IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 31, 2019, accepted February 19, 2019, date of publication March 4, 2019, date of current version March 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901342

Visual JND: A Perceptual Measurement in Video Coding

DI YUAN, TIESONG ZHAO, (Member, IEEE), YIWEN XU, HONG XUE, AND LIQUN LIN

College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China

Corresponding author: Liqun Lin (lin_liqun@fzu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Project 61671152.

ABSTRACT Humans cannot perceive the minimal level of difference in the pixel variation. To overcome the problem, the concept of just-noticeable difference (JND) was proposed. JND measures the minimal amount that must be changed for the variation to be detectable by humans. However, JND characteristics were not considered in the traditional perceptual measurements. In this paper, we provide a comprehensive survey of the latest JND-related studies. First, we provide an extensive overview of JND models. JND models comprise human visual system characteristics and masking effects. Next, we introduce the applications of JND models in the perceptual quality evaluation and video compression coding, especially in applying machine-learning techniques to JND prediction. In addition to a thorough summary of the recent progress and applications of JND, we summarize some unsolved technical challenges. We believe that our overview and findings can provide some insights into the related issues and future research directions in video coding.

INDEX TERMS Just noticeable difference (JND), human visual system (HVS), machine learning, perceptual measurement, video coding.

I. INTRODUCTION

In recent years, there has been a considerable increase in the demand for multimedia, and streaming media has dominated a considerable part of our lives. According to the Cisco visual networking index [1], the global IP video traffic is expected to increase up to 82% of all consumer internet traffic in 2021, an increase from 73% in 2016. It is expected that more than one million minutes of video content will be transmitted over the network every second in 2021. Meanwhile, people watch videos from computers, mobile phones, and other multimedia display terminals. With continuous improvements in display terminal resolution, video quality requirements are also increasing. Consequently, the increase in multimedia demand prompts a disrupting request for optimizing video encoding systems, improving transmission bit rate, and saving bandwidth.

Many concepts regarding the quality of user experience have been proposed in the past decade. They include the peak-signal-to-noise-ratio (PSNR) [2], structural similarity (SSIM) index [3], visual information fidelity (VIF) [4], etc. These metrics are utilized to evaluate the perceived quality of videos. However, it is hoped that perception measurement can be more consistent with real sensation observed by human visual system (HVS), and the effects of human visual characteristics can be considered in perception measurement.

In psychology, the concept of just noticeable difference (JND) was presented. JND determines how accurate human sense are. As Weber's law [5] first suggested, JND was the change of a threshold value in order for a difference to be noticeable. Then, the concept of JND in visual perception was introduced. According to [6], JND represented the least detectable difference sensory stimulus between two levels. Similarly, in a previous study, JND referred to the maximum tolerable distortion for the HVS in perceiving images/videos [7]. Studies on JND have been performed, and several surveys have been presented. Zhang *et al.* [8] surveyed and drew a comparison between the existing JND models. Chen and Liu [9] presented a comprehensive review of the basic JND models and JND applications in quality measurement and video coding. According to [10], Lin *et al.* summarized the perceptual visual quality metrics (PVQMs) and discissed JND calculation modules based on PVQMs. Lin and Zhang [11] argued a JND-based method for allocating compression bit rate according to coding distortion.

As mentioned, previous surveys focused on traditional JND models and their developments. Through the reviews, we observe that the technologies combined with machine learning in JND modeling and application have attracted

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Nappi.

TABLE 1. Information of three typical databases.

Database	Sample	Resolution	Duration	Frame Rate	Format	Distortion Type
MCL-JCI	5050	1920 × 1080	/	/	/	JPEG
MCL-JCV	1560	1920 × 1080	5sec	24, 30	YUV420p	H.264/AVC
VideoSet	45760	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360	5sec	24, 30	YUV420p	H.264/AVC

FIGURE 1. Representative thumbnail images of 50 source images in the MCL-JCI dataset.

attention over the past two to three years. However, few literatures have presented JND surveys regarding machine learning. Consequently, performing a new and thorough overview of JND in perceptual measurement and perceptual visual coding is necessary.

The rest of this paper is organized as follows. We first introduce the proposed JND databases in Section II. Then, JND models are discussed in Section III. Section IV presents an overview of the application of JND models in perceptual quality assessment. Section V provides an overview of the application of JND models in perceptual video coding. We also analyze new challenges and future research directions. Finally, Section VI concludes the paper.

II. JND DATABASE

The traditional subjective test for images/videos is typically performed by a small number of experts known as golden eyes. However, encoding image/video statistics on visual perception experience by non-experts are scarce.

JND point is a measurement of users' maximum unnoticeable differences. In addition, JND points can be used to estimate the quality of the encoded images/videos [12]. As an evaluation metric, JND is more suitable than other metrics to evaluate human subjective perception. Therefore, the video quality can be better gauged by estimating JND points. However, JND points are obtained from databases. The benchmark datasets regarding perception are scarce.

To solve the problem of insufficient databases, several image/video quality assessment datasets based on JND have been developed in recent years, such as MCL-JCI [13], MCL-JCV [14] and VideoSet [15]. MCL-JCI is a JND dataset for JPEG images, and MCL-JCV is a JND dataset for H.264/Advanced Video Coding (AVC) videos. VideoSet is a large-scale compressed video quality dataset based on JND measurement. Table 1 provides an overview of the three databases.

