J. Vis. Commun. Image R. 98 (2024) 104034

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Review

A survey on just noticeable distortion estimation and its applications in video coding[☆,☆☆]

Guoxiang Wang [a], Hongkui Wang [b,c,*], Hui Li [e], Li Yu [d], Haibing Yin [b,c], Haifeng Xu [b,c], Zhen Ye [a], Junfeng Song [a]

[a] Lishui University, Lishui, China
[b] School of Communication Engineering, Hangzhou Dianzi University, China
[c] Lishui Institute of Hangzhou Dianzi University, China
[d] College of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China
[e] Zhejiang Dahua Technology co., LTD, Hangzhou, China

ARTICLE INFO

Keywords:
JND estimation
JND-based perceptual video coding
Human visual system

ABSTRACT

With the developing explosion in video data delivery, perceptual video coding (PVC) plays an increasingly significant role in video compression. The just noticeable distortion (JND) reflects the tolerance limit of human visual system (HVS) to coding distortion directly, resulting in that JND-based PVC is the most important branch of video coding. This paper provides an extensive overview of JND estimation and JND-based PVC, so as to make interested readers aware of the status quo. The main contribution of this article can be briefly outlined as follows. Firstly, the general description of JND concepts and most existing computation models for JND are to be reviewed systematically. Secondly, most related works about JND-based perceptual image and video coding schemes are introduced, including JND-based coding preprocessing and JND-based codec embedding. Thirdly, in addition to a thorough summary of JND estimation and JND-based PVC, possible future directions and opportunities are analyzed and discussed.

1. Introduction

With the increasing demand for high-quality and high-throughput video services, video data is exploding towards multi-dimensional development, and the storage and transmission of video signals are facing tremendous pressure. In order to achieve rapid transmission and efficient storage of video data, video compression technology continues to innovate and more coding standards have been released, such as the High Efficiency Video Coding (HEVC) [1] and the Versatile Video Coding (VVC) [2]. However, existing coding standards based on rate–distortion optimization strategies remove spatial and temporal redundancy in an almost exhaustive manner, and the video compression performance is nearing optimal.

Actually, as the final receiver of videos, the human visual system (HVS) cannot perceive small variation in pixel difference, which indicates that video data could be further compressed by eliminating perceptual redundancy of videos [3–5]. Further research indicates that HVS can only detect a limited number of subjective quality levels for a series of compressed videos with different quality [6,7]. Thus, perceptual optimization of video coding may provide a choice for the best video compression scheme [8,9]. In fact, the visual sensitivity is always affected by the video content, and the distortion below a certain threshold cannot be detected by the HVS. The certain distortion threshold is the so-called just noticeable distortion (JND) [10]. Since JND threshold directly reflects the perceived critical noise of HVS, JND estimation has been deeply studied and widely introduced into perceptual optimization of video coding [11,12].

In the past twenty years, scholars have presented numerous JND estimation models which were mainly constructed by employing multiple visual masking effects with the statistical modeling approach [13–17]. The visual masking effect is evoked by interaction and interference among visual stimuli [10]. According to different application scenarios, existing early JND models could be mainly divided into two types,

☆ This work was supported in part by the Pioneer and Leading Goose R and D Program of Zhejiang Province under Grant 2023C01149 and 2022C01068; in part by the NSFC under Grant 62202134, 62031009 and 61972123; in part by the ZJNSF under Grant LDT23F01014F01.
☆☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail addresses: wanggx_bupt@163.com (G. Wang), wanghk@hdu.edu.cn (H. Wang), li_hui@dahuatech.com (H. Li), hustlyu@hust.edu.cn (L. Yu), haibingyin@163.com (H. Yin), xuhaifeng@hdu.edu.cn (H. Xu), yezhen@lsu.edu.cn (Z. Ye), dachan@126.com (J. Song).

https://doi.org/10.1016/j.jvcir.2023.104034
Received 26 May 2023; Received in revised form 11 November 2023; Accepted 20 December 2023
Available online 4 January 2024
1047-3203/

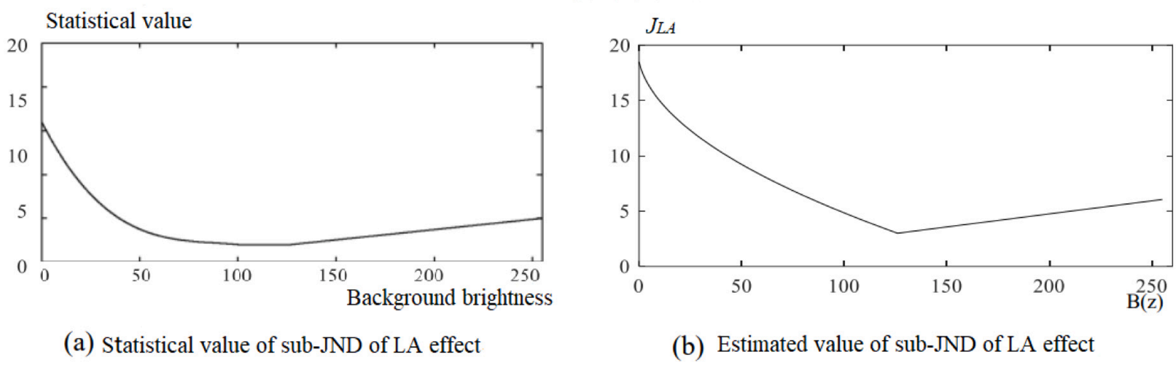


Fig. 1. Sub-JND threshold of LA effect.

i.e., pixel-wise JND models and subband-wise JND models. In the pixel-wise JND estimation, the inferences of various visual masking effects to HVS visual sensitivity are usually analyzed independently, and the masked distortion by each masking effect is determined according to the visual perception experiments. The sub-JND model of each visual masking effect is usually built according to the experimental data with the statistical modeling approach. The total JND model is usually obtained by combining all sub-JND models with a non-linear additivity model [13]. As for the subband-wise JND model, it mainly refers to the DCT-wise JND models because most video data is compressed in the DCT domain. By comparison, only a few JND models are built in other subband domains, such as in the Wavelet domain [14–16]. In the DCT domain, the visual sensitivity of HVS has the band-pass property with the spatial frequency [17]. This property is usually described with the spatial contrast sensitivity function (CSF) and its inference to the HVS is modeled as a base JND threshold. The inferences of other masking effects are usually modeled as multiple masking factors. The total subband-based JND model could be expressed as the product of the base threshold and multiple masking factors. In addition to pixel-wise JND and subband-wise JND, scholars have explored block-wise [18], picture-wise [19,20] and video-wise JND estimation [21,22] in recent years. Most of these models are built using machine learning and deep learning approaches. In order to obtain sufficient training samples, some compressed image or video datasets have been released successively [6,7].

The JND threshold reflects the tolerance limit of HVS to coding distortion directly, resulting in that JND-based image/video compression is one of the main target applications in JND-based image/video perceptual processing. Vast literatures have shown that JND-based perceptual image/video coding schemes perform promising bitrate reduction at similar perceptual visual quality compared to traditional video coding [23]. The existing JND-based perceptual image/video coding schemes could be classified into two categories, i.e., JND-based coding preprocessing and JND-based codec embedding. In JND-based coding preprocessing, perceptual redundancy is estimated in advance according to the JND model. On the premise of similar perceptual quality, the original image or video will be smoothed with a JND-based filter to remove perceptual redundancy. Instead, JND-based codec embedding introduces the JND estimation process into image/video coding process directly, and greatly saves the bitrate through perceptual optimization for different coding modules.

In recent years, scholars have introduced many works of JND modeling and perceptual video coding from different perspectives. For example, Zhang et al. analyzed physiological and psychological perceptual redundancies [24]. Yuan et al. mainly focused on visual JND modeling, briefly introduced existing JND datasets, JND modeling process, JND application in perceptual quality assessment and video coding [25]. Lin et al. introduced computational models for visual JND and briefly discussed JNDs for audio, smell, haptics and gustatory signals, as well as cross-modality/media [12].

In this article, we aim to provide a comprehensive review of the reports about JND estimation and its applications in perceptual video coding, so as to make interested readers aware of the status quo. The main contribution of this article can be briefly outlined as follows. Firstly, the general description of JND concepts and most existing computation models for JND are to be reviewed systematically. Secondly, most related works about JND-based perceptual image and video coding schemes are introduced, including JND-based coding preprocessing and JND-based codec embedding. Thirdly, in addition to a thorough summary of JND estimation and JND-based PVC, possible future directions and opportunities are analyzed and discussed.

The remainder of this paper is organized as follows. Sections 2 and 3 provide reviews of JND estimation and JND-based perceptual video coding. Section 4 discusses future directions and opportunities. Section 5 concludes this paper finally.

2. Review of JND estimation for image and video

2.1. Pixel-wise JND estimation

In the research of pixel-wise JND estimation, scholars independently analyzed the impact of various visual masking effects on HVS visual sensitivity, and determined the distortion intensity that different masking effects can mask according to the biological vision experiments. The total JND threshold is generally predicted by nonlinear superposition of each sub-JND threshold. The early pixel-wise JND models mainly considered the effects of luminance adaptation (LA) and contrast masking (CM). According to Weber–Fechner's law [26], the sensitivity of the human visual system to the signal is inversely proportional to the strength of the background signal. The greater the strength of the background signal, the lower the visual sensitivity. Therefore, for digital images, HVS is more sensitive to noise in gray background, but not to noise in brighter and darker backgrounds. This visual perception effect is called the LA effect. In general, the sub-JND determined by the LA effect (J_{LA}) can be predicted using a 'U' segmented function as shown in Fig. 1 [27].

The CM effect reflects the weakening effect of other signals on the current signal visibility. For example, in the area with complex texture, due to the complexity of the scene, other visual stimuli often reduce the visibility of the target excitation, resulting in that the noise (or the distortion) is difficult to be noticed. Therefore, the JND threshold determined by the CM effect is mostly predicted based on the texture complexity. In the pixel domain, the texture complexity is usually expressed by the maximum brightness difference of the background. In [28], the maximum output of the convolution of the background area of the current pixel and the four high-pass filters represents the texture complexity of the current pixel. The four high-pass filters are shown in Fig. 2. Because of the selective response mechanism of the visual nerve, the visual sensitivity of the edge region is much higher than that of the

$$g_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 8 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -3 & -8 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad g_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 8 & 3 & 0 & 0 \\ 1 & 3 & 0 & -3 & -1 \\ 0 & 0 & -3 & -8 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \quad g_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 8 & 0 \\ -1 & -3 & 0 & 3 & 1 \\ 0 & -8 & -3 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \quad g_4 = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 8 & 0 & -8 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix}$$

Fig. 2. High-pass filters in four directions.

texture region. Therefore, Liu et al. [29] divided the image into a structural component and a texture component through a total variation model (TV), and divided the contrast masking effect into edge masking effect and texture masking effect. The JND threshold determined by the edge masking effect and the texture masking effect is predicted on the structural component and texture component respectively, and finally the sub-JND of CM effect (J_{CM}) is expressed as the sum of sub-JND of edge masking and sub-JND of texture masking. However, the interaction between LA and CM effects is not considered in this model, so it is easy to cause the overestimation issue of JND estimation. In order to improve the accuracy of pixel-wise JND estimation, Yang et al. [13] designed a nonlinear additive model for masking (NAMM) to reduce the overlap effect caused by two masking effects. Specifically, the NAMM model is expressed as follows.

$$J_S = J_{LA} + J_{CM} - C^{gr} \cdot min(J_{LA}, J_{CM}) \tag{1}$$

J_S is the pixel-wise JND threshold. C^{gr} is the reduction factor and is set to 0.3 generally [13].

