

# Experimental Design and Analysis of JND Test on Coded Image/Video

Joe Yuchieh Lin<sup>a</sup>, Lina Jin<sup>a</sup>, Sudeng Hu<sup>a</sup>, Ioannis Katsavounidis<sup>b</sup>, Zhi Li<sup>b</sup>, Anne Aaron<sup>b</sup> and C.-C. Jay Kuo<sup>a</sup>

<sup>a</sup>University of Southern California, Los Angeles, California, USA

<sup>b</sup>Netflix, Los Gatos, California, USA

## ABSTRACT

The visual Just-Noticeable-Difference (JND) metric is characterized by the detectable minimum amount of two visual stimuli. Conducting the subjective JND test is a labor-intensive task. In this work, we present a novel interactive method in performing the visual JND test on compressed image/video. JND has been used to enhance perceptual visual quality in the context of image/video compression. Given a set of coding parameters, a JND test is designed to determine the distinguishable quality level against a reference image/video, which is called the anchor. The JND metric can be used to save coding bitrates by exploiting the special characteristics of the human visual system. The proposed JND test is conducted using a binary-forced choice, which is often adopted to discriminate the difference in perception in a psychophysical experiment. The assessors are asked to compare coded image/video pairs and determine whether they are of the same quality or not. A bisection procedure is designed to find the JND locations so as to reduce the required number of comparisons over a wide range of bitrates. We will demonstrate the efficiency of the proposed JND test, report experimental results on the image and video JND tests.

**Keywords:** Just-noticeable distortion, video compression, image compression, subjective quality assessment

## 1. INTRODUCTION

The mean-squared-error (MSE) has been widely used in measuring the distortion of compressed image and video content. It offers a continuous-scale distortion measure as a function of coding parameters, such as the bitrate, the quality parameter (QP) in video coding or the quantization factor (QF) in image coding, since it is computed based on pixel differences. Recent efforts have been made to develop new objective quality/distortion metrics that are better correlated with the human visual system (HVS).<sup>1</sup> All of them provide a quality measure of continuous-scale. However, it is well known that the HVS cannot perceive small changes in pixel differences. In reality, humans cannot perceive continuous-scale but discrete-scale quality changes over a range of coding bitrates. In this work, we quantify this phenomenon based on the notion of just-noticeable difference (JND),<sup>2</sup> and propose a new methodology to characterize the human visual experience on coded visual content.

JND is a statistical quantity that accounts for the maximum difference unnoticeable to a human being. It has been extensively studied to understand human visual sensitivity.<sup>3</sup> In the context of image/video compression, Watson<sup>4</sup> proposed a way to use JND for video quality measurement. His work adopts “pair comparison” or “two-alternative forced-choice”. That is, subjects are asked to determine which of two videos (*i.e.* the original source and the compressed one) is more distorted. Then, the distortion level of compressed video can be derived from the JND test result. Although Watson’s pioneering work offers a statistical relationship between visual quality assessment and JND, his result is difficult to apply to a real-world video coding system. Furthermore, the test duration required for each subject is long. Recently, several JND estimators based on HVS properties were investigated in.<sup>5,6</sup> It was also shown in<sup>7-9</sup> that JND-guided coding schemes can achieve perceptually similar quality with lower bitrates.

---

Further author information: (Send correspondence to J. Y. Lin)  
J. Y. Lin: E-mail: lin650@usc.edu

In this work, we study the problem of coded image/video quality assessment from a new angle based on JND. Given coded image/video content with densely sampled QF or QP values, we would like to measure the number of JND points and locations, and analyze the inter-person variance of measured quantities. One major contribution of this work is to develop a new methodology to achieve these tasks. It is confirmed by a small-scale subjective test that human perceived quality of coded images can be characterized by a piecewise constant function of the QF/QP with discontinuities at JND locations. Although these locations are content-dependent and statistically distributed, they do provide consistent and useful information in understanding the human visual experience.

The rest of this paper is organized as follows. The JND-based quality assessment problem is defined and its solution methodology is described in Section 2. The process of JND data collection is presented and its statistics are given in Section 3. The JND data post-processing technique and the final output quality plot as a function of the QF/QP are detailed in Section 4. Finally, concluding remarks are given in Section 5.

## 2. PROBLEM STATEMENT AND SOLUTION METHODOLOGY

### 2.1 Problem Formulation

To conduct the JND test on coded images, the first step is to collect a proper set of diverse, high-resolution images. Diversity can be characterized by their spatial complexity index (called the ‘‘Spatial Information’’ index and denoted by SI), colorfulness index (denoted by CF),<sup>10</sup> and semantic properties (e.g. a human object, etc.). We select five source images in our preliminary test as shown in Fig. 1. The SI index increases (or become more complex) from DB, RP, FT, CC to HS while the CF index increases (or become more colorful) from HS, RP, DB, CC to FT.