A. MCL-JCI

The quality factor (QF) in JPEG is utilized to control the quality of a coded image. A higher QF value is indicative of better quality. However, humans cannot distinguish every difference in the image QF values from 1 to 100. Consequently, building a large-scale dataset of human perception for image quality accessment is critical.

The MCL-JCI dataset contains 50 source images, each having a resolution of 1920 × 1080. The representative thumbnail images are shown in Fig.1. Each source image was encoded 100 times with the JPEG encoder. The QF value was set between 1 and 100, for a total of 5050 images. Further, 150 subjects participated in the subjective experiment. The test environment was consistent. The viewing distance was 2 m (1.6 times the picture height) from the center of the monitor to the seat. The image pair was displayed on a 65″ television with a resolution of 3840 × 2160. During the experiment, JND was obtained by a dichotomy search. By analyzing and post-processing the original JND data, the staircase quality function (SQF) was accessed. Subsequently, the relationship between image content and SQF was analyzed. Machine-learning technology was adopted to predict the SQF curves based on image content. It is a new methodology for human visual experience measurement.

B. MCL-JCV

Humans cannot perceive the quality difference between two video sequences with the same content and extremely close quantization parameters (QPs), even though their PSNR is relatively different. Consequently, a compressed video quality assessment dataset based on the JND model (MCL-JCV) was established.

The MCL-JCV database consists of 30 source video sequences with a resolution of 1902×1080. The representative thumbnail images are shown in Fig.2. The H.264/AVC

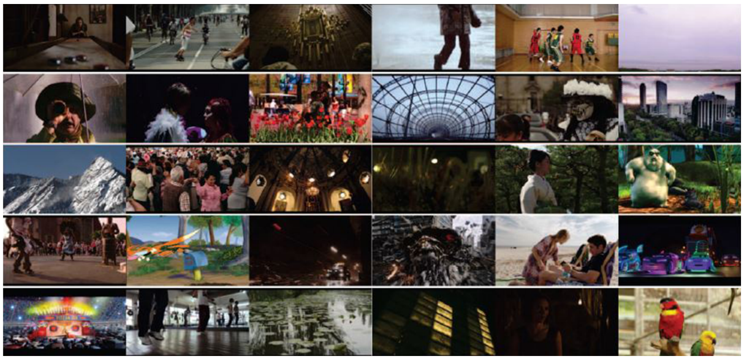


FIGURE 2. Representative thumbnail images of 30 source sequences in the MCL-JCV dataset.

TABLE 2. Video sequences formats in the videoSet.

Source(Resolution)	Total	Frame Rate	Pixel Format	Down-Sampling Resolution
EI Fuente(4096 × 2160)	31	30	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Chimera(4096 × 2160)	59	30	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Ancient Thought(3840 × 2160)	11	24	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Eidorado(3840 × 2160)	14	24	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Indoor Untouched(3840 × 2160)	5	24	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Lift Untouched(3840 × 2160)	15	30	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Lifting Off(3840 × 2160)	13	24	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Moment of Intensity(3840 × 2160)	10	30	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Skateboarding(3840 × 2160)	9	24	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Unspoken Friend(3840 × 2160)	13	24	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360
Tears of Steel(4096 × 1714)	40	24	YUV420p	1920 × 1080, 1280 × 720, 960 × 540, 640 × 360

encoder was employed to compress the video sequences with QP values from 1 to 50. Thus, there were 1560 encoded sequences. Fifty subjects participated in the subjective test. In the testing phase, all subjects were asked to watch two video sequences in succession and compare the noticeable differences between them. The bisection method was utilized to obtain the JND points. With the MCL-JCV dataset, the SQF curves were obtained from the collected JND data. In particular, the SQF of the video content may be predicted by machine-learning techniques in the near future.

C. *VideoSet*

Although the PSNR is widely accepted and the most popular metric in video quality measurement, it is not completely related to a human's subjective visual experience. However, JND measurements require considerable efforts in subjective evaluation tests. If large-scale datasets exist in perceptual video coding (PVC), it is possible to adopt machine-learning technology to predict the JND values in a short time and establish a suitable JND model. It is imperative to collect large-scale subjective test datasets on the perceived video quality. Hence, VideoSet was constructed.

The VideoSet comprises 220 5-s sequences, each with four resolutions (1920 × 1080, 1280 × 720, 960 × 540, and 640 × 360). Thus, there were 880 uncompressed video sequences in the dataset. Each video was encoded with 52 QP values from 0 to 51. Table 2 provides a detailed introduction for the database sequences. More than 30 subjects participated in the subjective experiment. The first three JND points were measured. The subjective experimental environment was established according to the ITU-R BT.2022 [16] recommendation. Each subject compared the quality of two consecutive clips and determined whether the two sequences were significantly different. Thus, the JND points were observed by a more robust binary search process. Unreliable subjects were removed based on the statistical procedures for subject screening provided in ITU-R BT.1788 [17]. Subsequently, Grubbs' test was employed to detect and remove the outliers. The distributions of 3 JND points in 220 sequences were provided in VideoSet. It is challenging to apply machine- learning technology to VideoSet for accurate and effective JND prediction in a short time. However, it is an essential step to make data-driven perceptual coding practical for real-world applications.