In fact, visual perception is a derivation of HVS to external visual stimulation through the internal generation mechanism (IGM). Friston et al. [30] proposed a free-energy principle to simulate the IGM in the human brain, and tried to make a unified explanation for human behavior and perception, etc. The free-energy principle indicates that the HVS extracts as much visual information as possible in the process of visual perception to minimize the uncertainty of the input signal. This principle reflects that the HVS tends to process the structural information in the visual signal while ignoring the specific details. This visual perception characteristic is called the disorderly concept masking (DCM) effect [31] or the entropy masking (EM) effect [32]. In [31], Wu et al. designed an autoregressive (AR) prediction model to simulate the IGM of HVS, and defined the difference between the original signal and the predicted signal as the visual free-energy. According to this AR model, the sub-JND of EM effect (J_{EM}) is estimated according to the absolute value of the prediction residual. Then, Wu et al. integrated the CM effect and the EM effect into the pattern masking effect (PM), and proposed a more accurate pixel-wise JND estimation model in [33].

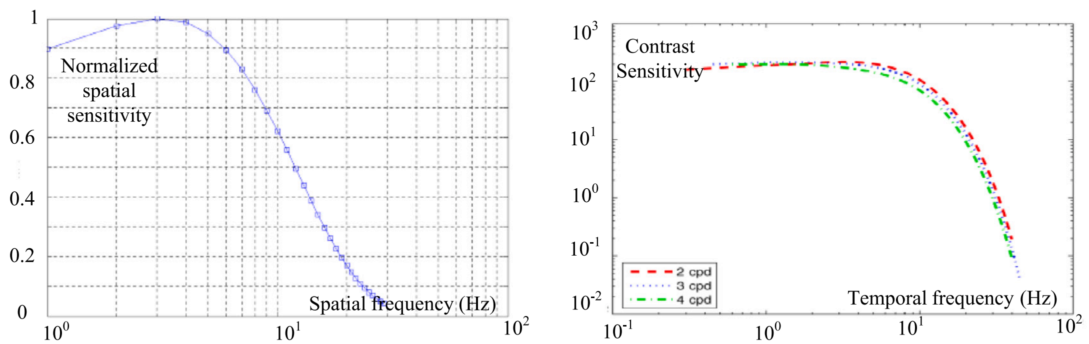
Due to the non-uniform distribution of photoreceptors in the retina of the human eye, the visual sensitivity near the fixation point of the human eye is much higher than that in other areas, and it decreases with the increase of the distance from the fixation point. Therefore, in the process of visual perception, the JND threshold near the fixation point is far lower than that of other regions. This visual masking effect is generally called the foveated masking (FM) effect [28]. In order to improve the visual perception efficiency, the human eye often finds the regions of interest (ROIs) in the image by scanning process and looks at these ROIs to obtain more interesting visual information. Thus, the JND threshold of the ROI is generally lower than that of the non-interest region. The fixation point of the human eye can be estimated according to the visual saliency, which to some extent reflects the degree of interest of the observer in the visual content. Therefore, in recent years, many scholars have introduced the foveated masking effect and the influence of saliency on visual sensitivity into the prediction of JND through saliency detection technology [28,34,35].

In [36,37], Wang et al. analyzed the visual perception process from the perspective of visual nerve response characteristics and believed that the perception process of HVS is the result of the combined effect of positive and negative perception effects. According to the effective coding theory [38] and the hierarchical prediction coding theory [39], the residual signal of visual excitation is the effective input of HVS. The residual signal, on the one hand, induces the response of visual neurons and the visual positive perception effect. On the other hand, the residual signal of the neighborhood will inhibit the expression of the neurons to the current stimulus, thus triggering the negative visual perception effect, namely the visual masking effect proposed previously. Therefore, Wang et al. [36,37] modeled the HVS as a hierarchical perceptual system, defined perceptual self-information and average self-information from the perspective of information theory to measure positive and negative perceptual effects respectively, and introduced them into the estimation of pixel-wise JND. The final JND was expressed with a nonlinear weighted superposition model with sub-JND thresholds and other perceptual adjustment factors. Recently, they analyzed the feedforward and feedback modulatory behaviors in the HVS, and summarized visual perception effects as modulatory effects and masking effects. They incorporated the modulatory mechanism into JND estimation and proposed a hierarchical modulation-based JND estimation framework in [35]. For video signals, the temporal masking (TM) effect [28] should also be considered in pixel-wise JND estimation. In general, human visual sensitivity decreases with the increase of inter-frame difference. Chen et al. [28] constructed a TM adjustment factor based on the video inter-frame residual. The pixel-wise JND model of video could be expressed as the product of the image-JND model and the TM adjustment factor.

2.2. Subband-wise JND estimation

Visual sensitivity presents a band-pass characteristic for visual signals of different frequencies. The relationship between visual sensitivity and spatial frequency is shown in Fig. 3(a) [40]. This visual characteristic is generally described by contrast sensitivity function (CSF). In addition, under the same spatial frequency, Kelly et al. found that the contrast sensitivity is nearly constant for the temporal frequencies less than 10 Hz as shown in Fig. 3(b) [41]. Specifically, this figure provided three temporal CSF curves in the case that the spatial frequency is set to 2 cycle/degree (cpd), 3 cpd and 4 cpd respectively. Because most of image/video coding algorithms are based on discrete cosine transform (DCT), most of subband-based JND models are proposed in the DCT domain. Based on the CSF effect, Ahumada and Peterson [42] established the first DCT-based JND model using the contrast sensitivity function to predict the JND threshold of each DCT component. Then, Watson et al. introduced the LA and CM effects into DCT domain JND estimation, and proposed the DCTune JND model [43]. In this model, the influence of CSF effect on visual sensitivity was modeled as the basic threshold of JND (J_{base}), the LA and CM effects were modeled as two adjustment factors (F_{LA} and F_{CM}). The final DCT-based JND model was represented by the product of the basic threshold and two adjustment factors. In addition to the research on JND estimation in DCT domain, some scholars have explored the JND estimation method in the wavelet transform domain and introduced it into image perceptual compression in JPEG2000 [15,16,44].

For 8-bit images, the LA effect indicates that the visual sensitivity of pixel values around 128 will be higher than that of other pixel values. Thus, the LA adjustment factor in the frequency domain can be expressed as the ratio of the JND of other pixels to the JND of 128. There is a nonlinear relationship between the brightness value and the pixel value in the display process. Wei et al. [45] corrected this non-linear relationship through gamma correction, and established a segmented 'U' model to estimate the LA adjustment factor. In image/video compression, video coding takes the image block as the basic processing unit. The contrast masking adjustment factor needs to analyze the texture complexity of the current image block during the modeling process. Wei et al. carried out edge detection of spatial image based on the Canny operator. By calculating the edge pixel density of each image block,



(a) Spatial CSF curve (b) Spatial-temporal visual sensitivity (cpd:cycle/degree)

Fig. 3. Relationship between visual sensitivity and frequency.

the image block was divided into three types: flat, edge and texture. Wang et al. [46,47] divided image blocks into flat, texture and contour by calculating block-level average absolute difference. To avoid cross-domain operations, Zhang et al. [48] also classified image blocks into flat, edge and texture in DCT domain by calculating the texture energy (TE) of each image block. Wen et al. [49] introduced the EM effect into the DCT-JND estimation. In Wen's contrast masking adjustment factor modeling, according to the complexity and order of texture in the image block, the image block was divided into five categories: flat, ordered edge, disordered edge, ordered texture and disordered texture. Based on the classification of image block, the CM factor is generally modeled as several constants, resulting in the discontinuity of JND estimation and affecting the accuracy of JND estimation. To avoid this defect, Bae et al. proposed the structural contrast index (SCI) measurement method to measure the structure of image blocks, and proposed a series of DCT-JND estimation models [50–52]. Wang et al. analyzed the distribution of pixel values in image blocks through perceptual LBP algorithm, proposed a block-level disorder measurement method to measure the texture disorder intensity of each image block, and introduced the EM effect into the DCT-JND modeling [47].

In the current mainstream coding standards, variable block coding method is widely used, i.e., the size of image processing unit is variable. For example, the size of DCT block is 4×4, 8×8, 16×16 and 32×32 in HEVC. Thus, scholars have begun to explore variable-block JND estimation models in recent years [50,52]. Ma et al. [53] proposed a JND estimation model for 8×8 and 16×16 image block. Bae et al. proposed a DCT-JND model suitable for 4×4 to 32×32 image blocks [50]. However, although scholars have explored the JND estimation models of different size blocks, it is still difficult to accurately estimate the JND in frequency domain for different sizes and different transform coefficient components. Moreover, the latest coding standards (VVC and AVS3) support non-square transformation, which brings new challenges to frequency domain JND estimation. The existing frequency domain JND estimation model is not fully compatible with the latest coding standards.

For video signals, the JND threshold in the frequency domain is not only related to the spatial frequency, but also to the temporal frequency. Jia et al. [54] and Wei et al. [45] deduced the expression of temporal frequency based on the spatial frequency and the motion information in the video signal. The contrast sensitivity under different spatial and temporal frequency combinations is shown in Fig. 3 (b). Based on the results of vision experiments, Wei et al. [55] proposed a TM adjustment factor in DCT-JND estimation. In recent years, some scholars have begun to explore the comprehensive masking effect of different visual effects. For example, Bae et al. [52] combined the FM effect and the TM masking effect, and proposed a foveated DCT-based JND estimation model suitable for video signals. In general, the DCT-JND estimation model of video signals is usually described by a combination of a basic threshold determined by CSF and multiple visual masking effect adjustment factors, such as the LA factor, the CM factor and the TM factor, etc. Thus, the subband JND model could be formulated with the multiplicative form as follows [12].

$$J_F = J_{base} \prod_{i=1}^{n} F_i \tag{2}$$

J_F is the subband-wise JND threshold. J_{base} is the base visibility and F_i denotes other masking factors, such as F_{LA} and F_{CM}.

2.3. Block-wise, picture-wise and video-wise JND estimation

The latest in JND estimation research reveals that human only perceive discrete-scale quality levels over a wide range of coding birates [6,7]. As shown in Fig. 4, for a series of images and videos compressed by JPEG and HEVC, human only perceive limited qualities of these compressed images and videos. In this figure, the JND point of each quality is measured by the quality factor (QF) for compressed images and the quantization parameter (QP) for compressed video.

