

文章编号:1001-9081(2007)07-1654-03

一种基于人眼视觉特性的 ROI 渐进图像传输算法

周大伟, 耿金玲, 郑继明

(重庆邮电大学 计算机科学与技术学院, 重庆 400065)

(no1_david@126.com)

摘要:一般基于 ROI 的渐进传输算法都未充分考虑人眼视觉特性,利用小波变换的特点并结合人眼视觉的掩蔽效应,提出了一种改进的 ROI 渐进图像传输算法。首先传输图像的低分辨率形式,供用户选择是否继续传输图像。进行 ROI 图像传输时,在不影响 ROI 主观质量的前提下,可以对 ROI 视觉效果影响不明显的小波系数延后传输,而利用节省下的带宽传输视觉上重要的背景系数。此外,还可以根据网络带宽情况设置扩展因子,控制背景图像进行分级扩展传输,从而保证接收到的图像始终具有较好的整体视觉效果。仿真实验证明了该算法的有效性及其扩展因子的控制作用。

关键词:人眼视觉特性;小波变换;感兴趣区;渐进图像传输

中图分类号: TN919.8 **文献标志码:** A

ROI progressive image transmission algorithm based on human visual specialties

ZHOU Da-wei, GENG Jin-ling, ZHENG Ji-ming

(College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: In conventional region of interest (ROI) progressive image transmission algorithms, human visual specialties are not taken into account. An improved algorithm based on the properties of wavelet transform and human visual specialties was proposed. Firstly, an image with low resolution was transmitted so that user could decide if the image was needed. On the condition that there is no impact on the subjective quality of ROI, the transmission of ROI coefficients which are not important for visual quality could be delayed. And the bandwidth could be utilized to transmit the important back ground (BG) coefficients. Furthermore, depending on the bit rates, an expansion factor was evaluated. To guarantee the image with better visual effects, the transmission of BG image was controlled by the expansion factor. The experimental results show that the improved algorithm is effective and the expansion factor can accommodate image transmission.

Key words: Human Visual Specialties (HVS); wavelet transform; Region Of Interest (ROI); progressive image transmission

0 引言

随着多媒体技术的发展,图像已经成为网络上必不可少的信息类型,这就需要大量的网络带宽来传输图像信息。为了提高图像传输的效率,人们提出渐进图像传输方法:首先传输图像的轮廓,然后传输图像细节信息,渐进改善图像质量,直到满足用户要求即停止传输。然而,在低比特率下或者用户想通过渐进传输获取大型图像时,可能还是要耗费大量的网络流量和时间,针对这一问题,人们又将图像感兴趣区域 (Region Of Interest, ROI) 的处理引入到渐进图像传输中^[1]。在进行图像传输时,可以优先高质量传输 ROI 图像,等待 ROI 图像传输完毕,再进行背景 (Back Ground, BG) 图像的传输。从而让接收者在极短的时间内获取高质量的感兴趣区图像内容,这极大地提高了图像的传输效率。

本文利用小波变换和人眼视觉特性,提出了一种基于感兴趣区的渐进图像传输算法,对一般的感兴趣区渐进图像传输算法进行了改进。传统的算法^[2]首先是将 ROI 数据全部传输完毕,然后才开始传输背景图像数据。这种方法虽然获

得了较高的 ROI 质量,可是由于没有考虑人眼的视觉特性,图像背景的质量得不到保证,尤其在低比特率下,图像的整体视觉效果不佳。而本文的算法在确保 ROI 图像达到较高主观质量的同时,充分结合了人眼视觉特性,利用有限的带宽资源去改善图像整体视觉效果,使得在渐进传输过程中,接收者看到的总是平滑且具有良好视觉效果的图像。

1 小波变换和人眼视觉特性

多分辨率分析可以将图像信号分解成不同分辨率和包含不同频率成分的小波子带。图像的每一级小波变换都是将低频图像进一步分解为更精细的频带,得到更低分辨率的图像。图像经过多级 (N 级) 小波变换就可以得到一个低频子带 LL_N 和多个高频子带 $HL_i, LH_i, HH_i (i = 1, 2, \dots, N)$ 。图像小波变换后的子带对应了一个最低分辨率、最低频率下的图像的近似信息和不同分辨率、不同频率的细节信息。图像的主要内容信息都包含在这个低频子带中,它是原始图像的最佳逼近。而那些高频子带主要包含了图像的边缘、轮廓、纹理等细节信息,但是它们所表示的细节信息方向不同,子带 HL_i 表示水平方

