

利用反向投影的 flash 场景自适应 视频编码算法

梁传君, 徐亮

(新疆工程学院 计算机工程系, 乌鲁木齐 830011)

摘要: 针对闪光造成的光照变化会导致视频帧之间巨大的强度差异问题, 提出利用反向投影的 flash 场景自适应视频编码算法; 根据直方图差异提取闪光和非闪光帧, 相应地为每个帧分配适当的编码类型, 并在加权预测 (WP) 参数集确定中采用运动向量导数, 通过反向投影保证 flash 场景的全局一致性; 实验结果显示, 提出的算法在 Lena、Peppers、Building、Baboon、Nestling 5 个视频上的峰值信噪比 (PSNR) 值分别可高达 32.31 dB、34.14 dB、34.76 dB、34.94 dB、35.05 dB, 非常接近原始图像的 PSNR; 相比其他几种加权预测算法, 提出的算法在 PSNR 及计算复杂度方面均获得了更加优越的编码性能。

关键词: Flash 场景; 自适应编码; 加权预测; 峰值信噪比; 反向投影; 计算复杂度

Adaptive Video Coding Algorithm for Flash Scene Using Back Projection

Liang Chuanjun, Xu Liang

(Department of Computer Engineering, Xinjiang Institute of Engineering, Urumqi 830011, China)

Abstract: Aiming at the huge intensity difference between video frames caused by the illumination change of flash, adaptive video coding algorithm based on flash scene is proposed in this paper. The flash and non flash frames are extracted according to the histogram difference, corresponding coding types are allocated for each frame correspondingly, and the motion vector derivative is used in the weighted prediction (WP) parameter set determination. And the back projection is used to keep the global consistency of flash scene. The experimental results show that the peak signal to noise ratio (PSNR) achieved by the proposed algorithm on the five videos Lena, Peppers, Building, Baboon and Nestling can arrive at 32.31 dB, 34.14 dB, 34.76 dB, 34.94 dB and 35.05 dB, respectively, which is nearly to PSNR of the primary images. And compared with the traditional weighted prediction algorithm, the proposed algorithm achieves better coding performance in terms of PSNR and computational complexity.

Keywords: flash scene; adaptive coding; weighted prediction; peak signal to noise ratio; back projection; computational complexity

0 引言

在新闻发布会、体育比赛、新闻采访等场合上闪光造成的光照变化会导致视频帧之间巨大的强度差异^[1], 这在运动估计和视频编码补偿中很容易被认错^[2]。

基于帧的加权预测 (Weighted Prediction Based on Frame, WPBF) 是一种加权预测 (Weighted prediction, WP) 方法^[3], 它仅能对具有全局亮度变化的场景进行编码, 例如有效地淡入和淡出, 但不能编码具有 flash 场景局部亮度变化的场景^[4]。基于宏块 (Macroblock, MB) 的方法用于解决局部亮度变化问题^[5], 但是在执行运动估计 (Motion Estimation, ME) 中增加了计算复杂度^[6]。有文献基于宏块概念扩展到多视点的视频以解决视图中光照和焦点的不匹配^[7]。还有文献采用二次搜索算法, 第一次搜索中, 使用平均删除搜索计算窗口中的视差向量, 第二次平均删除搜索通过过滤参考量找出每个 MB 的最佳匹配, 该算法也需要很高的计算复杂度, 且仅适合多视点视

频编码^[8]。自适应稀疏域选择算法 (Adaptive sparse domain selection algorithm, ASDS) 通过差值和修复非闪光帧来解决 flash 场景编码的问题, 然而闪光帧的客观质量下降较严重^[9]。

基于上述分析, 为了解决闪光造成的光照变化会导致视频帧之间巨大的强度差异问题, 提出一种利用反向投影的 flash 场景自适应视频编码算法, 根据直方图^[10]差异提取闪光和非闪光帧, 相应地为每个帧分配适当的编码类型, 并在 WP 参数集确定中采用运动向量导数^[11]。实验结果表明, 提出的算法在编码 flash 场景中相比传统算法具有更好的性能。

