

文章编号:1007-5321(2017)02-0079-05

DOI:10.13190/j.jbupt.2017.02.012

显著图联合改进的 JND 模型

肖 贺, 涂 钦, 门爱东

(北京邮电大学 通信工程学院, 北京 100876)

摘要: 提出一种改进的基于显著性检测图联合估计恰可失真(JND)阈值的视觉感知模型,将人眼注意力机制引入 JND 模型,通过感知特点建模得到更为精确的 JND 模型. 首先通过改进的显著性检测算法得到相应的显著图,在计算 JND 阈值的过程中,使用显著图来分配不同的权重给 JND 模型,并针对色度和亮度的不同给予不同的权重. 基于空域的 JND 模型主要用在计算图像中的平坦区域;而基于 DCT 域的 JND 模型更加适合计算纹理区域的阈值,新的模型还同时考虑加入对比敏感度函数和各种掩蔽效应因子. 将改进的 JND 模型融合到新的视频编码软件 HM 16.4 中,实验结果表明,与 HEVC 标准的数据对比,视觉感知质量没有明显下降.

关 键 词: 感知视频编码; 人眼视觉系统; 显著性检测; 恰可失真模型

中图分类号: TP391.4

文献标志码: A

Improved Just Noticeable Model with Saliency Map

XIAO He, TU Qin, MEN Ai-dong

(Department of Telecommunications, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Based on existing saliency detection algorithm, a model of joint estimation just noticeable distortion(JND) was proposed. This model not only makes full use of the advantage of the JND model, based on pixel domain and transform domain, but also employs the character of human visual system. The saliency detection algorithm used for saliency map, is employed to assign different weights to JND model, based on different brightness and color. The JND model based on pixel domain is mainly used on plain area, and the JND model, with masking effect and contrast sensitivity function, based on transform domain is used on texture area. The proposed model is introduced into the HEVC sample platform HM16.4. Experiments show that the proposed JND model, compared to HEVC standard, have the similar subjective quality.

Key words: perceptual video coding; human visual system; contrast sensitivity function; just noticeable distortion

视频编码技术主要针对空间、时间和统计冗余进行压缩编码. 近年来,研究人员针对人类视觉系统(HVS, human visual system)的一些特性,将感知编码技术引入传统的视频编码框架,以获得更高的编解码性能^[1]. 所形成的感知视频编码(PVC, perceptual video coding)技术,着重于通过挖掘人眼对

于图像的主观感知状态和人类天生的视觉选择性注意机制等来消除人类主观视觉冗余^[2]. 其中,恰可失真(JND, just noticeable distortion)和 SM 都是当前的研究热点. 目前常见的 JND 模型可以分为两类:基于空域的 JND 模型^[3]和变换域的 JND 模型(如 DCT 域或小波域 JND 模型).

收稿日期: 2016-07-11

基金项目: 国家自然科学基金项目(61671077)

作者简介: 肖 贺(1984—), 男, 博士生, E-mail: xiaoh804@gmail.com; 门爱东(1961—), 男, 教授.

基于空域的 JND 模型直接给出具体每一像素的最小可察觉失真值. 如文献[1]分别计算亮度自适应因子和纹理掩蔽效应因子,然后将2个因子的最大值作为 JND 阈值. 在此基础上,Yang 等^[2]提出了扩展的 JND 模型(非线性叠加掩蔽模型),将亮度掩蔽和纹理融合在一起用于运动估计等内容.

Ahumada 等^[4]在提出的 DCT 域的 JND 模型中融合了对比敏感函数,其还包括亮度自适应因子和各种掩蔽效应因子. Watson 等^[5]提出了 DCTune 模型,通过融合了亮度 and 对比度掩蔽效应到基本阈值中. Ahumada 等^[6]提出了将时空域 JND(ST-JND)的模型嵌入感知视频编解码器中,并使其兼容当前的 H.264 编解码框架. Naccari 等^[7]通过估计频域的扭曲模型调整频域变换参数,得到有效的 DCT 域 JND 模型.

显著性模型用于描述人眼视觉注意力区域. Itti 等^[8]在神经生理和解剖的基础上结合 Treisman 的特征整合理论提出了自底向上建模的视觉注意模型. 由于现实中大量的视频以压缩形式被获取,Goferman 等^[9]提出了面向压缩域的视频显著检测算法,利用一个移动窗口中的离散余弦变换(DCT)系数和运动信息修缮显著图模型,取得了较好的效果.

