VLLCVD Subjective Image Fidelity Criterion and Its Application to Evaluation of Visual Models

Day-Fann Shen, Jen-Hsing Sung, W.C. Fan, M.Z. Lin, J.G. Fun, Z.H. Shen, F.K. Tsai and Y.Z. Shin
Department of Electrical Engineering
Yunlin University of Science and Technology, TAIWAN
E-mail: shendf@pine.yuntech.edu.tw

Abstract

The emerging JPEG2000 image coding standard includes the requirements for visually lossless images, however, "visually lossless" is not clearly defined. Many researchers claim that their encoded/decoded images are visually lossless, yet if we examine those images closely under a strict viewing condition, such as in a dark room at night at very short distance, differences between the original and the decoded images can be detected. For this reason, serious image users, like physician and radiologist, are reluctant to use so call visually lossless compressed images. In this paper, we propose a more strict and precise subjective image fidelity criterion based on two techniques (1) the same position swapping (SPS) technique between the decoded image and the original and (2) the measurement of Visually Lossless Critical Viewing Distance (VLLCVD). The VLLCVD based SFC provides two scores: (1) the ratio of testers who perceive the decoded image as absolutely visually lossless (at any viewing distance) and (2) the average critical viewing distance that the image is perceived as visually lossless. The proposed VLLCVD SFC can be used to examine many visual models proposed for image coding as well as invisible image watermarking.

1. Introduction

Since human visual system (HVS) is the final receiver of image and video, it is very important to incorporate human visual properties (HVP) in various image processing tasks, such as image coding, image segmentation and invisible image watermarking. In image coding application, the perceptual image quality can be optimized by effective bits allocation according to HVP. As a result, same amount of bits can achieve better image quality [1]-[10]. Image segmentation is a process required by MPEG4 as well as image analysis and computer vision tasks. The HVP based pixel/region merge criterion in image segmentation is proven to be a simple and effective alternative to the more complicated statistically based merge criteria in image segmentation [11]. In invisible image watermarking, sufficient amount of watermark signals can be added to an image without being noticed by human perception, while increasing the robustness to attacks from pirates [12]~[16].

HVS and therefore HVPs are quite complicated [19], a complete description of the HVS and HVPs is not practical and unnecessary. Instead, an efficient and effective representation of HVPs in terms of a human visual model (HVM) is normally sufficient for a particular image processing task. Among the HVPs, Just Noticeable Difference (JND) property is the most frequently used. JND property states the facts that HVS's sensitivity to visual stimuli is limited, it cannot detect the difference between two visual components, if the difference is smaller than a certain threshold. Further, HVS's sensitivity is not linear to all visual stimuli, that is, visual contents of different luminances and frequencies are weighted differently. In addition, other factors including ambient lighting condition, quality of the image display monitor, viewing distance and personal eye sight capability may influence HVS's sensitivity to image quality.

Many JND based HVMs have been proposed [1-16]. Various HVMs are derived by researchers using different approaches and under different viewing conditions. For example, Watson et al. proposed HVMs in the form of a DCT quantization matrices for individual images [9-1], visibility of DCT basis function [9-2] and the visibility of wavelet quantization noises [9], which were derived from DWT (Discrete Wavelet Transform) basis function stimuli and DWT uniform quantization noise stimuli. While Chou et al. [8] derived a JND/MND profile in spatial domain for an image by considering the local property for each pixel, including the background luminance as well as the texture masking effects. The JND/NMD profile is then trasnformed into (DCT or Wavelet) subbands and assigned different weights for each subband according to HVS's sensitivity to each subband. Shen et al. [6] derived a JND model based on measurements of JND threshold on square waves of different frequencies and directions [6][7]; the result is a set of JND thresholds for each wavelet subband.

Two problems related to HVS remain unresolved in image coding. Firstly, given two decoded lossy images, how do we judge which image has better image fidelity? Secondly, given two HVMs [1-16], which HVM should we choose for image coding (or certain image processing task)?