D. *SUMMARY AND DISCUSSION*

MCL-JCI and MCL-JCV are two small-scale JND-based image/video quality datasets. They are targeted the JPEG image and the H.264/AVC video, respectively. MCL-JCI contains 50 source (or uncompressed) images. MCL-JCV includes 30 source (uncompressed) video sequences covering a wide diversity of video characteristics. VideoSet is a large-scale JND-based coded video quality dataset. All source/coded video clips as well as measured JND data are included in the VideoSet. It comprises two most dominant video formats (1080p and 720p). While 540p and 360p are included to capture the viewing experience on tablets or mobile phones. It is challenging to apply machine-learning technology to VideoSet for accurate and effective JND prediction in a short time. As a result, there is a call for JND databases. Currently, the lack of databases is mainly in two aspects: only indirect verification possible and no learning-based method. Once a sufficient amount of data are collected, it is possible to use the machine-learning technique to predict the JND value in a short time.

III. JND MODELING WITH DIFFERENT FACTORS

HVS provides a number of features, including contrast sensitivity [18], brightness masking [19], contrast masking [20], foveation features [21], step perception of stimuli and JND. Some new models have been formed in image/video perceived compression by combining JND with some features of the HVS. These models can optimize image/video encoding parameters, reduce transmission bit rate, and evaluate video perceived quality from different perspectives.

According to the concept of JND, the influence of JND on perceptual measurement cannot be ignored. Furthermore, determining JND is complicated as it is affected by multiple factors. In general, JND modeling can be classified into two categories: pixel domain (also known as spatial domain JND) and transform domain [10].

A. PIXEL DOMAIN JND

Chou and Li [22] proposed the pixel domain JND and analyzed the background luminance and edge masking effect. Yang *et al.* [23] reported that Chou's model overestimated the JND threshold of the edge region. Consequently, they applied the Canny edge detection to protect the JND estimates of the edge areas.

Two factors are considered in the JND model of the pixel domain: luminance masking effect and contrast masking effect. A new method of adding the edge mask effect and texture mask effect to JND model was previously proposed in [24]. The research demonstrated that JND thresholds were often underestimated in textured areas; hence, they decomposed the textured areas and non-textured areas for JND estimation. Wu *et al.* [25] introduced the effect of disorderly concealment on JND estimation. In summary, all proposed JND models were designed to improve the accuracy of JND estimation. However, they were still less successful than the human perception.

In [26], the mathematical expressions of various factors for JND models in the pixel domain were described in detail. The overall JND model was constructed. Its basic form is as follows:

$$JND = LA + CM + CSF - p \times min(LA, CM, CSF) \tag{1}$$

where LA denotes the luminance adaptation. CM represents contrast masking. CSF represents the contrast sensitivity function. p is the gain reduction factor that can compensate for the overlapping between masking factors. Its value is set as 0.3.

B. TRANSFORM DOMAIN JND

Video encoding often combines transformation and quantization, therefore, JND models in the transform domain are more popular than those in the pixel domain. In frequency-domain JND, images/videos should be converted to the frequency domain (such as DCT [27] or wavelet transform [28]), before subsequent analysis and processing. The basic JND model in the DCT domain was introduced in [29]. The formula is as follows.

$$T_{JND}(k, n, i, j) = T_{JNDs}(k, n, i, j) \times F_T(K, N, I, J) \tag{2}$$

where k is the index of a frame in video sequences. n is the index of a block in the kth frame. i and j are the DCT coefficients' indices. T_{JNDs} (k, n, i, j) is the spatial JND. F_T (k, n, i, j) is the temporal modulation factor.

In addition to the aforementioned basic JND models, many other JND models are affected by other factors, which are described in detail below. The traditional JND models are definitely unsuitable for the perceptual quality measure for stereo videos. There are many JND models combining the stereo video features and binocular visual features. In [30], it was argued that the transmission bandwidth of stereoscopic videos/images was closely related to binocular perception. The study of Aflaki *et al.* [31] indicated that stereoscopic perception was dominated by high-quality views, that it was not the same as two-dimensional (2D) videos. Thus, new JND models were developed for stereo vision perception. Zhao *et al.* [32] first built a three-dimensional (3D)-JND model, also called the BJND model. The BJND model added some unique binocular visual features to the traditional models, such as binocular combination and rivalry. Qi *et al.* [33] created a stereo JND (SJND) model. The SJND model regarded the intra-view and inter-view masking effects as well as luminance adaptation and spatial masking. In addition to stereo video features and human visual characteristics, some other models also take into account the influence of external environment factors. Silva *et al.* [34] established the JND in depth (JNDD) model. The JNDD model argued the human perception threshold for the depth variation of 3D videos, and considered the affected factors. The factors influenced the assessment of depth, such as viewing distance and display depth-level.

