

基于帧间相关性的高动态视频色调映射研究^①

李如春^{②*} 俞 楷^{*} 虞 露^{**}

(* 浙江工业大学信息工程学院 杭州 310023)

(** 浙江大学信息电子工程系 杭州 310027)

摘 要 目前高动态范围视频的色调映射基本采用逐帧处理的方式,效率不高。本文利用视频的帧间相关性,摒弃每帧做色调映射处理,研究讨论了一种新的处理方法。基于颜色直方图提取视频序列的关键帧,将视频序列分组并设置每组参考帧,对参考帧做双边滤波的色调映射处理,将组内每帧图像分块,与参考帧进行块匹配的运动估计,块匹配运动估计采用自适应十字模式搜索法和平均绝对误差匹配准则,利用参考帧中匹配块的信息得到对应的低动态图像,由此避免了对视频序列进行逐帧色调映射处理。通过结构相似性和处理速度两个维度对处理方法进行评估,仿真实验结果表明,与逐帧进行双边滤波压缩处理结果有约 89% 的结构相似度时,处理速度约有 2.7 倍的提高。

关键词 高动态范围视频,色调映射,双边滤波,帧间相关性,运动估计

0 引 言

高动态范围 (high dynamic range, HDR)^[1] 视频是展示真实自然世界中亮度大范围变化的一种高清视频类型,广泛应用于卫星遥感、医学成像、虚拟现实、游戏制作等各种领域^[2]。普通低动态范围 (low dynamic range, LDR) 视频的图像通常采用 RGB 色彩模型,每个色彩通道使用 8bit 整型存储,只能区分 256 个亮度级别,而 HDR 视频图像使用 10bit ~ 16bit 记录数据,支持浮点数运算。在 HDR 视频显示方面,尽管已经出现了一些专业的 HDR 图像显示设备,但价格昂贵,无法大范围普及,而普通显示器一般是 8 位存储,仅能显示 2 个数量级的亮度范围。因此,要在普通显示设备上显示 HDR 视频,必须对 HDR 视频进行色调映射处理 (tone mapping operator, TMO),压缩其动态范围,同时最大程度地保留图像的细节信息 and 对比度。HDR 色调映射研究已经开展多年,研究重点集中在 HDR 静态图像的处理。

对于 HDR 视频,常规的方法是逐帧对其进行 TMO 处理。Jinno^[3] 等人通过对每帧 HDR 图像的高亮度与低亮度范围使用不同的 GAMMA 映射函数进行压缩处理,保持了图像中的局部对比度。Chen^[4] 等人提出利用亮度直方图分布和运动估计算法进行图像分类处理。Donald^[5] 等人提出基于对比度受限的自适应直方图均衡 HDR 视频压缩算法,通过指数加权移动平均值将每帧 HDR 视频图像的映射结果综合,最小化闪烁效应。Shahid^[6] 等人采用两种色调映射方法处理 HDR 视频,在需要保持细节的区域采用 iCAM06 算法,其余部分采用快速高效的直方图均衡算法,并进行去闪烁处理,以防止帧间亮度变化过大。Ozcinar^[7] 等人探究了受时域限制的内容自适应色调映射法,为更好保护映射视频序列的时域相关性,提高编码效率,考虑了输入 HDR 帧的统计特性。Gommelet^[8] 等基于率失真优化设计了一个全局的色调映射算法,使 HDR 图像能够向后兼容。Youssfi^[9] 等为缩短 HDR 视频色调映射的处理时间,

① 国家自然科学基金(61431015)资助项目。
② 女,1968 年生,博士,硕士生导师,副教授;研究方向:多媒体信号处理等;联系人,E-mail: lrc@zjut.edu.cn
(收稿日期:2018-01-06)

一方面简化算法,同时在硬件上采用 GPU 并行进行处理。上述文献主要通过对 HDR 视频的每帧图像做色调映射处理完成压缩的目的,同时还需考虑视频的闪烁效应,导致计算量非常大。基于此,本文摒弃常规的逐帧色调映射方法,利用视频的帧间相关性,通过运动估计匹配,实现对 HDR 视频的压缩处理。

