# JND Modeling: Approaches and Applications

*(Invited Paper)*

Zhenzhong Chen and Hongyi Liu
School of Remote Sensing and Information Engineering
Wuhan University, Luoyu Road 129
Wuhan, 430079, China
{zzchen, liuhongyi}@whu.edu.cn

***Abstract*—Just-Noticeable-Difference (JND) measures the smallest detectable difference between two signals therefore can be utilized to quantify the perceivable distortion in the noise contaminated image. In this paper, a comprehensive review of JND models is provided. Specifically, spatial domain and frequency domain JND models are introduced, together with the relevant information about the human visual system. Moreover, the progress from the local JND to global JND is also discussed which aims to exploit the perceptual redundancy in the image. The applications are also discussed in the paper.**

***Index Terms*—Just-Noticeable-Difference; Human Visual System; Foveation;**

## I. Introduction

Nowadays, a great amount of information is conveyed in the form of digital images or videos, which will be presented to human visual system (HVS) after various kinds of processing, including compression, watermarking, transmission, reconstruction, enhancement and so on. The properties of HVS have been extensively studied due to its role in image and video processing. Just-noticeable-difference (JND) refers to the visibility threshold below which any change cannot be detected by the HVS [1]. The determination of JND is sophisticated as it can be affected by multiple factors such as the characteristics of target images, the viewing conditions and even observers' experience and preference. Based on operating domains, existing JND models can be classified into two categories, namely, spatial domain and frequency domain. Compared with the former category, the latter one is relatively well investigated with DCT decomposition [2]–[5]. Recent progress from local JND to global JND have been made in order to further exploit perceptual redundancy in images. The local JND assumes that the visual acuity remains constant over the whole image while global JND takes the foveation effect of HVS into consideration and thus its corresponding thresholds at periphery areas are of higher amplitudes than local JND. The JND model can be widely applied to visual signal processing tasks, since a JND profile provides localized information on perceptual sensitivity to the changes introduced by various processes.

The rest of the paper is organized as follows. In Section II, we present the introduction of HVS and the masking effects. Detailed discussion of JND models is provided in Section III. Section IV presents the applications of JND, and Section V concludes the paper.

## II. Human Visual System

### A. HVS

As the ultimate receiver of visual signal stimulus, the functioning of HVS involves complex and sophisticated physiological and psychophysical mechanisms. The sensitivity of HVS acts towards luminance contrast, spatial and temporal masking effects has been studied in dedicated efforts to design corresponding JND models. It is known the fine-scale variation of visual signal can not be perceived by human due to the psycho-visual properties of HVS, and the visual sensitivity of HVS can be measured by a spatio-temporal contrast sensitivity function (CSF) [6]. Nevertheless, the HVS is not only related to spatio-temporal frequency, but also highly space-variant. Due to the nonuniform distribution of photoreceptor cell on human retina, the human visual system has different visual acuity to perceive image details. According to [7], when the visual stimulus is projected on the fovea, where exists highest density of senors cells, it is perceived at the highest resolution. The density of cone and retinal ganglion cells decreased gradually as the eccentricity increases, which results in descending visual acuity.

### B. Masking Effects

Masking effect is complicated and generally refers to the perceptibility of one signal in the presence of another signal in its spatial, temporal,or spectral vicinity [1]. The perceptual redundancy of images or videos is mainly based on the sensitivity of the HVS due to spatial masking and temporal masking effects. The spatial masking effects can be further classified into luminance masking and contrast masking effects.

According to Weber's law, the perceptible luminance difference of a stimulus depends on the surrounding luminance level, which means the human visual perception is sensitive to luminance contrast rather than the absolute luminance value. This can be explained by the fact that the visibility threshold in very dark or bright areas tends to be higher than thresholds in regions of medium gray levels around 127.

The contrast masking effects reflect the fact that increased spatial non-uniformity of the background luminance will lead to the reduction in the visibility of stimuli. For example, it has been found that HVS can tolerate more noises in textured regions of digital images, where frequent spatial activities lowers the capability of detecting image differences.

