LUMA-AWARE MULTI-MODEL RATE-CONTROL FOR HDR CONTENT IN HEVC

Karina R. Perez-Daniel and Victor Sanchez

Dept. of Computer Science, University of Warwick, UK (vsanchez@dcs.warwick.ac.uk)

ABSTRACT

Rate control (RC) is an important tool that helps to control the bit rate of the compressed media. Current HEVC implementations include an RC algorithm, based on a $R-\lambda$ model, that does not consider the wide range of luma values depicted in HDR content. This paper proposes a multi-$R-\lambda$-model approach for RC in HEVC that accurately attains a target bit rate while improving the reconstruction quality of HDR content, especially for very bright regions. Results, in terms of the mPSNR and HDR-VDP-2 metrics, confirm the advantages of our approach compared to the current RC algorithm in HEVC.

***Index Terms*—** HDR, HEVC, Rate Control, $R-\lambda$ model.

1. INTRODUCTION

The Consumer Electronics Association (CEA) [1] has adopted the High Efficient Video Coding (HEVC) standard to deliver High Dynamic Range (HDR) [2] content due to its 1) ability to support high bit depth data, and 2) outstanding performance on Standard Dynamic Range (SDR) content [3] compared to its predecessor, H.264/AVC [4].

Rate Control (RC) stands out as one of the most useful tools in HEVC for applications that are constrained by bandwidth and storage resources, such as real-time transmission. RC establishes a trade-off between Rate (R) and Distortion (D) according to the nature of the data in order to produce a compressed bit stream at a specific bit rate with minimum distortion. To this end, several models have been proposed to characterize the RD relationship of videos [5, 6]. These models include the *quadratic* model, the ρ model, and the $R-\lambda$ model. The *quadratic* model [5] relies on a quadratic relation between R and D by considering the bits resulting from encoding the transform coefficients; i.e., Discrete Cosine Transformation of residual data. Due to its relatively low accuracy in characterizing the actual rate, this model is rarely adopted. Hence, the ρ model [6] is proposed to provide a more accurate RD relationship [6, 7, 8]. Unfortunately, the ρ model assumes a fixed-size transform block, which is not compatible with the flexible quad-tree structure of HEVC.

Both, *quadratic* and ρ models assume a direct relation between R and the Quantization Parameter (QP). However, in HEVC the bit rate depends not only on the QP, but also on the Lagrangian multiplier, λ. Based on this fact, an $R-\lambda$ model is proposed to closely approximate the relationship among R, λ and QP, while keeping the computational complexity relatively low [9]. Hence, current implementations of HEVC employ this $R-\lambda$ model as part of the RC algorithm [10].

An important aspect of the $R-\lambda$ model is the set of parameters used to approximate the slope of the RD curve of the sequence to be compressed. These parameters are usually adapted during the encoding process in order to compute the QP values that most accurately attain a target bit rate with the minimum possible distortion. However, du-ring bit allocation, the current $R-\lambda$-based RC algorithm in HEVC does not fully exploit the fact that certain regions may need more bits than others due to their distinctive RD relationships. This is the case of HDR content, where not only luminance levels greatly differ from one region to another within the same picture, but also high contrast regions can be found in a single Coding Unit (CU); e.g., a CU depicting a lamp bulb filament in a darkroom. For this reason, it is important to consider the brightness of different regions when allocating a target bit budget during RC, in order to better preserve HDR contents even after compression at low bit rates. Although some solutions that employ RC to compress sequences on a per-region basis have been proposed recently in [11, 12], RC approaches for HDR content are scarce.

This paper improves RC in HEVC for HDR content by using a multi-$R-\lambda$ model approach to encode regions according to their luminance values. The specific $R-\lambda$ model used for a CU is selected according to their luminance values. This results in an efficient bit budget allocation that assigns more bits to regions that are costly to encode, while accurately attaining a target bit rate. Performance evaluations over several HDR sequences show that the proposed approach attains important improvements on the reconstruction quality of very bright regions, thus allowing to preserve details depicted by the high range of pixels values of HDR content.