收稿日期:2007-01-22;修回日期:2007-03-14。 **基金项目:**重庆市自然科学基金资助项目(2005BB2063)。

作者简介:周大伟(1984-),男,安徽六安人,硕士研究生,主要研究方向:图像处理、图像编码;耿金玲(1979-),女,山东临沂人,硕士研究生,主要研究方向:小波分析、图像处理;郑继明(1963-),男,四川简阳人,副教授,主要研究方向:小波分析、多媒体技术。

向, LH_i 表示垂直方向, HH_i 表示对角方向, 水平和垂直方向的细节信息要比对角方向的细节信息更加重要。

表 1 Lena 图各小波子带能量比统计

分辨率	小波子带	能量比 (%)	各分辨率高频子带能量比之和 (%)
16 × 16	LL_4	94.9896	2.7663
	HL_4	0.7650	
	LH_4	1.5796	
	HH_4	0.4217	
32 × 32	HL_3	0.3413	1.2541
	LH_3	0.7221	
	HH_3	0.1907	
64 × 64	LH_2	0.1805	0.6726
	HH_2	0.3827	
	HL_2	0.1094	
128 × 128	HL_1	0.0882	0.3174
	LH_1	0.1897	
	HH_1	0.0395	

我们用 db4 小波对 Lena 图像(256 × 256 × 8) 进行 4 级小波分解, 并统计各个小波子带所占的能量比, 根据统计结果来分析上述的小波变换特性。计算子带能量比的公式定义如下:

$$\text{小波子带能量比} = \frac{\text{小波子带系数模平方和}}{\text{所有小波系数模平方和}}$$

从表 1 的统计结果可以看出低频子带包含了图像的大部分能量, 而高频子带只占图像能量的很小一部分, 且分辨率越低, 高频子带所含的能量越高。就算在同一分辨率下, 各高频子带所含的能量高低也不同, 垂直方向 LH_i 的能量最高, 水平方向 HL_i 的能量要比对角方向 HH_i 高。这些统计数据进一步说明了小波变换的特性, 以及各个子带对图像重构的重要性程度。依据这些特性, 我们可以按子带的重要性程度设计出合理的渐进图像传输码流结构, 以保证首先传输的是图像重构所需的最重要数据。

人眼视觉特性结合人眼的视觉生理和心理对人类的视觉现象进行研究, 通过研究结果发现人眼具有很多视觉掩蔽效应:

- 1) 人眼对图像边缘与轮廓信息的失真比较敏感;
- 2) 人眼对不同方向的细节信息敏感程度不同, 对水平和垂直方向细节信息要比对角方向的细节信息更加敏感;
- 3) 人眼对低频信息的敏感度高于高频端。

目前, 人们已经将人眼视觉特性和图像处理相结合^[4,5], 在各个领域中广泛应用。本文也充分结合了人眼视觉特性, 不过这里所关注的是人眼对图像的可见敏感度有一定的门限, 即人眼对图像主观视觉效果的掩蔽效应。当一幅图像达到一定的主观质量后, 继续改善图像的质量, 人眼是感觉不出来的。因此可以看出, 经过小波变换的图像与人眼的视觉特性十分吻合, 如果可以将小波图像这些特性应用于渐进图像传输之中, 可以有效地提高传输效率, 并且图像主观质量也会得到较大改善。



(a) Lena 小波图像的完全重构 (b) 少了 HH_1 子带重构的图像
图 1 人眼视觉特性测试

为了说明小波变换和人眼视觉的这种特性, 我们以上述

4 级小波分解的 Lena 图重构图像为例, 如图 1 所示。图(a) 是 Lena 小波图像的完全重构, 而图(b) 是少了 HH_1 子带进行重构的图像。尽管客观上图(b) 所用的小波系数比图(a) 少了一个高频对角方向的细节信息, 只有图(a) 所用小波系数的四分之三, 可是从主观视觉上来看, 已经分不出两幅图像有什么质量差别。实验从直观上印证了小波变换特性和人眼视觉的掩蔽效应。

2 ROI 掩模的产生

ROI 是在原始图像上定义的, 经过小波变换后我们需要知道哪些小波系数是重构 ROI 所必需的即与 ROI 相关, 这就需要产生 ROI 掩模。

ROI 掩模是随着小波变换一起产生的, 而具体的 ROI 掩模是由小波逆变换获得, 并且不同的小波涉及到的相关小波系数也不同, 主要是与所用小波的分解合成滤波器长度有关。