Figure 1: Five images selected for the JND test: (a) Color Checker (CC),<sup>11</sup> (b) Dark Building (DB),<sup>11</sup> (c) Food Truck (FT),<sup>12</sup> (d) Houses (HS),<sup>11</sup> and (e) Railway Platform (RP).<sup>11</sup>

Each source image is encoded by the JPEG encoder<sup>13</sup> with densely sampled bitrates. In the experiment, we encode each image 100 times with  $QF = 1, 2, \dots, 99, 100$ , where QF controls the scaling factor (denoted by  $S$ ) via

$$S = \begin{cases} \frac{5000}{QF}, & \text{if } QF > 50 \\ 200 - 2QF. & \text{otherwise} \end{cases} \quad (1)$$

The default quantization table (denoted by  $T$ ) and the matrix of quantization step (denoted by  $Q$ ) is related by the scaling factor  $S$  as

$$Q(i, j) = \frac{T(i, j) \times S + 50}{100}, \quad (2)$$

where  $Q(i, j)$  and  $T(i, j)$  are the  $(i, j)^{th}$  entry of  $Q$  and  $T$ , respectively. Therefore, a larger QF value results in a smaller quantization step size and, hence, a higher coding bitrate. Since the quality of a coded image is a monotonically increasing function of its coding bitrate, a higher QF will not yield a lower quality image. There are 101 images in total for each source by including the original one. It is expected that humans are not able to differentiate all 101 levels.

For the JND test on coded video, we collected five high-resolution and diverse video sequences. Again, their diversity can be characterized by their spatial and temporal information<sup>14</sup> and semantic properties.<sup>15</sup> The five video sequence selected in the preliminary test are shown in Fig. 2.

Each source video is encoded by two video encoders: x264<sup>18</sup> and x265.<sup>19</sup> They are implementations of the H.264 and the H.265 video coding standards. For video coding, the quality is controlled by quality parameter



Figure 2: Five video sequences selected for the JND test: (a) Bunny and Butterfly (BB),<sup>16</sup> (b) Basketball Drive (BD), (c) City Sky (CS),<sup>17</sup> (d) Fountain Boy (FB), and (e) Inside Church (IC).

(QP). Since the QP value can be dynamically adjusted by a rate control algorithm so that rate control methods may affect visual quality. In the experiment, we compare the following two rate control methods for both H.264 and H.265 encoders.

- Fixed QP (FQP). Each video is encoded with  $QP = 1, 2, \dots, 51$ .
- Variable Bitrate (VBR). We enable the default variable bitrate (VBR) of x264 and x265 encoders. To get the same number of encoded video as that in the FQP setting, we set the target bitrates to those obtained from FQP. For example, if the bit rate is 3000 kbps with  $QP = 20$ , we set the target bitrate to 3000 kbps to get a corresponding data point under the VBR setting.

As a result, each source video is encoded 204 times, and there are 205 video contents in total for each source by including the source. A larger QP value leads to a larger quantization step size and a lower coding bitrate in video coding. Again, humans are not able to differentiate all levels. We would like to ask the following three questions.

- Q #1: How many quality levels can a person discern?
- Q #2: What are the JND locations in the perceived quality level versus QP plot?
- Q #3: Are the above two results stable among multiple subjects?

In the following section, we attempt to develop a methodology to measure the quantities raised in Questions #1 and #2. Then, we would like to analyze the obtained results so as to answer Question #3.

## 2.2 Solution Methodology

We represent  $N$  coded images/videos with  $N$  nodes and arrange them in a bitrate ascending order from the left to the right. Under the assumption that quality is a monotonically increasing piecewise constant function of the bitrate, we would like to find JND locations. This can be done efficiently using the bisection search algorithm by starting from the two ends of the entire interval (denoted by nodes  $a$  and  $b$ ) as shown in the top case of Fig. 3a. If images at the two end nodes exhibit noticeable difference (ND), we expect that at least one JND point exists in that interval and will check the middle position (or the closest rounded integer) of the interval. In this example, the quality at node  $(a + b)/2$  is not distinguishable from that at node  $b$ . Then, the search terminates for the interval  $[(a + b)/2, b]$ . However, there still exists notice difference between nodes  $a$  and  $(a + b)/2$ , so that we will continue the bisection search in the corresponding interval. This process is repeated until the finest resolution is reached. Then, we can find the JND location between two adjacent nodes in the finest level.

For some cases, the process is terminated earlier before hitting the finest level. That is, the JND location may appear between nodes at a coarser level. An example is shown in Fig. 3b. We observe noticeable differences (ND) in image/video quality at nodes  $m$  and  $n$ . However, there is unnoticeable difference (UD) in visual quality between nodes  $m$  and  $(m + n)/2$  and between nodes  $(m + n)/2$  and  $n$ . This happens when the quality levels at nodes  $m$  and  $n$  are perceivable but very close. Then, one cannot tell the differences furthermore by going to the next finer level.

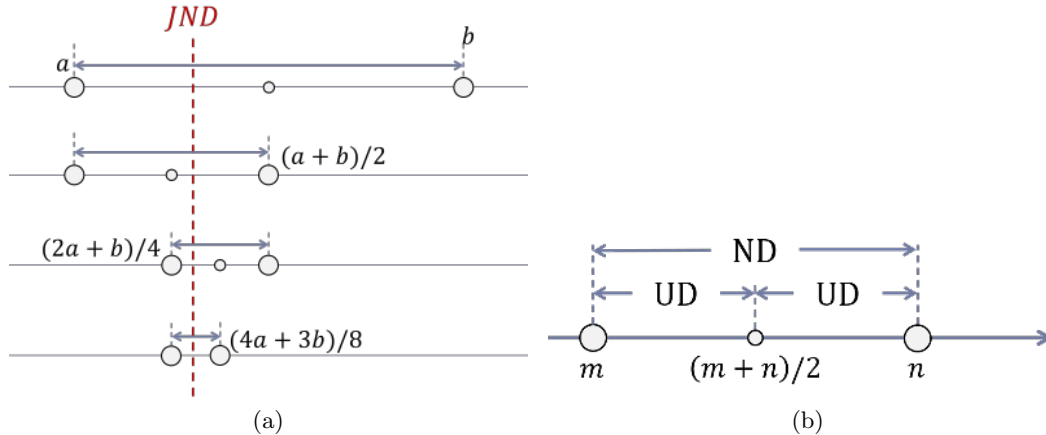


Figure 3: Illustrations of (a) the bisection search process to determine the JND location in the finest level and (b) the early termination condition of a bisection search process.