The rest of the paper is organized as follows. Section 2 explains the $R-\lambda$ model used in HEVC. Section 3 describes the proposed multi-$R-\lambda$ model approach. Section 4 presents evaluation results and Section 5 concludes the paper.

2. RC IN HEVC BASED ON AN $R-\lambda$ MODEL

RC algorithms are designed to attain, as accurately as possible, a target bit rate, R_{target}, for a bit stream with a minimum

978-1-5090-2175-8/17/$31.00

distortion, D [3]:

$$\min D \quad \text{subject to } R \leq R_{target}. \tag{1}$$

This constraint can also be handled as the unconstrained Rate Distortion Optimization (RDO) formulation by minimizing the total RD cost function, J [3]:

$$J(T_0, T_1, \ldots, T_{N-1}) = \sum_{i=0}^{N-1} D_i + \lambda * \sum_{i=0}^{N-1} R_i, \tag{2}$$

where T_i is the optimal number of bits for entity i (e.g., a CU) within a coding level (e.g., a picture), which minimizes distortion D_i between the original and the reconstructed entity subject to a corresponding bit budget, R_i. The Lagrangian multiplier, λ, controls the trade-off between rate, R, and distortion, D, by adjusting the amount of bits spent in each entity. For camera-caputed video sequences, it is shown in [9] that the relationship between λ and the quantization parameter, QP, which is in charge of reducing the bit rate of the compressed bit-stream, may be given as:

$$QP = a \ln \lambda + b, \tag{3}$$

where a and b are parameters related to the RD characteristics of the video sequence. The linear relationship in (3) reduces the computational complexity by avoiding the exhaustive search needed to determine the most suitable QP which minimizes J in (2). HEVC implementations profit from the linear relationship in (3) and adopt an $R - \lambda$ model to compute λ in terms of R and D.

The $R - \lambda$ model in HEVC considers that λ not only represents the slope of the RD curve (see (2)), but is also a critical parameter for bit allocation. It has been shown in [13] that the RD relationship for camera-captured video sequences may be approximated as a hyperbolic function:

$$D(R) = C * R^{-K}, \tag{4}$$

where C and K are determined by parameter fitting on the experimentally obtained RD curve of the sequence. Slope λ in (2) can then be obtained by differentiating (4) w.r.t. R,

$$\lambda = -\frac{\partial D}{\partial R} = C * K * R^{(-K-1)} = \alpha R_{bpp}^{\beta}, \tag{5}$$

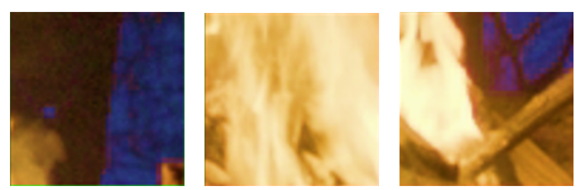
where $\alpha = C * K$, $\beta = -K - 1$ and R_{bpp} is the target bit rate in *bits per pixel* (bpp). Once λ is computed by (5), QP values at frame and CU levels are obtained according to (3).

In order to accurately reflect the RD characteristics of the whole video sequence, the current $R - \lambda$-based RC algorithm in HEVC uses the actual number of encoded bits and actual λ values to update the model parameters, α and β, as pictures and CUs are encoded. A consistent quality is then achieved by clipping all λ and QP values in a narrow range [9].

3. PROPOSED MULTI-$R - \lambda$ MODEL RC

Similarly to the current HEVC RC algorithm based on a single $R - \lambda$ model, our proposed RC approach can be divided into two steps. Step one allocates a bit budget to each coding level (GOP, pictures and CUs), while the second step computes the best set of QP values to attain the target bit rate.

Our RC approach uses a set of three distinct $R - \lambda$ models whose parameters are adapted, during encoding, to the RD characteristics of three distinct classes of regions depicted in the current picture. These classes are defined according to their range of luma values. Since the smallest entity considered by RC in HEVC is the largest CU (LCU), each $R - \lambda$ model is updated according to the RD characteristics of a particular class of LCUs. *Class A* LCUs are those with luma values associated with SDR content; *Class B* LCUs are those with luma values associated with very bright regions; and *class C* LCUs are those with luma values associated with both very bright regions and very dark regions. Therefore, for HDR content encoded at 10-bit precision with the ST2084 specification [14], the luma values are mostly in the range [0, 500] for *Class A* LCUs, mostly above 650 for *Class B* LCUs, and mostly span the whole range of values for *Class C* LCUs. LCUs are classified *a priori* the encoding of the current picture. These classes are exemplified in Fig 1.



