南開大學 Nankai University

电子信息与光学工程学院

Pathology Image Compression Based on JPEG2000, Multi-Resolutional Human Perception and the Region of Interest Predictions

Yinghai Jiang[1], Feng Liu[1,2,*], Rongsheng Cui[1], Xinzhuo Zhang[1], and Xianyuan Zhang[1]

[1]College of Electronic Information Science and Technology, Nankai University, Tianjin, China; [2]Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Nankai University, Tianjin, China. *Correspondence: liuf@nankai.edu.cn

Introduction

- The transmission of medical images is a vital part of telemedicine. Ever since last century, especially in recently years, more and more research efforts have been attracted to the development of the lossy image compression techniques based on the HVS perception.
- In [1], a model was introduced by Watson et al to calculate the visibility of quantization distortions of DWT, and was later implemented for gray-scale image coding.
- Han et al proposed a visually lossless image coding method using visibility thresholds (VTs) [2]. The model was later developed to support multi-resolution visually lossless display [3].
- In this work, we propose a coding scheme for the WSIs, with different resolutions and visual qualities with the visual masking effect sufficiently considered. Moreover, using the image region of interest (ROI) prediction results using the deep neural network (DNN), we further retain more important features in coding results of the lung squamous cell carcinoma (LSCC) WSIs.

Models of Wavelet Coefficients and Quantization Distortions in JPEG2000

- In a JPEG2000 Part-I encoder, we firstly convert the RGB color components into the YCrCb components with the multi-component transform (MCT). Then the 2D-DWT is used to generate wavelet coefficients, which are divided into code blocks and quantized by a deadzone quantizer. To produce the final codestream, the code blocks are arithmetically encoded.
- 4 orientations of the wavelet subbands exist: LL, HL, LH, HH.
  - For LL subbands, we model the wavelet coefficients with Gaussian distributions;
  - For HL, LH, and HH subbands, we model the wavelet coefficients with Laplacian distributions.
- The deadzone quantizer and the corresponding dequantizer using mid-point reconstruction are shown in Fig. 1.

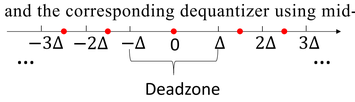


Fig. 1: Deadzone Quantizer and Mid-Point Reconstruction

- The quantization distortions $d = y - \hat{y}$ are the main source of the compression artifacts in JPEG2000, where y and $\hat{y}$ are the original and reconstructed coefficients, respectively.
- The quantization distortions in the subbands are modeled analytically as illustrated in Fig. 2.

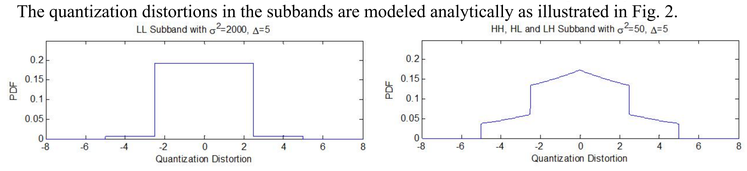


Fig. 2: Deadzone Quantization Distortion Models

Human Visual Sensitivity of Individual Wavelet Coefficients

- The probability to detect a single wavelet coefficient in the gray background is modeled using the Weibull psychometric function:
  - $W_T(y) = 1 - (1-\gamma)\exp\left(-\left|\frac{y}{T}\right|^{\beta}\right), \beta = 2$ (1)
  - $\gamma = 0$ for a "Yes/No test" ; $\gamma = \frac{1}{3}$ for a 3-Alternative-Forced-Choice (3AFC) test.
- T in Eq. (1) is the value of the wavelet coefficient that results in 75% probability of detection in a 3AFC test and denotes a parameter which determines sensitivity to the stimulus under study.
- 4 human subjects were recruited to measure the T values. With the QUEST toolbox [4], the 3AFC tests were conducted for each subband using 32 trials. In a normal office environment, the measurements were conducted on an ASUS PA328Q monitor at a 60 cm viewing distance, shown in Fig. 3. In this case, we define the display as the full resolution. The stimulus in these measurement is briefly shown in Fig. 4.
- The lowest T got from the 3 subjects was taken at each subband. At the full display resolution, the resulting T values for different subbands and components are shown in Tab. 1 and defined as $T_{c,(b,l),l=1}$

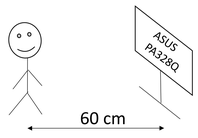


Fig. 3: Setup of the Psychovisual Experiments

Fig. 4: Stimulus in Measuring Single Coefficient Visibility

- Consider the images in the Wavelet domain and the spatial frequency domain:
  - If the full resolutional image is decreased by a factor of 2^{-u} (u is a positive integer), the spatial frequencies corresponding to the subband (b, l) in the original image will move into the subband $(b, l-u)$, i.e. $T_{c,(b,l-u),2^{-u}} = T_{c,(b,l),1}$