Compressed Image and Video Databases: As shown in Table 1, some compressed image or video databases have been released in recent years. In 2015, Lin et al. [56] compressed five images and five videos using JPEG and x264/x265, released the first JND points database. Then, they successively released MCL-JCI [57] image database and MCL-JCV video database [7] which have 50 and 30 original images and videos respectively. Huang et al. [58] compressed 40 videos using reference software of HEVC for the first time in 2017. At the same time, one large-scale compressed video database had been published in which 220 original videos generate 880 videos by downsampling [59]. In 2019, Fan et al. released the compressed stereoscopic images and videos using JPEG2000 and HM 16.7 [60]. In the same year, Liu et al. [63] created a panoramic compressed image database in which 40 source images were compressed with JPEG. Recently, with the release of the latest coding standard VVC, the video databases compressed by VTM have been published [62] in 2020. In this database, 202 source images are compressed by VTM-5.0 with 39 QPs. In order to increase the number of participants and compressed images or videos, Lin et al. [63] compressed 1008 original images using JPEG and BPG, released the KonJND-1k database in 2022.

Block-wise JND Estimation: Ki et al. initially explored the block-wise just noticeable quantization distortion (JNCD) estimation using the linear regression method and the CNN network in the DCT domain [18]. In [62,64], Wang et al. and Zhang et al. built block-wise JND models through local pooling and CNN network. Shen et al. [65] designed a block-wise JND model based on structural visibility. Tao et al. [66] designed a block-wise JND estimation model using the OTSU method and incorporated it into JPEG for image compression.

Picture-wise JND Estimation: Based on MCL-JCI dataset [57], Tian et al. designed a CNN model to predict the picture-wise JND label and applied the support vector regression to determine the number

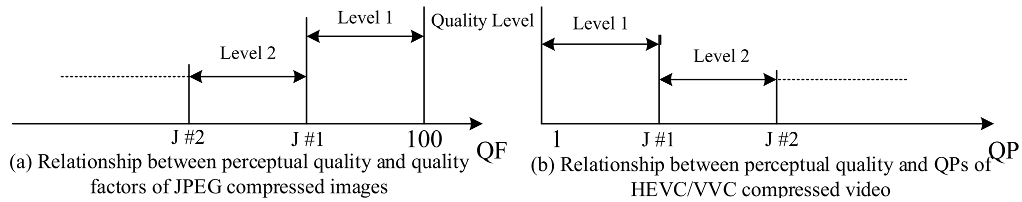


Fig. 4. Relationship between perceptual quality and compression parameter.

Table 1
Compressed image and video databases.

Database	Type	Original simples	Resolution ratio	QF/QP/VBR	Encoder	Compressed simples	Testers	Time
[56]	Image	5	1920 × 1080	QF 1–100	JPEG	5 × 100	20	2015
	Video	5	1920 × 1080	QP:1-51/VBR	x264/x265	5 × 51 × 2 × 2	20	2015
MCL-JCI [57]	Image	50	1920 × 1080	QF:1–100	JPEG	50 × 100	20	2015
MCL-JCV [7]	Video	30	1920 × 1080	QP:1–51	x264	30 × 51	50	2016
[58]	Video	40	1920 × 1080	QP:1–51	HM16.0-RA	40 × 51	30	2017
VideoSet [59]	Video	220 × 4 = 880	1920 × 1080 1280 × 720 960 × 540 640 × 360	QP:1-51	H.264/AVC	880 × 51	30+	2017
SIAT-JSSI [60]	Stereoscopic Image	12	1920 × 1080	QF:1–300	JPEG2000	7020	50	2019
SIAT-JASI [60]	Stereoscopic Image	12	1920 × 1080	QP:1–51	HM16.7-AI	7020	50	2019
JND-Pano [61]	Panoramic image	40	5000 × 5000	QF:1–100	JPEG	4040	24	2019
[62]	Image	202	1920 × 1080	QP:13–51	VTM5.0-AI	39 × 202	20	2020
KonJND-1k [63]	Image	1008	640 × 480	QF:1–100	JPEG or BPG	77 112	42	2022

of picture-wise JND levels for images. Liu et al. [20] formulated the task of predicting picture-wise JND as the binary classification problem and constructed a deep learning based binary classifier to predict the JND for images. The Satisfied User Ratio (SUR) curve for a lossy image compression scheme characterizes the probability distribution of the JND level. Thus, in [67–69], the picture-wise JND estimation is transformed to the prediction problem of SUR curve for image.

Video-wise JND Estimation: Huang et al. found that JND points of first quality level follow a normal distribution and proposed a video-wise JND model using the machine learning approach [58]. Based on MCL-JCV database, we analyzed JND points of the first three quality levels and proposed a multi-level JND estimation model using statistical approach [62]. Similar with [67–69], in [70–72], the JND prediction problem of video is transformed to the prediction problem of SUR curve. Bondžulić et al. [73] found the correlation between mean gradient magnitude and the PSNR of the first JND point, proposed a JND prediction method in which the JND point is measured by the PSNR value.

2.4. JND estimation for other videos

With the increasing demand for videos, other application videos have flourished in recent years, such as screen content videos, 3D videos and panoramic videos. Compared with traditional 2D natural video, these application videos have their unique characteristics. In order to improve the compression efficiency of these application videos, the academia and industry have begun to explore the modeling methods of JND models for these application videos recently.

JND Estimation of Screen Content Video: Screen content video is a mixture of computer-generated images, documents, images and videos captured by cameras. Compared with natural images/videos, screen content video is often characterized by sharp edges, high-purity colors and strong contrast [74]. MPEG of ISO/IEC and VCEG of ITU-T began the development of screen content coding standards in 2014, and formally released HEVC extended standard for screen content coding in December 2016 (i.e, HEVC-SCC). For the sharp edge of screen content, Wang et al. [75] adopted a local parameter edge model to describe the JND threshold of edge distortion, and adjusted the parameter selection of LA-JND estimation. HEVC-SCC supports video input in RGB and YCbCr color formats at the same time. Prangell et al. [76] proposed a JND model for Y, Cb and Cr components, which was used for screen content video compression with perceptual adaptive quantization.

JND Estimation of 3D Video: Compared with 2D video, the biggest difference of 3D video is that it contains depth information. Thus, a series of JND estimation models consisting of texture-based JND model and depth-based JND model are proposed. De Sliva et al. proposed a JND prediction model for depth images [77]. Xue et al. introduced the saliency of depth image into the JND estimation and proposed a Disparity JND model [78]. Zhong et al. [79] presented a Hybrid JND model considering geometric distortion in depth image. Lian et al. [80] proposed a multi-view JND model which is applicable to 3D video compression. For stereo video, Zhao et al. [81] built a single view JND model . Qi et al. [82] proposed a JND model considering left and right viewpoints. Li et al. [83] presented JND models for occluded and non-occluded regions respectively. The effect of parallax on visual sensitivity is considered in the JND estimation of occluded areas.

JND Estimation of Panorama Video: With the development of acquisition and display system technology in recent years, the demand for immersive panoramic image and video content is increasing. In panoramic image/video processing, images and videos often convert spherical images into 2D images at different frequencies for processing according to different projection formats, such as rectangular projection (ERP), cube map (CMP), and truncated square pyramid projection (TSP). The JND estimation of panoramic image/video is not only related to image content, but also closely related to the projection mode of image/video.

Despite the gradual rise of panoramic video applications in recent years, there are few panoramic images and video data sets for open research. Therefore, there is not enough training data to meet the training needs of JND models. Secondly, there are many kinds of projection methods for panoramic video, distortion such as image boundary and image deformation will be introduced in the process of projection [84]. Therefore, in order to improve the accuracy of JND estimation for panoramic video, more panoramic image or video datasets should be

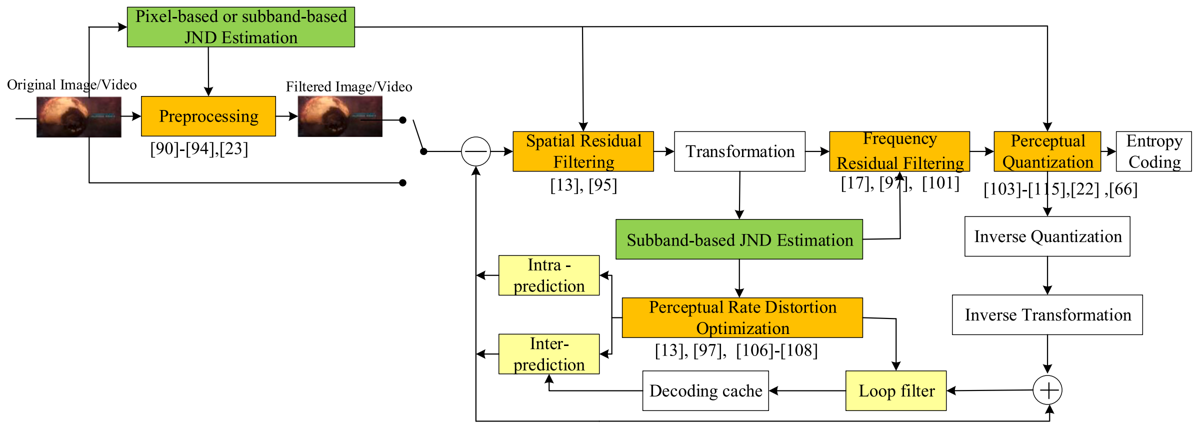


Fig. 5. JND-based perceptual image and video coding schemes.

released on the one hand. On the other hand, JND estimation models are designed for different projection methods to reduce the impact of distortion introduced by projection process [85]. In addition, the literature [86] improved the accuracy of saliency detection by predicting the motion of human head when people watch panoramic images and videos.

Azevedo et al. [87] carried out exploratory experiments on the JND estimation of panoramic video, and found that the HVS is more sensitive to block effect in panoramic video, but not to fuzzy distortion. Jabalah et al. [88] proposed a 360-JND estimation framework for panoramic video projected by ERP. In short, for the ERP image, the rectangular block is extracted and projected to the sphere, and then re-projected to the rectangle by linear projection. The JND model of 2D image is used to estimate the JND of the re-projected rectangular image, and the JND map is mapped to the ERP plane. The final JND is obtained by fusing JND maps of all ERP planes. In the fusion process, the JND threshold of overlapping boundary pixels is obtained by weighted summation. Guan et al. [89] believed that the JND of panoramic video is not only related to the video content, but also related to the view movement. Thus, he proposed an adjustment factor of the JND caused by the movement speed of the object relative to the view, the brightness relative to the view 5 s ago, and the depth difference. The JND model of the final panoramic image/video is expressed as the product of the 2D model and the special masking factor of the panoramic video.

3. Review of JND-based perceptual image and video compression

As mentioned in Section 1, existing JND-based perceptual image and video coding schemes could be divided into JND-based coding preprocessing and JND-based codec embedding. Combining the image and video coding process, all JND-based perceptual coding schemes could be summarized in Fig. 5 .