### 3. JND DATA COLLECTION AND STATISTICS

#### 3.1 Subjective Test

In the preliminary study, we invited 20 subjects to participate in the image and the video subjective tests, respectively. The five source images and videos were of resolution  $1920 \times 1080$  and coded by the JPEG, the H.264 and the H.265 encoders. The QF values of JPEG are  $1, 2, \dots, 100$ , and the QP values of x264 and x265 are  $1, 2, \dots, 51$ . Among all subjects, only eight of them had previous experience in image/video coding. All subjects were seated in a controlled environment. The viewing distance was 2 meters (1.6 times of the picture height) from the center of the monitor to the seat. For the image JND test, the image pair was displayed on a 65" TV with native resolution of  $3840 \times 2160$ . A subject compared two images displayed side by side. For the video JND test, the video pair was played sequentially with monitor resolution of  $1920 \times 1080$ . The subjects are asked to determine whether these two images/videos are noticeably different (ND) or unnoticeably different (UD). The JND location can be identified by tracking the transitional position from ND to UD. The bisection search methodology described in Section 2 was adopted in the test procedure for speed-up. Due to the efficiency of this search algorithm, the test duration of each subject for five images and videos was about 15 and 45 minutes, respectively.

#### 3.2 Statistics of Collected JND Data

The JND points in the domain of QF and QP values are collected for each subject. First, we want to find the total number of JND points for a given source, which in turn determines the number of perceived quality levels. That is, the number of quality levels equals the number of points plus one. The mean and the standard deviation of JND points for each image and video source are given in Table 1, 2 and 3. Furthermore, the box plot of image JND numbers is shown in Fig. 4 (a) and the box plots of video JND are shown in Fig. 5, where the bottom and the top of each box indicate the 25th and 75th percentiles of the samples, respectively, and the middle line is the mean value. We see from the table and the figure that the number of JND points does not vary significantly among subjects. This is especially true by focusing on the interquartile range (i.e. between the tops and bottoms).

Furthermore, we show the box plots of the highest and lowest JND locations in Figs. 4 (b) and (c), respectively. By comparing Figs. 4 (b) and (c), we see that there is correlation between the number of JND points and the highest JND location. That is, if the highest JND point occurs at a higher QF, the subjects will see more JND-based quality levels. For the lowest JND locations, since the images are heavily distorted and can be easily detected, the opinions are close among subjects. In contrast, the highest JND locations depend on subject's experience. Trained subjects can find differences at higher QF more easily than inexperienced ones. Thus, as compared with Fig. 4 (b), the ranges of the highest JND points from experienced subjects are significantly narrower as shown in Fig. 4 (d).

Table 1: Statistics of the number of JND points for JPEG images.

	Mean	Stdev
CC	5.20	1.47
DB	3.75	1.41
FT	7.05	1.63
HS	6.00	1.62
RP	5.00	1.41

Table 2: Statistics of the number of JND points for x264 coded video.

	FQP		VBR	
	Mean	Stdev	Mean	Stdev
BB	4.50	1.32	4.60	1.39
BD	3.15	0.59	3.20	0.77
CS	3.50	0.95	3.65	0.93
FB	3.15	0.99	2.80	1.01
IC	3.50	0.69	3.30	0.86

## 4. JND DATA POST-PROCESSING AND QUALITY PLOTS

### 4.1 JND Histogram and Its Post-Processing

Since some subjects may have more JND points than others, we need to normalize the weight of each JND point across different subjects. The normalization process is straightforward. That is, if a subject has  $K$  JND points, we give a weight of  $K^{-1}$  to each of his/her selected JND points. After normalization, we can aggregate the JND data from all subjects, and plot the histogram. The result for the Color Checker (CC) image is shown in Fig.