笔者提出改进的基于显著性引导的 JND 模型(SJ-JND 模型):联合考虑空域的 JND 模型和 DCT 域的 JND 模型,并分别对视频图像纹理、边界和平坦区域分别估计出更为精确的显著性区域;在此基础上,得到图像不同区域的重要程度,对于较为显著的区域给予更小的 JND 权重,并用于视频编码.

1 DCT 域和空域的 JND 模型

空域的 JND 模型主要由亮度自适应因子和对比度掩蔽效应组成. 而现有的 DCT 域的 JND 模型通常是在 DCT 域下计算 JND 阈值,然后再反 DCT 变换到空域 JND 阈值. DCT 域的 JND 模型可以由基本阈值和各种调节因子组成. 亮度的空域 JND 阈值^[3]为

$$J_{\text{DCT}}(x, y) = T_{\text{CSF}}(x, y) A_{\text{lum}} F_{\text{contrast}}(x, y) \quad (1)$$

其中 (x, y) 为对应的元素坐标. $T_{\text{CSF}}(x, y)$ 对应于亮度和色度的基本阈值. F_{contrast} 为由空间域对比度敏感函数生成的对比度掩蔽因子. A_{lum} 为亮度自适应效应,CSF 反映了对比度敏感与空间频率之间的关系.

2 SJ-JND 模型框架

在文献[10]基础上,提出的 SJ-JND 模型主要由2个影响因子组成:联合估计 JND 模型和显著性检测阈值. 有

$$J_{\text{SJ-JND}}(x, y) = J_{\text{JE}}(x, y) S(x, y) \quad (2)$$

图1所示为 SJ-JND 模型生成流程的3个部分. 第1部分是图像颜色空间转换. 第2部分是图像的显著性检测. 第3部分是图像进行 DCT 变换,在 DCT 域下生成 JND 模型,然后把 JND 阈值反变换回空域. 最后,利用式(2)计算得到 SJ-JND 阈值.

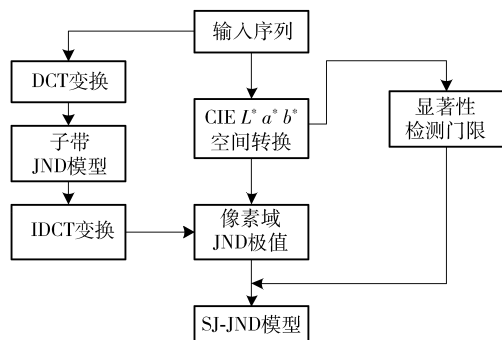


图1 SJ-JND 基础架构

2.1 改进的亮度和色度的基本阈值

SJ-JND 模型中式(3)为亮度的空域 JND 阈值,式(4)为色度的空域 JND 阈值:

$$J_{L\text{-DCT}}(x, y) = T_{L\text{-Basic}}(x, y, p) A_{\text{lum}} F_{L\text{-contrast}}(x, y, p) \quad (3)$$

$$J_{C\text{-DCT}}(x, y) = T_{C\text{-Basic}}(x, y) F_{C\text{-contrast}}(x, y) \quad (4)$$

变换块的第 (x, y) 坐标处对应的空间频率 $\lambda_{x, y}$ 计算如下:

$$\lambda_{x, y} = \frac{\sqrt{(x/\delta_i)^2 + (y/\delta_j)^2}}{2N} \quad (5)$$

$$\delta_i = \delta_j = 2 \arctan \left(\frac{1}{2KH} \right) \quad (6)$$

参数 δ_i 和 δ_j 分别表示为像素的水平和垂直视角. 变换系数块的维数是 N . 等式(6)中的 K 表示观看距离与图像高度的比值, H 表示图像高度.

经过实验,得到色度分量的 CSF 函数曲线的色度基本阈值公式为

$$T_{C\text{-Basic}}(x, y, p) = \beta \left[(1 - a) \exp(b\omega^c) + \alpha \right] \quad (7)$$

根据实验,其中参数 α 为 $1/256$,当色度分量为 a^* 时, $\beta = 280$, $b^* = -31.9795$ 和 $c = 2.06$. 对于色度分量为 b^* 时, $\beta = 320$, $b = -22.4413$ 和 $c = 1.74$.