Currently, there are three categories of criterions for assessing the fidelity of a decoded image, they are (1) Subjective fidelity criterion (SFC) (2) Objective fidelity criterion (OFC) (3) Visually weighted Objective fidelity criterion (VWOFC) [17] [18]. Given a decoded image and its original, an OFC estimates the fidelity based on the arithmetic errors between the pair. Examples of common used fidelity measure are SNR (Signal to Noise Ratio), PSNR (Peak Signal to Noise Ratio), MSE (Mean Square Error) or RMSE (Root Mean Square Root), the result is a number indicating the degree of fidelity between the decoded image and the original. OFC is easy to compute and is widely adopted in the area of image compression, however, it is well known that results by OFC do not always consistent with human perceptions.

To overcome this shortcoming, VWOFC assigns errors in different image components with different weights according to a human visual model. It is expected that the results from VWOFC are consistent with human perception while without involving the time consuming efforts from a group of testers.

Nevertheless, human eyes are the final receivers of an image and are naturally the best judges of the image fidelity. For serious image fidelity evaluations, in spite of requiring human efforts and time consuming, SFC is often adopted instead of OFC and VWOFC. A SFC calls for a group of human testers (or observers, umpires) who are responsible to rate the fidelity at their subjective judgements. For example rating 1 to 5 may corresponds to (1)Absolutely unable to distinguish, (2) Almost unable to distinguish, (3) difficult to distinguish, (4) Easy to distinguish (5) Very easy to distinguish}. The average rating from all testers indicates the degree of fidelity. FSC has at least two drawbacks, firstly, it requires efforts from a group of human testers and it is time consuming, secondly, SFC by rating is vague and not precise. Average rating from a SFC reflects the true fidelity of a decoded image perceived by average human beings.

Although not perfect, many criterions are available for assessing the fidelity of a decoded image as described above; Currently, there is no criterion for assessing a HVM! In this paper, we attack the two problems at one time. First, we propose a more accurate SFC as the image fidelity criterion. The proposed SFC image fidelity criterion is based on the SPS (Same Position Swapping) technique as well as the VLLCVD (Visually LossLess Critical Viewing Distance) measurements. The SPS technique utilizes the contrast principle of human perception, while the VLLCVD yields a score in distance which is more precise (in cm or inch) than the rating system. Secondly, we proposed a procedure for comparison of VHMs using the VLLCVD criterion.

In the second section, we define the observation conditions for the SFC measurements; In the third section, the SPS technique and VLLCVD measurement are illustrated. In the fourth section, the two subjective fidelity scores are defined based on the VLLCVD measurements. In the fifth section, we demonstrate the application of the proposed SFC in evaluation and comparison of two visual models in the context of image coding. Conclusion is made in Section 6.

2. Conditions for Subjective experiments

Many parameters (factors) may affect the observations and consequently the result of the proposed SFC. These parameters are adjusted to minimize the uncertainties as well as for most strict results.

1.The surrounding lighting condition: All observations are conducted in a dark room at night with no other light sources except the PC monitor in use. This guarantees the smallest variation in surrounding lighting condition as well as increases the sensitivity of human visual perception.

2.Monitor: a SONY-20 inches CRT PC monitor (Model GDM-20) is used with Contrast and Brightness knobs adjusted to yield the sharpest and most clear images.

1. Testers: 6 volunteers from the image processing class, ages range from 21 to 24 and eyesight from 0.8 to 1.2. Their personal data is listed in Table I. For unbiased judgement, testers are not told anything about the images.

4.Observation method: SPS and VLLCVD to be described in the next section.

Table 1. Data of the six testers

No.	Name	Age	Eye sight		ID. number
			L	R	
1	Fan . W. C	24	0.8	0.8	H122058688
2	Lin . M . Z	21	0.8	0.8	R123024423
3	Fun . J . G	21	1.0	1.0	R122882136
4	Shen . Z . H	21	1.0	1.0	S122880044
5	Tsai . F . K	24	1.0	1.0	L122310327
6	Shin . Y . Z	22	0.8	0.8	L122873481

3. SPS technique and VLLCVD measurements

The SPS technique

Instead of placing two images side by side, the SPS (Same Position Swapping) technique places the decoded image on top of its original i.e. both are displayed in the same position on the center of the monitor. The occluded image in the bottom can be swapped to the top by clicking the mouse. Since human perception is more sensitive to contrast changes in time, the contrast created by swapping the two images can be detected. For better focus of the testers' attention, both the decoded and the original images are divided into 128*128 sub-blocks using thin lines (one pixel in width, 180 gray levels).