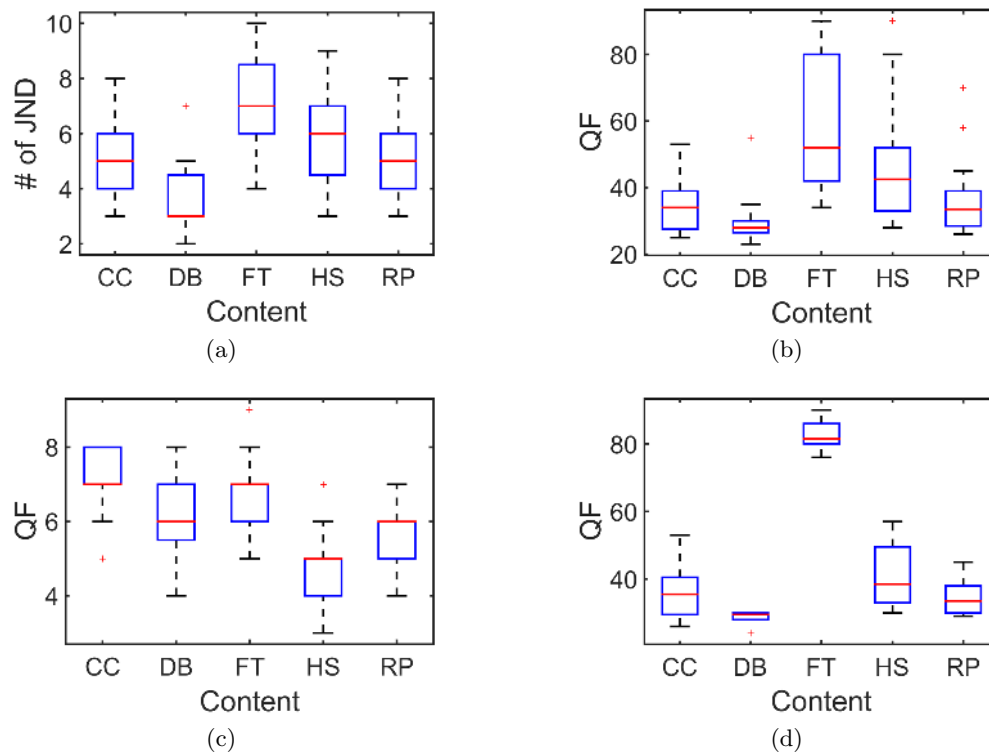


Figure 4: The box plots of (a) JND numbers, (b) the highest JND locations of all subjects, (c) the lowest JND locations of all subjects, and (d) the highest JND locations from experienced subjects only.

Table 3: Statistics of the number of JND points for x265 coded video.

	FQP		VBR	
	Mean	Stdev	Mean	Stdev
BB	3.83	1.04	3.85	0.75
BD	3.00	0.86	2.55	1.10
CS	3.30	0.80	3.43	0.68
FB	2.85	0.49	2.65	0.59
IC	3.65	1.04	3.74	1.05

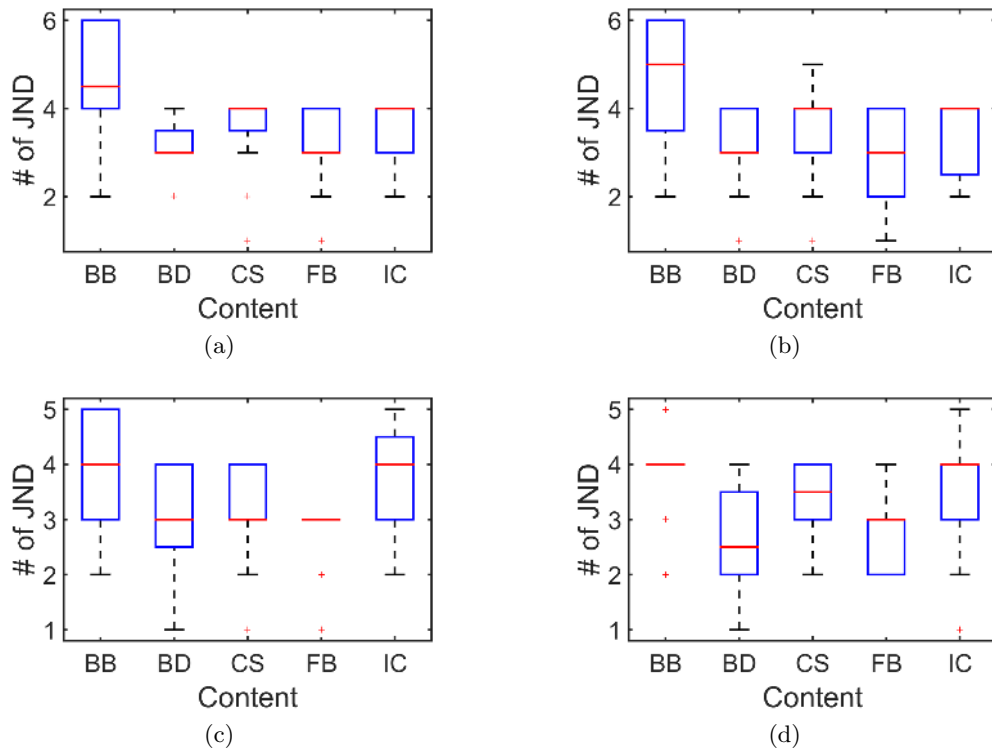


Figure 5: The box plots of JND numbers encoded by (a) x264 with FQP, (b) x264 with VBR, (c) x265 with FQP, and (d) x265 with VBR.

6(a). We see that there is rarely any JND point for QF greater than 50. The quality of these images is high enough that they are not differentiable from the original. The same observation applies to other test images. Thus, we will focus on the interval where QF ranges from 1 to 50.

Next, we describe a procedure to derive the JND-based quality level plot from the histogram plot. Since JND is a random variable, we need to account for its statistical variation. We use the rounded mean number in Table 1 as the target number of JND points. For example, we choose 5 JND points for the CC image. To cluster the JND histogram into 5 sub-intervals, we run the k-means algorithm with  $k = 5$  on the JND histogram and partition the whole QF interval into 5 sub-intervals. The JND location bars in these 5 sub-intervals are shown in 5 different colors in Fig. 6(a). After that, we select a high peak that is close to the mean of each sub-interval as the desired JND point. The selected JND location bar is labeled by a circle on its top. The reason of selecting the high peak rather than the mean value as the JND location is that we would like to align it with the JND experimental data.