1 提出的自适应 flash 场景编码算法

闪光灯的显著特点是在很短的时间内同一场景帧之间的亮度突变。采用提出的基于 MB 的加权预测对闪光帧进行编码工作, 根据输入场景中的闪光自适应改变编码顺序, 以提高编码的效率^[12]。非闪光帧 f_{t+2} 可以通过其他非闪光帧 f_t 进行预测, 因此 f_{t+2} 不会产生大量的比特。算法流程图如图 1 所示。

1.1 基于 flash 的自适应编码顺序

要达到这一效果, 需要进行闪光场景检测, 因此可以改变预测结构中的编码顺序, 从而更易于得到具有闪光灯效果的视频序列的编码增益。闪光灯检测方法在视频自动索引、浏览和检索方面已被广泛研究, 研究表明^[13], 可通过计算平均强度或直方图与相邻图像之间比值, 以完成对一个 flash 帧的定位。提出的算法通过改进文献 [13] 提出的基于直方图差异化

收稿日期: 2017-01-13; 修回日期: 2017-02-21。

基金项目: 新疆维吾尔自治区高校科研计划青年教师科研启动基金项目 (XJEDU2014S074); 新疆工程学院科研基金项目 (2015xgy101712)。

作者简介: 梁传君 (1980-), 女, 硕士, 讲师, 主要从事算法设计、图像处理等方向的研究。

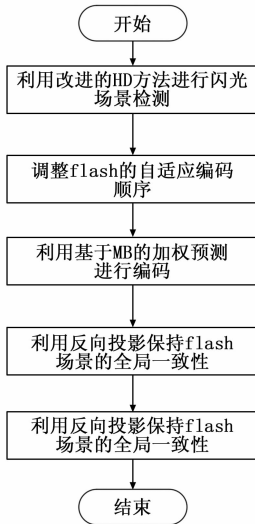


图 1 提出的算法的流程图

(Histogram differentiation, HD) 的方法, 使其更加适用于视频编码。

对于编码一个 m 帧的 flash 场景, 两个非闪光帧和 $m-2$ 闪光帧应该考虑在内。令 f_{FLS} 和 f_{FLE} 分别代表一个 flash 场景的开始和结束帧, f_{FLS} 和 f_{FLE} 之间的关系可以写为:

$$Luma(f_{FLE}) \approx Luma(f_{FLS}) \quad (1)$$

式中, $Luma(f)$ 表示所有像素 $p(x, y)$ 的平均亮度值, 帧 f 的大小为 w 和 h 时, 可以表示为:

$$Luma(f) = \frac{1}{w \times h} \sum_{(x, y) \in f} p(x, y) \quad (2)$$

闪光时亮度水平开始从 f_{FLS-1} 到 f_{FLS} 有一个较大的增加, 即 $Luma(f_{FLS-1}) \ll Luma(f_{FLS})$ 。随后是一段恒定从 f_{FLS} 到 $f_{FLS+m-3}$ 的高亮度水平, 并结束于亮度水平从 $f_{FLS+m-3}$ 到 $f_{FLS+m-2}$ 一个较大的下降, 即 $f_{FLS+m-3} \ll f_{FLS+m-2}$ 。然后根据帧之间的直方图差异, flash 场景检测的问题可以转换为确定 f_{FLS-1} 和 f_{FLE+1} 。帧 f_a 和帧 f_b 之间的直方图差异可以表示为:

$$HD(f_a, f_b) = \sum_0^{63} |Hist(f_a)_x - Hist(f_b)_x| \quad (3)$$

其中, $Hist(f_a)_x$ 和 $Hist(f_b)_x$ 分别表示 f_a 和 f_b 的第 x 位的归一化亮度, 每个直方图划为 64 位用于噪声抑制和快速计算。因此, f_{FLS-1} 的直方图 $Hist(f_{FLS-1})$, f_{FLS} 的直方图 $Hist(f_{FLS})$ 从左偏移到右。 $f_{FLS+m-2}$ 的亮度水平 $Luma(f_{FLS+m-2})$ 保持与闪光灯效果后 f_{FLS-1} 的亮度水平 $Luma(f_{FLS-1})$ 相似的水平, 即 $Luma(f_{FLS+m-2}) \approx Luma(f_{FLS-1})$ 。因此, f_{FLS-1} 和 $f_{FLS+m-2}$ 之间的直方图差异 $HD(f_{FLS-1}, f_{FLS+m-2})$ 是非常小的, 得到以下的关系:

$$HD(f_{FLS-1}, f_{FLS+i-3}) > k \cdot HD(f_{FLS-1}, f_{FLS+m-2}), 3 \leq i \leq 5 \quad (4)$$

k 是一个控制 flash 场景检测器敏感度的整数。

为了检测 flash 场景, 采用一个 n 帧的滑动窗口, 其第一帧是被编码的当前帧 f_c 。计算 f_c 和滑动窗口中其他帧之间的直方图差异 $HD(f_c, f_{c+j}), 1 \leq j \leq n$ 。如果存在满足关系式(4)的 $f_{FLS+m-2} (m < n)$, f_c 预计为 f_{FLS-1} 。通常情况下, 闪光灯效果持续时间不超过 5 帧。因此, 滑动窗口的大小 n 设置为 6。因为

f_{FLS-1} 和 $f_{FLS+m-2}$ (或 f_{FLE+1}) 均是非闪光帧, 相关性高且通过采用 f_{FLS-1} 作为参考帧能够有效地将 $f_{FLS+m-2}$ 编码为 P 帧。此外, 对于 f_{FLS-1} 和 $f_{FLS+m-2}$ 之间的闪光帧, 在加权预测的帮助下被编码为 B 帧。

1.2 基于 MB 的运动向量加权预测

在 $f_{FLS+m-2}$ 编码后, 使用基于宏块的加权预测, 编码所检测的 flash 场景内所有闪光帧 ($f_{FLS+i}, 0 < i < m-2$)。一个闪光帧的典型特点是在整个图像上强度变化是非均匀分布的, 即图像的一些区域可能具有比其他区域更大的强度增益。基于宏块的加权预测技术允许每个宏块有子集的 W 和 O , 它更适合于编码闪光帧用于处理不同区域中不同数量的亮度变化。然而, 在基于宏块的加权预测中, 每个宏块的加权预测参数集不能正确预测, 而没有精确的加权预测参数集就不能得到真正的运动向量。对每个搜索点所有可能的加权参数进行一个详尽的评估过程, 以解决检测困境, 但这会大大增加计算复杂度。因此, 通过计算 W 和 O 得到闪光帧初始的向前和向后运行向量 MV , 分别是 \overline{MV}_{FW} 和 \overline{MV}_{BW} 。这些初始 MV 是基于 $f_{FLS+m-2}$ 指向 f_{FLS-1} 的前向 MV , 并保证有一个以上的 MV 通过当前闪光帧宏块进行跟踪。为了选择最具代表性的加权参数集生成 MV , $f_{FLS+m-2}$ 的前向运动向量 \overline{MV}_F 通过闪光帧当前编码的宏块进行跟踪, 然后采用最大的重叠区域用于生成 \overline{MV}_{FW} 和 \overline{MV}_{BW} , 定义如下:

$$\overline{MV}_{FW} = \frac{TD_i}{TD_m} \overline{MV}_P \quad (5)$$

其中, TD_i 是当前 B 帧与 f_{FLS-1} 之间的距离, TD_m 是两个非闪光帧 f_{FLS-1} 和 $f_{FLS+m-2}$ 之间的距离。这些生成的 \overline{MV}_{FW} 和 \overline{MV}_{BW} 指定的区域用于精确定 W 和 O , 使得在不增加较大复杂度的情况下将目标运动考虑在内。另外, \overline{MV}_F 的可靠性对所提出的方案是至关重要的, 因为 \overline{MV}_{FW} 和 \overline{MV}_{BW} 影响闪光帧加权参数的精度。在自适应编码顺序机制的帮助下, f_{FLS-1} 和 $f_{FLS+m-2}$ 之间的相关性较高, $f_{FLS+m-2}$ 的 \overline{MV}_F 也更加可靠。也就是说, 适当缩放的 \overline{MV}_F 可以更好的生成 flash 场景中闪光帧宏块的 \overline{MV}_{FW} 和 \overline{MV}_{BW} , 且只需要宏块的额外信息用于指示加权预测, 然后使用 H. 264 中定义的 Exp-Golomb 编码^[14] 方式将差异进行比较来获得更高的编码效率。

