Learning-based Satisfied User Ratio Prediction for Symmetrically and Asymmetrically Compressed Stereoscopic Images

Chunling Fan, Yun Zhang, *Senior Member, IEEE* Raouf Hamzaoui, *Senior Member, IEEE* Djemel Ziou, Qingshan Jiang

Abstract—The Satisfied User Ratio (SUR) for a given distortion level is the fraction of subjects that cannot perceive a quality difference between the original image and its compressed version. By predicting the SUR, one can determine the highest distortion level which allows to save bit rate while guaranteeing a good visual quality. We propose the first method to predict the SUR for symmetrically and asymmetrically compressed stereoscopic images. Unlike SUR prediction techniques for 2D images and videos, our method exploits the properties of binocular vision. We first extract features that characterize image quality and image content. Then, we use gradient boosting decision trees to reduce the number of features and train a regression model that learns a mapping function from the features to the SUR values. Experimental results on the SIAT-JSSI and SIAT-JASI datasets show high SUR prediction accuracy for H.265 All-Intra and JPEG2000 symmetrically and asymmetrically compressed stereoscopic images.

Index Terms—stereoscopic image quality assessment, satisfied user ratio, picture-level just noticeable difference, symmetric stereoscopic compression, asymmetric stereoscopic compression

1 Introduction

Image compression is essential to reduce storage and transmission costs. To achieve high compression ratios, lossy compression schemes such as JPEG and JPEG2000 are used. However, lossy compression introduces distortion in the reconstructed image, which degrades the viewer experience. Subjective Image Quality Assessment (IQA) tests have shown [1] that the distortion can be perceived only when it reaches a given level known as the Picture-level Just Noticeable Difference (PJND). This distortion level is a random variable that depends on the subject, image, and compression scheme. The Satisfied User Ratio (SUR) is the Complementary Cumulative Distribution Function (CCDF) of the PJND. It gives the proportion of the population for which the PJND is greater than a given distortion level. The p% SUR corresponds to the largest distortion level for which the SUR is greater than or equal to $p/100$. For example, the 75% SUR is the largest distortion level for which 75% of the population is satisfied in the sense that it cannot perceive any difference between the original image and its distorted version. Formal definitions of the PJND, SUR and p% SUR are given in [2]. The PJND and SUR are normally determined by statistical analysis following subjective IQA tests for a group of subjects.

Since subjective visual quality tests are time-consuming and expensive, it is desirable to develop computer-based methods that can predict the PJND and SUR in a fully automated way. Except for our recent conference paper [3], previous research on PJND and SUR prediction has focussed on either 2D images [2], [4], [5] or videos [6]. The methods proposed for 2D images and videos are not suited for 3D images as they do not consider the characteristics of binocular vision such as binocular fusion and binocular rivalry. Moreover, stereoscopic images have the distinct feature that they can be compressed asymmetrically. Note also that traditional visual quality assessment methods for 3D images aim to predict the visual quality of a distorted image on a continuous grading scale (e.g., on the Differential Mean Opinion Score (DMOS)) [7], while SUR methods use fine-grained quantization to generate a large number of gradually distorted images and aim to predict the largest distortion level that cannot be perceived by a given percentage of the population.

In this paper, we propose a learning-based SUR prediction model for symmetrically and asymmetrically compressed stereoscopic images. The main contributions of our work are as follows.

1) We extract various features that characterize the quality and content of the original and compressed stereoscopic images.
2) We use Gradient Boosting Decision Trees (GBDT) [8] to select the best features and learn a mapping function from the feature space to the SUR values.
3) For a given reference stereoscopic image and its compressed versions, we use the trained decision

- *Chunling Fan, Djemel Ziou, and Qingshan Jiang are with Shenzhen Key Lab for High Performance Data Mining, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China, 518055. Djemel Ziou is also with Département Faculté d'informatique, Université de Sherbrooke, Québec, Canada. E-mail: fancl@siat.ac.cn, Djemel.Ziou@USherbrooke.ca, qs.jiang@siat.ac.cn*
- *Yun Zhang (Corresponding author) is with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China, 518055. Email: yun.zhang@siat.ac.cn.*
- *Raouf Hamzaoui is with the School of Engineering and Sustainable Development, De Montfort University, UK. Email: rhamzaoui@dmu.ac.uk*

Manuscript received April 19, 2005; revised August 26, 2015.

trees to predict the SUR value for each distortion level. We then use the method of least squares to fit a Gaussian CCDF to the predicted SUR values.