## 4.2 JND-based Quality Plots

To demonstrate the robustness of our experimental design in Section 3 and the post-processing technique described above, we aggregate the JND data from those subjects whose JND numbers are in the interquartile range

(with 10 subjects in total) to obtain the corresponding JND histogram plot and the quality level plot as shown in Figs. 6(c) and (d), respectively. The JND-based quality level plots are quite similar for the two cases with 10 and 20 subjects. Finally, the JND-based quality level plot for other four test images with all 20 subjects are shown in Fig. 7. We apply the same procedure to the collected video data. The final output of the five video sequences encoded by x264 with FQP are shown in Fig. 8. In video coding, a smaller QP value gives better quality, so  $QP = 0$  corresponds to  $Level = 1$ . The results obtained by other codec settings are shown in Fig. 9, 10 and 11.

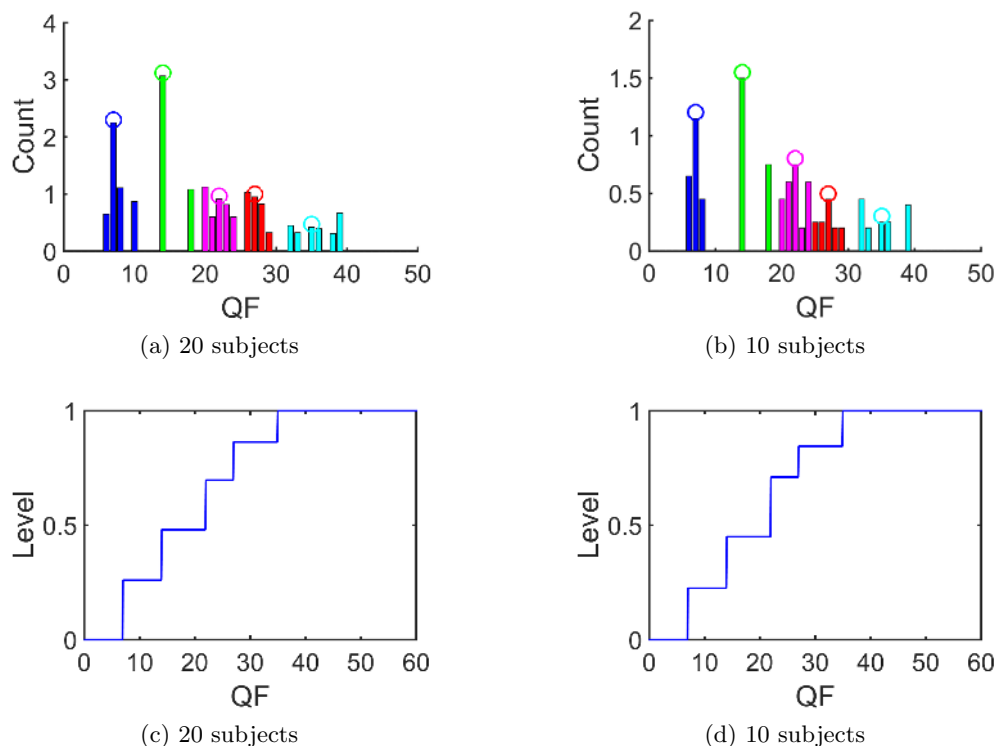


Figure 6: Comparison of the JND location histogram plots in (a) and (b) and the JND-based quality level plots in (c) and (d), where (a) and (c) are based on all 20 subjects and (b) and (d) are based on the 10 subjects whose JND numbers are in the interquartile range.

## 5. CONCLUSION AND FUTURE WORK

A new methodology for human visual experience measurement was proposed in this work. A preliminary subjective test was carried out to collect the JND data to demonstrate the feasibility of this idea. The collected raw JND data was analyzed and post-processed to derive the JND-based quality level plot. It was shown that this quality level plot is robust by shrinking the number of test subjects from 20 to 10. Currently, we are working on large-scale JND-based compressed image/video quality assessment datasets to gain a deeper understanding of the relationship between the JND-based quality level plot and the underlying image/video content. Then, for a given image/video content, we would like to be able to predict its JND numbers and locations based on extracted features.

## ACKNOWLEDGMENTS

Computation for the work described in this paper was supported by the Center for High-Performance Computing at the University of Southern California (<http://hpcc.usc.edu>).