如一维 5/3 整数小波变换的小波逆变换公式如下^[6]:

$$X(2n) = L(n) - \frac{H(n-1) + H(n)}{4} \quad (1)$$

$$X(2n+1) = \frac{L(n) + L(n+1)}{2} + \frac{-H(n-1) + 6H(n) - H(n+1)}{8} \quad (2)$$

其中, $X(2n)$ 、 $X(2n+1)$ 代表原始图像中的像素, 而 $L(n)$ 、 $L(n+1)$ 是小波变换后低频子带小波系数, 而 $H(n-1)$ 、 $H(n)$ 和 $H(n+1)$ 代表高频子带小波系数。从(1)(2) 两式可以看出, 像素 $X(2n)$ 与 $X(2n+1)$ 的重构与小波系数 $L(n)$ 、 $L(n+1)$ 、 $H(n-1)$ 、 $H(n)$ 和 $H(n+1)$ 有关, 用同样的方法分别对 ROI 每个像素分别计算其相关小波系数, 得到的这些小波系数的集合就构成了 ROI 的掩模。用类似的方法可以获得其他小波产生的 ROI 掩模, 只是可能涉及到的具体相关系数略有差别。

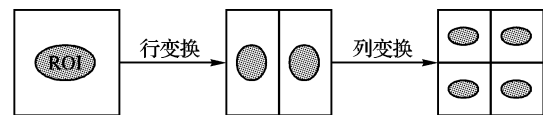


图 2 ROI 掩模的产生

一般 ROI 掩模的计算比较复杂, 如果 ROI 是矩形, 掩模的计算复杂度将大大降低。因为无须逐个计算矩形中的每个像素, 只需将矩形左上角和右下角像素的相关系数求出, 得出矩形在各子带内新的左上角和右下角坐标位置, 位于这两个新坐标点之间的系数都包含在掩模内, 从而极大地简化了掩模的计算, 提高了掩模产生的效率。假设矩形的左上角和右下角坐标为 (X_L, Y_L) 和 (X_R, Y_R) , 小波分解后子带中这两点的新坐标为:

$$X'_L = \frac{X_L}{2} - neg, Y'_L = \frac{Y_L}{2} - neg \quad (3)$$

$$X'_R = \frac{X_R}{2} + pos, Y'_R = \frac{Y_R}{2} + pos \quad (4)$$

常量 neg 与 pos 的取值一方面取决于两个顶点是在低频子带还是高频子带, 另一方面取决于原始信号 (X_L, Y_L) 和 (X_R, Y_R) 是应用低通滤波器还是应用高通滤波器处理的。

例如, 对于 5/3 整数小波变换而言, 若由低频带重建原始信号, 则有 $neg = 0$ 和 $pos = 1$; 利用高通带重建原始信号, 则 $neg = 1$ 和 $pos = 1$ 。

3 改进算法

根据以上所述小波变换系数的特点和人眼视觉系统对于

图像主观质量的掩蔽效应,本文对一般的 ROI 渐进图像传输算法提出了改进。由小波变换可知,图像经过小波变换后,主要的内容信息和能量都集中在低频子带上,虽然此子带对图像的重构非常重要,但这个低频子带却具有极少的系数,它与小波变换的级数 N 成反比, N 级小波变换后,此子带的系数只占有所有系数的 $1/4^N$,同时此图像的分辨率也只有原图像的 $1/4^N$ 。因此在渐进图像传输过程中,可以首先将最低分辨率的图像(小波变换后原始图像的近似逼近图像)进行传输显示,让用户从整体上预览低分辨率的原始图像内容,从而决定是否要继续获取更高分辨率和质量的图像信息。

如果此图像不是用户所需,可以停止对图像数据的传输,而去获取其他图像信息,有效地节约了带宽。若仍需继续改善图像质量,则可以优先传输 ROI 图像的细节信息,使 ROI 率先达到较高的质量。一般基于 ROI 的图像传输方法都是将 ROI 数据全部传输完毕,才开始传输背景图像数据,从人眼视觉特性可以看出这是没有必要的,因为当图像质量改善到一定程度时,人眼已经无法察觉质量的进一步提高。利用这一特性,当 ROI 图像达到较好的主观质量时(一般与原始图像已经看不出明显的差别),就停止对 ROI 图像小波系数的传输,而开始传输背景图像的小波系数。目的是在不影响 ROI 图像主观视觉效果的前提下,用有限的带宽去改善背景图像质量,从而提高图像的整体视觉效果。