(a) A: SDR-like (b) B: Very bright (c) C: High contrast

Fig. 1. LCUs illustrating three classes based on luma values.

3.1. Bit allocation

Bit allocation is done at GOP, picture and CU levels [10]. Our proposed approach works for inter-coded pictures.

GOP level: the target number of bits of the current GOP, T_G, is computed based on the frame rate, f, the number of frames in the current GOP, N_G, the number of pictures already encoded in the current GOP, N_c, their respective bit cost, T_c, and a smoothing window, S_w:

$$T_G = N_G * \left(\frac{\frac{R_{target}}{f}(N_c + S_w) - T_c}{S_w} \right) \tag{6}$$

Picture level: for inter-coded frames, bit allocation can be done equally, hierarchically or adaptively. In hierarchical bit allocation, which is used in the Random Access (RA) profile, the hierarchical level, ω_p, of the current picture, p, is used to determine the corresponding target bits, T_{Pic_p}:

$$T_{Pic_p} = \frac{T_G - \bar{T}_G}{\sum_{j>p} \omega_j} * \omega_p, \tag{7}$$

where $\bar{T}_G$ is the number of bits already spent in the current GOP. Class $x \in \{A, B, C\}$ is then assigned $T_{Pic_{p_x}}$ bits based on T_{Pic_p}, the number of LCUs in the class, denoted by n_x, and $\mathcal{C}$, the total number of LCUs in the current picture, p:

$$T_{Pic_{p_x}} = \left(\frac{T_{Pic_p} * n_x}{\mathcal{C}}\right) \quad x \in \{A, B, C\}. \tag{8}$$

CU level: the $T_{Pic_{p_x}}$ bits are initially distributed among the n_x LCUs of class x using weights. Specifically, the target number of bits for the current LCU i of class x, denoted by $T_{LCU_{i_x}}$, is:

$$T_{LCU_{i_x}} = \frac{T_{Pic_{p_x}} * \omega_{LCU_{i_x}}}{\sum_{n_x} \omega_{LCU_x}} \quad x \in \{A, B, C\}, \tag{9}$$

where $\sum_{n_x} \omega_{LCU_x}$ is the total weight of LCUs of class x. The weight of the current LCU i of class x, $\omega_{LCU_{i_x}}$, with N_{pix} pixels, is:

$$\omega_{LCU_{i_x}} = N_{pix} \left(\frac{\lambda_{i_x}}{\alpha_{i_x}}\right)^{1/\beta_{i_x}} \quad x \in \{A, B, C\} \tag{10}$$

$$\lambda_{i_x} = \alpha_x \left(\frac{T_{Pic_{p_x}}}{n_x N_{pix}}\right)^{\beta_x} \quad x \in \{A, B, C\} \tag{11}$$

where α_{i_x} and β_{i_x} are the $R-\lambda$ model parameters at the LCU level, and λ_{i_x} is defined at the LCU level using $R-\lambda$ model parameters at the class level, α_x and β_x.

In order to guarantee the target number of bits for class x, i.e., $T_{Pic_{p_x}}$, is met, the number of bits allocated to the current LCU i of class x is refined taking into account the number of bits of already encoded LCUs of class x, denoted by $\bar{T}_{Pic_{p_x}}$, and a smoothing window, $\delta = 4$:

$$\hat{T}_{LCU_{i_x}} = T_{LCU_{i_x}} - \frac{\sum_{j>i} T_{LCU_{j_x}} - \left(T_{Pic_{p_x}} - \bar{T}_{Pic_{p_x}}\right)}{\delta} + 0.05 \quad x \in \{A, B, C\} \tag{12}$$