978-1-4799-4612-9/14/$31.00 © 2014 IEEE

The temporal masking is related to the change of visual contents over time. It can be added to scale the spatial JND for videos. The sensitivity of HVS to spatial details is lowered after rapid scene change or large temporal difference. Moreover, it can be found that temporal masking due to high-to-low luminance changes is more prominent than that due to low-to-high luminance changes.

## III. JND Models

The JND estimation can be accomplished through modeling the relationship between visual sensitivity and spatial-temporal masking effects. Various computational models have been proposed in both spatial and frequency domains to exploit the perceptual redundancy of HVS. Based on the space-variant properties of HVS, global JND has also been developed to measure the reduction of visual acuity due to increased viewing eccentricity.

### A. Spatial domain JND

Typical spatial domain JND considers two major factors for each pixel in image: luminance masking and contrast masking. Several models have been proposed to respectively measure these masking effects and their combination [8], [9]. In [8], the visibility threshold of luminance masking effects is estimated based on the average luminance of nearby pixels of a certain point. According to the experimental results, the corresponding threshold to low background luminance (less than 127) is modeled by a root equation while the other part is approximated by a linear function. The contrast masking effect is modeled by a function of the background luminance and contrast of a certain pixel. The visibility threshold presents a linear relationship with pixel contrast, whose slope tends to increase slightly as the background luminance increases. Finally, the spatial JND is determined by the dominant factor of the two masking effects.

Yang *et al.* [9] developed a new spatial JND estimator with the nonlinear additivity model for masking (NAMM) to count for the overlapping effect of luminance masking and contrast masking. The proposed model specifies the difference between edge regions and nonedge regions where the masking in edge regions is less significant than in nonedge regions. The visibility threshold is thus higher in regions with both two masking effects compared with the case of an individual masking effect.

Yang's model is further improved by [10] and [11]. In [10], the edge masking(EM) and texture masking(TM) are distinguished and measured separately to provide more precise estimation of contrast masking effect. The input image is decomposed into structural image and textural image for the estimation of EM and TM, respectively. The contrast masking is then represented as the combination of EM and TM. In [11], the disorderly concealment effect is measured to predict higher JND threshold in disorderly region. An auto-regressive model is deduced to predict the orderly contents of input images. The disorderly contents can be obtained as the prediction residual of orderly contents. The JND thresholds for orderly and disorderly contents are then calculated, respectively. Finally, the overall JND threshold is deduced as the combination of the two kinds of JND.

### B. Frequency domain JND

Some JND models measure the visibility threshold in frequency domain such as sub-band [12]–[14], discrete cosine transform(DCT) [2]–[5] and wavelet domains [15]–[18].

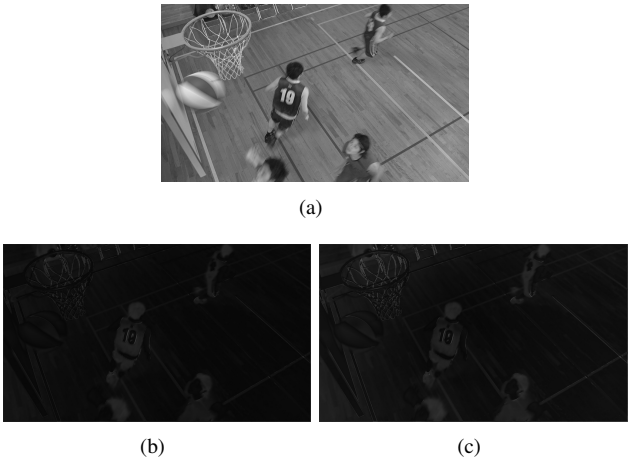
In [12], an input image is passed through a separable Generalized Quadrature Mirror Filter (GQMF) in both horizontal and vertical dimensions, and then decomposed into 16 total sub-bands. For each sub-band, the base sensitivity is first obtained using mid-gray images. The sensitivity adjustments for luminance and contrast maskings are then determined based on perceptual experiments. The final perceptual threshold for each sub-band is the combination of the base sensitivity, luminance and contrast adjustments. In [19], the JND threshold for local noise tolerance in each subband is determined through two main components, namely luminance-adjusted contrast sensitivity and contrast masking adjustment. The first factor is obtained by measuring the base sensitivity and brightness correction adjustment. The contrast masking adjustment is defined as the combination of intraband and interband masking adjustment.