1.3 使用反向投影保证全局一致性

利用反向投影^[15] 保证全局一致性。设 X_i 为加权后的视频帧, 为了最小化 X_i 与当前 B 帧之间的距离, 通过梯度下降法求解以下最小化问题:

$$\min_{TD_i} \|X_i - X_{i-1}\|_2^2 + c \|X_{i-1} - B\|_2^2 \quad (6)$$

式中, $c > 0$ 为权衡参数。修正后的梯度下降方法为:

$$X_{i+1} = X_i - v [(X_i - X_{i-1}) + c(X_i - B)] \quad (7)$$

式中, $v > 0$ 为步长参数, 将优化后的 X 作为具有全局一致性的 flash 场景。

2 实验结果

在两个具有 flash 场景的 720p 视频序列上进行实验来评估各种加权预测算法的性能, 测试序列均为 25 帧每秒。每个实验分别执行 100 次, 每次实验使用不同的观察序列, 这些序列

在每个实验中通过特定过程生成。仿真实验在配置为 Win7 操作系统、intel 酷睿 II 处理器、1.86 GHz 主频、2.95 GB RAM 的 PC 机上进行, 使用 MATLAB 7.1 编程。除此之外, 还在 3 个标准序列的每帧中间应用了合成的闪光灯, 进行实验来表明运动活动、flash 持续时间和闪光灯强度在不同算法上的影响。将提出的算法与 3 种算法进行比较, 分别为 WPBF、二次搜索算法和 ASDS。

为了比较算法的性能, 使用峰值信噪比 (Peak Signal to Noise Ratio, PSNR)^[16] 指标, PSNR 定义如下:

$$PSNR[dB] = 10 \log_{10} \frac{255^2}{MSE} \quad (6)$$

式中, MSE 为原视频序列与估计视频序列的均方误差, 且 PSNR 越高表明性能越好。

在 Lena、Peppers、Building、Baboon、Nestling 5 个视频上进行编码实验, 计算几种算法的 PSNR 值, 测量结果如表 1 所示, 分辨率增强因子设为 2, PSNR 值越高表示图像的高频成分越多。

表 1 几种算法的 PSNR 值 (dB)

视频	原 PSNR 值	WPBF	二次搜索算法	ASDS	提出的算法
Lena	32.90	27.85	29.04	29.74	32.31
Peppers	34.50	26.96	34.21	30.75	34.14
Building	35.19	26.65	29.41	29.89	34.76
Baboon	35.47	25.47	29.98	28.94	34.94
Nestling	35.45	26.12	31.10	28.47	35.05

从表 1 可以看出, 提出的算法在 Lena、Peppers、Building、Baboon、Nestling 5 个视频上的 PSNR 值分别可高达 32.31 dB、34.14 dB、34.76 dB、34.94 dB、35.05 dB, 非常接近原始图像的 PSNR, 相比其他几种算法, 提出的算法在大多数情况下均能获得最高的 PSNR 值, 表明该算法的编码效果更好。

表 2 所示为各个算法在实验平台保持一致情况下的平均运行时间。

表 2 各个算法的平均运行时间比较

算法	WPBF	二次搜索算法	ASDS	提出的算法
运行时间/s	1.4	1.7	3.5	1.5

从表 2 可以看出, WPBF 算法的运行时间最少, 因为 WPBF 算法最简单。通过对 MATLAB 中常见的视频序列分别进行上述 4 种算法运算处理, 提出的算法运行时间仅略高于 WPBF, 而少于其他几种比较的算法。