3.1. JND-based coding preprocessing

Because of interaction and interference among visual stimuli, the HVS could not perceive all detailed information from images and videos. Perceptual redundancy in the images and videos could be removed according to the JND-based perceptual filter. According to the type of JND estimation, JND-based coding preprocessing could be divided into spatial-JND coding preprocessing and subband-JND coding preprocessing.

Spatial-JND Coding Preprocessing: According to the pixel-based JND model, perceptual redundancy could be estimated and filtered directly in the spatial domain. Specifically, Ding et al. [90] adopted a linear iterative clustering to group pixels into perceptually meaningful regions. A Gaussian filter controlled by JND thresholds was designed to smooth the original pixels in each perceptually meaningful region. Vidal et al. [91] introduced a bilateral filter and a weighted filter to smooth the original video. The filter strength is controlled by the estimated JND threshold. Wu et al. [92] divided the image into several 8×8 blocks, pixels in each block are filtered by:

$$\check{I}_o(z) = \begin{cases} I_o(z) + J_S(z), & I_o(z) - \overline{I}_o(z) < -J_S(z), \\ \overline{I}_0(z), & |I_o(z) - \overline{I}_o(z)| \le J_S(z), \\ I_o(z) - J_S(z), & I_o(z) - \overline{I}_o(z) > J_S(z). \end{cases} \tag{3}$$

Here, $I_o(z)$ and $\check{I}_o(z)$ are the original spatial residue and the filtered spatial residue, respectively. $\overline{I}_o(z)$ is the average value of the 8×8 block. $J_S(z)$ is the estimated JND threshold. According to the spatial-JND based filter, the image block will be smoothed and visual details are removed.

Subband-JND Coding Preprocessing: Similar with spatial coding preprocessing, the original image/video could be transformed and filtered with the subband-JND model based filter in the subband domain. In [93], Xiang et al. designed a JND-based filter (JNDF) in DCT domain to remove unnoticeable information for each frame. In addition, they also designed an adaptive bilateral filter (ABF) to smooth the filtered frame by the JNDF. The filter parameter of ABF is determined by the estimated spatial JND threshold and the quantization parameter (QP). Bae et al. presented a just-noticeable-quantization-distortion (JNQD) estimation model in the DCT domain. On the basis of [94], Ki et al. [23] recently proposed two JNQD-based filters in DCT domain based on linear regression (LR) and convolutional neural network (CNN), respectively. Usually, the JND-based filter in the frequency domain could be expressed simply by:

$$\check{C}_o(z) = \begin{cases} 0, if\ |C_o(z)| < w_p \cdot J_F(z), \\ sign\left(C_o(z)\right) \cdot \left(|C_o(z)| - w_p \cdot J_F(z)\right). \end{cases} \tag{4}$$

Here, $C_o(z)$ and $\check{C}_o(z)$ represent the DCT coefficients of the original image/video and the filtered image/video, respectively. $sign(\cdot)$ is the symbol operation. $J_F(z)$ is the estimated DCT-based JND threshold. w_p is the weighted filter parameter which is applied to control the filtering strength.

Obviously, the JND-based coding preprocessing approach eliminates perceptual redundancy directly before image and video compression. However, the existing JND estimation models are not accurate enough and the JND threshold is easy to be overestimated or underestimated. Additionally, image/video coding is a serial process and the quantization module usually introduces quantization distortion, which means that the total coding distortion (i.e., the sum of filtering distortion and quantization distortion) may be larger than the JND threshold. Therefore, the JND-based coding preprocessing scheme is easy to cause subtle loss of perceptual quality.

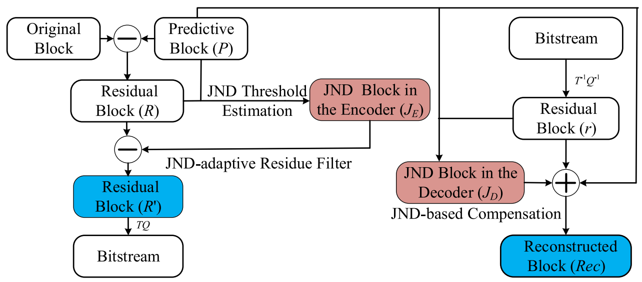


Fig. 6. The procedure of JND compensation based PVC scheme.

3.2. Perceptual residual filtering

After intra or inter prediction, the residue will be transformed and quantized in image and video coding. The residue whose amplitude is smaller than the JND threshold need not to be encoded. Distortion caused by the uncoded residue could not be perceived by the HVS. Similar with the preprocessing schemes, the residue could be filtered to reduce residual amplitude directly. This method also could be divided into spatial-JND residual filtering and subband-JND residual filtering.

Spatial-JND Residual Filtering: According to the pixel-based JND estimation model, the residue could be filtered according to the following filter [13,95]:

$$\breve{R}(z) = \begin{cases} 0, & if\ |R(z)| < J_S(z), \\ R(z), & otherwise. \end{cases} \tag{5}$$

Here, $R(z)$ and $\breve{R}(z)$ are the original residue and the filtered residue. As shown in this filter, when the amplitude of residue is smaller than the pixel-based JND threshold, this residue is filtered to zero directly, otherwise this residue is not changed.

Recently, we proposed a JND compensation based PVC scheme for the HEVC codec in [96]. In this work, the block-wise JND threshold is estimated according to the predictive block in the encoder and decoder sides. In the encoder, the residue is suppressed by the JND-adaptive residue filter. In the decoder, the distortion of reconstructed block will be compensated based on the estimated JND threshold. The procedure of this scheme is shown in Fig. 6.

Subband-JND Residual Filtering: In the DCT domain, the transform coefficient can be suppressed with a similar JND-based filter. This suppression process is usually described by

$$\check{C}(z) = \begin{cases} 0, if\ |C(z)| < w_f \cdot J_F(z), \\ sign\big(C(z)\big) \cdot \big[|C(z)| - w_f \cdot J_F(z)\big]. \end{cases} \tag{6}$$

$C(z)$ and $\check{C}(z)$ are the original transform coefficient and the suppressed transform coefficient, respectively. w_f is a block-wise weighted filter parameter which is determined by the quantization distortion [97–100]. In [97], Bae et al. defined a ratio function of local distortion to the subband-JND threshold to determine the weighted parameter. For an $N \times N$ block, this function is defined as follows:

$$D_B = \left[\frac{1}{N^2} \cdot \sum_{z=0}^{N^2-1} \left(\frac{C(z) - \mathbf{Q}^{-1}\big(\mathbf{Q}(\check{C}(z))\big)}{J_F(z)}\right)^{\gamma}\right]^{1/\gamma}. \tag{7}$$

Here, $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ represent the quantization operation and the inverse quantization operation. γ is a model parameter and is usually set to 2. In practical coding, the filter distortion of the block should be as close as possible to the estimated JND threshold. So, the weighted parameter can be determined by

$$w_f = argmin|D_B - 1|. \tag{8}$$

As described above, perceptual residual filtering reduces the amplitude of residue on the one hand, and increases the probability of the occurrence of the all-zero block on the other hand. However, partial perceptual redundancy has been predicted, resulting in that relatively limited bitrate saving compared to the preprocessing schemes. After residual preprocessing, the suppressed residue will be quantized, resulting into that the total coding distortion may be larger than the JND threshold. To a certain extent, perceptual residual filtering may reduce the perceptual quality of compressed image or video slightly.

In addition, the prevailing image/video coding standards support variable block coding technology. For example, the transform block size is 4×4 to 32×32 in HEVC. The VVC standard not only has square transform units (TUs), but also introduces rectangular TUs. So, some scholars have explored the subband-based JND estimation for variable block-sized transforms [17,101]. Nevertheless, these estimation models were proposed for square transform blocks. As for as we know, there are no subband-based JND models for rectangular TUs. In other word, subband-JND residual filtering may be not fully compatible with the latest coding standard.

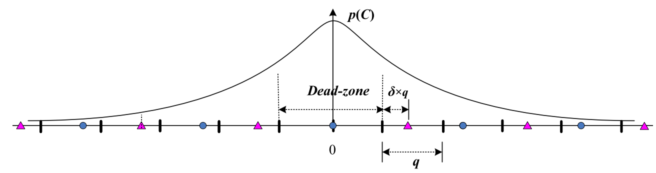


Fig. 7. The diagram of USQ+DZ quantization.

3.3. Perceptually adaptive quantization

In video coding, the DCT coefficient is usually quantized by the uniform scalar quantization with Dead-zone (USQ+DZ). The diagram of USQ+DZ is shown in Fig. 7 and the output of USQ+DZ is expressed by [102]:

$$I_q = sign(C) \cdot floor(|C|/q + \delta). \tag{9}$$

q is the quantization step (Qstep) size which is roughly equal to $2^{(QP-4)/6}$ in HEVC. δ is the normalized quantization offset which is used to control the size of Dead-zone. In intra-prediction and inter-prediction, the offset δ is set to 1/3 and 1/6, respectively. As shown in Fig. 7, Dead-zone is equal to $2(1-\delta) \cdot q$. $floor(\cdot)$ is the round down operation and I_q is the quantization result. According to the relationship between the quantization parameter and the quantization step size, perceptually adaptive quantization could be achieved by determining a suitable QP or Qstep for each DCT coefficient [103].

Perceptually Adaptive Qstep: Statistical experiments have shown that the transform coefficients obey Laplace distribution approximately. So, the transform coefficient distribution (TCD) could be fitted by the Laplace probability density function. The probability of the DCT coefficient could be expressed by [104,105]:

$$f(C) = \frac{1}{2\Lambda} \cdot e^{-\frac{|C-\mu|}{\Lambda}} \tag{10}$$

Λ is the distribution parameter which can be estimated by $\sigma/\sqrt{2}$. σ is the standard deviation of coefficients and μ is set to zero. For the input signal obeying Laplace distribution, the coding distortion measured by mean squared error (MSE) can be estimated as follows for USQ+DZ quantization [106]:

$$\begin{aligned} D_{Lap}(q, \delta, \Lambda) &= \sum_{k=-\infty}^{\infty} \int_{(k-\delta)\cdot q}^{(k+1-\delta)\cdot q} (C - k \cdot q)^2 f(C) dC \\ &= 2\Lambda^2 - \frac{2\Lambda q e^{-\alpha/\Lambda} e^{-q/(2\Lambda)}}{1 - e^{-q/\Lambda}} \cdot \left[\frac{\alpha}{\Lambda} + 1\right] \end{aligned} \tag{11}$$

where $\alpha = (1-\delta) \cdot q - \frac{q}{2}$ [106].