1 算 法

本文给出的算法流程如图 1 所示。首先将需处理的 HDR 视频序列按帧分成若干图片组,每个图片

组的长度 $N \geq 2$; 选取图片组的参考帧,设置为 I 帧,剩余的为此图片组的非参考帧,设为 P 帧;对 I 帧进行 TMO 处理,TMO 的处理方法可以任选合适的,获取 LDR 信息;同组的 P 帧被分成大小相同的块,与 I 帧进行块匹配运动估计,在 I 帧中查找与之相似的匹配块;采用平均绝对误差 (mean absolute difference, MAD) 匹配准则,确定当前块与匹配块的相似度阈值,将运动估计所获得的匹配结果分成三种不同的情况,分别进行处理,最终获取相应的 LDR 信息。

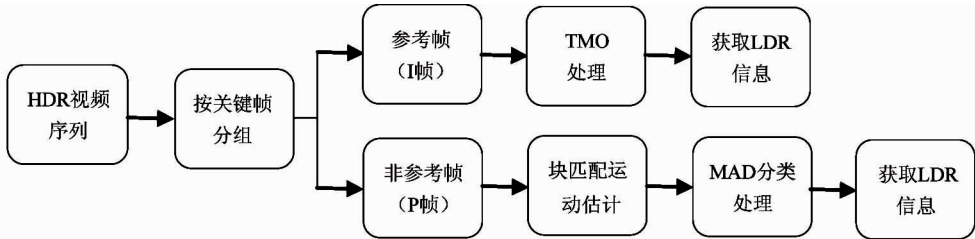


图 1 算法流程示意图

1.1 分组及关键帧提取

视频分组时,每组的帧数大小会影响视频压缩的质量与处理时间。如果每组的视频帧数过小,则一个视频序列中需要进行完整色调映射的帧数过多,处理时间较长;如果每组的视频帧数过大,则会影响每组末尾图像的映射效果,造成较为明显的“残影”现象,同时组与组之间的过渡会比较突兀,所以需综合考虑。在参考帧确定方面,首先定义每组的第 1 帧为参考帧,对于场景切换较快的视频,有可能在组内会出现场景切换点,此时相邻的两帧会有较大的变化,需重新定义能够代表视频主要内容的图像帧为 I 帧,同时提前结束分组,以此帧为首帧重新开始下一个分组。

1.1.1 依据重建帧质量决定分组帧数

分别取两个分辨率为 1920×1080 、总长度为 100 帧的视频序列 Students 和 Hallway,其中 Students 场景切换较快,Hallway 场景切换较慢。根据前述图 1 描述的算法,分块处理的大小取 40×40 ,相似度阈值取 20,TMO 采用双边滤波处理方法,块匹配运动估计采用自适应十字模式 (adaptive rood pattern

search, ARPS) 搜索法^[10],选取视频序列的分组长度分别为 6 帧、11 帧,基于 Matlab 软件平台进行 HDR 视频压缩仿真处理,实验结果如图 2 所示。由图 2 可知,在场景切换较快的 Students 视频序列中,距离参考帧较远的第 6 重建帧图像开始出现稍微残影,第 11 重建帧图像则出现明显残影;场景切换较慢的 Hallway 序列也在第 11 重建帧图像出现比较明显的块效应。因此,视频序列的分组长度不宜取得太长,一般取 4 ~ 6 帧为宜。

1.1.2 关键帧提取方法

视频是由一系列图像帧组合而成的,其中可以代表视频主要内容和运动变化的图像帧称为关键帧。关键帧提取的方法有多种^[11,12],常用的有:基于镜头边界的提取方法,只提取视频镜头序列中的第一帧和最后一帧作为关键帧;基于运动分析的提取方法,分析和计算光流得到视频序列的运动量,比较各帧运动量的值,选取局部最小值的帧为关键帧;基于聚类的提取方法,依据图像间相似度的大小,将视频帧序列进行聚类,然后依次从每个聚类簇中选取关键帧。折中考虑关键帧的选取效果和处理时



图2 不同分组长度的重建帧图像

间,本文采用基于颜色直方图的关键帧提取方法^[13]。

将视频帧图像从 RGB 空间转换到 HSV 空间,将 H 、 S 、 V 三个分量按照人类的颜色感知进行非等间隔量化。HSV 颜色空间的量化方案有很多,本文将色度 H 空间分成 16 份,饱和度 S 和亮度 V 空间分别分成 4 份,计算颜色直方图。通过计算相邻两帧的颜色直方图相似度,同设定的阈值比较,提取关键帧。颜色直方图相似度的计算方法有多种,主要有卡方系数、相关系数、相交系数、巴氏距离等。本文采用一种 X^2 直方图相似度计算方法^[14],公式如下:

$$S(H_k, H_{k+1}) = \sum_{m=1}^n \frac{[H_k(m) - H_{k+1}(m)]^2}{H_k(m)} \quad (1)$$