综合表 1 和表 2 中的数据可知, 提出的算法在保证较高的编码效率情况下, 仍能保持较低的计算复杂度, 可见其优越性。

3 结论

针对闪光造成的光照变化会导致视频帧之间巨大的强度差异问题, 提出一种利用反向投影的 flash 场景自适应视频编码算法。该算法根据直方图差异提取闪光和非闪光帧, 相应地为每个帧分配适当的编码类型, 并在加权预测参数集确定中采用运动向量导数, 通过反向投影保证了全局一致性。在两个具有

flash 场景的 720p 视频序列上进行实验来评估几种加权预测算法的性能, 在 PSNR 和平均运行时间方面对几种算法进行比较, 验证了提出的算法的有效性。实验结果表明, 与传统的加权预测算法相比, 提出的算法在 PSNR 及运行时间方面获得了更优越的编码性能。

未来会将提出的算法应用于其他的 flash 场景, 同时引入其他的编码技术, 从而更好地改善 PSNR 和降低算法执行时间。

参考文献:

- [1] 孙中廷, 华钢, 徐永刚. 基于双边信息的分布式视频压缩感知模型研究 [J]. 计算机测量与控制, 2015, 23 (10): 3545-3548.
- [2] Zhang Z, Hou J, Ma Q, et al. Efficient video frame insertion and deletion detection based on inconsistency of correlations between local binary pattern coded frames [J]. Security & Communication Networks, 2015, 8 (2): 311-320.
- [3] 李鹏, 彭宗举, 李持航, 等. 基于纹理特性与空域相关的高效视频编码帧内分级快速算法 [J]. 计算机应用, 2016, 36 (4): 1085-1091.
- [4] Lim S W, Moon J H. Weighted Prediction considering Global Brightness Variation and Local Brightness Variation in HEVC [J]. 2015, 20 (4): 489-496.
- [5] 王风随, 王冠凌, 瞿成明, 等. 基于宏块多相关性的多视点视频编码方法 [J]. 计算机工程, 2015, 41 (2): 258-262.
- [6] Strelow D W. Motion Estimation from Image and Inertial Measurements [J]. International Journal of Robotics Research, 2015, 24 (5): 463-482.
- [7] 王风随, 沈庆宏, 都思丹. 多视点视频编码快速帧间模式选择算法 [J]. 计算机应用, 2014, 34 (1): 167-170.
- [8] 孙新领, 马绍惠, 徐平平. 基于 SIR 粒子滤波和局部搜索算法的运动目标跟踪方案 [J]. 湘潭大学学报, 2016, 38 (2): 84-88.
- [9] 卢迪, 林雪. 一种改进自适应权重稀疏区域立体匹配算法 [J]. 哈尔滨理工大学学报, 2016, 21 (1): 83-88.
- [10] 童学锋, 李书萌, 沈广策, 等. 基于直方图对无损数据隐藏的立体图像编码 [J]. 同济大学学报 (自然科学版), 2015, 43 (8): 1226-1233.
- [11] 张雨浓, 劳稳超, 丁玮翔, 等. 基于 ARIMA 与 WASDN 加权组合的时间序列预测 [J]. 计算机应用研究, 2015, 32 (9): 2630-2633.
- [12] Wang Z, Dong S, Wang R, et al. Dynamic macroblock wavefront parallelism for parallel video coding [J]. Journal of Visual Communication & Image Representation, 2015, 28 (3): 36-43.
- [13] Wu C J, Wang Q, Li H, et al. DWI-associated entire-tumor histogram analysis for the differentiation of low-grade prostate cancer from intermediate-high-grade prostate cancer [J]. Abdominal Radiology, 2015, 40 (8): 3214-3221.
- [14] Li S, Fu J, Liu P, et al. An Information Hiding Approach Based on Integer Transform Coefficient and Virtual Space Encoding for H. 264/AVC [J]. Circuits, Systems, and Signal Processing, 2015, 34 (11): 3573-3594.
- [15] 骆岩红, 李公平, 文茜, 等. 基于投影缺失正弦图的 CT 图像重建研究 [J]. 计算机应用研究, 2015, 32 (2): 609-612.
- [16] 黄璇, 杨晓梅. 基于低秩及全变分的视频超分辨率重建 [J]. 计算机应用研究, 2015, 32 (3): 938-941.