HVS perception demonstrated the highest spatial resolution and sensitivity at the point of fixation, called the fovea area [35]. The resolution/sensitivity decreased significantly with increasing eccentricity. Thus, the foveal provided precise and detailed visual content. In [36], a foveation model and foveated JND (FJND) model were described. Subsequently, a feature based on the non-uniform sampling of the human retina was suggested to provide high-quality videos at a reduced bit rate. In [37], it is proposed to add visual attention model to the FJND model. The new model utilizes

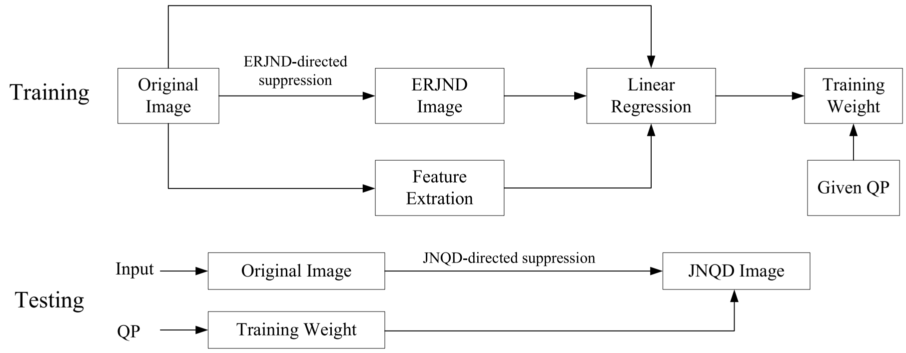


FIGURE 3. Block diagram of LR-JNQD model.

visual attention and foveated masking characteristics to form a saliency map based on the Gaussian Mixed Model. The visual attention model can reveal saliency areas where our eyes focus when viewing a video. Wang *et al.* [38] developed a new saliency-JND model. It combined visual attention model and visual sensitivity model. The JND threshold in the salient area was decreased, and the threshold in the non-salient area was increased. The proposed model reduced the bit rate in high efficiency video coding (HEVC), and had nearly the same subjective perceptual quality.

C. JND MODEL BASED ON MACHINE LEARNING

In addition to the aforementioned JND models, other JND models are combined with machine-learning techniques. A new DCT-based energy JND model (ERJND) was recommended [39]. The prototype of ERJND is a DCT-based contrast masking (CM)-JND model. The CM-JND model established the correlation between the DCT coefficient suppression amplitude estimation and the DCT block texture complexity by regression. ERJND reduced the energy of DCT coefficients that affected quantification. In the ERJND model, the scale factor was added to obtain the just noticeable quantization distortion (JNQD) model. The scale factor adjusted the ERJND-directed suppression levels according to the quantization step size during preprocessing. ERJND was extended by two quantitative distortion JNQD models, namely LR-JNQD and CNN-JNQD.

LR-JNQD involves a linear regression based on JNQD that adjusts the boundary between the two cases. These two cases are the true perceptual distortion threshold value for a non-compression case and a suppression level in preprocessing for a compression case. The adjustment yielded the encoder quantization of the JNQD level bit rate optimization. Figure. 2-C shows the training and testing blocks of LR-JNQD. However, LR-JNQD has its limitations; for example, it cannot conduct precise modeling. To overcome this limitation, CNN-JNQD was modeled to directly suppress images using adaptive quantization step-size. The CNN-JNQD is a convolutional neural network training JNQD with given QPs. CNN-JNQD model includes three layers of CNNs; each convolution layer is followed by a rectified linear unit (ReLU) layer. Figure. 3-A shows the training and testing blocks of CNN-JNQD. CNN-JNQD preprocessed the compressed input with lower perceptual distortion. The proposed LR-JNQD and CNN-JNQD exhibit better bit rate reductions than Bae's PVC scheme, and exhibit better performance. In addition, the CU block encoded by the CNN-JNQD model is larger than that by the LR-JNQD model with fewer bits. The models further optimize the PVC scheme, and effectively remove the perceived redundancy. The experiments showed that both LR-JNQD and CNN-JNQD models were applied to HEVC, and the maximum(average) bitrates respectively reduced by 38.51% (10.38%) and 67.88% (24.91%), with little subjective video quality degradation. Therefore, the JND models based on machine learning perform better than traditional JND models.

In addition, visual saliency detection has attracted increasing attention of researchers in recent years. In [40], the CNN-based saliency estimation was proposed. It adopted deep CNN to extract features predicting the saliency of objects. The combination model of saliency and JND has been introduced. In future research, the CNN-based saliency prediction may be combined with the JND model to get a better performance.

D. SUMMARY AND DISCUSSION

In this section, we discuss the JND models in the pixel and transform domains. Although our review may not include all the recent investigations in this area, the basic framework and research status are separately introduced. Through a literature review, we observe that most improved JND models are based on the DCT domain, and barely few are based on the pixel domain. The JND models of pixel domain can give the JND threshold intuitively. Thus, they are mainly applied in motion estimation, image preprocessing and image enhancement. However, the JND threshold was underestimated in the edge regions and texture regions. In transform domain, the estimation of contrast masking and texture characteristics was more accurate. Since the DCT is widely used in video processing, the JND models in the transform domain can be better applied in compression coding. Therefore, the JND model in transform domain is more effective in quality evaluation,