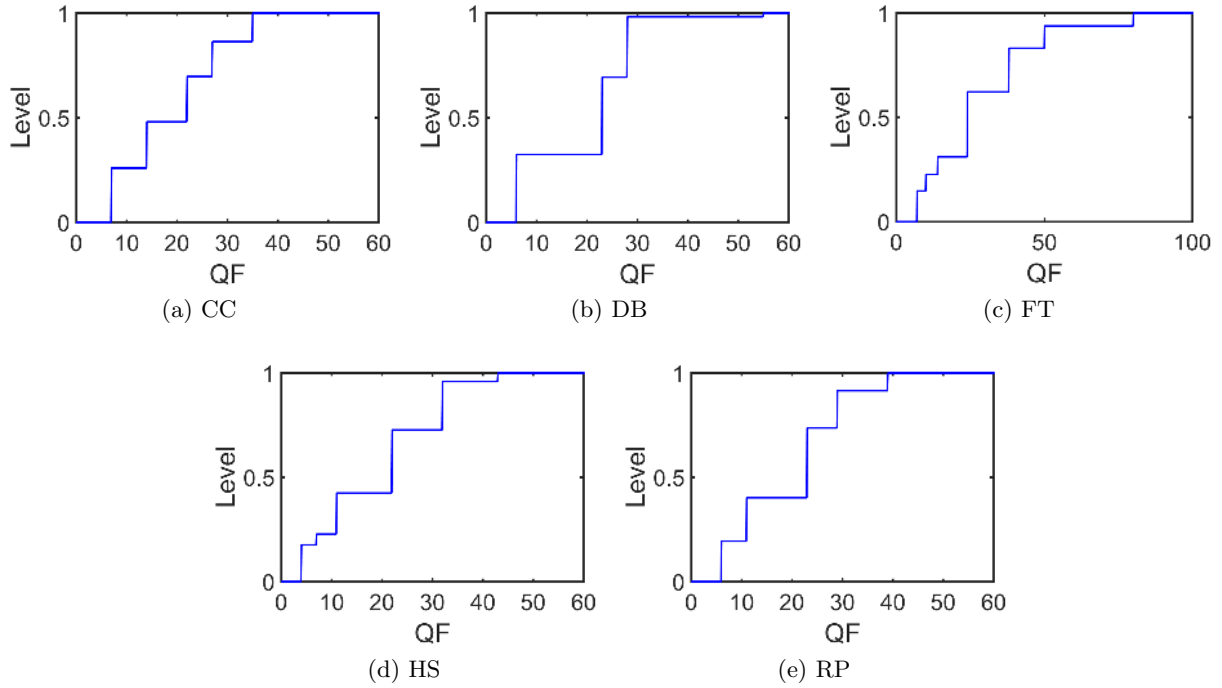


Figure 7: The image JND level versus the QF plot for (a) CC (b) DB, (c) FT, (d) HS and (e) RP.

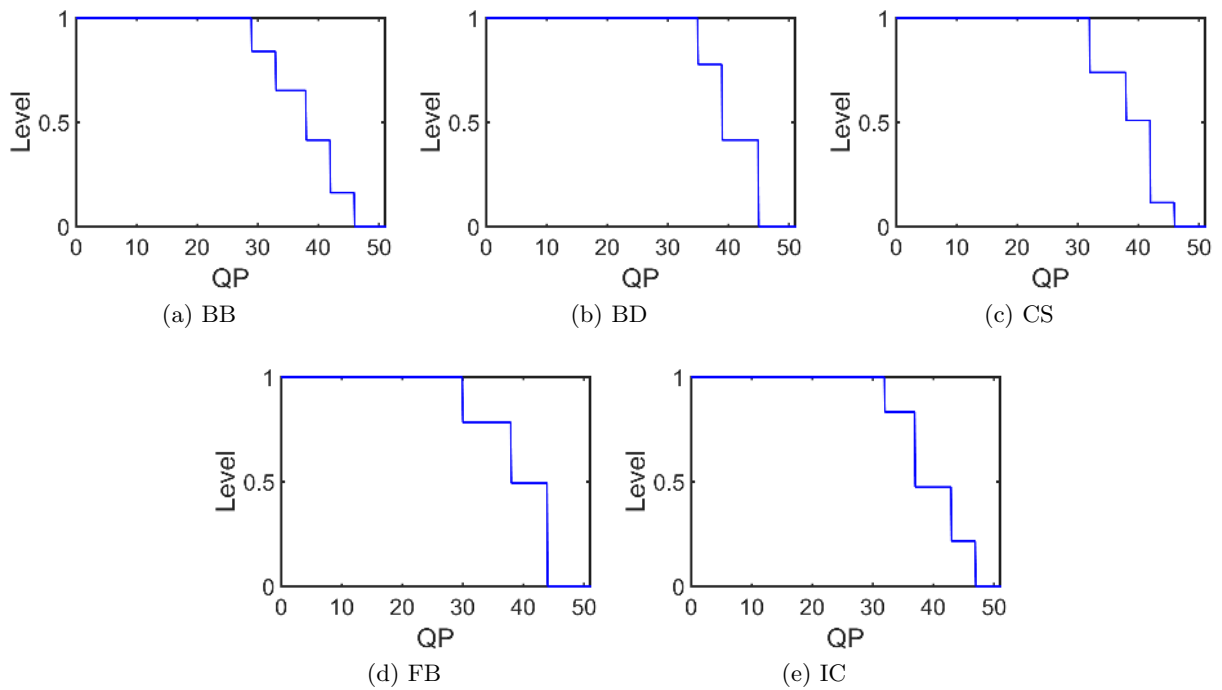


Figure 8: The JND-based quality level plot for x264 with FQP: (a) BB (b) BD, (c) CS, (d) FB and (e) IC.