As we all know, the residue whose amplitude is smaller than the JND threshold, this residue could be set to zero directly. Therefore, Luo et al. [106] adjusted the quantization offset δ to control the Dead-zone as follows:

$$\breve{\delta} = \begin{cases} \delta, & T \cdot J_F \leq (1-\delta)\cdot q \\ 1-(T \cdot J_F)/q, & T \cdot J_F > (1-\delta)\cdot q. \end{cases} \tag{12}$$

$\breve{\delta}$ is the JND-based quantization offset. T is the normalized JND threshold adjustment parameter. For a 16×16 macroblock, it is divided into 16 4×4 blocks to transform. The video quality is mainly determined by 16 DC coefficients. For the JND-based offset $\breve{\delta}$, the average distortion of macroblock could be estimated by

$$\overline{D}_{Lap}^{pro} = \sum_{k=1}^{16} D_{Lap}(\breve{\delta}(k), q, \sigma_F(k))/16 \tag{13}$$

where $\sigma_F(k)$ is the standard deviation for DC coefficient of kth 4×4 block. It can be estimated by the standard deviation of all coefficients of this block. $\breve{\delta}(k)$ is the JND-based offset for DC coefficient in kth block. For the original offset δ, the average distortion of macroblock is expressed by

$$\overline{D}_{Lap}^{ori} = \sum_{k=1}^{16} D_{Lap}(q, \sigma_F(k), \delta)/16. \tag{14}$$

Finally, for the original quantization step size (q_{ori}), its perceptual Qstep ($\breve{q}$) can be searched out in a set of quantization step sizes (S_Q) with the following formula:

$$\breve{q} = \arg\min_{q \in S_Q} |\overline{D}_{Lap}^{pro}(q) - \overline{D}_{Lap}^{ori}(q_{ori})|. \tag{15}$$

Perceptually Adaptive QP: The purpose of perceptual quantization is to make the quantization distortion as close as possible to the target distortion guided by JND threshold. Thus, the perceptual QP could be obtained as follows:

$$\breve{Q}P = argmax(D_c(QP) < D_{tar}, QP \in [0, 51]) \tag{16}$$

where $D_c(QP)$ is the distortion from the real encoding process. D_{tar} is the target distortion which could be determined based on the estimated JND threshold. Shen et al. proposed [107] a two pass perceptual coding strategy, in which the coding information will be collected in the first pass coding to derive the appropriate QP for the second pass coding process. Except for this method for perceptual QP determination, Chen et al. [108,109] proposed a weighting QP determination method according to the estimated spatial JND threshold. Briefly, with a JND-based adjustable parameter, the macroblock having smaller JND threshold will be quantized with a smaller QP, whereas it will be quantized with a larger QP. In [108], the perceptual QP of the current macroblock is defined as follows:

$$\breve{Q}P_b = \sqrt{\frac{w_r}{w_b}} \cdot QP. \tag{17}$$

w_r is a constant and set to 1. w_b is the weighted parameter which is usually defined as a sigmoid function about the average JND thresholds of the current macroblock and the current frame as follows:

$$w_b = a + b\frac{1 + m \cdot exp(-c\dfrac{J_S^b - \overline{J}_S}{\overline{J}_S})}{1 + n \cdot exp(-c\dfrac{J_S^b - \overline{J}_S}{\overline{J}_S})} \tag{18}$$

where $a = 0.7, b = 0.6, m = 0, n = 1, c = 4$ defined empirically [108]. J_S^b and $\overline{J_S}$ are the average spatial JND thresholds of the macroblock and the frame, respectively. As described above, the perceptual Qstep or QP for each macroblock is determined by fine adjustment on the basis of given Qstep or QP. The bitrate saving of adaptive perceptual quantization is limited. In addition, the perceptual adaptive Qstep determination method is mainly designed for H.264/AVC codec, it may not be directly used in the codecs of the latest video coding standards.

Actually, human only perceive discrete-scale quality levels over a wide range of coding bitrate [110]. In other word, perceptual quantization could be transformed into the problem of how to determine the maximum QP under the same perceptual quality level. Thus, some scholars began to explore the perceptual quantization for multi-level coding quality. In our previous work [111], the perceptual QP for each quality level was determined based on the approach of mathematical statistics. Literatures [112–114] predicted the perceptual QP based on the machine learning approach or the deep learning approach. Liu et al. [115] built a CNN-based JND model to predict the quality factor of each quality level for JPEG image compression. In [66], block-wise QF is deduced based on a block-level JND model. Wu et al. [22] designed lightweight network to predict the QP value of the first JND point which is used for perceptual quantization in VTM 12.0.

3.4. Perceptual rate distortion optimization

In image and video coding, bitrate and quality are a pair of contradiction, which are determined by coding algorithms and coding parameters. A good encoder must be a compromise strategy for this contradiction. Due to the constraints of storage capacity and transmission bandwidth, video compression often has an average rate constraint. The problem of image/video coding optimization is to find the minimum coding distortion under the constraint of average bitrate, which can be expressed with the following formula:

$$\min_{m \in M} D(CP_m), \ subject\ to\ Bit(CP_m) \leq Bit_t. \tag{19}$$

Here, D and Bit represent coding distortion and bitrate. Bit_t is the target bitrate. CP_m represents coding parameter combination. Usually, this optimization problem could be solved with the Lagrange multiple method, so that the formula (19) could be rewritten by

$$\min_{m \in M} RDO_{cost} = \min_{m \in M} \left(D(CP_m) + \lambda \cdot Bit(CP_m)\right) \tag{20}$$

RDO_{cost} is the RDO cost, λ is the lagrange multiple which is applied to balance coding distortion and bitrate. In the view of this computational formulation, the perceptual rate distortion optimization could be divided into perceptual distortion measurement and lagrange multiple perceptual optimization.

Perceptual Distortion Measurement: As mentioned above, the distortion below the JND threshold cannot be perceived. So, Yang et al. [13] proposed a perceptual distortion factor as follows to filter the distortion that could be ignored.

$$w_q(z) = \begin{cases} 1, & |I_o(z) - I_c(z)| > J_S(z) \\ 0, & Otherwise. \end{cases} \tag{21}$$

I_o represents the original image or video, I_c represents the reconstructed image or video. As we all know, the coding distortion is usually measured with sum of absolute difference (SAD), sum of squared error (SSE) or mean squared error (MSE), etc. Yang et al. [13] proposed the sum of absolute perceptual difference (SAPD) which is used in motion estimation (ME) to determine the optimal motion vector in the sense of perceptual distortion. The SAPD of the current block and the reference block is calculated by

$$SAPD = \sum_{z=1}^{N^2} (|I_o(z) - I_r(z)| - J_S(z)) \cdot w_q(z) \tag{22}$$

I_r is the reference block. Based on the factor w_q, Chen et al. [108] proposed a perceptual distortion measurement named the peak signal-to-perceptual-noise ratio (PSPNR). The PSPNR of reconstructed image or video can be calculated as follows:

$$PSPNR = 10log_{10}\frac{w \times h \times 255^2/w_q(z)}{\sum\sum_{z=1}^{w\times h}(|I_o(z) - I_c(z)| - J_S(z))^2} \tag{23}$$

where $w \times h$ is the resolution of the image/video. With the similar method, Yang et al. [116,117] proposed the perceptual SSE to measure coding distortion and introduced it into the sample adaptive offset (SAO) filter. Zhu et al. [118] proposed a JND-based weighted MSE to optimize the AV1 encoder.

Lagrange Multiple Optimization: In rate distortion optimization, the lagrange multiple λ is equal to $-\partial D/\partial Bit$ and could be computed based on the Bit-Q model and the D-Q model. In the H.264/AVC and HEVC reference softwares, the value of λ depends on the Qstep and the dependence between λ and the QP, which can be expressed by

$$\lambda = \beta \cdot q^2 = \beta \cdot 2^{(QP-12)/3} \tag{24}$$

where β is a constant which is determined by the prediction method and the software. As described in Section 2.3, the perceptual Qstep and QP for each macroblock can be determined according to the estimated JND threshold. Thus, the JND-based lagrange multiple for each macroblock can be obtained according to the perceptual QP of each macroblock (QP_b).

$$\check{\lambda}_b = \beta \cdot \check{q}_b^2 = \beta \cdot 2^{(\check{Q}P_b-12)/3} \tag{25}$$

In addition, similar with the perceptual QP, Chen et al. [108] suggested the lagrange multiple should be adjusted considering the HVS visual sensitivity. Thus, they proposed a JND-based weighed factor to adjust the formula (24) and the perceptual lagrange multiple for each macroblock is expressed by

$$\check{\lambda}_b = \beta w_b \cdot \check{q}_b^2 = \beta w_b \cdot 2^{(\check{Q}P_b-12)/3} \tag{26}$$

As described in this subsection, perceptual RDO can be performed by perceptual distortion measurement and lagrange multiple perceptual optimization. Additionally, perceptual RDO has been widely used in image/video coding modules, such as inter-prediction and loop filter, etc. However, RDO is applied to determine the optimal coding parameters in image and video coding. So, compared with other JND-based PVC schemes, JND-based perceptual RDO makes a small contribution to bitrate reduction [97].

4. Discussion and future work

4.1. Discussion about JND estimation

As mentioned in Section 2, related works about JND estimation have been divided into five types, i.e., pixel-wise, subband-wise, block-wise, picture and video-wise JND estimation. In the JND estimation, the main issue that needs to be addressed is bridging the gap between the perceptual theory and the JND estimation modeling. From the perspective of modeling methods, existing JND models are built mainly using the statistical modeling method, the theoretical modeling method and the learning-based modeling method.

Statistical JND Modeling: Actually, as we all know, visual perception characteristics can be qualitatively analyzed but difficult to quantify. Therefore, researchers typically conduct complex visual experiments to obtain experimental data. Based on experimental data, expression of JND models can be constructed with the statistical modeling method and it is relatively simple and low complexity. However, on the one hand, visual experiments are susceptible to experimental conditions. On the other hand, it is difficult to analyze the coupling among visual perception effects. Thus, JND models constructed by this way generally have the problem of overestimation and underestimation [33].

Theoretical JND Modeling: By simulating the visual perception process, JND models constructed by the theoretical modeling method improve the prediction accuracy of JND thresholds by introducing more complex visual perception effects and analyzing the correlation between visual perception effects. In recent years, the free-energy theory [119] and the hierarchical prediction coding theory [39] have been introduced into lots of pixel-wise and subband-wise JND modeling [31,33,36,37,47]. Due to the lack of modeling representations of visual perception processes, such methods cannot analyze the coupling relationships between different dimensional visual perception effects, and related research needs to be expanded and in-depth.

Learning-based JND Modeling: To avoid the complex coupling analysis among visual perception effects, lots of learning-based JND models have been built in recent years. In terms of learning-based JND modeling, there some compressed image/video databases are released in recent years. Yet, the testing samples and tester are always insufficient. On the one hand, most existing JND databases are constructed based on compressed images or compressed videos. Yet, in real life, images and videos do not only contain encoding distortion, but may also contain other noise such as Gaussian noise and bipolar noise. On the other hand, since the label of databases is typically picture-wise or video-wise, the proposed JND models using learning-based modeling method may not be fully suitable for block-wise image or video compression.

4.2. Development trend of JND estimation

HVS can only identify limited subjective quality levels for a series of compressed videos. Thus, JND estimation is developing from single-level estimation to multi-level estimation. In addition, in order to meet the optimization requirements of multi-granular video coding, JND estimation is also developing from pixel-wise to video-wise. Actually, With the multi-dimensional development of images and videos, JND estimation can be further studies for human vision and machine vision.