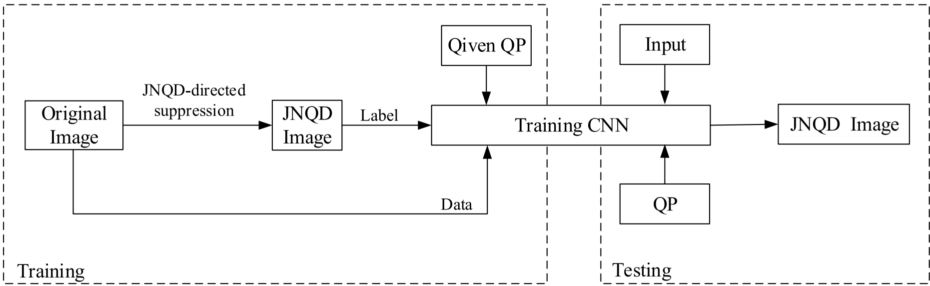


FIGURE 4. Block diagram of CNN-JNQD model.

quantization control for compression and data hiding. However, almost all DCT-based JND models fail to take quantization effects into account in the JND modeling. As a result, JND suppression cannot be accurately performed during the quantization process. Finally, we review machine-learning applications in JND. The energy-reduced JND (ERJND) is a new DCT-based model by taking into account the quantization effects. It performs effective perceptual redundancy removal. We believe that more relevant investigations will be conducted in the near future. It is crucial to improve the performance of JND models and develop perceptual coding.

IV. JND APPLICATION IN PERCEPTUAL QUALITY ASSESSMENT

The JND model is critical in video quality perception measurement. A variety of quality assessment methods for perceptually compressed video were mentioned in [41]. The characteristics of visual masking and contrast sensitivity are typically measured by the JND model. Thus, the relationship between JND and perceptual measurement is particularly important.

Based on extracted video features, Huang *et al.* [42] studied a machine-learning method to predict the mean of the JND distribution. It utilized the masking effect and designed a spatial-temporal sensitive map to capture the unique characteristics of the source content. A novel JND measurement and prediction methodology was explored, realizing perceptual lossless coding of HEVC coding sequences. JND prediction can be applied to perceptual lossless coding and subjective evaluation.

Some latest literatures were based on the VideoSet database. According to [43], a flexible user model was established. It considered the subject and content factors in the JND framework. The JND data was from the VideoSet during the experiment. The proposed model can predict SUR distribution of a specific user group, and provide some insights on the quality assessment problem. According to the conclusion of [44], video quality was evaluated by the distribution of the satisfactory user ratio (SUR) curve. Furthermore, in [45], a machine-learning framework was used to predict SUR based on the VideoSet. Figure. 4 illustrates the framework of SUR prediction. First, the quality of video clips was evaluated by using the video multi-method assessment fusion (VMAF) quality index. Then, combined with masking effect, the SUR curves were predicted by using the support vector regression (SVR). Thereby, JND points were predicted. Another study [46] reported the improvement in the VMAF framework. The VMAF included multiple quality perception features. The VideoSet dataset was employed to train the model and predict the video quality.

Quality of experience (QoE) represents the user-perceived quality experience of videos. JND indicates the visual sensitivity of eyes. JND is usually utilized as a metric for QoE to estimate whether a user can perceive differences in video quality. At present, there have been a lot of researches on CNN-based QoE prediction. Moreover, the JND is as a metric to predict QoE.

Zhang *et al.* [47] explored a framework for predicting the video QoE based on machine learning, called DEEPQoE. Figure. 4 shows the basic structure of the framework. First, a 3D convolutional neural network was used to extract the video features, whereas GloVe and linear filters were used to process the text and value information. Subsequently, learning by the deep neural network, a linear function and softmax function were obtained, and the functions were adopted to predict the video QoE. The performance evaluation of the DEEPQoE framework was based on VideoSet, and JND was used as a QoE metric. The result illustrated that the DEEPQoE model performed better at 90.94% accuracy, which was higher than that obtained using the comparison methods.

The current investigations on the combination of quality evaluation and machine learning are primarily based on the aforementioned databases. The predicted JND points were obtained by training the aforementioned datasets. Compared with the traditional JND models, the prediction results are closer to the human subjective reality perception.

V. JND APPLICATION IN PERCEPTUAL VIDEO CODING

The HVS is highly complex [48]. It is a challenging task to combine human visual perception with the coding system in perceptual coding. The primary purpose is to minimize the coding bit rate under the condition of guaranteeing a certain video quality or to minimize the coding distortion under the condition of a certain coding bit rate limitation. The perceived redundancy of the HVS can be effectively reduced

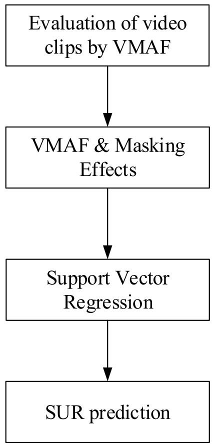


FIGURE 5. Illustration of SUR prediction framework.