在一些特定的应用领域,用户在获取较好视觉质量的 ROI 图像之后,不一定对整个背景图像都感兴趣,有时可能位于 ROI 周围的背景信息相对比较重要。比如医学图像,医生首先肯定对病变区域是最感兴趣的,此外对病变周围的图像内容要比其他背景内容更加重视。尤其在低比特率的情况下,如果将整个背景作为整体进行图像传输,则可能需要传输大量的数据才能达到较好的视觉质量,用户可能在这段时间内获得的背景图像非常模糊,且能感觉到在 ROI 边界有明显的质量跃迁^[3],这极大地影响了整体图像的视觉效果。

为了解决此种问题,本文提出了一种更加灵活的背景传输方式,即背景以 ROI 为中心进行分级扩展渐进传输,由所谓的扩展因子 $S(S \geq 0)$ 来控制背景图像以一定的扩展半径分级传输。扩展因子 S 的设置与 ROI 的感兴趣程度和网络带宽情况相关联,一般来说用户对 ROI 越重视, S 应越小;比特率越高, S 也应越大。

为了简化掩模的计算,一般基于 ROI 的算法都只考虑矩形和椭圆的 ROI,这里我们也只考虑这两种形状的 ROI,并说明 ROI 为这两种形状情况下扩展因子的定义。首先,分别用 (x_L, y_L) 和 (x_R, y_R) 代表矩形 ROI 左上角和右下角的坐标,对于椭圆 ROI,这两个坐标分别为包含椭圆最小矩形的左上角和右下角坐标,如图 3 所示。并且我们定义 $a = \min(|x_L - x_R|, |y_L - y_R|)$, $b = (|x_L - x_R|, |y_L - y_R|)$ 为矩形长与宽或椭圆的短轴和长轴。这样扩展因子 S 的计算公式如下:

$$S = 2\Delta r/a \quad (5)$$

上式中的 Δr 为一级背景扩展的半径。

由上式可知,当 S 取得最大值 S_{\max} 时,也就是说以整个背景作为整体进行图像传输。 S_{\max} 的取值不仅与 ROI 的大小有关,还与 ROI 在图像中所处的位置有关,因为当背景图像扩展时,如果背景的任何一边达到原始图像的边缘,即表明达到最后一级的扩展,就以上一级扩展后剩下的背景图像作为最后一级扩展进行图像传输。 S_{\max} 的具体计算公式如下:

$$S_{\max} = \frac{2\min(x_L, y_L, M - x_R - 1, N - y_R - 1)}{a} \quad (6)$$

其中, $M \times N$ 为原始图像的大小。

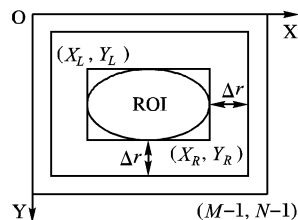


图3 矩形和椭圆形状的 ROI 表示

不同的扩展因子将会产生不同的图像传输方式,图 4 展示了几种不同的扩展因子所导致的背景分级传输效果。从图中可以看出, S 越小($S > 0$),背景图像分级越多。 $S = 0$ 和 $S = S_{\max}$ 时,都以整个背景图像为整体进行图像传输,可是两者却有不同之处。 $S = 0$ 表示用户对 ROI 图像极为重视,需要无损地传输,此时的图像传输方式和一般的基于感兴趣区渐进图像传输方式一样:首先将全部的 ROI 系数传输完毕,才开始传输背景图像的数据。而 $S = S_{\max}$ 时,首先只需传输对视觉质量比较重要的 ROI 系数,当 ROI 达到一定的主观质量后就开始传输对视觉质量影响比较大的 BG 系数。因此,一般基于 ROI 的渐进图像传输算法只是本文算法的一个特例,而且本文算法还提供了更加灵活的图像传输方式。

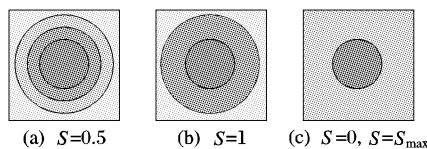


图4 不同扩展因子产生的分级效果

综合以上所述,具体的渐进图像传输算法步骤如下:

- 1) 对图像进行小波变换;
- 2) 首先传输低分辨率图像,等待用户决定是否要继续传输图像,如继续则进行下一步,否则停止图像传输;
- 3) 计算 ROI 掩模,并传输视觉上重要的 ROI 系数;
- 4) 依据扩展因子,决定下一步传输的图像数据,若 $S = 0$,继续传输 ROI 剩余系数;否则,依据 S 计算分级 BG 的掩模,并按顺序传输 BG 系数;
- 5) 将剩余的 ROI 和 BG 系数传输完毕。

4 仿真实验分析

为了验证本文算法的有效性,下面就以 Lena ($256 \times 256 \times 8$) 标准图像为例,在 Matlab 环境下对其进行三层小波分解,所用的滤波器为常用的双正交 9/7 小波滤波器,并指定 Lena 的脸部区域为 ROI,正好是图像正中的四分之一正方形区域。实验对小波变换后的 Lena 图用本文算法和一般的 ROI 传输算法进行渐进图像传输,并对传输过程中的四个阶段进行了比较。由上述可知, S 的取值与具体的 ROI 图像有关,不同的 S 将产生不同的传输效果, S 较小,算法的复杂度可能较大; S 较大,分级传输的效果可能不明显。根据这些特性,本文实验采用的扩展因子为 $S = 0.5$,实验结果如图 5 所示。

从图中可以看出,改进算法首先传输图像的低频子带,给接收者提供一个比较清晰的原始图像低分辨率形式,由其决定此图像是否为自己所需,若不是,则立刻停止图像的传输,既有效地节省了用户的时间又节约了网络带宽。如果需要继续传输图像,则根据以上所述的算法进行渐进传输,此实验是将 ROI 图像传输到还剩最后一个高频子带 HH_1 时,停止对

(下转第 1673 页)

的异步性稍短。

表 3 句子“Sil-two-nine -Sil”的部分音素视素时间切分结果

音素	音频流/s	视频流(帧数, 25 fps)	视素流/s	视素
T	0.43 ~ 0.54	9 ~ 9	0.36 ~ 0.36	Vm
uw1	0.55 ~ 0.84	10 ~ 16	0.4 ~ 0.64	vg
n	0.83 ~ 0.97	17 ~ 17	0.68 ~ 0.68	vp
ay1	0.98 ~ 1.20	18 ~ 29	0.72 ~ 1.16	Vc
n	1.21 ~ 1.34	30 ~ 30	1.20 ~ 1.20	vp

3 结语

通过构建新的单流单音素 DBN 模型,我们对语音的音视频数据流进行了语音识别和音素时间切分研究,得到了基于音频特征和视频特征的词识别结果、音素识别结果和音素时间切分序列。基于音频特征的结果表明,DBN 模型比 HMM 模型具有更好的噪声鲁棒性;而且基于单音素 DBN 模型的音素切分时间和基于三音素 HMM 的音素切分时间很接近。而基于视频特征的结果表明,DBN 模型有更好的识别性能和视素切分结果。最后从音频和视频特征得到的音素时间切分的结果比较看到,语音信号和视频之间的异步关系已经远远超出音素的边界,而且异步关系随着时间和音素类型而变,因此体现和得到音视频流的准确异步关系对于多流音视频连续语音识别和逼真的视觉语音合成具有很重要的意义。

参考文献:

[1] BREGLER C, KONIG Y. Eigenlips for robust speech recognition

- [C]// Proceedings of International Conference on Acoustics, Speech and Signal Processing. Adelaide, Australia: IEEE, 1994: 669 - 672.
- [2] POTARMIANOS G, NETI C, GRAVIER G, *et al.* Recent advances in the automatic recognition of audiovisual speech [J]. Proceedings of IEEE, 2003, 91(9): 1306 - 1326.
- [3] DUPONT S, LUETTIN J. Audio-visual speech modeling for continuous speech recognition [J]. IEEE Transactions on Multimedia, 2000, 2(3): 141 - 151.
- [4] BILMES J, ZWEIG G. The graphical models toolkit: An open source software system for speech and time series processing[C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Orlando, Florida: IEEE, 2002.
- [5] BILMES J, BARTELS C. Graphical Model Architectures for Speech Recognition [J]. IEEE Signal Processing Magazine, 2004, 22(5): 89 - 100.
- [6] BILMES J, ZWEIG G, RICHARDSON T, *et al.* Discriminatively structured graphical models for speech recognition[EB/OL]. [2006 - 12 - 31]. <http://www.clsp.jhu.edu/ws2001/groups/gmstr/GM-RO-final-rpt.pdf>.
- [7] ZHOU Y, GU L, ZHANG H-J. Tangent Shape Model: Estimating Shape and Pose Parameters via Bayesian Inference[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2003). Wisconsin, USA: IEEE, 2003.
- [8] 蒋冬梅, 谢磊, RAVYSE I, 等. 基于 Viseme 的连续语音识别系统及 Talking Head[J]. 电子与信息学报, 2004, 26(3): 375 - 381.

