

doi:10.3969/j.issn.1001-893x.2014.05.008

引用格式:王晓龙.基于感兴趣区域的自适应帧内更新编码算法[J].电讯技术,2014,54(5):569-573.[WANG Xiao-long. Adaptive Intra Update Coding Algorithm Based on Region of Interest Video Coding[J]. Telecommunication Engineering, 2014, 54(5):569-573.]

## 基于感兴趣区域的自适应帧内更新编码算法\*

王晓龙\*\*

(中航工业洛阳电光设备研究所,河南 洛阳 471009)

**摘要:**针对视频压缩码流对信道差错异常敏感的问题,提出了一种基于感兴趣区域的自适应帧内更新编码算法。该算法利用人眼的视觉感知特性设计感兴趣区域提取模型,根据信道差错累积情况自适应调整帧内更新编码策略,将更少的失真分配给人眼感兴趣区域,提高差错信道下传输视频图像的主客观质量。实验结果表明,与基于端到端失真的帧内更新算法相比,在不同的信道丢包条件下,所提算法可以获得更好的主客观质量,感兴趣区域的峰值信噪比(PSNR)平均提高0.87 dB左右,提高了差错信道下视频通信的鲁棒性。

**关键词:**视频通信;视频压缩;帧内更新编码;感兴趣区域

**中图分类号:**TN919.81 **文献标志码:**A **文章编号:**1001-893X(2014)05-0569-05

## Adaptive Intra Update Coding Algorithm Based on Region of Interest Video Coding

WANG Xiao-long

(Luoyang Institute of Electro-Optical Equipment, Aviation Industry Corporation of China, Luoyang 471009, China)

**Abstract:** In view of the sensitivity of video coding stream to channel error, an intra refresh algorithm based on region of interest video coding is proposed. The algorithm first estimates the model for extracting region of interest using human visual perception, and then it chooses intra coding strategy adaptively according to the accumulation of errors for every macroblock, allocating less distortion to region of interest to improve the subjective and objective visual quality under lossy channel. Simulation shows that under various packet loss rates, the proposed algorithm can obtain a better subjective and objective quality of the reconstructed video and outperforms 0.87 dB in region of interest than the end-to-end intra update algorithm, thus improving the robustness of video communications under lossy channel.

**Key words:** video communication; video coding; intra update coding; region of interest

### 1 引言

视频信源压缩编码和视频信道编码历来是差错信道下视频通信研究热点,其中,视频压缩编码的目的是使压缩码流满足信道传输带宽的要求。为此,现有的视频编码算法广泛采用了运动预测/补偿和熵编码等技术,而这会导致视频码流对信道误码或丢包异常敏感,造成严重的误码扩散现象,大大降低

重建视频的质量<sup>[1]</sup>。帧内更新是一种简单有效的抗差错手段,它通过在模式选择阶段增加帧内编码宏块的数量,从而阻止由于信道差错造成的帧间误码扩散,提高视频码流的抗差错鲁棒性。

人们在观看视频图像的时候,并不是对视频图像的所有区域都具有相同的注意力,而是对不同的区域有不同的注意力,该区域被称为感兴趣区域

\* 收稿日期:2013-12-23;修回日期:2014-03-07 Received date:2013-12-23;Revised date:2014-03-07

\*\* 通讯作者:heluoxiaolong@126.com Corresponding author:heluoxiaolong@126.com

(Region of Interest, ROI)<sup>[2]</sup>。其中感兴趣区域的信源编码失真和信道传输失真对观察者的主观感受影响最大,而基于感兴趣区域的帧内更新算法的主要是将信道信源失真更多的转移至非感兴趣区域,从而提高端到端视频的质量。文献[3]提出基于感兴趣区域中的随机帧内更新算法,该算法根据信道丢包情况来更新感兴趣区域的采用帧内编码宏块的数量,而宏块位置的选择是随机的。该算法简单,但是抗差错效果有限。此外,文献[4]提出了通过引入信源信道联合失真提出基于端到端率失真模型的感兴趣区域帧内更新编码算法,该算法性能较文献[3]性能更优,但是由于每帧编码要进行像素级别的端到端失真计算,故运算量太大。为此,本文提出了一种低复杂度的基于感兴趣区域的帧内更新编码算法,该算法通过统计宏块的信道差错累计情况来自适应调整编码策略,提高差错信道下传输视频图像的主客观质量。