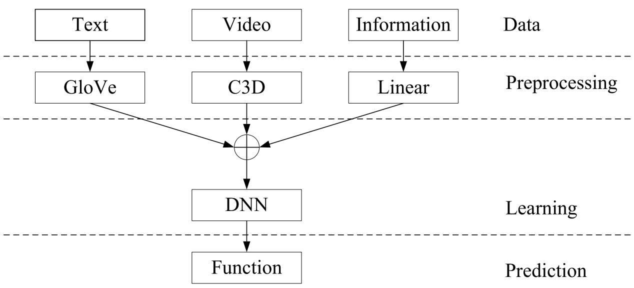


FIGURE 6. Illustration of DEEPQoE framework.

by perceptual coding. By removing the perceptual redundancy information to the JND levels, compression gain can be further realized, compared with traditional video coding methods [49].

A. RATE-DISTORTION OPTIMIZATION

Rate-distortion (RD) optimization is crucial in video compression algorithms [50]. RD optimization technology is the primary method of ensuring the coding efficiency of the encoder. In [51], the JND perception model was applied to the H.264/AVC codec to improve the RD performance and overall compression efficiency. The JND model was employed to control the quantization of redirecting residuals at the encoder, and the JND threshold for inverse quantization was estimated at the decoder. The proposed perceptual codec improved the RD performance, and the overall compression efficiency, which are the primary purposes of the proposed JND models.

B. QUANTIZATION PROCESS

The JND model is often applied to guide the quantization process in video coding. In [52], the idea of local JND was presented and introduced into the HEVC system. This method can adapt to the size of various coding units (CUs), realizing better bit allocation of the video coding system and the bit rate reduction based on ensuring the perceived quality. Meanwhile, Luo *et al.* [53] improved the H.264/AVC framework based on the JND-directed suppression tool, and realized precise quantification adjustment using the JND normalization error model.

The study of Bai *et al.* [54] first introduced the spatial-temporal JND model into multiple description coding, and only encoded the imperceptible visual information within the JND threshold. Finally, it obtained better bit rate and perceptual visual distortion performance.

The DCT-based JND profiles aforementioned have been widely applied in human perception-based video coding. The study of Vidal *et al.* [55] utilized perceptual pre-filtering for the rate-quality optimization of video encoding. Subsequently, a JND model was explored to adaptively control the filtering intensity and smooth the imperceptible visual details. This method reduced the bit rate without affecting the quality of video perception. However, two problems were discovered: 1) the estimation of the JND threshold was for fixed size blocks; 2) the DCT coefficient of the residual was too small to be sufficiently suppressed [56]. Thus, the DCT-based local distortion detection probability model was constructed. Subsequently, Xiang *et al.* [57] developed an adaptive preprocessing method to solve the problem of the coarse quantization of block DCT coefficients, and established a new filter model based on the JND filter (JNDF) and adaptive bilateral filter.

C. SUMMARY AND DISCUSSION

In this section, we discuss the application of JND in video coding. Some visual information cannot be perceived in the JND threshold, and can be regarded as redundant information eliminated during preprocessing. More efficient coding can be achieved by better redundancy allocation. Using JND threshold values to adjust quantization parameters, lower bit rate can be obtained with rarely visible distortions compared with traditional quantitative methods. Although there are many studies regarding JND models in coding modules, methods combining machine learning are few. ER-JNQD model and CNN-JNQD model combine with machine learning techniques, but both models are applied to image coding. At present, relevant applications of video coding have not been proposed. By combining machine-learning technology with the JND model, the video coding scheme will become complex. Thus, time cost is a problem to be considered. If the existing coding framework applying machine-learning technology is improved, the perceptual coding performance will be improved significantly. It is the future research trend and challenge in perceptual video coding.

VI. CONCLUSION

In this paper, we performed an extensive review on JND. JND facilitates decision making in our work and study. Utilizing JND can turn human imperfectness to system advantages, and make human visual experience more realistic. Our survey reviewed the studies performed in two areas:

perceptual coding and quality assessment system, especially in research methods combined with machine-learning techniques. Therefore, it is necessary to built the JND databases for the application of machine-learning techniques. JND can further improve the algorithm accuracy and performance. In addition, Augmented Reality (AR), Virtual Reality (VR), and panoramic video are also constantly evolving. They remain challenging tasks in JND applications for future research. Combining visual JND with other senses (*i.e.* sound, touch, smell, and taste) to form multi-sensory JND is an extension for new possibilities. To represent the most recent developments in JND, we reviewed nearly 60 technical papers. It could provide popular research trends and the most contributive studies for both newcomers and current researchers. We expect that our survey can not only promote the research activities in JND perceptual measurement, but also provide inspirations for future research in the next generation video perceptual coding standard.