Extension of Visual JND Estimation: As we all know, the video continues to evolve in terms of resolution, frame rate and color gamut. Yet, visual JND estimation for these changes have not been fully studied [120]. Besides, types of images of videos are constantly increasing. Thus, JND estimation for specific images or videos needs urgently further exploration, such as JND estimation for chart images [121]. In the view of visual perceptual theory, the audio signal can effectively enhance human vision [122]. Thus, multi-modal collaborative JND estimation may improve the accuracy of multimedia JND models. As shown in Table 1, several JND datasets have been released. These datasets consist of compressed images and videos, noise in these compressed data only comes from the coding distortion. Thus, datasets contained other noises (e.g., the Gaussian noise and the bipolar noise) should be built and released. In addition, The distortion distribution in compressed images and videos with different encoders is not completely consistent. Thus, building more JND datasets is very important for improving JND estimation, especially building more uncompressed images and videos datasets.

JND Estimation for Machine Vision: Inspired by the concept of JND of human perception, scholars conducted comprehensive experiments on various machine vision tasks, proving that machine vision also has thresholds similar to JND of HVS [65,123]. In [123], Zhang et al. presented the concept of just recognizable distortion (JRD) which is defined as the maximum distortion caused by data compression that will reduce the machine vision model performance to an unacceptable level. In order to estimate the JRD, they built a JRD-annotated dataset containing 340 000 images and established an ensemble-learning-based framework to predict the JRD for diverse vision tasks. Subsequently, except for the JRD, they also proposed the concept of Satisfied Machine Ratio (SMR), which is used to evaluate image quality from a machine perspective, and designed a deep learning-based model to predict the SMR of compressed images [124]. Similarly, Jin et al. constructed a JND model for the image classification task named by DMV-JND-NET through unsupervised learning [125].

As is mentioned above, research of JND estimation for machine vision is in the early exploratory stage. On the one hand, most existing JRD or SMR models are full-reference or half-reference models which cannot be directly applied to image or video compression for machine vision tasks. On the other hand, JND estimation for machine vision is coupled with machine vision tasks and specific algorithms. The universality of estimation models needs to be improved to be suitable for different machine vision tasks.

4.3. Discussion about JND-based video coding

As mentioned in Section 3, JND-based image/video coding has been deeply investigated in decades. Most of these studies focus on perceptual coding optimization of a single coding module for human vision. Thus, we will discuss JND-based video coding in the view of multi-module collaborative PVC for human vision and the exploration of PVC for machine vision.

Multi-module Collaborative PVC for Human Vision: As we all know, video compression is a serial multi-module collaborative coding process. Thus, some scholars attempted to explore multi-module collaborative perceptual coding optimization. For example, Yang et al. [13] designed a perceptual residual filter and a perceptual prediction scheme for video coding. Luo et al. [106] optimized the quantization module and the motion vector prediction module at the same time. In [62], we designed a perceptual residual filter controlled by quantization distortion at the encoder. Besides, a JND-based compensation scheme was proposed for decoder.

These works have proven that collaborative perceptual coding optimization could further improve coding efficiency, but these works have not fully investigated the coupling effect and the distortion transfer effect between coding modules. The coupling effect mainly comes from the following two aspects. (1) Multi-module perceptual optimization may cause that total coding distortion exceeds the JND threshold, resulting in the quality reduction of compressed video. (2) Multi-module perceptual optimization may affect each other, resulting in a decrease in overall coding gain. As for the distortion transfer effect, it is mainly reflected in the serial distortion transfer between modules and the accumulation of distortion transfer between frames. In [126–128], scholars have discussed the problem of distortion transfer in traditional video coding optimization. Yet, multi-module perceptual coding optimization on the one hand leads to serial distortion transfer between modules, on the other hand, it exacerbates the distortion accumulation between frames. The distortion transfer effect is more prominent in video perceptual coding.

Exploration of PVC for Machine Vision: Based on JRD prediction, the coding process can be optimized by using more appropriate coding parameters (such as the quantization parameter) to achieve the better balance between bit-rate and the performance of machine vision tasks. Just like the research of JRD prediction, the research of PVC for machine vision is also in its initial stage. But, experimental results have shown that JRD prediction has great potential in image and video compression for machine vision tasks. In [124], Zhang et al. had found that 40.47% and 22.77% of bit-rate can be saved under the same level of image classification and object detection performance.

In summary, the development of perceptual coding optimization is relatively mature for human vision. Due to the problems of coupling and distortion transfer between modules, multi-module cooperative perceptual coding optimization needs further in-depth research. Exploring decoupling multi-module cooperative perceptual coding optimization strategies is expected to further improve video compression efficiency. As for the development of perceptual coding optimization for machine vision, it has shown significant potential but requires more in-depth study to diverse vision tasks.

5. Conclusion

In this paper, we reviewed JND estimation and its application in perceptual video coding systematically. Specifically, we introduced most existing JND computation models comprehensively and discussed their possible future directions. In addition, we conducted a summary about JND-based perceptual video coding schemes and analyzed the necessity of perceptual coding optimization for human vision and machine vision.

CRediT authorship contribution statement

Guoxiang Wang: Writing – review & editing for "Review of JND Estimation for Image and Video". **Hongkui Wang:** Writing – review & editing for "Review of JND-based Perceptual Image and Video Compression". **Hui Li:** Writing – review & editing for "Review of JND Estimation for Image and Video". **Li Yu:** Writing – review & editing for "Review of JND-based Perceptual Image and Video Compression". **Haibing Yin:** Writing – review & editing for "Review of JND-based Perceptual Image and Video Compression". **Haifeng Xu:** Writing – review & editing for "Review of JND-based Perceptual Image and Video Compression". **Zhen Ye:** Writing – review & editing for "Review of JND Estimation for Image and Video". **Junfeng Song:** Writing – review & editing for "Review of JND Estimation for Image and Video".

Declaration of competing interest

No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

Data availability

No data was used for the research described in the article.

References

[1] G.J. Sullivan, J.R. Ohm, W.J. Han, T. Wiegand, Overview of the high efficiency video coding (HEVC) standard, IEEE Trans. Circuits Syst. Video Technol. 22 (12) (2013) 1649–1668.
[2] B. Bross, J. Chen, S. Liu, Versatile Video Coding (Draft 2), Document JVET-K1001 of JVET, Ljubljana, SI, 2018.
[3] Z. Pan, F. Yuan, W. Yu, J. Lei, N. Ling, S. Kwong, RDEN: Residual distillation enhanced network-guided lightweight synthesized view quality enhancement for 3D-HEVC, IEEE Trans. Circuits Syst. Video Technol. 32 (9) (2022) 6347–6359.
[4] Z. Pan, H. Zhang, J. Lei, Y. Fang, X. Shao, N. Ling, S. Kwong, DACNN: Blind image quality assessment via a distortion-aware convolutional neural network, IEEE Trans. Circuits Syst. Video Technol. 32 (11) (2022) 7518–7531.
[5] Z. Pan, F. Yuan, J. Lei, Y. Fang, X. Shao, S. Kwong, VCRNet: Visual compensation restoration network for no-reference image quality assessment, IEEE Trans. Image Process. 31 (2022) 1613–1627.
[6] Y.C. Lin, Experimental design and analysis of JND test on coded image/video, in: Applications of Digital Image Processing XXXVIII, 2015.
[7] H. Wang, W. Gan, S. Hu, J.Y. Lin, L. Jin, L. Song, P. Wang, I. Katsavounidis, A. Aaron, C.-C.J. Kuo, MCL-JCV: A JND-based H.264/AVC video quality assessment dataset, in: 2016 IEEE International Conference on Image Processing, ICIP, 2016, pp. 1509–1513.
[8] X. Zhu, T. Guijing, H.266/VVC: A new generation general video coding international standard, J. Nanjing Univ. Posts Telcommun. 41 (2) (2021) 1–11.
[9] Y. Zhang, L. Zhu, G. Jiang, S. Kwong, C.J. Kuo, A survey on perceptually optimized video coding, CoRR, abs/2112.12284, 2021, arXiv:2112.12284.
[10] S.L. Macknik, M.S. Livingstone, Neuronal correlates of visibility and invisibility in the primate visual system, Nature Neurosci. 1 (2) (1998) 144–149.
[11] Y. Di, T. Zhao, Y. Xu, X. Hong, L. Lin, Visual JND: A perceptual measurement in video coding, IEEE Access 7 (99) (2019) 29014–29022.
[12] W. Lin, G. Ghinea, Progress and opportunities in modelling just-noticeable difference (JND) for multimedia, IEEE Trans. Multimed. 24 (2022) 3706–3721.
[13] X. Yang, W. Lin, Z. Lu, E. Ong, S. Yao, Just noticeable distortion model and its applications in video coding, Signal Process., Image Commun. 20 (2005) 662–680.
[14] H. Oh, A. Bilgin, M. W.Marcellin, Visually lossless encoding for JPEG2000, IEEE Trans. Image Process. 22 (1) (2013) 189–201.
[15] J.-R. Leandro, A.L. Francesc, W.M. Michael, Visually lossless strategies to decode and transmit JPEG2000 imagery, IEEE Signal Process. Lett. 21 (1) (2014) 35–38.
[16] H. Feng, M.W. Marcellin, B. A., A methodology for visually lossless JPEG2000 compression of monochrome stereo images, IEEE Trans. Image Process. 25 (2) (2015) 560–572.
[17] S. Bae, M. Kim, A novel generalized DCT-based JND profile based on an elaborate CM-JND model for variable block-sized transforms in monochrome images, IEEE Trans. Image Process. 23 (8) (2014) 3227–3240.