(上接第 1656 页)

ROI 数据的传输,而开始根据扩展因子分级传输背景图像系数。图中的每个阶段,两种不同的算法所传输的系数数量相同,第一阶段只传输了低频子带的系数数量,而此时一般方法只传输了一些 ROI 系数,且图像相当模糊;第二阶段,一般方法的 ROI 图像达到原始图像质量,而改进算法在保证 ROI 质量与一般方法无明显差别的同时,已经开始改善周围的背景图像;第三阶段,改进算法显示的图像已经具有较好的视觉效果,而一般方法的图像边缘模糊且 ROI 边界有明显质量跃迁;最后一个阶段,改进算法的整幅图像已经达到良好的视觉效果,而一般方法的背景图像在边缘细节方面还有些许失真,这影响了图像的整体视觉效果。

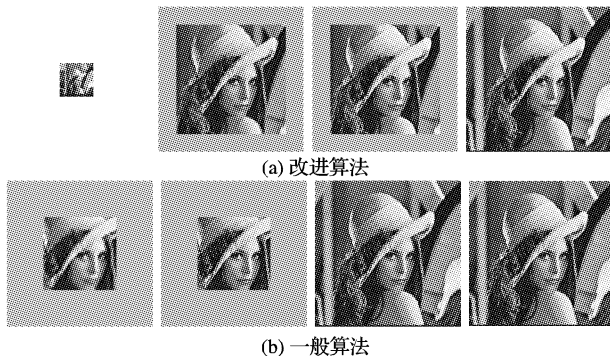


图 5 两种渐进图像传输算法效果对照

5 结语

本文首先分析了小波变换的特点,并结合人眼视觉特性,提出了一种基于感兴趣区渐进图像传输改进算法。该算法不

仅保证 ROI 图像具有较好视觉效果,还充分利用有限的带宽去改善图像的背景区域质量。然而,本文算法也有一定的局限性,首先需要指出的是改进算法要求 ROI 一般位于图像中间部分,否则扩展因子可能失去应有的作用;其次,虽然本文算法比较灵活,但它是以更高复杂度换取的,因为与传统算法相比,它还需要计算背景的掩模。可是如果 ROI 为矩形,则这部分的复杂度可以忽略。从总体上来说,该算法有效地利用了网络带宽,改善了图像的整体主观视觉效果,极大地提高了渐进图像传输效率。

参考文献:

- [1] RAUSCHENBACH U. Progressive image transmission using levels of detail and regions of interest[C]// Proc'98 IASTED Conference on CGIM. Halifax, Nova Scotia, Canada: [s. n.], 1998.
- [2] CHRISTOPOULOS C, ASKELOF J, LARSSON M. Efficient methods for encoding regions of interest in the upcoming JPEG2000 still image coding standard[J]. IEEE Transactions on Signal Processing Letters, 2000, 7(9): 247 - 249.
- [3] SIGNORONI A, LAZZARONI F, LEONARDI R. Exploitation and extension of the region-of-interest coding functionalities in JPEG2000 [J]. IEEE Transactions on Consumer Electronics, 2003, 49(4): 818 - 823.
- [4] 梁亚玲, 杨春玲, 余林英. 基于人眼视觉特性的 ROI 编码[J]. 计算机应用, 2005, 25(7): 1598 - 1601.
- [5] 黄贤武, 李家骅. 一种改进的基于人眼视觉特性的零树图像编码算法[J]. 计算机研究与发展, 2001, 38(8): 977 - 981.
- [6] 张旭东, 卢国栋, 冯健. 图像编码基础和小波压缩技术——原理、算法和标准[M]. 北京: 清华大学出版社, 2004.