The VLLCVD (Visually LossLess Critical Viewing distance) procedure

Task 1: Set up the observation environment as described in section 2. The decoded image f and its original are displayed on the screen as required by the SPS technique.

Task 2: Begin at the initial viewing distance (IVD) (from the center of the screen to the center of the tester's eye) of 60 cm, a tester *t* pays full attention on examining a sub-block at a time using the SPS technique for all sub-blocks. If any noticeable difference at any sub-block is found, the tester *t* reports the areas where noticeable difference occur and go to Task 3. Otherwise, if no difference is found, then go to Task 4.

Task 3: The tester gradually increases the viewing distance (VD) (away from the screen) until no noticeable difference is found for all sub-blocks. Go to Task 5.

Task 4: The tester gradually decreases the viewing distance (VD) (close to the screen) until no noticeable difference are found in all sub-blocks, then go to Task 5. If for any distance includes viewing distance of zero (the tester's forehead and nose touch the screen), no difference can be found, then go to Task 6.

Task 5: At this point, the decoded image is regarded as Visually LossLess (VLL). This critical viewing distance is denoted as the VLLCVD. For image f and test *t*. It is noted that the VLLCVD depends only on the image under examination f (and of course its original) and the tester *t*, all other parameters that affect the observation are predefined in the previous section, this critical viewing distance is recorded as $cd_{VLL}(f,t)$. Since image f is not VLL for VD $cd_{VLL}(f,t)$, it is not Absolutely Visually LossLess , we record $A_{VLL}(f,t)=0$.

Task 6: The image f is regarded as Absolutely Visually LossLess or AVLL and is denoted as $A_{VLL}(f,t)=1$.

Figure 1 shows the flowchart of the proposed subjective Fidelity Criterion Based on VLLCVD and AVLL. Experiences show that for any VD > $cd_{VLL}(f,t)$, image f is always VLL to tester *t*. It is noted that the smaller the $cd_{VLL}(f,t)$, the better fidelity of the decoded image f for tester *t*.

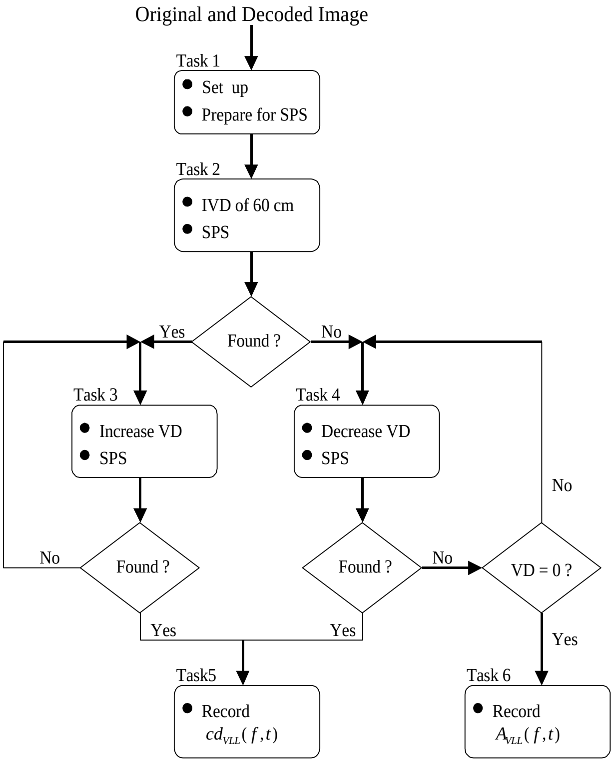


Figure 1: The Process of the Subjective Fidelity Criterion based on VLLCVD.