JND can also be measured in DCT domain [2]. The image is first divided into blocks of size $8 \times 8$, and then transformed into DCT. To propose an image-dependent perceptual model, luminance masking threshold matrix for each block is determined. The luminance threshold presents a positive relation with background luminance. In addition, the contrast masking of each block is modeled as an combination of corresponding DCT coefficients and luminance masking threshold. The quantization error is further computed in JND form. Finally, the error in each block is pooled using a Minkowski metric to obtain the visibility threshold.

The visual threshold for samples of uniform quantization noise of a DWT based on linear phrase 9/7 wavelet is measured in [15]. The threshold is detected in Y, Cb and Cr color channels using perceptual experiments in which stimulus are modulations of either channel while two other unmodulated channels are set to zero. A mathematical model is then constructed for detected DWT noise threshold as a combination of transform level, orientation and display visual resolution. In [20], JND thresholds for each DWT subband$(\lambda, \theta)$ are obtained through modeling three visual phenomena : contrast sensitivity, luminance masking and contrast masking. The contrast sensitivity function is built by detection experiments and corresponding parameters are determined. The effect of luminance masking adjustment is approximated using a power function. The contrast masking effect is estimated by modeling the self contrast masking factor and neighborhood contrast masking factor.

### C. From Local JND to Global JND

It is assumed by SJND (Spatial domain JND) that each pixel of the image is projected on the fovea area of HVS,



(a) (b) (c)

Fig. 1. (a) Orginal image. (b) Map of $SJND$ ($\times 4$ for display). (c) Map of $FJND$ ($\times 4$ for display) [7].

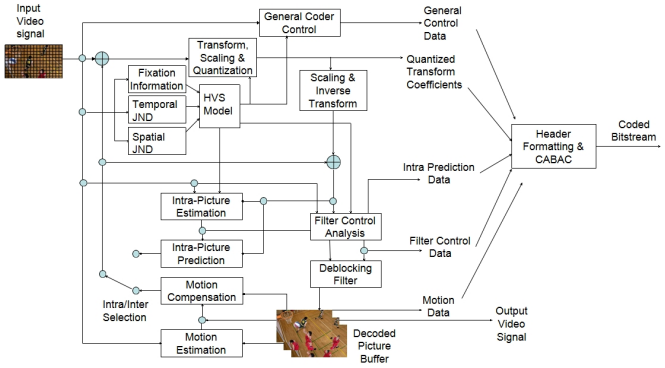



Fig. 2. Video coding frame incorporating JND models.

where it can be perceived under the highest visual acuity. However, the visual acuity will decrease when visual stimulus is projected out of the fovea region. Thus the SJND could only provide a local visibility threshold. In [7], a global JND (FJND) is proposed to measure the global visibility threshold of the whole image by incorporating the viewing eccentricity to SJND model. A Foveated model $F$ is developed based on the experiment which obtains the visibility threshold due to the foveation property of the HVS. The $F$ is both background luminance and eccentricity dependent, and it tends to increase as the eccentricity increases. Considering the fact that different visual content of images will lead to different fixation points, saliency models can be implemented to obtain fixation points by locating the conspicuous visual area of an image. Given the viewing distance and fixation points, $F$ can be calculated by simply considering the nearest fixation point which results in the smallest viewing eccentricity and the minimum foveation weight $F_i$. Then FJND model is formulated by combining the SJND model with the foveation model $F$.

Fig. 1 compares the differences between SJND and FJND. Compared with SJND (1(b)), FJND (1(c)) injects more noise at the periphery area of the image. This is due to the fact that those area is projected outside the fovea region and the visual acuity decreases accordingly, hence more noise could be tolerated.

## IV. JND Applications

JND measures the smallest detectable difference between two signals and it can be applied in watermarking, video coding and streaming by exploiting the visibility of the noise in the image.