[18] S. Ki, S. Bae, M. Kim, H. Ko, Learning-based just-noticeable-quantization-distortion modeling for perceptual video coding, IEEE Trans. Image Process. 27 (7) (2018) 3178–3193.
[19] T. Tian, H. Wang, L. Zuo, C. Kuo, S. Kwong, Just noticeable difference level prediction for perceptual image compression, IEEE Trans. Broadcast. 66 (3) (2020) 690–700.
[20] H. Liu, Y. Zhang, H. Zhang, C. Fan, X. Fan, Deep learning-based picture-wise just noticeable distortion prediction model for image compression, IEEE Trans. Image Process. PP (99) (2019) 1.
[21] Y. Zhang, H. Liu, Y. Yang, X. Fan, S. Kwong, C. Kuo, Deep learning based just noticeable difference and perceptual quality prediction models for compressed video, IEEE Trans. Circuits Syst. Video Technol. (2022).
[22] Y. Wu, Z. Wang, W. Chen, L. Lin, H. Wei, T. Zhao, Perceptual VVC quantization refinement with ensemble learning, Displays: Technol. Appl. (70) (2021) 1–8.
[23] S. Ki, S.H. Bae, M. Kim, H. Ko, Learning-based just-noticeable-quantization-distortion modeling for perceptual video coding, IEEE Trans. Image Process. 27 (7) (2018) 3178–3193.
[24] Y. Zhang, L. Zhu, G. Jiang, S. Kwong, C.C.J. Kuo, A survey on perceptually optimized video coding, 2021, arXiv:2112.12284.
[25] D. Yuan, T. Zhao, Y. Xu, H. Xue, L. Lin, Visual JND: A perceptual measurement in video coding, IEEE Access 7 (2019) 29014–29022.
[26] Digital pictures:representation and compression.
[27] Y.-J. Chin, T. Berger, A software-only videocodec using pixelwise conditional differential replenishment and perceptual..., IEEE Trans. Circuits Syst. Video Technol. 9 (3) (1999) 438.
[28] Z. Chen, C. Guillemot, Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model, IEEE Trans. Circuits Syst. Video Technol. 20 (6) (2010) 806–819.
[29] A. Liu, W. Lin, M. Paul, C. Deng, F. Zhang, Just noticeable difference for images with decomposition model for separating edge and textured regions, IEEE Trans. Circuits Syst. Video Technol. 20 (11) (2010) 1648–1652.
[30] K. Friston, Friston, K.J.: The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11 (2) 127–138.
[31] J. Wu, G. Shi, W. Lin, A. Liu, F. Qi, Just noticeable difference estimation for images with free-energy principle, IEEE Trans. Multimed. 15 (7) (2013) 1705–1710.
[32] A. Houmansadr, H. Pirsiavash, S. Ghaemmaghami, Robust contentbased video watermarking exploiting motion entropy masking effect, in: Proc. Int. Conf. Signal Process. Multimedia Appl., 2006, pp. 252–259.
[33] J. Wu, L. Li, W. Dong, G. Shi, W. Lin, C. Kuo, Enhanced just noticeable difference model for images with pattern complexity, IEEE Trans. Image Process. PP (6) (2017) 1.
[34] H. Wang, L. Yu, S. Wang, G. Xia, H. Yin, A novel foveated-JND profile based on an adaptive foveated weighting model, in: 2018 IEEE Visual Communications and Image Processing (VCIP), 2019.
[35] H. Yin, H. Wang, L. Yu, J. Liang, G. Zhai, Feedforward and feedback modulations based foveated JND estimation for images, ACM Trans. Multimed. Comput. Commun. Appl. 19 (5) (2023) 1–23.
[36] H. Wang, S. Wang, T. Li, H. Yin, L. Yu, Surprise-based JND estimation for images, IEEE Signal Process. Lett. 27 (2020) 181–185.
[37] H. Wang, Hierarchical predictive coding-based JND estimation for image compression, IEEE Trans. Image Process. 30 (2020) 487–500.
[38] H.B. Barlow, Possible Principles Underlying the Transformation of Sensory Messages, MIT Press, 2012.
[39] R. Rao, Predictive coding in the visual cortex : a functional interpretation of some extra-classical receptive-field effects, Nat. Neurosci. 2 (1999).
[40] N.B. Nill, A visual model weighted cosine transform forimage compression and quality assessment, IEEE Trans. Commun. 33 (6) (1985) 645–652.
[41] D.H. Kelly, Motion and vision. II. Stabilized spatio-temporal threshold surface, J. Opt. Soc. Amer. 69 (1979) 1340–1349.
[42] A.J. Ahumada, H.A. Peterso, Luminance-model-based DCT quantization for color image compression, 1997.
[43] A.B. Watson, Visual optimization of DCT quantization matrices for individual images, 2013.
[44] H. Oh, A. Bilgin, M.W. Marcellin, Visually lossless encoding for JPEG2000, IEEE Trans. Image Process. 22 (1) 189–201.
[45] Z. Wei, S. Member, IEEE, Fellow, IEEE, Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain, IEEE Trans. Circuits Syst. Video Technol. 19 (3) (2009) 337–346.
[46] H. Wang, L. Yu, T. Li, M. Fan, H. Yin, A DCT-JND profile for disorderly concealment effect, in: Advances in Multimedia Information Processing - PCM 2018 - 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part III, in: Lecture Notes in Computer Science, vol. 11166, Springer, 2018, pp. 188–199.
[47] H. Yang, H. Wang, L. Yu, J. Liang, T. Li, An adaptive quantization based PVC scheme for HEVC, in: 2020 Data Compression Conference, DCC, 2020.
[48] X. Zhang, W. Lin, P. Xue, Improved estimation for just-noticeable visual distortion, Signal Process. (2005).
[49] W. Wan, J. Wu, X. Xie, G. Shi, A novel just noticeable difference model via orientation regularity in DCT domain, IEEE Access (2017) 1.
[50] S. Bae, M. Kim, A novel generalized DCT-based JND profile based on an elaborate CM-jnd model for variable block-sized transforms in monochrome images, IEEE Trans. Image Process. (8) (2014) 23.
[51] S.-H. Bae, M. Kim, A novel DCT-based JND model for luminance adaptation effect in dct frequency, IEEE Signal Process. Lett. 20 (9) (2013) 893–896.
[52] S.-H. Bae, M. Kim, A DCT-based total JND profile for spatiotemporal and foveated masking effects, IEEE Trans. Circuits Syst. Video Technol. 27 (6) (2017) 1196–1207.
[53] M. Lin, K.N. Ngan, Z. Fan, S. Li, Adaptive block-size transform based just-noticeable difference model for images/videos, Signal Process. Image Commun. 26 (3) (2011) 162–174.
[54] Y. Jia, W. Lin, A.A. Kassim, Estimating just-noticeable distortion for video, IEEE Trans. Circuits Syst. Video Technol. (7) (2006) 16.
[55] Z. Wei, K.N. Ngan, Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain, IEEE Trans. Circuits Syst. Video Technol. 19 (3) (2009) 337–346.
[56] J. Lin, L. Jin, S. Katsavounidis, et al., Experimental design and analysis of JND test on coded image/video, in: Applications of Digital Image Processing XXXVIII. International Society for Optics and Photonics, 2015.
[57] L. Jin, J.Y. Lin, S. Hu, H. Wang, P. Wang, I. Katsavounidis, A. Aaron, C.J. Kuo, Statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis, in: Image Quality and System Performance XIII, Electronic Imaging 2016, IQSP, San Francisco, CA, USA, 14-18 February 2016, Ingenta, 2016, pp. 1–9.
[58] Q. Huang, H. Wang, S.C. Lim, Y.K. Hui, C. Kuo, Measure and prediction of HEVC perceptually lossy/lossless boundary qp values, in: Data Compression Conference, 2017.
[59] H. Wang, I. Katsavounidis, J. Zhou, J. Park, S. Lei, X. Zhou, M.O. Pun, X. Jin, R. Wang, X. Wang, VideoSet: A large-scale compressed video quality dataset based on JND measurement, 2017, arXiv e-prints.
[60] B. Cfa, Z.A. Yun, B. Hza, C. Rh, A. Qj, Picture-level just noticeable difference for symmetrically and asymmetrically compressed stereoscopic images: Subjective quality assessment study and datasets - ScienceDirect, J. Vis. Commun. Image Represent. 62 (C) (2019) 140–151.
[61] X. Liu, Z. Chen, X. Wang, J. Jiang, S. Kowng, JND-pano: Database for just noticeable difference of JPEG compressed panoramic images, in: Pacific Rim Conference on Multimedia (PCM), Springer, 2018, pp. 458–468.
[62] H. Wang, L. Yu, X. Tang, H. Yin, J. Liang, A qd&jnd compensation based PVC scheme for HEVC, in: 2020 Data Compression Conference, DCC, 2020.
[63] H. Lin, G. Chen, M. Jenadeleh, V. Hosu, U. Reips, R. Hamzaoui, D. Saupe, Large-scale crowdsourced subjective assessment of picturewise just noticeable difference, IEEE Trans. Circuits Syst. Video Technol. 32 (9) (2022) 5859–5873.
[64] X. Zhang, H. Wang, T. Tian, Perceptual video coding with block-level staircase just noticeable distortion, in: 2019 IEEE International Conference on Image Processing, ICIP, 2019, pp. 4140–4144.
[65] X. Shen, Z. Ni, W. Yang, X. Zhang, S. Kwong, Just noticeable distortion profile inference: A patch-level structural visibility learning approach, IEEE Trans. Image Process. 30 (2021) 26–38.
[66] T. Tian, H. Wang, S. Kwong, C.-C.J. Kuo, Perceptual image compression with block-level just noticeable difference prediction, ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) 16 (4) (2021) 1–15.
[67] H. Wang, X. Zhang, C. Yang, C.-C.J. Kuo, Analysis and prediction of JND-based video quality model, in: Picture Coding Symposium, PCS, 2018.
[68] C. Fan, H. Lin, V. Hosu, Y. Zhang, D. Saupe, SUR-net: Predicting the satisfied user ratio curve for image compression with deep learning, in: 11th International Conference on Quality of Multimedia Experience, QoMEX 2019, 2019.
[69] C. Fan, Y. Zhang, R. Hamzaoui, D. Ziou, Q. Jiang, Learning-based satisfied user ratio prediction for symmetrically and asymmetrically compressed stereoscopic images, IEEE MultiMedia 28 (3) (2021) 8–20.
[70] H. Wang, I. Katsavounidis, Q. Huang, X. Zhou, C.J. Kuo, Prediction of satisfied user ratio for compressed video, 2018, pp. 6747–6751.
[71] X. Zhang, C. Yang, H. Wang, W. Xu, C. Kuo, Satisfied-user-ratio modeling for compressed video, IEEE Trans. Image Process. PP (99) (2020) 1.
[72] J. Zhu, P. Le Callet, A.-F. Perrin, S. Sethuraman, K. Rahul, On the benefit of parameter-driven approaches for the modeling and the prediction of satisfied user ratio for compressed video, in: 2022 IEEE International Conference on Image Processing, ICIP, 2022, pp. 4213–4217.
[73] B. Bondžulić, N. Stojanović, V. Petrović, B. Pavlović, Z. Mili?ević, Efficient prediction of the first just noticeable difference point for JPEG compressed images, Acta Polytech. Hung. 18 (8) (2021) 201–220.
[74] Z. Xiuchang, T. Guijing, Techniques and standards for screen content video coding, J. Nanjing Univ. Posts Telecommun. 37 (3) (2017) 11.
[75] S. Wang, L. Ma, Y. Fang, Weisi, Siwei, W. Gao, Just noticeable difference estimation for screen content images., IEEE Trans. Image Process. (2016).
[76] L. Prangnell, V. Sanchez, JND-based perceptual video coding for 4:4:4 screen content data in HEVC, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2018.
[77] D.V.S.X. De Silva, E. Ekmekcioglu, W.A.C. Fernando, S.T. Worrall, Display dependent preprocessing of depth maps based on just noticeable depth difference modeling, IEEE J. Sel. Top. Sign. Proces. 5 (2) (2011) 335–351.