## 2 感兴趣区域提取

本节提出的感兴趣区域提取模型是以人眼的视觉感知为基础,将肤色感知特性、运动感知特性、空间位置特性 3 个要素结合在一起,并进行不同的加权组合,得到视觉感知权重图,最后即可得到符合人眼感知的 ROI 区域。在本节中以宏块为基本单元,为每个宏块计算感知权重以确定宏块是否属于感兴趣区域。

在视频通信应用中,人往往是通信的主体,头肩部也是最多的场景模式,因此本节采用文献[5]提出的人脸肤色检测模型,该模型具有精确性高和计算复杂度低的优点,适用于实时性要求较高的应用。该算法基于 YCbCr 的颜色空间,模型的具体数学表达式如式(1)和式(2)所示:

$$\frac{(x-e_{c_x})^2}{a^2} + \frac{(y-e_{c_y})^2}{b^2} = 1 \quad (1)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} Cb'-c_x \\ Cr'-c_y \end{bmatrix} \quad (2)$$

其中, $c_x = 109.38$ ,  $c_y = 152.02$ ,  $\theta = 2.53$  (单位:rad),  $e_{c_x} = 1.60$ ,  $e_{c_y} = 2.41$ ,  $a = 25.39$ ,  $b = 14.03$ 。由于该算法得到的人脸肤色是以像素为单位的,而 H.264/AVC 是以宏块(16 pixel×16 pixel)为基本编码单位的,所以需要对其结果进行后处理,变为以宏块为单位的人脸肤色视觉感知权重值。定义宏块的肤

色感知权重因子  $SA_n(i, j)$  来描述第  $n$  帧第  $(i, j)$  个宏块的肤色感知权重,如式(3)所示:

$$SA_n(i, j) = \frac{\sum_{x=1}^{16} \sum_{y=1}^{16} Sk_{i,j}(x, y)}{A} \quad (3)$$

其中,  $A$  是一个宏块中的像素值的个数,  $Sk_{i,j}(x, y)$  是检测出来的肤色-非肤色的二值图表达式,当其值为 1 表示该像素属于肤色区域,值为 0 表示为非肤色区域。 $SA_n(i, j)$  可见的数值位于闭区间  $[0, 1]$  当中。

另外,视频图像中的运动区域也容易引起人眼的关注,在有全局运动的场景下,运动剧烈的区域比运动缓慢的区域更容易受到人眼的关注,因此应该赋予运动剧烈的区域更高的感知权重。定义宏块的运动权重因子  $MA_n(i, j)$  来描述第  $n$  帧第  $(i, j)$  个宏块的运动剧烈程度,  $MA_n(i, j)$  的表达式如式(4)所示:

$$MA_n(i, j) = \frac{\sqrt{[MV_x^n(i, j)]^2 + [MV_y^n(i, j)]^2}}{\text{Max}(\sqrt{MV_{x, \max}^2 + MV_{y, \max}^2})} \quad (4)$$

其中,  $MV_x^n(i, j)$  和  $MV_y^n(i, j)$  分别表示第  $(i, j)$  个宏块水平方向和垂直方向的运动矢量,  $MV_{x, \max}$  和  $MV_{y, \max}$  分别是当前帧中  $MV_x^n(i, j)$  和  $MV_y^n(i, j)$  的最大值,假设当前帧由  $M \times N$  个宏块组成,那么则有  $i = 0, 1, 2, \dots, N-1$  和  $j = 0, 1, 2, \dots, M-1$ 。如果  $MA_n(i, j)$  的数值越大,则表示当前宏块运动越剧烈,那么该宏块也就越容易引起人眼关注。

研究表明位于视频图像中央区域的部分也最容易引起人眼的注意,另外在常见的视频场景中,摄像人员总是把观众最关注的主要目标放在镜头的中心区域,所以位于中央区域的图像要比边缘区域的图像具有更高的重要性,定义空间位置权重因子  $PA_n(i, j)$  来描述第  $n$  帧第  $(i, j)$  个宏块的位置重要程度,如式(7)所示:

$$\omega_{px}^{(i, j)} = \sqrt{2} + (1 - \sqrt{2}) \frac{|i - CX|}{CX} \quad (5)$$

$$\omega_{py}^{(i, j)} = \sqrt{2} + (1 - \sqrt{2}) \frac{|j - CY|}{CY} \quad (6)$$

$$PA_n(i, j) = \omega_{px}^{(i, j)} \times \omega_{py}^{(i, j)} - \delta \quad (7)$$