2.2 改进的对比度掩蔽效应模型

2.2.1 改进的亮度对比度掩蔽因子

得到的亮度对比度掩蔽效应因子计算如式下.

$$F_{L_const}(x, y, p) = \begin{cases} \psi, & (x^2 + y^2) \leq 16 \text{ 边线与平面} \\ \psi \min \left(4, \max \left(1, \frac{C_L(x, y, p)}{T_{L_Basic}(x, y, p) A_{lum}} \right) \right), & \text{其他} \end{cases} \quad (8)$$

$$\psi = \begin{cases} 1, & \text{边线与平面} \\ 2.25, & (x^2 + y^2) < 16 \text{ 纹理区} \\ 1.25, & (x^2 + y^2) > 16 \text{ 纹理区} \end{cases} \quad (9)$$

其中,式(8)中参数 $C_L(x, y, p)$ 代表亮度变换块的系数. 而式(9)说明了亮度对比度掩蔽效应在不同类型块所得到的掩蔽因子不同.

2.2.2 改进的色度对比度掩蔽因子

在这里,每个变换块的第 (x, y) 坐标的对比掩蔽因子计算公式如下:

$$F_{C_contrast}(x, y, p) = \max \left(1, \frac{C_c(x, y, p)}{T_{C_Basic}(x, y, p)} \right)^{0.36} \quad (10)$$

其中: $C_c(x, y, p)$ 表示第 P 帧的 (x, y) 位置变换块的系数, $T_{C_Basic}(x, y, p)$ 表示为第 P 帧的 (x, y) 位置色度分量的基本阈值.

3 改进的显著性探测算法

3.1 改进的局部和全局相结合的显著性探测

根据 Goferman 等^[9]提出的基于上下文感知的显著性算法,当图像中的一个像素值所在的小块区域(图像块)与其周围的小块区域的特征存在较大差异的时候,当前的小区域块可以被认为是显著性区域. 因此,位置欧氏距离可以用来检测两块的远近程度,如果当前显著的图像块 p_i 与图像块 p_j 距离较远,则认为 p 不是显著块. 块之间的相异度计算公式如下:

$$d(p_i, p_j) = \frac{d_{ED}(p_i, p_j)}{1 + cd_{P_ED}(p_i, p_j)} \quad (11)$$

这里的 c 为 3. 块之间的 $d_{P_ED}(p_i, p_j)$ 越大和 $cd_{P_ED}(p_i, p_j)$ 越小,则当前块越显著. 在单一的 t 尺度下,像素值的显著性计算公式为

$$S_i^t = 1 - \exp \left\{ - \frac{\sum_k d(p_i^t, (p')_k^t)}{k} \right\} \quad (12)$$

3.2 改进的多尺度显著性检测方法

在选择图像块时,要把尺度因素考虑进来. 则

式(19)可以改写为

$$S_i^t = 1 - \exp \left\{ - \frac{\sum_k d(p_i^t, (p')_k^{tk})}{k} \right\} \quad (13)$$

在不同尺度下,像素 i 的平均显著性值为

$$S_i = \sum_{t \in T} \frac{S_i^t(t \in T)}{M} \quad (14)$$

参数 S_i^t 定义在(14). 实验中,首先把输入的图像进行调整到同一个尺度,然后计算其显著性值. 图像块的划分是以 7×7 为单位的,块之间存在重叠. M 设置为 49.

3.3 改进的显著性上下文信息

常见的混合视频编码系统并未考虑人类视觉特性中注意力对于压缩的影响. 根据上述特性,在 SJ-JND 中,对于注意力焦点远点应该尽量压缩,对于注意力焦点周围应该减少压缩权重. 利用这一原理,像素 i 的显著性重新定义为

$$S_i' = \sum_{t \in T} S_i^t (1 - d_{foci}^t(i)) / M \quad (15)$$

其中 $d_{foci}^t(i)$ 表示在尺度为 t 时,像素 i 与最近的显著性像素的位置欧氏距离.