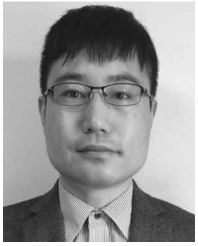
REFERENCES

[1] *Cisco Visual Networking Index: Forecast and Methodology, 2016–2021*. Accessed: 2018. [Online]. Available: https://www.cisco.com /c/en/us/solutions/collate-ral/service-provider/visual-networking-index-vni/complete-white-paperc11-481360.html
[2] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.
[3] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.
[4] H. R. Sheikh and A. C. Bovik, "A visual information fidelity approach to video quality assessment," in *Proc. 1st Int. Workshop Video Process. Qual. Metrics Consum. Electron.*, 2005, pp. 23–25.
[5] *Weber's Law of Just Noticeable Difference*. [Online]. Available: http://apps.usd.edu/coglab/WebersLaw.html
[6] D. C. Knill and A. Pouget, "The Bayesian brain: The role of uncertainty in neural coding and computation," *Trends Neurosci.*, vol. 27, no. 12, pp. 712–719, 2004.
[7] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, no. 10, pp. 1385–1422, Oct. 1993.
[8] X. H. Zhang, W. S. Lin, and P. Xue, "A new DCT-based just-noticeable distortion estimator for images," in *Proc. 4th Int. Conf. Inf., Commun. Signal Process.*, vol. 1, Dec. 2003, pp. 287–291.
[9] Z. Chen and H. Liu, "JND modeling: Approaches and applications," in *Proc. 19th Int. Conf. Digit. Signal Process.*, Aug. 2014, pp. 827–830.
[10] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.
[11] Y.-B. Lin and X.-M. Zhang, "Recent developments in perceptual video coding," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit.*, Jul. 2013, pp. 259–264.
[12] Y. C. Lin, "Experimental design and analysis of JND test on coded image/video," *Proc. SPIE*, vol. 9599, Sep. 2015, Art. no. 95990Z.
[13] L. Jin *et al.*, "Statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis," *Electron. Imag.*, vol. 13, pp. 1–9, Feb. 2016.
[14] H. Wang *et al.*, "MCL-JCV: A JND-based h.264/AVC video quality assessment dataset," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1509–1513.
[15] H. Wang *et al.*, "VideoSet: A large-scale compressed video quality dataset based on JND measurement," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 292–302, Jul. 2017.
[16] *General Viewing Conditions for Subjective Assessment of Quality of SDTV and HDTV Television Pictures on Flat Panel Displays*, document ITU-R BT.2022, 2012.
[17] *Methodology for the Subjective Assessment of Video Quality in Multimedia Applications*, document ITU-R BT.1788, 2007.
[18] D. H. Kelly, "Visual contrast sensitivity," *Optica Acta, Int. J. Opt.*, vol. 24, no. 2, pp. 107–129, 1977.
[19] M. A. Paradiso and K. Nakayama, "Brightness perception and filling-in," *Vis. Res.*, vol. 31, nos. 7–8, pp. 1221–1236, 1991.
[20] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1458–1471, Dec. 1980.
[21] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 129–132, Mar. 2002.
[22] 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 Technol.*, vol. 5, no. 6, pp. 467–476, Dec. 1995.
[23] X. K. Yang, W. S. Ling, 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, Aug. 2005.
[24] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010.
[25] J. Wu, G. Shi, W. Lin, A. Liu, and F. Qi, "Just noticeable difference estimation for images with free-energy principle," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1705–1710, Nov. 2013.
[26] M. Uzair and R. D. Dony, "Estimating just-noticeable distortion for images/videos in pixel domain," *IET Image Process.*, vol. 11, no. 8, pp. 559–567, Aug. 2017.
[27] Y. Jia, W. Lin, and A. A. Kassim, "Estimating just-noticeable distortion for video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 820–829, Jul. 2006.
[28] X. Zhang, W. Lin, and P. Xue, "Just-noticeable difference estimation with pixels in images," *J. Vis. Commun. Image Represent.*, vol. 19, no. 1, pp. 30–41, Jan. 2008.
[29] 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 Technol.*, vol. 19, no. 3, pp. 337–346, Mar. 2009.
[30] L. B. Stelmach, W. J. Tam, D. V. Meegan, A. Vincent, and P. Corriveau, "Human perception of mismatched stereoscopic 3D inputs," in *Proc. Int. Conf. Image Process.*, vol. 1, Sep. 2000, pp. 5–8.
[31] P. Aflaki, M. Hannuksela, J. Häkkinen, P. Lindroos, and M. Gabbouj, "Subjective study on compressed asymmetric stereoscopic video," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4021–4024.
[32] Y. Zhao, Z. Chen, C. Zhu, Y.-P. Tan, and L. Yu, "Binocular just-noticeable-difference model for stereoscopic images," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 19–22, Jan. 2011.
[33] F. Qi, D. Zhao, X. Fan, and T. Jiang, "Stereoscopic video quality assessment based on visual attention and just-noticeable difference models," *Signal, Image Video Process.*, vol. 10, no. 4, pp. 737–744, 2016.
[34] D. V. S. X. De Silva, E. Ekmekcioglu, W. A. C. Fernando, and S. T. Worrall, "Display dependent preprocessing of depth maps based on just noticeable depth difference modeling," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 2, pp. 335–351, Apr. 2011.
[35] A. Liu, M. Verma, and W. Lin, "Modeling the masking effect of the human visual system with visual attention model," in *Proc. 7th Int. Conf. Inf., Commun. Signal Process. (ICICS)*, Dec. 2009, pp. 1–5.
[36] Z. Chen and C. Guillemot, "Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 806–819, Jun. 2010.
[37] H. Wang *et al.*, "A novel foveated-JND profile based on an adaptive foveated weighting model," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process.*, Dec. 2018, pp. 12–17.
[38] H. Wang, L. Wang, X. Hu, Q. Tu, and A. Men, "Perceptual video coding based on saliency and just noticeable distortion for H.265/HEVC," in *Proc. Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Sep. 2014, pp. 106–111.
[39] S. Ki, S.-H. Bae, M. Kim, and H. Ko, "Learning-based just-noticeable-quantization-distortion modeling for perceptual video coding," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3178–3193, Jul. 2018.
[40] A. Azaza and A. Douik, "Deep CNN features for visual saliency estimation," in *Proc. 15th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Mar. 2018, pp. 688–692.
[41] F. Zhang and D. R. Bull, "Quality assessment methods for perceptual video compression," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 39–43.
[42] Q. Huang, H. Wang, S. C. Lim, H. Y. Kim, S. Y. Jeong, and C.-C. J. Kuo, "Measure and prediction of hevc perceptually lossy/lossless boundary Qp values," in *Proc. Data Compress. Conf. (DCC)*, Apr. 2017, pp. 42–51.