其中,  $i = 0, 1, 2, \dots, N-1$  和  $j = 0, 1, 2, \dots, M-1$ , ( $CX$ ,  $CY$ ) 为位于图像中心宏块的位置,  $M$  和  $N$  分别是水平和垂直方向宏块的数量,  $\delta$  是修正因子。可见,宏块的位置感知权重位于闭区间  $[0, 1]$ 。其中,位于图像边缘的宏块的权重为 0,而位于中心位置宏块的权重为 1。

综合式(3)、(4)和(7),对不同的感知特征进行不同的加权组合,得到最终的视觉感知权重图,如式(8)所示:

$$VS_n(i,j) = k_s \times SA_n(i,j) + k_m \times MA_n(i,j) + k_p \times PA_n(i,j) \quad (8)$$

其中, $k_s$ 、 $k_m$ 和 $k_p$ 分别表示以上3个感知特性所对应的权重因子,在对视频序列进行提取检测的时候,预先设定阈值 $T$ ,当宏块的 $VS_n(i,j)$ 大于该值就认为该宏块为感兴趣区域,否则是非感兴趣区域。图1给出了测试序列Foreman的感兴趣区域提取结果。

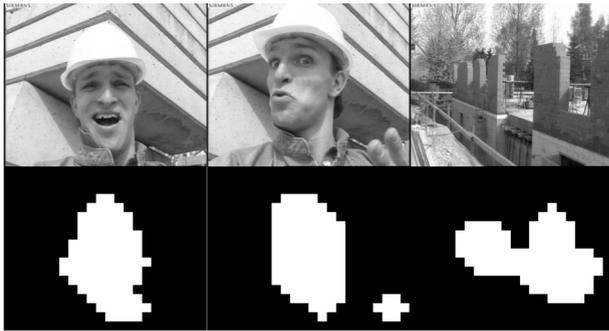


图1 Foreman序列第26、89和244帧的原始图像对应的感兴趣区域二值图对比( $k_s=1, k_m=1, k_p=1, k_t=1, T=1.96$ )

Fig. 1 The comparison between original image frames and their ROI binary image for Foreman sequence ( $k_s=1, k_m=1, k_p=1, k_t=1, T=1.96$ )

### 3 基于差错估计的帧内更新编码算法

现有的视频编码框架是基于块的编码结构,H.264标准中支持多种块大小的编码模式,对宏块最优编码模式的选择是通过Lagrange乘子法的率失真优化算法实现的<sup>[1]</sup>,即在给定的码率限制 $R_c$ 下选择最优编码模式,使得总的编码失真 $D$ 最小:

$$\min \{ D(o) \}, \text{ s. t. } R(x) < R_c \quad (9)$$

其中, $R_c$ 是信道速率, $D(o)$ 宏块为编码模式为 $m$ 时的编码失真,对应码流为 $R(x)$ , $\lambda$ 为Lagrange乘子。

与帧间编码模式相比,帧内更新编码会增加额外编码比特,使得压缩性能下降。因此,帧内更新编码技术的关键在于如何确定帧内更新编码宏块的数量和位置。其中随机帧内更新算法根据信道丢包情况决定每一帧中进行帧内更新宏块的个数,根据随机或固定的方式决定宏块更新的位置<sup>[6]</sup>。该种算法简单但是不能够根据视频内容特性采用帧内编码,因此抗差错性能较差;而基于端到端失真的率失真优化帧内更新编码算法,将信道失真引入率失真优化的模式选择过程中,对失真严重的宏块进行帧

内编码,因此可得到最优的端到端的差错控制效果,但是复杂度很高<sup>[7]</sup>。为此,本文提出基于感兴趣区域的帧内更新编码算法,根据信道丢包情况估计每个宏块的差错扩散失真,并结合其对应的视觉感知权重,若当前宏块的失真(包括差错扩散失真和感知权重失真)超过一定的更新阈值,则进行帧内更新编码,而该阈值是根据当前编码比特消耗情况自适应选取。