Subband	Y	Cb	Cr	Subband	Y	Cb	Cr
LL0	15.00	86.58	32.20	HL/LH3	0.57	2.55	1.28
HH1	20.24	72.92	49.99	LL3	0.43	2.44	0.60
HL/LH1	5.17	38.99	13.97	HH5	0.39	2.42	0.85
LL1	3.68	25.78	6.19	HL/LH4	0.37	1.77	0.53
HH2	2.35	18.29	7.43	LL4	0.36	1.18	0.41
HL/LH2	1.23	7.86	3.84	HH5	0.30	1.09	0.40
LL2	1.13	7.40	2.49	HL/LH5	0.30	0.52	0.39
HH3	0.68	4.12	2.07	LL5	0.36	0.47	0.33

Tab. 1: T values for Single Wavelet Coefficients

- If the relative resolution r is not an integer power of 2 and $0 < r < \frac{1}{2}$, we can always find a positive integer m, so that $T_{c,(b,l+m),r\cdot 2^m} = T_{c,(b,l),r}$, with $\frac{1}{2} < r \cdot 2^m < 1$.
- For $\frac{1}{2} < r \cdot 2^m < 1$, according to the relationship between the wavelet domain and the frequency domain, we have: $T_{c,(b,l),r} = \left(\frac{\alpha_{b,LL,2,r}}{t^{\beta}_{c,(LL,l+1),1}} + \sum_{u=0}^{1}\sum_{b_0 \in LL}\left(\frac{\alpha_{b,b_0,u+1,r}}{t^{\beta}_{c,(b_0,l+u),1}}\right)\right)^{-\frac{1}{\beta}}$, where $\alpha_{b,b_0,k,r} = \sum_{q\in(b_0,k)} x_q^{\beta}$ denotes the sum of β powered resulting wavelet coefficients in the subband (b_0, k) from the unit stimulus in the subband (b, r).

The Proposed Image Coding Based on the Visibility Thresholds in JPEG2000

- To start with, we first define a visual quality parameter, denoted by Q, based on the probability of detection of quantization distortion within a field of view (FOV) of 2 degrees at a viewing distance of 60 cm. The relationship between the detection probability of such stimuli and the Q value can be calculated by:
  - $P_d = 1 - (1-\gamma)exp(-Q)$ (2)
- We refer to the quantization step size which results in the probability of detection given in Eq. 2 as the visibility threshold (VT) for that Q value.
- From Eq. (1) and (2) and the assumption that the visibility of each wavelet coefficient is independent, from a "Yes/No" test we have:
  - $exp(-Q) = \left\{\int_{-VT_{c,(b,l),r}(Q)}^{VT_{c,(b,l),r}(Q)} \exp\left[-\left|\frac{e}{T_{c,(b,l),r}}\right|^{\beta}\right] f(e)de\right\}^{N_{(b,l),r}}$ (3)
  - $f(e)$ is the distribution of quantization distortion discussed in Fig. 1, $N_{(b,l),r}$ is the number of wavelet coefficients in that subband, given the image display relative resolution r.
- **The value of VT** can be solved numerically from Eq. 3, given the Q and the T parameters.
- **The visibility masking** is considered from the following 2 aspects in this work:
  - Adjustment of the Q parameter from the aspect of wavelet decomposition direction:
    - $Q_{subband} = \left(\frac{\sigma_{b,l,c}}{\sigma_{LL,5,c}}\right)^{\rho_r} \times Q_{component}$, where $\sigma_{b,l,c}$ is the standard deviation of the coefficients in the wavelet subband, $\rho_r = 0.7,\ 0.85,\ 1$ for $r = 1,\ 0.75,\ 0.5$, respectively.
  - Adjustment of the Q parameter from the aspect of luminance and chrominance components:
    - $Q_{component} = m_{component}^{\frac{1}{r}} \times Q_{image}$, where $m_Y = 1$, $m_{Cb} = 0.7$ take the value of 0.7 and $m_{Cr} = 0.85$.
- Given that VT was defined as the quantization step size that results in quantization distortion with the probabilities of detection at the desired JND level, **the encoding process** is done as follows:
  - the coding passes from the MSB are included in the final codestream until the first coding pass, where the absolute quantization distortion within the deadzone falls below VT and the absolute quantization distortion out of deadzone falls below VT/2.
  - All subsequent coding passes are discarded.

The ROI-based Image Coding

- The DNN implemented in this paper to predict the probability map of the lesions is established based on the U-Net achitecture and the residual neural network modules.
- In contrast to the conventional U-Net, the convolutional layers of the encoding path in the conventional U-Net are replaced with residual convolutional modules.
- We take the LSCC WSI blocks as examples to conduct the lesion region (i.e. ROI) predictions and image coding. To train the proposed DNN, the training data are established based on the open-source US National Cancer Institute (NCI) Clinical Proteomic Tumor Analysis Consortium Lung Squamous Cell Carcinoma (CPTAC-LSCC) database [5]. The lesion regions in the WSIs were manually labeled with the help of the pathologists in Tianjin Chest Hospital.
- The taining was conducted using 3917 training image blocks sized 576×576, with 100 epochs and the batch size of 4. The cross-entropy was implemented as the loss function in DNN training.
- During the coding process, we can find the position of the pixel corresponding to the wavelet coefficients after decoding, and generate the probability weight of each wavelet coefficient:
  - $D' = \omega \cdot D$ (4)
  - D and D' are the actual maximum quantization distortion and the weighted maximum quantization distortion used for comparing with VT in the code block, respectively.
  - $\omega = \frac{1+e^{\lambda(\overline{P}-\zeta P_t)}}{\xi T+1+e^{\lambda(\overline{P}-\zeta P_t)}}$, where $\xi = 128$ and $\lambda = 100$. $\overline{P}$ is the average probability of a lesion pixel at the corresponding position of the wavelet coefficient. P_t is the average probability of the lesion pixels across the entire image. $\zeta = 0.5$ represents the selectivity of ROI and non-ROI.