### A. Watermarking / Data Hiding

The aim of watermarking and data hiding is to covertly embed some information in noise-tolerant signal, such as the ownership protection, authentication, access control and annotation [21], [22]. An effective watermarking scheme should be robust and transparent, which means the watermark will not be destroyed or transformed by editing or illegal tampering, while remain imperceptible to human visual system. JND models can be directly extended to the watermarking application by providing upper bounds on watermarking intensity level, which guarantee transparency while providing a robust watermark. It can also be utilized to determining the maximum number of watermarks that can be applied to a certain image with low probability error [21].

In [21], two JND-based watermarking techniques are presented: one where the frequency decomposition consists of block-based DCT (IA-DCT scheme) and one where the frequency decomposition consists of a wavelet decomposition (IA-W scheme). For the former one, images are decomposed into CT domain, and JND for each domain is computed using DCT coefficient, frequency sensitivity, luminance sensitivity and contrast masking component [23]. The watermarked DCT coefficient is obtained through DCT coefficient and weight sum of the sequence of watermark value and computed JND. The latter one follows the similar process. Given the wavelet-decomposed image, the watermarked wavelet coefficient is calculated based on wavelet coefficient, watermark sequence and the corresponding frequency weight.

### B. Image / Video Coding

The purpose of image(video) coding is to achieve highest perceptual quality at a given bit-rate, since digital signals will be ultimately processed by the human visual system. Since JND represents the minimum difference below which the change is imperceptible to HVS, the threshold could be used to determine optimum quantization step sizes or bit allocation for different parts of the image as determined by a model of the HVS and local image characteristics [7], [24]–[27].

In [9], a perceptually adaptive preprocessor for motion-compensated residues in video encoder has been proposed. In this method, a JND estimator is devised using nonlinear additivity model for masking(NAMM). The JND profile is then incorporated into a residue signal preprocessor to adjust motion compensated residue. The proposed method improves both objective coding quality (PSNR), and perceptual quality of the decoded images for a given bit rate.

The HVS is also space-variant as the retina in a human eye does have a nonuniform density of photoreceptor cells. As the density of the cone and retinal ganglion cells drops with the increased retinal eccentricity, the visual acuity decreases with increased eccentricity as well. This foveation property of HVS is measured by FJND, which is utilized by foveated video coding algorithms for MB quantization adjustments. Smaller quantization parameter is used where noticeable distortion weight is large, which is a function of FJND value in each MB. This algorithms can achieve better perceptual quality of the reconstructed video at the given bit rate.

### C. Image / Video Streaming

Researchers have focused their attention on image(video) streaming to protect transmitted data from channel errors [28]–[33]. A perceptually error-resilient method is proposed in [28] for video steaming based on the characteristics of HVS. A JND is employed to estimate the perceptual loss impact due to error propagation for allocating intra-refreshed macroblocks in the transcoded video. The computation of the perceptual loss impact is based on the expected distortion value that exceeds the JND threshold instead of the mean square error. Experimental results demonstrates the better performance of the proposed method in terms of visual quality of the received video from the end users.

## V. Conclusion

In this paper, we present a comprehensive review of JND models. Spatial domain and frequency domain JND models are specifically introduced, together with the relevant information about the human visual system. We also introduce the progress from local JND to global JND, which aims to exploit the perceptual redundancy in the image. Finally, the applications of JND models are discussed.

## References


[1] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE.*, vol. 81, no. 10, pp. 1385–1422, 1993.

[2] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," in *Soc. for Info. Display Dig. Tech. Papers.* Society for information display, 1993, vol. 24, pp. 946–946.

[3] X. H. Zhang, W. S. Lin, and P Xue, "A new dct-based just-noticeable distortion estimator for images," in *ICICS-PCM.* IEEE, 2003, vol. 1, pp. 287–291.

[4] H. A. Peterson, "DCT basis function visibility in RGB space," *Soc. for Info. Display Dig. Tech. Papers.*, 1992.

[5] H. A. Peterson, H. Peng, J. H. Morgan, and W. B. Pennebaker, "Quantization of color image components in the DCT domain," International Society for Optics and Photonics, 1991, pp. 210–222.