式中, $H_k(m)$, $H_{k+1}(m)$ 分别表示归一化后的颜色直方图 H_k , H_{k+1} 在第 m 个颜色上的值。

针对上述相似度计算方法,其关键帧提取阈值的设定,根据经验和对多段视频的仿真分析,令测试序列所有相邻帧颜色直方图差的平均值为 $\bar{D}$, 当阈值取 $T \approx 5\bar{D}$ 时^[15], 关键帧提取有较好的准确率。

1.2 I 帧色调映

在图 1 所示的算法流程中,需对选定的参考帧 I 帧进行色调映射处理,采用双边滤波的色调映射处理方法,以期获取视觉效果优越的 LDR 图像。双边滤波算法^[16]是在经典高斯滤波算法的基础上,由 Durand 等提出的图像滤波算法,其基本原理:先将图像由 RGB 颜色空间转到 YUV 颜色空间,获取图像的亮度分量 Y ,利用双边滤波技术将图像的亮度分量分层,得到动态范围变化较大的基本层和反应局部细节的细节层,然后将基本层进行对比度压缩衰减,而对细节层采取全部保留,合并基本层和细节层,得到处理后的亮度分量,最后还原成彩色图像实现色调映射的最终效果。算法流程如图 3 所示。

双边滤波器是一种边缘保持的滤波器,每个像素的权重由空间域中的高斯函数和强度域中的影响函数共同决定。对于像素 p , 双边滤波输出结果可用以下公式表示:

$$F^b(p) = \frac{1}{W} \sum_{p, q \in \Omega} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I(p) - I(q)|) I(q) \quad (2)$$

$$W_p = \sum_{q \in \Omega} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I(p) - I(q)|)$$

(3)

式中, q 为邻域 Ω 内一个像素点, $I(q)$ 为该像素点对应的像素值, G_{σ_s} 代表具有核参数为 σ_s 的空域高斯

函数, G_{σ_r} 代表具有核参数为 σ_r 的强度域高斯函数, W_p 为归一化系数, $F^b(p)$ 是像素 p 双边滤波的输出结果。通过选择合适的 σ_s 和 σ_r 值, 将图像的亮度分量分为基本层和细节层。

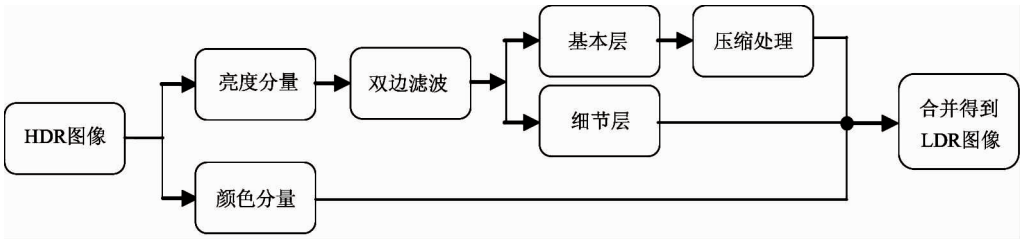


图 3 基于双边滤波的高动态范围图像色调映射算法流程图

对于图像的基本层,使用一比例因子在对数域内压缩衰减亮度分量 Y 的动态范围,公式如下:

$$Y_{\text{factor}} = \lg[g/(\max(Y_{\text{base}}) - \min(Y_{\text{base}}))]$$

(4)

式中, Y_{base} 为图像亮度的基本层信息, g 为可调节参数,其数值大小反映压缩的对比度,经验取值范围为 5 ~ 50,具体取值需根据视频图像的内容进行合适的仿真选取。若图像中有大面积的较亮区域或者较暗区域的细节丢失,则选择较小的 g 值;反之,选择较大的 g 值。

1.3 P 帧匹配处理

首先获取 P 帧图像的亮度分量,将 I 帧和 P 帧图像的亮度分量做线性映射,获取对应的低动态分量 I_l 和 P_l ,用于块匹配运动估计,这样能够有效缩短处理所需的时间。由于自适应十字搜索法相比于钻石搜索、十字搜索等方法快速和高效,本文采用自适应十字模式搜索法。将 P 帧图像划分成相等大小的不重复块,对于其中的每一块,在 I 帧中搜索相似的块,采用平均绝对误差匹配准则进行匹配处理。在匹配过程中,需要设定相似度阈值 T_{MAD} ,并将运动估计所获得的匹配结果分成 3 种不同的情况,分别进行处理。具体如下:

(1)在 I 帧中找到完美的匹配块,即 MAD 为 0,则 P 帧压缩后的低动态信息 P_{ldr} 选取 I 帧匹配块所对应的低动态信息 I_{ldr} ;

(2)在 I 帧中找到相似的匹配块,即当前块与匹配块的 MAD 小于等于相似度阈值 T_{MAD} ,则做如下

处理:

① 通过运动矢量,在 I 帧细节层 (I_{detail}) 中找到对应块。

② 将 I_{detail} 中获得的块信息与线性映射的 P_l 相结合,获得最终的低动态信息 P_{ldr} 。

(3)在 I_l 中没有找到合适的匹配块,即当前块与匹配块的 MAD 大于相似度阈值 T_{MAD} ,需对 P 帧的当前块单独做色调映射处理,以获取 P_{ldr} 。

2 实验结果及分析

为验证本文所提算法的可行性,本节将 HDR 视频序列基于帧间相关性与每帧均做色调映射的压缩方法进行比较。从结果图像与原始图像的结构相似性 (structural similarity, SSIM) 和处理所需要的时间两个维度分别进行比较,且取每帧双边滤波 TMO 处理的图片结果作为结构相似度 SSIM 的比较标准。

仿真实验基于 Matlab 软件平台,选择 4 个不同特点的 HDR 视频序列,具体如表 1 所示。

表 1 4 个测试视频序列的特点

测试序列	分辨率	内容、场景切换
Festival	1920 × 1080	细节多,场景切换慢
Students	1280 × 720	细节少,场景切换快
Hallway	1280 × 720	细节少,场景切换慢
Window	1280 × 720	细节多,场景切换快

仿真实验中,每个测试视频序列取 100 帧,分组长度为 5 帧,起始设定每组的第 1 帧为参考帧,检索遇到关键帧时,则设定此关键帧为参考帧,并另起重新分组;运动估计中的阈值设置范围为 4 ~ 20,块大小设置的范围为 10 × 10 ~ 120 × 120。

2.1 关键帧提取结果

根据前述 1.1.2 节所述的关键帧提取算法,针对表 1 所列视频序列进行测试。在提取关键帧的时候应该遵循关键帧间的相似度最小原则,从而保证关键帧代表的信息最大化。选取相邻帧中的后者为关键帧,获得视频序列关键帧提取结果如表 2 所示。

表 2 视频序列关键帧提取结果

测试序列	总帧数	关键帧数	处理时间(s)
Festival	100	2	6
Students	100	3	5
Hallway	100	2	4
Window	100	4	5

2.2 I 帧色调映射结果

I 帧的色调映射方法可以自由选择,论文选取效果较好的双边滤波色调映射进行处理。这一过程中,空间邻域标准差 σ_s 和像素亮度标准差的选取是关键。 σ_s 有两种效果相当的经验取值,分别是: $0.02 \times \max(\text{height}, \text{width})$, $0.02 \times \min(\text{height}, \text{width})$, 其中 height 和 width 代表了图像的分辨率; σ_r 的取值范围一般在 0 到 1 之间,其取值对基本层影响不明显,而对细节层影响较大,一般取 0.7 左右能达到较好的效果。

选取 Festival 序列中的 I 帧作为示例,其双边滤波处理后获得的图像如图 4 所示。

从图中可以观察到,HDR 图像在普通显示器上不能显示高亮度部分的背景细节信息,经过双边滤波的色调映射处理之后,图像的细节信息明显呈现,视觉效果远好于 HDR 图像直接在普通显示器上的显示。



图 4 I 帧色调映射处理结果

2.3 块大小和相似度阈值选取对压缩效果的影响

依据前述的帧间相关性 HDR 视频压缩算法,分别考虑运动估计时块大小选取和相似度阈值对压缩效果的影响,同时给出每帧都做双边滤波处理的结果,作为算法的比较标准。算法效果从结构相似度

和处理时间两个维度进行考量,仿真结果如表 3 和表 4 所示。为直观起见,以 Students 序列为例,将块大小选取和相似度阈值对压缩效果影响的仿真结果画成曲线图,得到图 5、图 6 的仿真结果图。图 6 中,纵坐标为本文算法的处理速度和每帧都做双边