4 实验结果和讨论

为了验证模型的效果,实际测试中使用测试序列分别为: BasketballDrive、BQTerrace、Cactus、Park-Scene、PeopleOnStreet. 序列 PeopleOnStreet 分辨率为 2560×1600 ,其他都为 1920×1080 . 采用 main profile,采用全 I 帧编码方式. 为了表明提出模型的有效性,跟标准 HM16.4 模型和 Wang^[10] 的模型做比较.

4.1 主观评估方案

为了确保主观测试的可靠性,采用 ITU-R BT. 500 标准^[11]中定义的测试方法 DSQS(double stimulus quality scale)主观视觉评测方案.

4.2 基于 SJ-JND 模型的测试结果

表 1 和表 2 显示了提出的 SJ-JND 模型用于视频压缩时比参考模型和标准有更多的比特率下降,其中比特率降低了 2.51% ~ 28.57%,对比标准软件的比特平均节省 10.7%,Wang 等提出模型的比特率平均节省 8.56%,SJ-JND 模型的比特率平均节省了 2.14% 左右. 从表 3 可以看出,SJ-JND 与 Wang 等提出的模型 PSNR 平均下降了 0.129 dB. 从实验结果可以看出,在主观视觉质量没有降低的前提下,使用 SJ-JND 模型可使图像压缩率得到极大提高.

表 1 各个模型的码率大小				
序列	Bitrate/(KB·s ⁻¹)			
	QP	HM 16.4	Wang	Our model
Cactus	21	134 987. 714	100 657. 475	96 425. 925
	26	58 828. 057	54 711. 600	53 252. 575
	31	32 140. 714	30 761. 000	30 090. 600
	36	17 294. 000	16 611. 275	16 405. 950
BasketballDrill	21	134 987. 714	100 657. 475	96 425. 925
	26	12 352. 975	11 248. 950	10 881. 325
	31	6 623. 325	6 227. 650	6 092. 900
	36	3 488. 475	3 378. 325	3 303. 300
BasketballDrive	21	60 203. 028	46 838. 400	46 055. 42
	26	23 356. 628	21 791. 675	21 599. 40
	31	12 190. 542	11 914. 375	11 884. 45
	36	6 735. 028	6 481. 575	6 462. 875
PeopleOnStreet	21	115 357. 05	109 099. 350	106 169. 310
	26	67 291. 59	65 398. 995	63 383. 220
	31	38 600. 04	37 639. 770	36 700. 605
	36	22 063. 50	21 826. 875	21 474. 855
ParkScene	21	55 388. 667	45 286. 332	43 468. 572
	26	30 046. 738	26 068. 932	25 125. 180
	31	15 923. 986	14 504. 868	13 984. 680
	36	7 815. 840	7 285. 200	7 086. 024
BQTerrace	21	198 802. 834	3 167 034. 84	161 823. 84
	26	109 458. 754	0 98 815. 05	96 566. 07
	31	59 331. 942	0 56 008. 20	55 004. 34
	36	32 531. 280	0 30 728. 64	30 405. 96

表 2 模型的码率节省			
序列	QP	Wang' s bitrate	Our Bitrate
		Saved/%	saved/%
Cactus	21	25. 43	28. 57
	26	7	9. 48
	31	4. 29	6. 38
	36	3. 95	5. 14
BasketballDrill	21	25. 43	28. 57
	26	8. 94	11. 91
	31	5. 97	8. 01
	36	3. 16	5. 31
BasketballDrive	21	22. 2	23. 5
	26	6. 7	7. 52
	31	2. 27	2. 51
	36	3. 76	4. 04

续表 2			
序列	QP	Wang' s bitrate	Our Bitrate
		Saved/%	saved/%
PeopleOnStreet	21	5. 42	7. 97
	26	2. 81	5. 8
	31	2. 49	4. 92
	36	1. 07	2. 67
ParkScene	21	18. 24	21. 52
	26	13. 24	16. 38
	31	8. 91	12. 18
	36	6. 79	9. 34
BQTerrace	21	15. 98	18. 6
	26	9. 72	11. 78
	31	5. 6	7. 29
	36	5. 54	6. 53
平均值		8. 56	10. 7