The target bit rate (bpp) of the current LCU i of class x is then:

$$R_{LCU_{i_x}} = \frac{\hat{T}_{LCU_{i_x}}}{N_{pixels}} \tag{13}$$

From (10), one can see that the number of bits assigned to each LCU depends on the parameters of the $R-\lambda$ model. If a single $R-\lambda$ model is used for all LCUs, the bit budget allocation tends to initially assign an equal number of bits to all LCUs within a frame (before this number is refined by (12)). This assignment may affect those LCUs that depict very bright details since they may be forced to meet a strict target bit budget by increasing the value of their QP, which may degrade their reconstruction quality. In compressed HDR sequences, it is highly desirable to maintain the visual details inherent to the HDR content, particularly in regions that are very bright (i.e., *class B* LCUs). Based on this observation, our proposed RC algorithm uses a distinct $R-\lambda$ model for each class in order to assign a bit budget to LCUs according to their luma values (see (10)-(11)). Our approach then attains an overall target bit-rate, R_{target}, by 1) updating parameters for each $R-\lambda$ model independently; i.e., α_x and β_x, with $x \in \{A, B, C\}$, as LCUs are encoded, and 2) computing the corresponding QP values (see (3)) using the target bit rate of each LCU (see (12)-(13)).

3.2. Updating process

Parameters α_{i_x} and β_{i_x} of the $R-\lambda$ model of the ith LCU of class $x \in \{A, B, C\}$, are updated by using the actual number of encoded bits and actual λ_{i_x} values of previously coded LCUs within the same class, as exemplified in Fig 2. For example, for the first LCU A_0 in the sample frame of Fig 2, α_0 and β_0, i.e., the initial set of parameters available at the encoder, are used as this is the first LCU of this class. After encoding this LCU, the updating process produces parameters α_{0_A} and β_{0_A}, which are used to encode the next LCU of *class A*; i.e., LCU A_1. The updating process performed after encoding LCU A_1, produces parameters α_{1_A} and β_{1_A}, to be used in the next LCU of *class A*. Specifically, after encoding LCU i of class x, α_{i_x} and β_{i_x} are updated, on a per-class basis, to produce $\alpha_{(i+1)_x}$ and $\beta_{(i+1)_x}$:

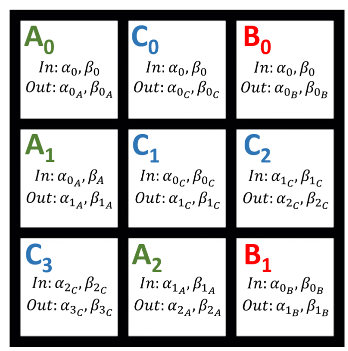


Fig. 2. Update of α_{i_x} and β_{i_x}, on a per-class basis, for a sample frame with nine LCUs. *In:* parameters to encode the current LCU. *Out:* updated parameters after encoding the current LCU.

$$\begin{aligned} \alpha_{(i+1)_x} &= \alpha_{i_x} + 0.1(ln(\lambda_{i_{x_{real}}}) - ln(\lambda_{i_x})) * \alpha_{i_x} \\ \beta_{(i+1)_x} &= \beta_{i_x} + 0.05(ln(\lambda_{i_{x_{real}}}) - ln(\lambda_{i_x})) * ln(bpp_{i_{x_{real}}}) \end{aligned} \tag{14}$$

where $\lambda_{i_{x_{real}}}$ and $bpp_{i_{x_{real}}}$ are the actual lambda and bit rate of LCU i of class $x \in \{A, B, C\}$. QP and λ values of all LCUs of each class are averaged in order to clip, on a per-class basis, QP and λ values at the LCU level in a narrow range in subsequent frames.