4. Subjective Fidelity Criterion based on VLLCVD measurements

For a decoded image f, we define two subjective fidelity scores $S_1(f)$ and $S_2(f)$ based on $cd_{VLL}(f,t)$ and $A_{VLL}(f,t)$ recorded by all testers.

$$S_1(f)=\frac{\sum_{\langle testers\rangle} A_{VLL}(f,t)}{N_T}, \quad S_1\in\{\frac{0}{N_T},\frac{1}{N_T},\frac{2}{N_T},\ldots,\frac{N_T}{N_T}\}$$

$$S_2(f)) = \frac{\sum_{\langle testers \rangle} cd_{VLL}(f,t)}{I_T}, \quad S_2 > 0$$

Where N_T is the number of testers, I_T is number of testers who execute Task 5 and record $cd_{VLL}(f,t)$ as well as setting $A_{VLL}(f,t)$ =0. $S_1(f)$ indicates the ratio of testers who perceive image f as AVLL. The larger the ratio, image f is more likely an AVLL. $S_2(f)$ is the average VLLCVDs in cm from all testers who record $cd_{VLL}(f,t)$, the smaller the value, the better the quality. S_1, S_2 together provide a more strict and precise subjective fidelity criterion.

5. Applying VLLCVD SFC to The Assessment of Human Visual Models

Visual model plays an important role in image coding as well as in digital image watermarking[1-16]. In this section, two wavelet-based visual models SY model [6] and WYSV model [9] are compared for their performances using the proposed VLLCVD subjective fidelity criterion. Both visual models are derived by different approaches under different conditions. Both are in the form of Subband Quantization Table or SQT as shown in Table 1 and 2.

Table 1. Basic SQT derived from SY JND model with display resolution=26.256 pixels/degree at 60 cm viewing distance

View distance = 60 cm

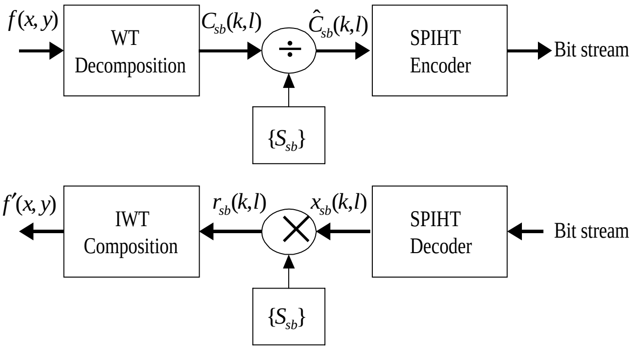
Subband	Step size	Subband	Step size
HH1	52.59	*LH3*	6.00
HL1	14.11	*HH4*	6.00
LH1	15.27	*HL4*	6.00
HH2	11.93	*LH4*	6.00
HL2	6.35	*HH5*	6.00
LH2	6.34	*HL5*	6.00
HH3	6.94	*LH5*	6.00
HL3	6.00	*LL5*	6.00

Table 2. Basic SQT derived from WYSV JND model with display resolution of 32 pixels/degree at 70 cm viewing distance

View distance = 70 cm

Subband	Step size	Subband	Step size
HH1	58.76	*LH3*	12.71
HL1	23.03	*HH4*	17.86
LH1	23.03	*HL4*	14.16
HH2	28.41	*LH4*	14.16
HL2	14.68	*LL4*	14.50
LH2	14.69	*HH5*	*
HH3	19.54	*HL5*	*
HL3	12.71	*LH5*	*

We examine these two visual models in the context of image compression. We adopt SPIHT with JND_SQ as shown in Figure 2 as the image compression algorithm.



$$\begin{cases} \hat{C}_{sb}(k,l) = round\left[\dfrac{C_{sb}(k,l)}{S_{sb}}\right] \\ C_{sb}(k,l) = x_{sb}(k,l) \cdot S_{sb} \end{cases}$$