滤波处理速度的比值。

表 3 固定阈值 $T_{MAD} = 20$ 时不同块大小的仿真结果

测试序列	Festival		Students		Hallway		Window	
块大小	SSIM	时间 (s)	SSIM	时间 (s)	SSIM	时间 (s)	SSIM	时间 (s)
10 × 10	0.839	156	0.903	164	0.873	159.4	0.919	150
20 × 20	0.811	65	0.896	69	0.869	68	0.902	69
40 × 40	0.797	40	0.891	44	0.867	40	0.895	42
100 × 100	0.781	30	0.890	30	0.866	34	0.892	32
120 × 120	0.778	30	0.885	28	0.864	32	0.890	30
每帧双边滤波	1	84	1	80	1	82	1	84

表 4 固定块大小为 (40 × 40) 时不同阈值的仿真结果

测试序列	Festival		Students		Hallway		Window	
阈值大小	SSIM	时间 (s)	SSIM	时间 (s)	SSIM	时间 (s)	SSIM	时间 (s)
4	0.799	58	0.894	72	0.867	48	0.896	54
7	0.798	48	0.892	54	0.867	44	0.895	48
10	0.798	44	0.892	48	0.867	43	0.895	46
20	0.797	40	0.891	44	0.867	40	0.895	42
每帧双边滤波	1	84	1	80	1	82	1	84

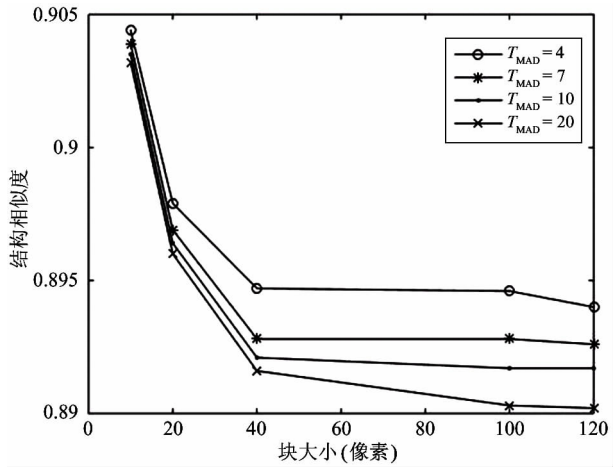


图 5 不同阈值下块大小与 SSIM 的关系

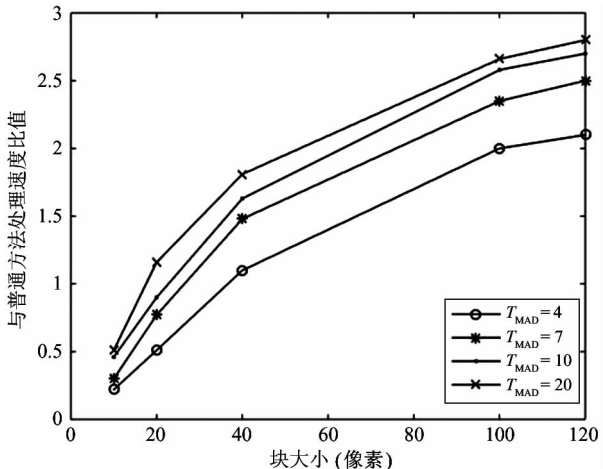


图 6 不同阈值下块大小与处理速度的关系

由上述表图可知,实验结果跟视频序列的内容细节和场景切换快慢没有直接关系,随着运动估计块大小增大,结构相似度略有下降,但处理时间缩短;当相似度阈值取值较小时,结构相似度较好,但处理时间较长;当块大小超过 100 × 100 时,结构相似度和处理时间趋于一定固定值。相比于每帧都做双

边滤波 TMO 压缩处理方法,本文算法在保持约 89% 左右的结构相似度时,压缩处理所需的时间大大缩短,当相似度阈值为 20,块大小取 100 × 100 时,处理速度大约有 2.7 倍的提高,充分说明本文算法的优越性。

3 结 论

本文研究讨论了一种基于帧间相关性的 HDR 视频压缩方法,在获得丰富细节内容的 LDR 视频的同时,提高了压缩处理所需的时间。本文算法利用视频帧间相关性,仅对每组的参考帧做色调映射处理,其余帧利用运动估计匹配法获取压缩处理结果,避免了对每帧图像都做色调映射压缩处理,从而大大节省了处理时间。仿真结果表明,本文算法在保持与每帧双边滤波 TMO 压缩处理结果约 89% 的结构相似度时,处理速度约有 2.7 倍的提高。