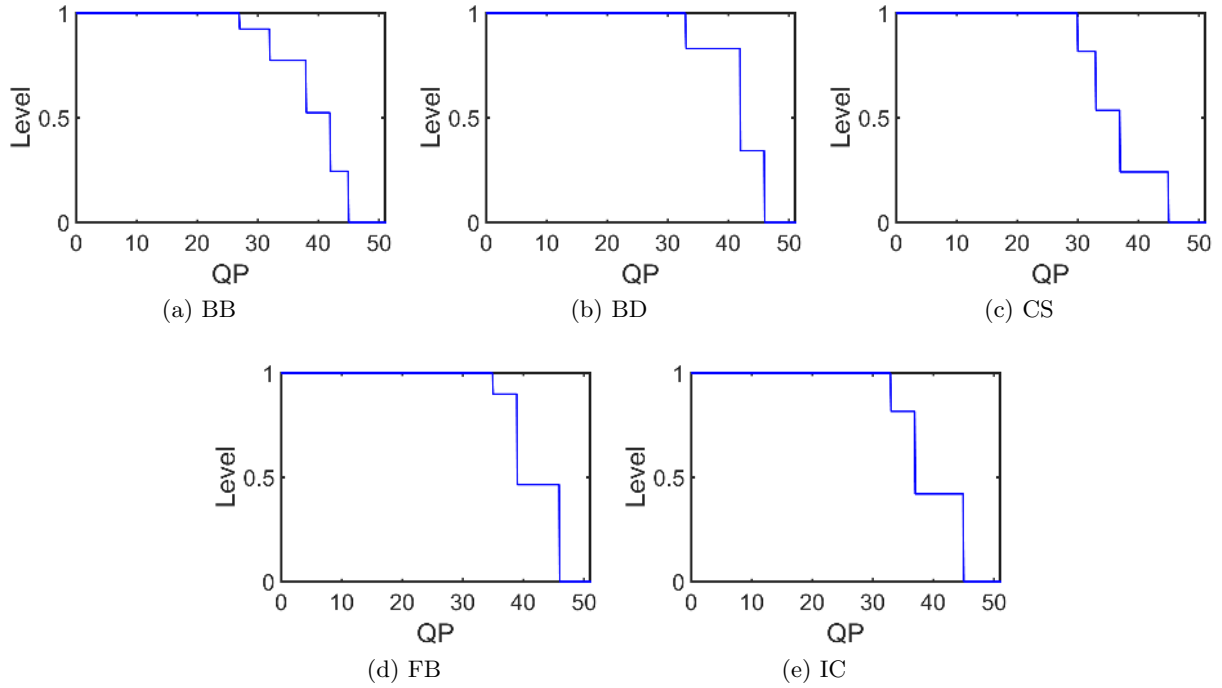


Figure 9: The JND-based quality level plot for x264 with VBR: (a) BB (b) BD, (c) CS, (d) FB and (e) IC.

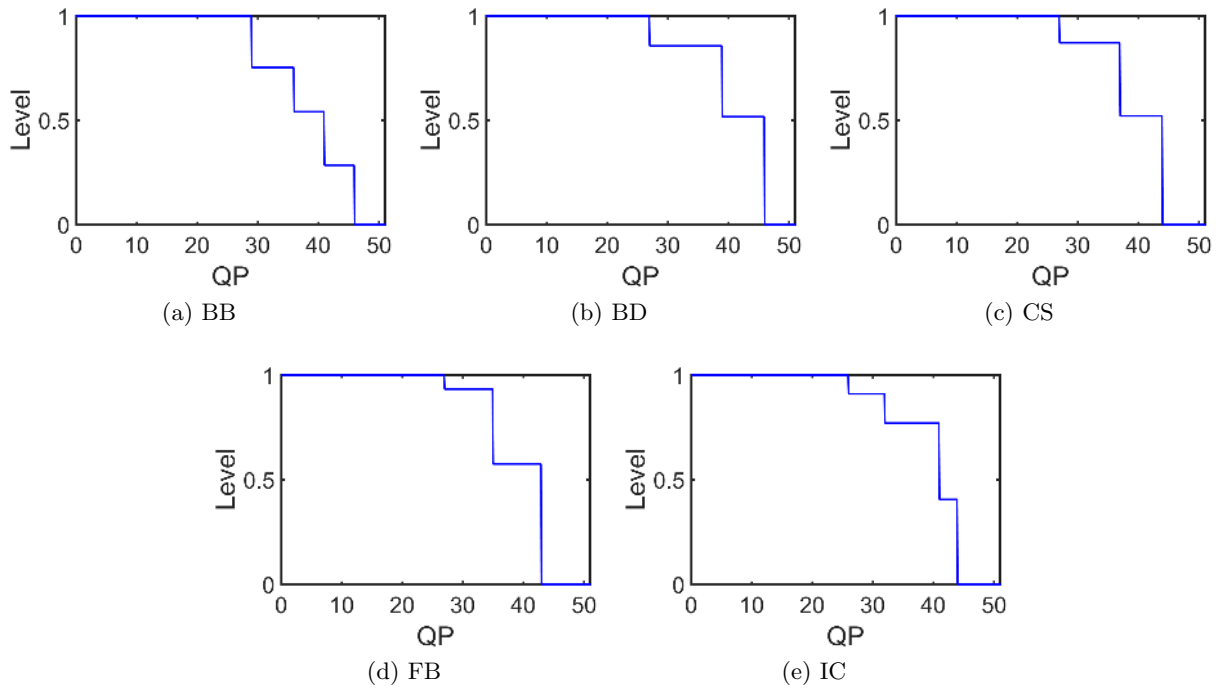


Figure 10: The JND-based quality level plot for x265 with FQP: (a) BB (b) BD, (c) CS, (d) FB and (e) IC.