对宏块的差错传播失真的推导是基于 $4 \times 4$ 子块进行的。设第 $n$ 帧的参考帧为第 $n-t$ 帧,第 $n-t$ 帧中第 $(x,y)$ 个子块的差错估计为 $d_{ep}((x,y), n-t)$ ,那么由于运动预测所造成的差错扩散可以表示为

$$d_{ep}((x,y), n) = d_{ep}((x+v_x, y+v_y), n-t) \quad (10)$$

其中, $(v_x, v_y)$ 为相应的运动矢量。

图2所示为差错传播失真估计。

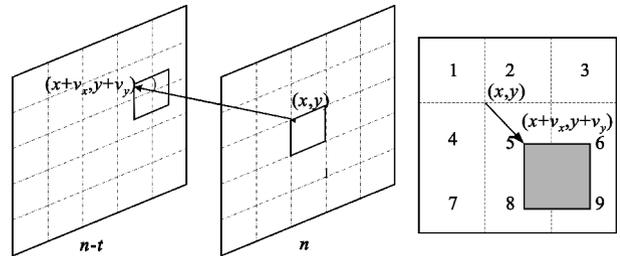


图2 差错传播失真估计

Fig. 2 The estimation of error propagation distortion

由图2可知,第 $n-t$ 帧的子块 $(x+v_x, y+v_y)$ (图中的灰色块)可以和划分的若干个子块重叠,因此, $d_{ep}((x,y), n)$ 可以用与之重叠的子块的加权和来表示:

$$d_{ep}((x,y), n) = \sum_{k=1}^K w_k d_{ep}((x,y)_k, n-t) \quad (11)$$

则第 $n$ 帧的第 $m$ 个宏块的差错失真为其所有 $4 \times 4$ 块的 $d_{ep}$ 值之和,即为 $D_{ep}(m, n)$ ,其中 $w_i$ 与重叠面积大小成正比。对于给定的阈值 $T$ ,结合各个宏块的视觉感知权重,当累积差错超过阈值时,即

$$\sum_{n=n_0}^N \alpha(m, n) \cdot D_{ep}(m, n) > T \quad (12)$$

则对该帧进行强制帧内编码,并对其差错估计累计值清零,其中:

$$\alpha(m, n) = \alpha(i \cdot N + j, n) = f(VS_n(i, j)) \quad (13)$$

$$f(VS_n(i, j)) = \begin{cases} \delta \times VS_n(i, j), & VS_n(i, j) \geq Th \\ 1, & VS_n(i, j) < Th \end{cases} \quad (14)$$

式中, $f(*)$ 是权重修正函数, $Th$ 是提取感兴趣区域

的门限,  $\delta$  为修正因子。设丢包率为  $p$ , 那么正确接收数据包的概率是  $1-p$ , 此时可得差错估计为

$$D_{ep,1}(m,n) = \begin{cases} 0, & \text{INTRA} \\ D_{ep}(m,n), & \text{INTER} \end{cases} \quad (15)$$

如果宏块丢失, 则对该宏块进行误码掩盖算法, 假设采用最简单的误码掩盖算法, 即当前帧中的丢失宏块由前一帧相应位置的宏块代替, 此时差错传播由运动估计所造成的差错传播失真和误码掩盖失真, 即

$$D_{ep,2}(m,n) = D_{ep}(m,n-1) + D_{ec}(m,n) \quad (16)$$

由式(15)和(16)可得平均意义下的差错传播失真

$$D_{ep}(m,n) = (1-p)D_{ep,1}(m,n) + pD_{ep,2}(m,n) \quad (17)$$

每帧编码时重复进行式(10)~(17), 从而确定每帧的帧内编码宏块。从式(12)可以看出, 当阈值  $T$  越大, 一个视频帧中容许帧内更新编码的宏块就越少, 因此最终生成的编码速率就会下降; 当阈值  $T$  减小, 一帧中容许帧内编码的宏块就增多, 视频编码速率也会随之提高, 因此阈值  $T$  的大小和编码速率有关, 其数值可以通过实验确定, 具体在编码过程中可以按照如下流程进行选取阈值  $T$ 。