表 3 PSNR 大小和 MOS 评分						
序列	PSNR/dB				MOS	
	QP	HM 16.4	Wang	Our model	Wang	Our
Cactus	21	41. 514	37. 332	37. 141	5	4. 8
	26	38. 288	36. 248	36. 113	4. 8	4. 8
	31	36. 044	34. 714	34. 629	4. 7	4. 6
	36	33. 578	32. 675	32. 640	4. 6	4. 7
BasketballDrill	21	42. 492	38. 236	37. 971	5	5
	26	38. 986	36. 871	36. 688	5	4. 9
	31	35. 993	34. 959	34. 867	4. 9	4. 9
	36	33. 320	32. 735	32. 679	4. 8	4. 5
BasketballDrive	21	41. 754	39. 349	39. 237	4. 8	4. 7
	26	39. 673	38. 167	38. 074	4. 7	4. 6
	31	38. 263	36. 998	36. 876	4. 5	4. 5
	36	36. 579	35. 406	35. 367	4. 4	4. 3
PeopleOnStreet	21	43. 958	38. 334	37. 971	5	5
	26	40. 456	37. 295	37. 059	4. 9	4. 8
	31	37. 312	35. 581	35. 471	4. 8	4. 7
	36	34. 409	33. 423	33. 3813	4. 6	4. 6
ParkScene	21	42. 186	36. 406	36. 282	5	4. 9
	26	39. 274	35. 687	35. 592	4. 8	4. 7
	31	36. 422	34. 326	34. 27	4. 6	4. 6
	36	33. 607	32. 384	32. 363	4. 6	4. 5
BQTerrace	21	43. 705	37. 222	36. 922	5	5
	26	38. 614	35. 625	35. 427	4. 9	4. 9
	31	35. 110	33. 476	33. 373	4. 9	4. 8
	36	32. 081	30. 856	30. 822	4. 6	4. 5
平均值		38. 067	35. 597	35. 468		

5 结束语

提出基于显著性图引导联合估计 JND 阈值的模型,对不同区域得到不同的 JND 阈值,并且除了由亮度自适应因子和对比度掩蔽因子组成外,还加入 CSF 函数. 将图像分解为结构和纹理分量,动态给予不同权重. 最后对不同区域进行 JND 阈值估计,还考虑了图像中感兴趣区域的分布. 实验结果表明,改进的 JND 模型与 HEVC 标准的数据对比,视觉感知质量没有明显的下降. 提出的模型与现有的模型对比存在更好的压缩性能,能够容忍更多的视频数据失真. 提出的模型在 DCT 域建模 JND 还存在分块不合理等问题,在未来的工作中,可以进一步改善.

参考文献:

- [1] Chou C H, Li Y C. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile[J]. IEEE Transaction on Circuits and Systems for Video Technology, 1995, 5(6): 467-476.
- [2] Yang X K, Lin W S, Lu Z, et al. Just-noticeable distortion profile with nonlinear additivity model for perceptual masking in color images[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. Hongkong: [s. n.], 2003: 609-612.
- [3] Xu Shengyang, Yu Mei, et al. Just noticeable coding distortion model for HEVC video coding[J]. Journal of Optoelectronics and Laser, 2015, 26(12): 2381-2392.
- [4] Ahumada A J, Peterson H A. Luminance-mode based DCT quantization for color image compression[J]. Human Vision, Visual Processing, and Digital Display III, 1992(1666): 365-374.
- [5] Watson A B, Yang G Y, Solomon J A, et al. Visibility of wavelet quantization noise[J]. IEEE Transaction on Image Processing, 1997, 6(8): 1164-1175.
- [6] Ahumada A J, Peterson H A. Luminance-mode based DCT quantization for color image compression[J]. Human Vision, Visual Processing, and Digital Display III, 1992: 365-374.
- [7] Naccari M, Pereira F. Advanced H. 264/AVC-based perceptual video coding: architecture, tools, and assessment[J]. Circuits and Systems for Video Technology, IEEE Transactions on, 2011, 21(6): 766-782.
- [8] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
- [9] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2012, 34(10): 1915-1926.
- [10] Wang Huiqi, Wang Lin, Hu Xuelin, et al. Perceptual video coding based on saliency and just noticeable distortion for H.265/HEVC[C]//Wireless Personal Multimedia Communications (WPMC), International Symposium on. Sydney: [s. n.], 2014: 106-111.
- [11] Recommendation I. 500-11—2002, Methodology for the subjective assessment of the quality of television pictures[S]. Geneva, Switzerland: International Telecommunication Union, 2002: 1-47.