参考文献

- [1] Debevec P E, Malik J. Recovering high dynamic range radiance maps from photographs[J]. *ACM siggraph computer graphics*, 1997, 97:369-378
- [2] Chalmers A, Debattista K. HDR video past, present and Future: a perspective [J]. *Signal Processing: Image Communication*, 2017, 54: 49-55
- [3] Jinno T, Mouri K, Okuda M. HDR video tone mapping based on gamma blending [C]. In: Proceedings of the IEEE International Conference on Image Processing, Hong Kong, China, 2010. 2521-2524
- [4] Chen T Y, Chen C H, Tseng T F, et al. HDR enhancement based on bi-curve mapping for image sequences [C]. In: Proceedings of the 10th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kitakyushu, Japan, 2014. 285-288
- [5] Natale D J, Baran M S, Tutwiler R L. High dynamic range (HDR) video processing for the exploitation of high bit-depth sensors in human-monitored surveillance [C]. In: Proceedings of the Applied Imagery Pattern Recognition Workshop, Washington, USA, 2015. 1-6
- [6] H. Shahid, D. Li, A. Fanaswala, et al. A new hybrid tone mapping scheme for high dynamic range (HDR) videos [C]. In: Proceedings of the 2015 IEEE International Conference on Consumer Electronics, Las Vegas, USA, 2015. 351-352
- [7] Ozcinar C, Lauga P, Valenzise G, et al. HDR video coding based on a temporally constrained tone mapping operator [C]. In: Proceedings of the 2016 Digital Media Industry & Academic Forum, Santorini, Greece, 2016. 43-47
- [8] Gommelet D, Roumy A, Guillemot C, et al. Gradient-based tone mapping for rate-distortion optimized backward-compatible high dynamic range compression [J]. *IEEE Transactions on Image Processing*, 2017, 26 (12):5936-5949
- [9] Youssfi Z, Hassan F. Speeding up tone mapping operators-exploiting parallelism for real-time, high dynamic range video [C]. In: Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, USA, 2017. 192-195
- [10] Ma K F, Qiu G. Unequal-arm adaptive rood pattern search for fast block-matching motion estimation in the JVT/H.26L [C]. In: Proceedings of the 2003 International Conference on Image Processing, Barcelona, Spain, 2003. 901-904
- [11] Chugh L, Gupta R, Kumar R. Techniques for key frame extraction: shot segmentation and feature trajectory computation [C]. In: Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Noida, India, 2016. 463-466
- [12] Piperagkas G S, Mariolis I, Ioannidis D. Key-frame extraction with semantic graphs in assembly processes [J]. *IEEE Robotics and Automation Letters*, 2017, 2 (3): 1264-1271
- [13] Zong Z K, Gong Q. Key frame extraction based on dynamic color histogram and fast wavelet histogram [C]. In: Proceedings of the 2017 IEEE International Conference on Information and Automation (ICIA), Macau, China, 2017. 183-188
- [14] 郝会芬. 视频镜头分割和关键帧提取关键技术研究: [硕士学位论文][D]. 武汉:华中师范大学计算机学院, 2015. 10-12
- [15] Jiang H, Helal A S, Elmagarmid A K, et al. Scene change detection techniques for video database system [J]. *Multimedia System*, 1998, 6(3):186-195
- [16] Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic-range images [J]. *ACM Transactions on Graphics*, 2002, 21(3):257-266

Tone mapping operator of high dynamic range video based on temporal correlation

Li Ruchun^{*}, Yu Kai^{*}, Yu Lu^{**}

(^{*} College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

(^{**} Department of Information Electron Engineering, Zhejiang University, Hangzhou 310027)

Abstract

At present, the tone mapping operator of high dynamic range video basically adopts the method of frame-by-frame processing, which is inefficient. In this paper, a new processing method is proposed, the inter-frame temporal correlation of video is used to avoid the tone mapping operator for each frame. Based on the color histogram, the key frames are extracted from the video sequence. After grouping the video sequence, the reference frames for each group are set to be processed by bilateral filtering tone mapping operator. The rest frames in the group are divided into different blocks and matched with the reference frame. To achieve the frame matching, the adaptive rood pattern search method and the mean absolute difference matching criteria are applied, the low dynamic range graphs obtained from frame matching effectively avoids the frame-by-frame tone mapping on the video sequence. The processing method is evaluated by structural similarity and processing speed. The simulation results show that the processing speed is about 2.7 times higher than the result of frame-by-frame bilateral filter compression processing, when the structural similarity is about 89%.

Key words: high dynamic range video, tone mapping operator, bilateral filter, temporal correlation, motion estimation