Image coding and Validation Experiments

- The method was verified with an ensemble of 100 LSCC WSI blocks sized 512×512 from the CPTAC_LSCC database.
- Without considerations of the ROI, the resulting bit rates are shown in Tab. 2. The averaged HDR-VDP-2 MOS scores and the representative HDR-VDP-2 distortion probability maps [6] from the HVS-optimized and the MSE-optimized encoding methods are shown in Fig. 5 and 6, respectively.
- With considerations of the ROI, the coding of 8 representative images at $Q = r = 1$ were performed. Bit rate comparisons and the decoding results between the ROI and non-ROI methods are shown in Tab. 3 and Fig. 7, respectively.

Q	Bit Rates (Bits/Pixel, maximum/average/minimum)		
	r=0.5	r=0.75	r=1
0.1	0.89/0.74/0.58	3.28/1.99/0.93	4.33/2.73/1.36
0.3	1.08/0.90/0.71	4.55/2.94/1.47	5.23/3.83/2.01
0.7	1.43/1.18/0.94	5.11/3.38/2.04	5.99/4.33/2.63
1	1.46/1.21/0.96	5.22/3.67/2.09	6.51/4.71/3.03

Tab. 2 Averaged Coding Bit Rates without Consideration of ROI

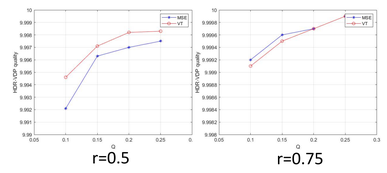
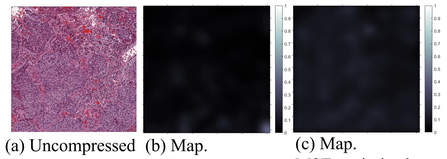
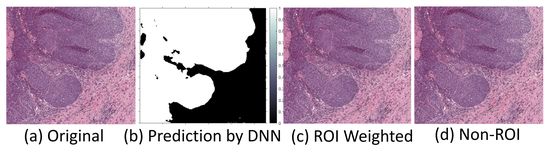


Fig. 5 Averaged MOS of the Two Encoders without Consideration of ROI



(a) Uncompressed Image (b) Map. HVS-optimized (c) Map. MSE-optimized

Fig. 6 Distortion Probability Maps on a Sample Image Using the Two Encoders without Consideration of ROI



(a) Original (b) Prediction by DNN (c) ROI Weighted (d) Non-ROI

Fig. 7 Representative Decoding Results from the ROI Weighted and Non-ROI Coding Methods

Image Index	ROI	Non-ROI
1	4.13	4.69
2	3.72	4.94
3	3.29	4.86
4	4.79	5.79
5	4.09	4.62
6	3.64	6.11
7	2.68	3.86

Tab. 3 Representative Bit Rate (Bits/Pixel) Comparisons between the RO Weighted and Non-ROI coding Methods

Conclusions

- In this paper, we proposed a multi-resolutional HVS sensitivity model for JPEG2000 compression.
- The proposed model was developed into a multi-resolutional and multi-perceptional-quality image coding method for the digital pathology whole slice images.
- The proposed method was further improved in the essential information retainment, with the help of the lesion region prediction from the DNN.

Acknowledgements: This work was supported by the Natural Science Foundation of China (grant no. 61910233) and Natural Science Foundation of Tianjin City, China (grant no. 19JCQNJC00900).

References

[1] A. Watson, G. Yang, J. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," IEEE Transactions on Image Processing, vol. 6, no. 8, pp. 1164-1175, 1997.
[2] H. Oh, A. Bilgin, and M. Marcellin, "Visually lossless encoding for JPEG2000," IEEE Transaction on Image Processing, vol. 22, no. 1, pp. 189-201, 2013.
[3] H. Oh, A. Bilgin and M. Marcellin, "Visually lossless JPEG2000 for remote image browsing," Information, vol. 7, no. 3, pp. 45, 2016.
[4] A. Watson, "QUEST: A Bayesian adaptive psychometric method," Perceptions & Psychophysics, vol. 33, no. 2, pp. 113-120, 1983.
[5] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium, "Radiology Data from the Clinical Proteomic Tumor Analysis Consortium Lung Squamous Cell Carcinoma Collection," https://doi.org/10.7937/k9/tcia.2018.6emub5l2, last accessed on March 5, 2021.
[6] 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," in ACM Transactions on Graphics (TOG), vol. 30, no. 4. ACM, 2011, p. 40.