[6] D. H. Kelly, "Motion and vision. ii. stabilized spatio-temporal threshold surface," *JOSA.*, vol. 69, no. 10, pp. 1340–1349, 1979.

[7] Z. Chen and C. Guillemot, "Perceptually-friendly H. 264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 20, no. 6, pp. 806–819, 2010.

[8] C.-H. Chou and Y.-C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 5, no. 6, pp. 467–476, 1995.

[9] X. Yang, W. S. Lin, Z. Lu, E.-P. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 15, no. 6, pp. 742–752, 2005.

[10] A. M. Liu, W. S. Lin, M. Paul, C. W. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 20, no. 11, pp. 1648–1652, 2010.

[11] J. J. Wu, G. M. Shi, W. S. Lin, A. M. Liu, and F. Qi, "Just noticeable difference estimation for images with free-energy principle," *IEEE Trans. Multimedia.*, 2013.

[12] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 1989, pp. 1945–1948.

[13] Y. Jia, W. S. Lin, and A. A. Kassim, "Estimating just-noticeable distortion for video," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 16, no. 7, pp. 820–829, 2006.

[14] X. H. Zhang, W. S. Lin, and P Xue, "Improved estimation for just-noticeable visual distortion," *Signal Process.*, vol. 85, no. 4, pp. 795–808, 2005.

[15] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1164–1175, 1997.

[16] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2091–2110, 1989.

[17] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 45, no. 5, pp. 485–560, 1992.

[18] Z. Wei and K. N. Ngan, "Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 19, no. 3, pp. 337–346, 2009.

[19] I. Hontsch and L. J. Karam, "Locally adaptive perceptual image coding," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1472–1483, 2000.

[20] Z. Liu, L. J. Karam, and A. B. Watson, "Jpeg2000 encoding with perceptual distortion control," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1763–1778, 2006.

[21] C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Trans. Select. Areas Commun.*, vol. 16, no. 4, pp. 525–539, 1998.

[22] M. Wu and B. Liu, "Data hiding in image and video. I. Fundamental issues and solutions," *IEEE Trans. Image Process.*, vol. 12, no. 6, pp. 685–695, 2003.

[23] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE.* International Society for Optics and Photonics, 1993, pp. 202–216.

[24] Z. Chen, W. S Lin, and K. N. Ngan, "Perceptual video coding: challenges and approaches," in *Proc. IEEE ICME.* IEEE, 2010, pp. 784–789.

[25] X. K. Yang, W. S. Lin, Z. K. Lu, E. P. Ong, and S. S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Process.:Image Commun.*, vol. 20, no. 7, pp. 662–680, 2005.

[26] W. S. Geisler and J. S. Perry, "Real-time foveated multiresolution system for low-bandwidth video communication," in *Proc. SPIE.* International Society for Optics and Photonics, 1998, pp. 294–305.

[27] C.-W. Tang, C.-H. Chen, Y.-H. Yu, and C.-J. Tsai, "Visual sensitivity guided bit allocation for video coding," *IEEE Trans. Multimedia.*, vol. 8, no. 1, pp. 11–18, 2006.

[28] V.-A. Nguyen, Z. Chen, and Y.-P. Tan, "Perceptually optimized error resilient transcoding using attention-based intra refresh," in *Proc. IEEE ISCAS.* IEEE, 2010, pp. 4217–4220.

[29] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Mag.*, vol. 17, no. 4, pp. 61–82, 2000.

[30] H.-J. Chiou, Y.-R. Lee, and C.-W. Lin, "Content-aware error-resilient transcoding using prioritized intra-refresh for video streaming," *J. Visual Commun. Image Represent.*, vol. 16, no. 4, pp. 563–588, 2005.

[31] J. Y. Liao and J. Villasenor, "Adaptive intra block update for robust transmission of H. 263," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 10, no. 1, pp. 30–35, 2000.

[32] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 966–976, 2000.

[33] Y. Eisenberg, F. Zhai, T. N. Pappas, R Berry, and A. K. Katsaggelos, "VAPOR: Variance-aware per-pixel optimal resource allocation," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 289–299, 2006.