4. PERFORMANCE EVALUATION

Performance evaluations are carried on a set of 10 HDR video sequences with the following characteristics: 4:2:0 format, 10-bit precision, resolution of 1920×1080 pixels, frame rate of 25 fps and

Table 1. Test HDR sequences

Source	Sequences
[15]	BeerFestTeaser, FireplaceTeaser, ShowGirl2Teaser
[16]	beerfest_lightshow, bistro, carousel_fireworks, cars_longshot, showgirl_01
[17]	04_EBU_HD_100p_HDR, 13_EBU_HD_200p_HDR

Rec.2020 container with the PQ (SMPTE ST.2084) EOTF [14], as tabulated in Table 1. Only the first 64 frames of each video are used for these experiments. Our approach is implemented on the reference software HM v16.9 [18] and compared with the corresponding RC algorithm based on a single $R-\lambda$ model. All sequences are encoded using the RA profile with hierarchical bit allocation enabled at different target bit rates (12.4416, 24.8832, 37.3248, 49.7664 and 62.208 Mbps). We first compute the average absolute Bit Rate Error (BRE), in percentage, for all tested sequences. The BRE indicates how accurately the target bit rate is attained. Fig. 3 plots the average absolute BRE at different bit rates. We can observe that our multi-$R-\lambda$ model approach helps to attain the target bit-rate more accurately than the current approach in HEVC. *Class B* LCUs tend to require a large number of bits in the encoding process compared to SDR-like areas. This inevitable demands more bits from the target bit budget, thus forcing the current RC algorithm in HEVC to spend more bits in these regions, resulting in overspending the overall bit budget. Our approach, on the other hand, assigns a bit budget to each class of LCUs according the RD characteristic of the whole class, thus minimizing overspending. As model parameters are updated on a per-class basis, our bit allocation guarantees that very costly regions are assigned more bits at the expense of reducing the bits assigned to regions that are easy to encode. All this, while guaranteeing that the overall target bit rate is met. Based on this fact, the quality of the reconstructed sequences can be measured on a per-class basis. To this end, we use the mPNSR metric [19], which is a PSNR version for HDR content that considers the exposure levels (see Fig. 4), and the HDR-VDP-2 metric [20] (see Fig. 5), for which the resulting probability map is summarized into a single dB value.

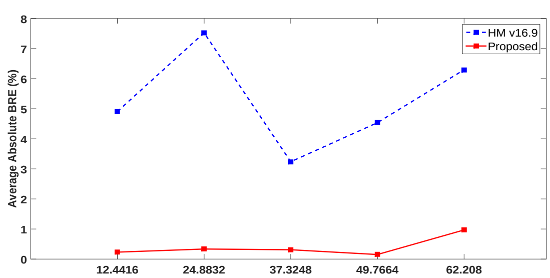


Fig. 3. Average absolute Bit Rate Error (%) at different bit rates.

Fig. 4 shows that the proposed approach attains a particularly higher reconstruction quality for *class B* than the current RC algorithm. This shows that details and the quality of very bright regions, which are inherent to HDR content, are better preserved by our proposed approach. A similar trend is observed for the HDR-VDP-2 metric in Fig. 5. Note that although for both metrics our proposed approach attains improvements for all classes, those attained for *class C* are relatively negligible compared to those attained for

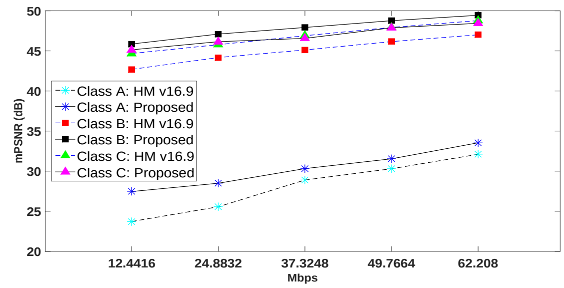


Fig. 4. Average mPSNR, per-class, at different bit rates.

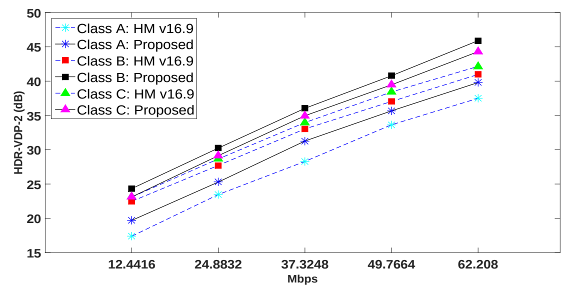


Fig. 5. Average HDR-VDP-2, per-class, at different bit rates.