[43] H. Wang *et al.*, "A user model for JND-based video quality assessment: Theory and applications," *Proc. SPIE, Appl. Digit. Image Process. XLI*, vol. 10752, Sep. 2018.
[44] H. Wang, X. Zhang, C. Yang, and C.-C. J. Kuo, "Analysis and prediction of JND-based video quality model," in *Proc. Picture Coding Symp. (PCS)*, Jun. 2018, pp. 278–282.
[45] H. Wang, I. Katsavounidis, Q. Huang, X. Zhou, and C.-C. J. Kuo, "Prediction of satisfied user ratio for compressed video," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2018, pp. 6747–6751.
[46] C. G. Bampis, Z. Li, and A. C. Bovik, "Spatiotemporal feature integration and model fusion for full reference video quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2018.
[47] H. Zhang, H. Hu, G. Gao, Y. Wen, and K. Guan, "Deepqoe: A unified framework for learning to predict video QoE," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.
[48] Z. Chen, W. Lin, and K. N. Ngan, "Perceptual video coding: Challenges and approaches," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2010, pp. 784–789.
[49] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 742–752, Jun. 2005.
[50] Z. Pan, J. Lei, Y. Zhang, X. Sun, and S. Kwong, "Fast motion estimation based on content property for low-complexity H.265/HEVC encoder," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 675–684, Sep. 2016.
[51] M. Naccari and F. Pereira, "Decoder side just noticeable distortion model estimation for efficient H. 264/AVC based perceptual video coding," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 2573–2576.
[52] W.-W. Chao, Y.-Y. Chen, and S.-Y. Chien, "Perceptual HEVC/H. 265 system with local just-noticeable-difference model," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2016, pp. 2679–2682.
[53] Z. Luo, L. Song, S. Zheng, and N. Ling, "H.264/advanced video control perceptual optimization coding based on JND-directed coefficient suppression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 935–948, Jun. 2013.
[54] H. Bai, W. Lin, M. Zhang, A. Wang, and Y. Zhao, "Multiple description video coding based on human visual system characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 8, pp. 1390–1394, Aug. 2014.
[55] E. Vidal, N. Sturmel, C. Guillemot, P. Corlay, and F.-X. Coudoux, "New adaptive filters as perceptual preprocessing for rate-quality performance optimization of video coding," *Signal Process., Image Commun.*, vol. 52, pp. 124–137, Mar. 2017.
[56] S.-H. Bae, J. Kim, and M. Kim, "HEVC-based perceptually adaptive video coding using a DCT-based local distortion detection probability model," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3343–3357, Jul. 2016.
[57] G. Xiang, H. Jia, J. Liu, B. Cai, Y. Li, and X. Xie, "Adaptive perceptual preprocessing for video coding," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2016, pp. 2535–2538.

DI YUAN received the B.S. degree in electronic information engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2017. She is currently pursuing the M.S. degree in signal and information processing with Fuzhou University, China.

Her research interests include visual quality assessment and perceptual video coding.



TIESONG ZHAO received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2006, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2011.

He was a Research Associate with the Department of Computer Science, City University of Hong Kong, from 2011 to 2012; a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, from 2012 to 2013; and a Research Scientist with the Ubiquitous Multimedia Laboratory, The State University of New York at Buffalo, until 2015. He is currently a Professor with the College of Physics and Information Engineering, Fuzhou University, China. His research interests include image/video signal processing, visual quality assessment, and video coding and transmission.

YIWEN XU received the Ph.D. degree from the Department of Electronic Engineering, Xiamen University, Xiamen, China, in 2012.

He has been an Associate Professor with the College of Physics and Information Engineering, Fuzhou University, China, since 2013. His research interests include multimedia information processing, video codec and transmission, and video quality assessment.

HONG XUE received the B.S. degree in electronic information engineering from Fujian Normal University, China, in 2017. She is currently pursuing the M.S. degree in communication and information system with Fuzhou University, China.

Her research interests include visual and tactile interaction.

LIQUN LIN received the M.S. degree in communication and information system from Fuzhou University, China, in 2007, where she is currently pursuing the Ph.D. degree in communication and information system.

She has been a Lecturer with the College of Physics and Information Engineering, Fuzhou University, since 2007. Her research interests include image/video signal processing, visual quality assessment, and video coding and transmission.