**第 1 步:** 根据设定的目标码率计算预分配的编码比特  $B_f$ ;

**第 2 步:** 由预分配编码比特  $B_f$  计算量化参数  $Q$ ;

**第 3 步:** 对当前第  $l$  个宏块进行编码, 此时当前编码的总比特为  $R_l$ ;

**第 4 步:** 按照式(18)和(19)更新阈值  $T$ :

$$F = \beta \times (R_l - \frac{lB_f}{L}) / l \quad (18)$$

$$T = T + F \times T_s \quad (19)$$

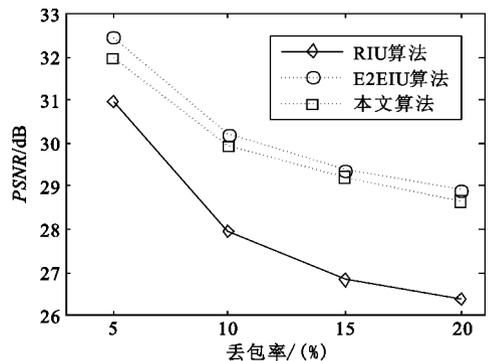
其中,  $T_s$  为调整步长,  $\beta$  为给定系数,  $L$  为视频帧中总的宏块数量。

## 4 实验仿真与分析

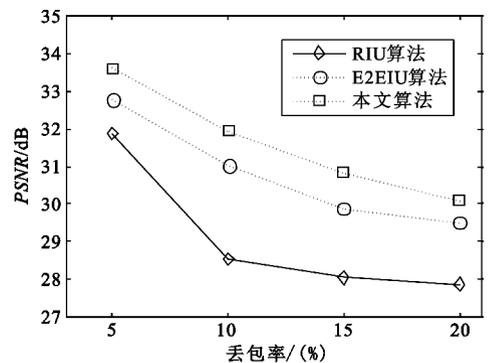
为了测试算法的性能, 采用 JM10.2 作为仿真测试平台, 将提出的基于感兴趣区域的帧内更新编码算法在该参考软件版本上进行实现。编码器采用如下配置: 采用所有帧间编码模式, 1 个参考帧, 采用 CABAC 熵编码, 帧率为 15 frame/s, GOP 结构为 IPPP。采用 Slice 编码打包模式, 1 帧分为 6 个 slice 包, P 帧分为 4 个 slice 包, 采用随机丢包, 丢包率分

别为 5%、10%、15% 和 20%。误码掩盖算法采用宏块替代的时域误码掩盖算法<sup>[8]</sup>。由于本算法主要针对中等运动及复杂度的视频场景, 故测试序列采用 Foreman 和 Coastguard 序列, 图像大小为 CIF 格式(352 pixel×288 pixel), 目标码率设为 350 kb/s。通过多次实验统计(本实验中统计次数为 50 次)重建视频的平均 PSNR 来评价算法的性能, 并与随机帧内更新编码算法(Random Intra Update, RIU)<sup>[9]</sup>和基于端到端率失真最优的帧内更新编码算法(End to End Intra Update, E2EIU)<sup>[10]</sup>进行对比。

图 3(a)和图 4(a)分别给出了采用 Foreman 和 Coastguard 序列时不同算法在不同信道丢包率下的重建图像的整体 PSNR 比较。相比于 RIU 算法, 本算法在同样的码率约束下, 对于 Foreman 序列重建视频的平均 PSNR 可以提高 1.01~2.28 dB, 而较 E2EIU 算法的平均 PSNR 性能损失控制在 0.3 dB 左右, 而 RIU 算法与其他算法的性能随着信道条件的恶劣差距越来越明显, 这主要是因为 RIU 算法没有考虑到信道差错造成的传播失真, 随着丢包率的增加, 此时由信道差错引入的传播失真占据了主导地位。

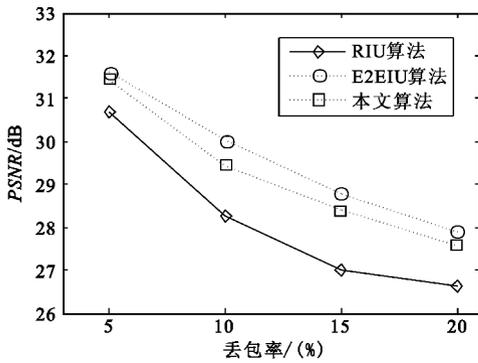


(a) Foreman 序列重建视频整体 PSNR

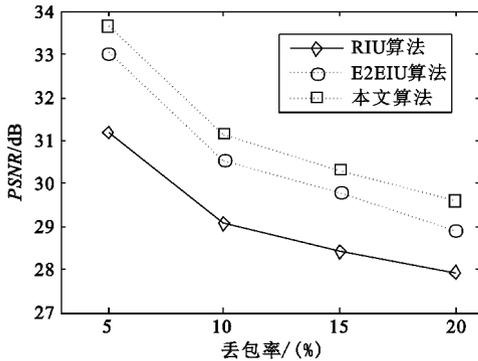


(b) Foreman 序列重建视频 ROI 区域 PSNR

图 3 Foreman 序列不同算法性能比较  
Fig. 3 The performance comparison between different algorithms for Foreman sequence