(a) HM v16.9 (b) Proposed

Fig. 6. Sample visual results for HDR regions - class B.

class A and *class B*. This shows that our approach distributes the bit budget efficiently by considering not only the RD characteristics of each class, but also the total number of of LCUs of each class in each picture (see (8)). Fig. 6 shows visual results for two frames of sequence *cars_longshot* [16] at 12.4416Mbps. We can observe that very bright regions are indeed reconstructed with less compression artifacts after using our approach.

5. CONCLUSIONS

This paper proposed a multi-$R-\lambda$ model RC approach in HEVC for HDR sequences, which uses three distinct $R-\lambda$ models to accurately approximate the RD characteristics of different regions in a HDR picture. Results, in terms of the mPSNR and HDR-VDP-2 metrics, show that our approach attains important reconstruction quality improvements compared to the current RC algorithm in HEVC, especially for very bright regions, while accurately attaining an overall target bit rate.

6. REFERENCES

[1] Consumer Technology Association, "CEA defines hdr compatible displays. [Online] Available," https://www.cta.tech/News/Press-Releases/2015/August/CEA-Defines-%E2%80%98HDR-Compatible%E2%80%99-Displays.aspx.

[2] T. Borer and A. Cotton, "A display independent high dynamic range television system," *BBC White Paper WHP309*, 2015.

[3] G. J. Sullivan, J. R. Ohm, F. Bossen, and T. Wiegand, "JCT-VC AHG report: HM subjective quality investigation," *JCTVC-H0022*, 2012.

[4] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264 avc video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[5] M. Naccari and F. Pereira, "Quadratic modeling rate control in the emerging HEVC standard," *Proc. Picture Coding Symp., 2012*, 2012.

[6] X. Liang, Q. Wang, and et. al, "A novel RQ model based rate control scheme in HEVC," *Vis. Commun. Image Process*, pp. 1–6, 2013.

[7] Z. He, Y. K. Kim, and S.K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 928–940, 2001.

[8] Y. Pitrey, M. Babel, and O. Deforges, "One-pass bitrate control for MPEG-4 Scalable video coding using ρ-domain," *IEEE Int. Sym.on Broadband Mult. Sys.* & *Broadcasting*, pp. 1–5, 2014.

[9] B. Li, H. Li, and L. Li, "Adaptive bit allocation for R-lambda model rate control in HM, in JCT-VC M0036," *13th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3*, 2013.

[10] B. Li, H. Li, and L. Li, "λ-domain rate control algorithm for High Efficiency Video Coding," *IEEE Transactions on Image Processing*, vol. 23, no. 9, 2014.

[11] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "Region-of-Interest-based rate control scheme for High-Efficiency Video Coding," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014.

[12] V. Sanchez, F. Auli-Llinas, R. Vanam, and J. Bartrina-Rapesta, "Rate control for lossless region of interest coding in HEVC intra-coding with applications to digital pathology images," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2015)*, pp. 1250–1254, 2015.

[13] M. R. Ardestani, A. A. B. Shirazi, and M. R. Hashemi, "Rate-distortion modling for scalable video coding," *In Proc. IEEE 17th Int. Conf Telecommun. (ICT).*, 2010.

[14] SMPTE, "Study group report High-Dynamic-Range (HDR) imaging ecosystem," *SMPTE 2015.*, 2015.

[15] Stuttgart University, "High Dynamic Range database Graded MPEG Teaser clips. [Online] Available," https://hdr-2014.hdm-stuttgart.de/.

[16] Stuttgart University, "High Dynamic Range database Full Footage Color Graded. [Online] Available," https://hdr-2014.hdm-stuttgart.de/.

[17] EBU-SVT, "Public HDR Test sequences. [Online] Available," https://tech.ebu.ch/EBU_SVT_Public_Test_Sequences.

[18] High Efficiency Video Coding (HEVC) Software., "HM v16.9. [Online] Available," https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.

[19] F. Banterle, K. Debattista, P. Ledda, and A. Chalmers, "A GPU-friendly method for high dynamic range texture compression using inverse tone mapping," *GI '08 Proceedings of Graphics Interface 2008.*, 2008.

[20] R. Mantiuk, K. J. Kim, a. G. Rempel, and W. Heidrich, "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2011*, vol. 30, no. 4, 2011.