Figure 2. Block Diagram of JND_SPIHT

Where $\{S_{sb}\}$ is the adjusted SQT obtained by multiplying the basic SQT (in Table 1 or 2) with a Compression Control Factor (CCF) ϕ. Images with various bit rates (therefore various fidelities) are obtained by adjusting ϕ. For bit rates of 0.7, 1.4 and 1.9 bpp on Lena, the corresponding $S_1(Lena)$ by SY/Watson visual models are 102/124, 56/82, 52/56 cm respectively, while $S_2(Lena)$ are $\frac{0}{6}/\frac{0}{6}$, $\frac{1}{6}/\frac{0}{6}$, $\frac{5}{6}/\frac{4}{6}$ respectively. For bit rate of 0.8, 1.5 and 2.0, the corresponding $S_1(F16)$ are 93/109, 58/60, 39/47 cm, while $S_2(F16)$ are $\frac{0}{6}/\frac{0}{6}$, $\frac{2}{6}/\frac{0}{6}$, $\frac{5}{6}/\frac{4}{6}$ respectively. For bit rates of 0.8 and 1.1 bpp, the corresponding $S_1(F16)$ are 93/109, 58/60, 39/47 cm and $S_2(F16)$ are $\frac{0}{6}/\frac{0}{6}$, $\frac{2}{6}/\frac{0}{6}$, $\frac{5}{6}/\frac{4}{6}$ respectively. For bit rates of 0.8 and 1.1 bpp, the corresponding $S_1(Salesman)$ are 62/78, 44/49 cm and $S_2(Salesman)$ are $\frac{0}{6}/\frac{0}{6}$, $\frac{4}{6}/\frac{3}{6}$ respectively. These data are summaried below in Table 3.

Table 3. Comparison SY and Watson's Visual model in terms of bpp vs. image fidelity measure (JND_PSNR, S_1 and S_2) for Lena and F16.

(a) for Lena

Bpp	JNDPSNR (dB)	$cd_{VLL}(Lena)(SY/Watson)\_in\_cm$						S_1 Mean $cd_{VLL}(Lena)$
		1	2	3	4	5	6	
0.7	46.5/45.7 +0.8	80/97	93/125	142/186	86/93	150/181	63/62	102/124 -22
1.4	59.0/57.0 +2.0	38/40	-/82	50/116	45/82	90/119	56/55	56/82 -26
1.9	81.8/67.3 +13.5	-/-	-/-	-/47	-/-	52/65	-/-	52/56 -4

(b) For F16

Bpp	JNDPSNR (dB)	$cd_{VLL}(F16)(SY/Watson)\_in\_cm$						S_1 Mean $cd_{VLL}(F16)$
		1	2	3	4	5	6	
0.8	46.5/45.7 +0.8	80/97	93/125	142/186	86/93	150/181	63/62	93/109 -22
1.5	59.0/57.0 +2.0	38/40	-/82	50/116	45/82	90/119	56/55	58/60 -26
2.0	81.8/67.3 +13.5	-/-	-/-	-/47	-/-	52/65	-/-	39/47 -4

The results from the proposed SFC indicate that reconstructed images by SY model consistently have better subjective fidelity than those reconstructed images by Watson's model at various fixed bit rates.

6. Conclusion

In this paper, we propose a subjective fidelity criterion called VLLVCD, which is based on the SPS (Same Position Swapping) technique and VLLCVD (Visually LossLess Critical Viewing Distance) process. The VLLCVD SFC provides two scores S_1 and S_2, that are more strict and precise for evaluating the fidelity of a decoded image. We also demonstrate the application of the proposed VLLCVD FSC in the comparison of visual models.

7. Acknowledgement

This work is supported by Taiwan Government under National Science Council contract NSC-89-2213-E-224-039.

8. References

[1] N.S. Jayant, J.D. Johnston and R.J. Safranek," Signal compression based on models of human perception".

[2] B. Zhu and A.H. Tewfik,"Low bit rate near-transparent image coding" Proceedings of SPIE international conference on wavelet applications for Dual use Vol. 2491, Orlando, FL, 1995, page 173-184.