(a) Coastguard 序列重建视频整体 PSNR



(b) Coastguard 序列重建视频 ROI 区域 PSNR

图 4 Coastguard 序列不同算法性能比较

Fig. 4 The performance comparison between different algorithms for Coastguard sequence

图 3 (b) 和图 4 (b) 分别给出了 Foreman 和 Coastguard 序列在不同的丢包率情况下重建视频感兴趣区域的平均 PSNR,可以看出,本算法在图像整体 PSNR 性能上比 E2EIR 算法略微下降,但是感兴趣区域的平均 PSNR 要高于后者 0.87 dB 左右。图 5 给出了在 10% 丢包的情况下对于 Foreman 序列各种算法的主观效果比较情况,可以看出,本文算法的主观感觉要明显好于其他算法。



(a)RIU算法重建视频 (b)E2EIR算法重建视频 (c)本文算法重建视频

图 5 Foreman 序列(第 40 帧)不同算法主观效果比较图

Fig. 5 The subjective effect comparison between different algorithms for Foreman sequence

## 5 结论

本文首先利用人眼对视频场景中的不同区域具有不同的视觉敏感性,从肤色感知特性、运动感知特性和位置感知特性入手建立了感兴趣区域提取模

型,随后提出了基于感兴趣区域的帧内更新编码算法,在差错信道下通过对差错的累积估计选择帧内编码宏块,从而显著改善了端到端视频通信的抗差错性能。但是感兴趣区域提取模型的应用场景还具有一定局限性,对于复杂场景中感兴趣区域提取算法的效果还不甚理想,而这将会影响算法最终的抗差错性能以及算法的应用场景,如何建立高效的感兴趣区域提取模型将是下一步的研究方向。

## 参考文献:

- [1] Richardson I E. The H. 264 advanced video compression standard[M]. New York:John Wiley & Sons,2011.
- [2] 路羊. 率失真最优的高效稳健视频编码与传输技术研究[D]. 北京:清华大学,2008.  
LU Yang. Research on Rate-distortion Optimized Video Coding and Transmits Technology[D]. Beijing:Tsinghua University,2008. (in Chinese)
- [3] Krishnamurthy R, Sethuraman S. Region-based refresh strategy for video compression; U. S. Patent 6304295 [P]. 2001-10-16.
- [4] Chen Q, Chen Z, Gu X, et al. Attention-based adaptive intra refresh for error-prone video transmission[J]. IEEE Communications Magazine,2007,45(1):52-60.
- [5] Hsu R, Mohamed A, Jain A. Face detection in color images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2002,24(5):696-706.
- [6] Vanam R, Reznik Y. Error-resilient video coding for wireless video telephony applications[C]//Proceedings of 2012 SPIE. San Diego, California, USA:SPIE,2012:1-9.
- [7] Stuhlmuller K, Farber N, Link M, et al. Analysis of video transmission over lossy channels[J]. IEEE Journal on Selected Areas in Communications,2000,18(6):1012-1032.
- [8] Wenger S. H. 264/AVC over IP[J]. IEEE Transactions on Circuits and Systems for Video Technology,2003,13(7):645-656.
- [9] Cote G, Kossentini F. Optimal intra coding of blocks for robust video communication over the Internet[J]. Signal Processing: Image Communication,1999,15(1):25-34.
- [10] Xiao J, Tillo T, Lin C, et al. Error-resilient video coding with end-to-end rate-distortion optimized at macroblock level[J]. EURASIP Journal on Advances in Signal Processing,2011(1):1-10.

## 作者简介:



王晓龙(1985—),男,河南洛阳人,2012年获博士学位,现为工程师。

WANG Xiao-long was born in Luoyang, Henan Province, in 1985. He received the Ph. D. degree in 2012. He is now an engineer.

Email:heluoxiaolong@126.com