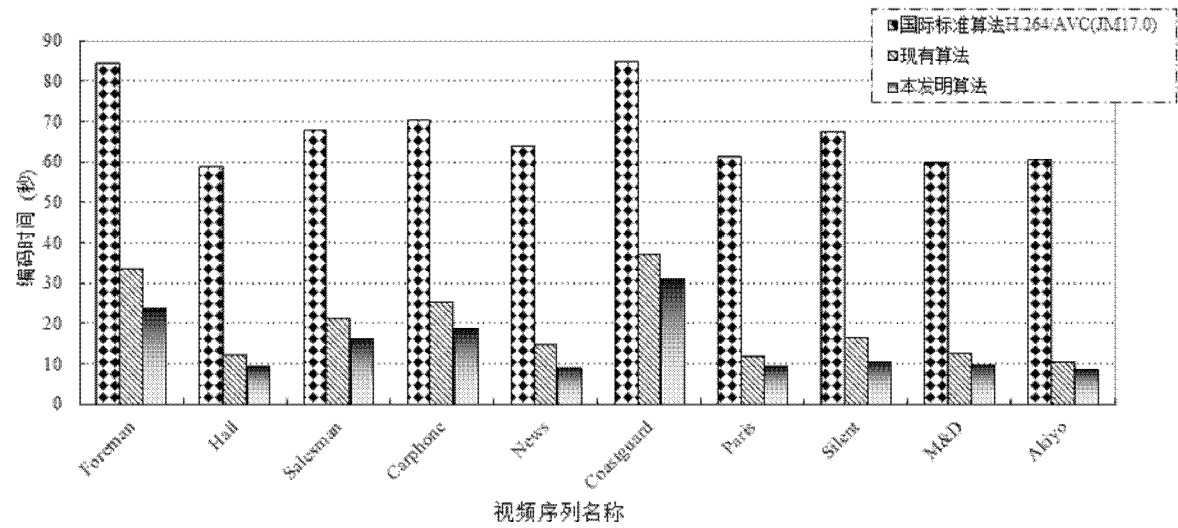


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/100056

A. CLASSIFICATION OF SUBJECT MATTER

H04N 19/30 (2014.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC, GOOGLE: visual perception, significance level, resource, distribute, predictive mode, rate-distortion, human visual system, HVS, feature, encoding, priority, layer, interest, ROI, fuse, time, predict+, mode, rate distortion, optimiz+, LIU, Pengyu; JIA, Kebin

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104539962 A (BEIJING UNIVERSITY OF TECHNOLOGY), 22 April 2015 (22.04.2015), claims 1-6	1-6
A	CN 103618900 A (BEIJING UNIVERSITY OF TECHNOLOGY), 05 March 2014 (05.03.2014), the whole document	1-6
A	CN 102186070 A (BEIJING UNIVERSITY OF TECHNOLOGY), 14 September 2011 (14.09.2011), the whole document	1-6
A	LIU, Pengyu; JIA, Kebin, "Video ROI Extraction Algorithm Based on Reconstructed Encoding Information", COMPUTER ENGINEERING, volume 37, number 24, 31 December 2011 (31.12.2011), ISSN: 1000-3428, the whole document	1-6
A	CN 101621709 A (ZHEJIANG UNIVERSITY), 06 January 2010 (06.01.2010), the whole document	1-6
A	US 8774272 B1 (GEO SEMICONDUCTOR INC.), 08 July 2014 (08.07.2014), the whole document	1-6

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 14 February 2016 (14.02.2016)	Date of mailing of the international search report 29 February 2016 (29.02.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer MENG, Jia Telephone No.: (86-10) 82245567

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/100056

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2007206674 A1 (STREAMING NETWORKS PVT. LTD.), 06 September 2007 (06.09.2007), the whole document	1-6

International application No.
PCT/CN2015/100056

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2015/100056

A. 主题的分类

H04N 19/30(2014.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

H04N

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNKI, CNPAT, WPI, EPODOC, GOOGLE: 视觉感知, 特征, 融合, 分层, 编码, 优先级, 感兴趣, 显著度, 资源, 分配, 时间, 人类视觉系统, 预测模式, 最优, 率失真, human visual system, HVS, feature, encoding, priority, layer, interest, ROI, fuse, time, predict+, mode, rate distortion, optimiz+, pengyu liu, kebin jia

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 104539962 A (北京工业大学) 2015年 4月 22日 (2015 - 04 - 22) 权利要求1-6	1-6
A	CN 103618900 A (北京工业大学) 2014年 3月 5日 (2014 - 03 - 05) 全文	1-6
A	CN 102186070 A (北京工业大学) 2011年 9月 14日 (2011 - 09 - 14) 全文	1-6
A	刘鹏宇 贾克斌. “基于重建编码信息的视频ROI提取算法” 《计算机工程》, 第37卷, 第24期, 2011年 12月 31日 (2011 - 12 - 31), ISSN: 1000-3428, 全文	1-6
A	CN 101621709 A (浙江大学) 2010年 1月 6日 (2010 - 01 - 06) 全文	1-6
A	US 8774272 B1 (GEO SEMICONDUCTOR INC.) 2014年 7月 8日 (2014 - 07 - 08) 全文	1-6

☒ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2016年 2月 14日

国际检索报告邮寄日期

2016年 2月 29日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10) 62019451

授权官员

孟佳

电话号码 (86-10) 82245567

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	US 2007206674 A1 (STREAMING NETWORKS PVT. LTD.) 2007年 9月 6日 (2007 - 09 - 06) 全文	1-6

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2015/100056

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	104539962	A	2015年 4月 22日	无			
CN	103618900	A	2014年 3月 5日	无			
CN	102186070	A	2011年 9月 14日	无			
CN	101621709	A	2010年 1月 6日	无			
US	8774272	B1	2014年 7月 8日	无			
US	2007206674	A1	2007年 9月 6日	WO	2007103713	A1	2007年 9月 13日

表 PCT/ISA/210 (同族专利附件) (2009年7月)