[3] B. Chitprasert and K.Rao, "Human visual weighted progressive image transmission", IEEE Transactions on communications, vol. COM-38, No. 7, page 1040-1044, July 1990.

[4] Van Ness & Bouman M.A. "Spatial modulation transfer in the human eye", Journal of the Optical Society of America Vol. 57, page 401-406, 1967

[5] E. Legge and J.M. Foley, "Contrast masking in human vision," Journal of Optical Society of American, Vol. 70, No. 12 page, 1458-1471, 1980.

[6] Day-Fann Shen and Loon Shan Yan, "JND measurements and wavelet based image coding", Proceedings of SPIE on input/output and Imaging technologies, vol. 3422-19. July, 1998.

[7] Zen-Fung Tan, "JND-based approach to color palette construction", Proceedings of SPIE on input/output and Imaging technologies, vol. 3422-40. July, 1998.

[8] C.H. Chou and Y.C. Li, "A perceptually tuned subband image coder based on the measure of just noticeable distortion prfile", IEEE Transactions of Circuits and Systems on Video Technologies. Vol. 5, No. 6, pp 467-476, Dec. 1995.

[9] Andrew B. Watson, Gloria Y. Yang, Joshua A. Solomon and John Villasenor, "Visibility of Wavelet Quantization Noise" IEEE Transactions on Image

Processing . Vol. 6, pp. 1164-1175, 1997.

[9-1] A. B. Watson, “DCT Quantization Matrices Visually Optimized For Individual Images”, Proceedings SPIE Conference on Human Vision, Visual Processing and Digital Display VI, 1995.

[9-2] A. B. Watson, J. A. Solomon, A. J. Ahumada, “Visibility of DCT basis functions: Effects of display resolution”, Proceedings Data Compression Conference , Los Alamitos, CA, IEEE computer Society Press, 1994, page 371-379.

[10] B. Girod, “The Information theoretical significance of spatial and temporal masking in video signals” Proceedings SPIE Human Vision, Visual Processing and Digital Display Vol. 1077, 1989, page 178-187.

[11] D.F. Shen and S. C. Wang, “Measurements of JND property of HVS and its applications to image segmentation, coding and requantization”, Proceedings EUROPT/SPIE Vol. 2952, Digital Compression, Technologies and Systems for Video Communications, Berlin, FRG, Oct. 1996.

[12] Christine I. Podilchuk and Wenjun Zeng, “Image Adaptive Watermarking Using Visual Models”, IEEE Journal on Selected Areas in Communications, Vol. 16, No. 4, May 1998. Page 525-539.

[13] Raymond B. Wolfgang, Christine I. Podilchuk and Edward J. Delp, “Perceptual Watermarks for Digital Images and Video”, Proceedings of the IEEE, VOL. 87, No. 7, July, 1999. Pages: 1197– 1207.

[14] S.W. Kim, S. Suthaharan, H.K. Lee and K.R. Rao, “Image atermarking scheme using visual model and BN distribution”, IEE Electronics Letters 4th Feb. 1999 Vol. 35, No. 3, page 212-214.

[15] Y.S. Kim, O.H. Kwon and R. H. Park, “Wavelet Based Watermarking Method For Digital Images Using The Human Visual System”, 1999 IEE Electronic Letters, 18th March 1999, Vol. 35, No. 6. Page 466-468.

[16] Z.H. Wei, P. Qin and Y.Q. Fu, “Perceptual Digital Watermark of Images Using Wavelet Transform”, IEEE Transactions on Consumer Electronics, Volumn. 44, No. 4, November 1998, page 1267-1272.

[17] Anil K. Jain, *Fundamentals of Digital Image Processing.* Englewood Cliffs, New Jersey, Printice-Hall, 1989.

[18] R.C. Gonzalez and R.E. Woods, Digital Image Processing. Addison-Wesley, 1992.

[19] Brain. A. Wandell, “*Foundations of Vision*” Sinauer Associates, Inc, Massachusetts, 1995.