[78] X. Fei, C. Jung, J. Kim, Disparity-based just-noticeable-difference model for perceptual stereoscopic video coding using depth of focus blur effect, Displays 42 (apr.) (2016) 43–50.
[79] Z. R., H. R., W. Z., 3D hybrid just noticeable distortion modeling for depth image-based rendering, Multimedia Tools Appl. 75 (2015) 10457–10478.
[80] F. Lian, S. Liu, X. Fan, D. Zhao, G. Wen, A new just-noticeable-distortion model combined with the depth information and its application in multi-view video coding, IEEE Comput. Soc. (2012).
[81] Y. Zhao, Z. Chen, C. Zhu, Y.-P. Tan, L. Yu, Binocular just-noticeable-difference model for stereoscopic images, IEEE Signal Process. Lett. 18 (1) (2011) 19–22.
[82] F. Qi, D. Zhao, X. Fan, T. Jiang, Stereoscopic video quality assessment based on visual attention and just-noticeable difference models, 2015.
[83] X. Li, W. Yue, D. Zhao, T. Jiang, Z. Nan, Joint just noticeable difference model based on depth perception for stereoscopic images, in: 2011 Visual Communications and Image Processing, VCIP, 2011.
[84] M. Startsev, M. Dorr, 360-Aware saliency estimation with conventional image saliency predictors, Signal Process. Image Commun. (2018) S0923596518302595.
[85] H.-T. Cheng, C.-H. Chao, J.-D. Dong, H.-K. Wen, T.-L. Liu, M. Sun, Cube padding for weakly-supervised saliency prediction in 360 videos, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 1420–1429.
[86] Y. Zhu, G. Zhai, X. Min, J. Zhou, The prediction of saliency map for head and eye movements in 360 degree images, IEEE Trans. Multim. 22 (9) (2020) 2331–2344.
[87] R. Azevedo, N. Birkbeck, I. Janatra, B. Adsumilli, P. Frossard, On the first JND and break in presence of 360-degree content: an exploratory study, in: International Workshop on Immersive Mixed and Virtual Environment Systems, 2019.
[88] S. Jaballah, A. Bhavsar, M.C. Larabi, Perceptual versus latitude-based 360-deg video coding optimization, in: 2020 IEEE International Conference on Image Processing, ICIP, 2020.
[89] Y. Guan, C. Zheng, X. Zhang, Z. Guo, J. Jiang, Pano: optimizing 360 video streaming with a better understanding of quality perception, in: Proceedings of the ACM Special Interest Group on Data Communication, 2019.
[90] L. Ding, G. Li, R. Wang, W. Wang, Video pre-processing with JND-based Gaussian filtering of superpixels, in: Proceedings of Spie the International Society for Optical Engineering, Vol. 9410, 2015.
[91] E. Vidal, N. Sturmel, C. Guillemot, P. Corlay, F.X. Coudoux, New adaptive filters as perceptual preprocessing for rate-quality performance optimization of video coding, Signal Process., Image Commun. 52 (2017) 124–137.
[92] J. Wu, L. Li, W. Dong, G. Shi, W. Lin, C..J. Kuo, Enhanced just noticeable difference model for images with pattern complexity, IEEE Trans. Image Process. 26 (6) (2017) 2682–2693.
[93] G. Xiang, H. Jia, J. Liu, B. Cai, Y. Li, X. Xie, Adaptive perceptual preprocessing for video coding, in: 2016 IEEE International Symposium on Circuits and Systems, ISCAS, 2016, pp. 2535–2538.
[94] S. Ki, M. Kim, H. Ko, Just-noticeable-quantization-distortion based preprocessing for perceptual video coding, in: 2017 IEEE Visual Communications and Image Processing, VCIP, 2017, pp. 1–4.
[95] H. Wang, L. Yu, J. Liang, H. Yin, T. Li, S. Wang, Hierarchical predictive coding-based JND estimation for image compression, IEEE Trans. Image Process. 30 (2021) 487–500.
[96] H. Wang, L. Yu, X. Tang, H. Yin, J. Liang, A QD&JND compensation based PVC scheme for HEVC, in: 2020 Data Compression Conference, DCC, 2020, pp. 396–396.
[97] S. Bae, J. Kim, M. Kim, HEVC-based perceptually adaptive video coding using a DCT-based local distortion detection probability model, IEEE Trans. Image Process. 25 (7) (2016) 3343–3357.
[98] J. Cui, R. Xiong, X. Zhang, S. Wang, S. Ma, Perceptual video coding based on visual saliency modulated just noticeable distortion, in: 2019 Data Compression Conference, DCC, 2019, pp. 565–565.
[99] J. Kim, S. Bae, M. Kim, An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels, IEEE Trans. Circuits Syst. Video Technol. 25 (11) (2015) 1786–1800.
[100] S. Ki, J. Do, M. Kim, Learning-based JND-directed HDR video preprocessing for perceptually lossless compression with HEVC, IEEE Access 8 (2020) 228605–228618.
[101] L. Ma, K.N. Ngan, F. Zhang, S. Li, Adaptive block-size transform based just-noticeable difference model for images/videos, Signal Process., Image Commun. 26 (3) (2011) 162–174.
[102] H. Yin, H. Wang, X. Huang, H. Yin, Efficient hard-decision quantization using an adaptive deadzone offset model for video coding, IEEE Access 7 (2019) 151215–151229.
[103] H. Wang, L. Yu, H. Yang, H. Xu, H. Yin, G. Zhai, T. Li, Z. Kuang, Surprise-based JND estimation for perceptual quantization in h.265/hevc codecs, Signal Process., Image Commun. 118 (2023) 117019.
[104] J. Si, S. Ma, S. Wang, W. Gao, Laplace distribution based CTU level rate control for HEVC, in: Visual Communications and Image Processing, 2013.
[105] H.-M. Huang, J.-J. Chen, Source model for transform video coder and its application .1. fundamental theory, IEEE Trans. Circuits Syst. Video Technol. 7 (2) (2002) 299–311.
[106] Z. Luo, L. Song, S. Zheng, N. Ling, H.264/advanced video control perceptual optimization coding based on JND-directed coefficient suppression, IEEE Trans. Circuits Syst. Video Technol. 23 (6) (2013) 935–948.
[107] X. Shen, X. Zhang, S. Wang, S. Kwong, G. Zhu, Just noticeable distortion based perceptually lossless intra coding, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2020, pp. 2058–2062.
[108] Z. Chen, C. Guillemot, Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model, IEEE Trans. Circuits Syst. Video Technol. 20 (6) (2010) 806–819.
[109] Y. Yan, G. Xiang, Y. Li, X. Xie, W. Yan, Y. Bao, Spatiotemporal perception aware quantization algorithm for video coding, in: 2020 IEEE International Conference on Multimedia and Expo, ICME, 2020, pp. 1–6.
[110] H. Wang, W. Gan, S. Hu, J.Y. Lin, L. Jin, L. Song, P. Wang, I. Katsavounidis, A. Aaron, C.-C.J. Kuo, MCL-jcv: A JND-based H.264/AVC video quality assessment dataset, in: 2016 IEEE International Conference on Image Processing, ICIP, 2016, pp. 1509–1513.
[111] H. Yang, H. Wang, L. Yu, J. Liang, T. Li, An adaptive quantization based PVC scheme for HEVC, in: 2020 Data Compression Conference, DCC, 2020, pp. 404–404.
[112] S. Nami, F. Pakdaman, M.R. Hashemi, Juniper: A jnd-based perceptual video coding framework to jointly utilize saliency and JND, in: 2020 IEEE International Conference on Multimedia Expo Workshops, ICMEW, 2020, pp. 1–6.
[113] X. Zhang, H. Wang, T. Tian, Perceptual video coding with block-level staircase just noticeable distortion, in: 2019 IEEE International Conference on Image Processing, ICIP, 2019, pp. 4140–4144.
[114] J. Kim, D.Y. Lee, S. Jeong, S. Cho, Perceptual video coding using deep neural network based JND model, in: 2020 Data Compression Conference, DCC, 2020, pp. 375–375.
[115] H. Liu, Y. Zhang, H. Zhang, C. Fan, X. Fan, Deep learning-based picture-wise just noticeable distortion prediction model for image compression, IEEE Trans. Image Process. 29 (2019) 641–656.
[116] K. Yang, S. Wan, Y. Gong, H.R. Wu, Y. Feng, Perceptual based SAO rate-distortion optimization method with a simplified JND model for H.265/HEVC, Signal Process., Image Commun. 31 (2015) 10–24.
[117] M. Zhou, X. Wei, S. Kwong, W. Jia, B. Fang, Just noticeable distortion-based perceptual rate control in HEVC, IEEE Trans. Image Process. 29 (2020) 7603–7614.
[118] C. Zhu, L. Song, R. Xie, J. Han, Y. Xu, JND-based perceptual rate distortion optimization for AV1 encoder, in: 2019 Picture Coding Symposium, PCS, 2019, pp. 1–5.
[119] K. Friston, J. Kilner, L. Harrison, A free energy principle for the brain, J. Physiol.-Paris 100 (1–3) (2006) 70–87.
[120] E. DoVale, High frame rate psychophysics: Experimentation to determine a JND for frame rate, SMPTE Motion Imaging J. 126 (9) (2017) 41–47.
[121] M. Lu, J. Lanir, C. Wang, Y. Yao, W. Zhang, O. Deussen, H. Huang, Modeling just noticeable differences in charts, IEEE Trans. Vis. Comput. Graphics 28 (1) (2022) 718–726.
[122] A. Haruvi, R.B. Kopito, N. Brande-Eilat, S. Kalev, E. Kay, D. Furman, Measuring and modeling the effect of audio on human focus in everyday environments using brain-computer interface technology, Front. Comput. Neurosci. 15 (2022).
[123] Q. Zhang, S. Wang, X. Zhang, S. Ma, W. Gao, Just recognizable distortion for machine vision oriented image and video coding, Int. J. Comput. Vis. 129 (10) (2021) 2889–2906.
[124] Q. Zhang, S. Wang, X. Zhang, C. Jia, J. Pan, S. Ma, W. Gao, SMR: Satisfied machine ratio modeling for machine recognition-oriented image and video compression, 2022, arXiv preprint arXiv:2211.06797.
[125] J. Jin, X. Zhang, X. Fu, H. Zhang, W. Lin, J. Lou, Y. Zhao, Just noticeable difference for deep machine vision, IEEE Trans. Circuits Syst. Video Technol. 32 (6) (2021) 3452–3461.
[126] J. Liu, Y. Cho, Z. Guo, J. Kuo, Bit allocation for spatial scalability coding of h.264/SVC with dependent rate-distortion analysis, IEEE Trans. Circuits Syst. Video Technol. 20 (7) (2010) 967–981.
[127] C. Pang, O.C. Au, F. Zou, J. Dai, An analytic framework for frame-level dependent bit allocation in hybrid video coding, IEEE Trans. Circuits Syst. Video Technol. 23 (6) (2013) 990–1002.
[128] S. Wang, S. Ma, S. Wang, D. Zhao, G. Wen, Rate-GOP based rate control for high efficiency video coding, IEEE J. Sel. Top. Sign. Proces. 7 (6) (2013) 1101–1111.