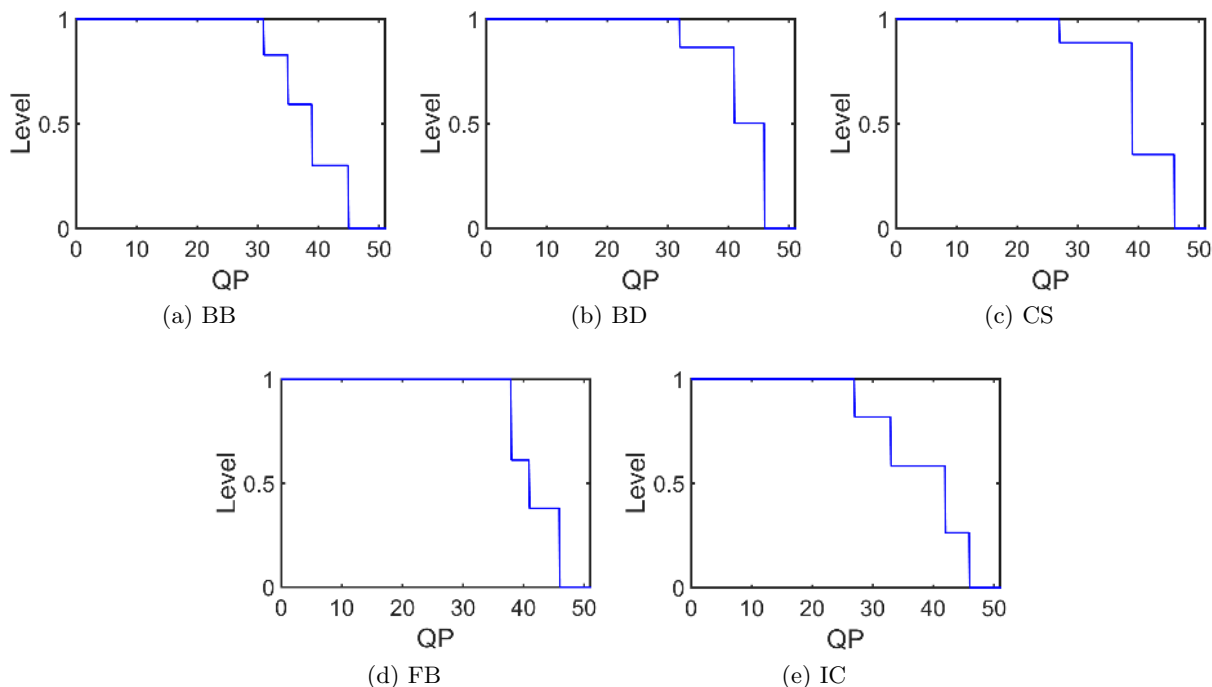


Figure 11: The JND-based quality level plot for x265 with VBR: (a) BB (b) BD, (c) CS, (d) FB and (e) IC.

## REFERENCES

- [1] Lin, W. and Kuo, C.-C. J., "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation* **22**(4), 297–312 (2011).
- [2] Jayant, N., Johnston, J., and Safranek, R., "Signal compression based on models of human perception," *Proceedings of the IEEE* **81**, 1385–1422 (Oct 1993).
- [3] Lin, W., "Computational models for just-noticeable difference," *Digital video image quality and perceptual coding* (2005).
- [4] Watson, A. B., "Proposal: Measurement of a JND scale for video quality," *IEEE G-2.1. 6 Subcommittee on Video Compression Measurements* (2000).
- [5] Jia, Y., Lin, W., and Kassim, A., "Estimating just-noticeable distortion for video," *Circuits and Systems for Video Technology, IEEE Transactions on* **16**, 820–829 (July 2006).
- [6] Wu, J., Shi, G., Lin, W., Liu, A., and Qi, F., "Just noticeable difference estimation for images with free-energy principle," *Multimedia, IEEE Transactions on* **15**, 1705–1710 (Nov 2013).
- [7] Lin, W., Dong, L., and Xue, P., "Visual distortion gauge based on discrimination of noticeable contrast changes," *Circuits and Systems for Video Technology, IEEE Transactions on* **15**, 900–909 (July 2005).
- [8] Luo, Z., Song, L., Zheng, S., and Ling, N., "H.264/Advanced Video Control Perceptual Optimization Coding Based on JND-Directed Coefficient Suppression," *Circuits and Systems for Video Technology, IEEE Transactions on* **23**, 935–948 (June 2013).
- [9] Zhang, X., Lin, W., and Xue, P., "Just-noticeable difference estimation with pixels in images," *Journal of Visual Communication and Image Representation* **19**(1), 30–41 (2008).
- [10] Winkler, S., "Analysis of public image and video databases for quality assessment," *Selected Topics in Signal Processing, IEEE Journal of* **6**, 616–625 (Oct 2012).
- [11] "RAW-Samples." <http://www.rawsamples.ch>. Accessed: 2015-01-04.
- [12] "PhotographyBLOG." <http://www.photographyblog.com>. Accessed: 2015-03-16.
- [13] Independent JPEG Group, "JPEG image compression software." <http://www.ijg.org>. Accessed: 2015-03-20.

- [14] ITU, "Recommendation ITU-T P.910, Subjective video quality assessment methods for multimedia applications," *International Telecommunication Union, Geneva, Switzerland* **910** (1999).
- [15] Lin, J. Y., Song, R., Wu, C.-H., Liu, T., Wang, H., and Kuo, C.-C. J., "MCL-V: A streaming video quality assessment database," *Journal of Visual Communication and Image Representation* **30**(0), 1 – 9 (2015).
- [16] Goedegebure, S., Goralczyk, A., Valenza, E., Vegdahl, N., Reynish, W., Lommel, B. V., Barton, C., Morgenstern, J., and Roosendaal, T., "Big Buck Bunny." <http://www.bigbuckbunny.org/> (2008).
- [17] CDVL, "The Consumer Digital Video Library." <http://www.cdvl.org/> (2010).
- [18] Aimar, L., Merritt, L., Petit, E., Chen, M., Clay, J., Rullgrd, M., Heine, C., and Izvorski, A., "x264-a free H.264/AVC encoder." <http://www.videolan.org/developers/x264.html>. Accessed: 2015-07-01.
- [19] "x265 HEVC Encoder/H.265 Video Codec." <http://x265.org>. Accessed: 2015-07-01.