2 Related Work

In recent years, the PJND concept has received growing attention. Several research teams have conducted subjective visual quality tests to build PJND datasets for images, stereoscopic images, and video sequences. Jin et al. [1] explored the PJND characteristics of 50 2D images for JPEG compression and built a PJND dataset called MCL-JCI. Wang et al. [9] studied the PJND characteristics of 220 2D videos in four resolutions for H.264/AVC compression and generated a dataset called VideoSet. In [10], the PJND characteristics of stereoscopic images for symmetric and asymmetric H.265 All-Intra coding and JPEG2000 compression were studied. It was found that the PJND of stereoscopic images is highly related to image content. However, as subjective tests are time-consuming and expensive, the scale of existing PJND datasets is very limited. To make the PJND more widely applied, a number of objective models have been proposed to predict the PJND and SUR.

To predict the mean PJND for a video clip, Huang et al. [6] extracted spatial and temporal local features and fed the combined feature vector into a Support Vector Regression (SVR) model. As features, they used a gradient map, spatial and temporal smoothness maps, a saliency map, and a spatial sensitivity map. Hadizadeh et al. [11] used sparse coding to extract a feature vector from reference and distorted versions obtained by adding noise whose amplitude is determined according to a JND model. Then, they fed the feature vector into a multi-layer neural network to predict whether the distorted image is perceptually different from the reference image. Fan et al. [5] proposed the first deep learning method to predict SUR curves for image compression. The method relies on a siamese Convolutional Neural Network (CNN) trained on pairs consisting of an original image and a compressed image. Transfer learning and data augmentation were used to address over-fitting. In [2], the performance of the method in [5] was improved by optimizing the network architecture and using feature learning instead of fine tuning. Moreover, instead of assuming a Gaussian distribution, maximum likelihood estimation and the Anderson-Darling test were used to select the JND statistical model. Liu et al. [4] converted the PJND prediction problem into multiple binary classification tasks and used a deep learning model as the binary classifier.

SUR and PJND prediction is more challenging for stereoscopic images than for 2D images due to the characteristics of binocular vision. The only previous work on SUR prediction for stereoscopic images is our conference paper [3]. Since the existing PJND datasets for stereoscopic images [10] are small, and deep learning methods require large datasets to outperform traditional machine learning techniques, we used SVR instead of deep learning. In this paper, we improve our previous work [3] by building more features, exploiting feature selection, and replacing SVR by a more powerful machine learning technique (GBDT). Moreover, we extend our work in [3], which focused on symmetrically compressed stereoscopic images, to both symmetrically and asymmetrically compressed images. Finally, while the work in [3] only predicts the SUR at a finite number of points, we build a smooth SUR curve by fitting a Gaussian model to the predicted values.

3 Proposed Method

The proposed model consists of three main parts as shown in Fig. 1. In the first part, features are extracted from the reference and distorted stereoscopic images. Two types of features are considered: image quality features and image content features. Image quality features reflect the quality of the compressed stereoscopic image, the quality of the cyclopean image, the quality of the difference map, and the rivalry quality. Image content features comprise both monocular and binocular visual features. In the second part, feature selection is applied to reduce the number of the features and improve the performance of the predictive model. The features are then concatenated and used to learn the mapping function from the feature space to the SUR values. In the third part, a Gaussian CCDF is fitted to the predicted SUR values to obtain any p% SUR value from the predicted SUR curve.

3.1 Problem Modeling

Let $(I_k^L[0], I_k^R[0])$, $k \in \{1, \ldots, K\}$ be a sequence of K pristine stereoscopic images, where $I_k^L[0]$ represents the left view and $I_k^R[0]$ represents the right view. Each view is compressed at the N different distortion levels $1, 2, \ldots, N$ to generate N distorted versions. Let $(I_k^L[i], I_k^R[j])$, $k \in \{1, \ldots, K\}$, $i, j \in \{1, \ldots, N\}$ be a distorted version of the pristine stereoscopic image $(I_k^L[0], I_k^R[0])$, where i is the distortion level of the left view and j is the distortion level of the right view. For a symmetrically compressed image, i=j. On the other hand, when the image is asymmetrically compressed, $i \neq j$. Given the kth pristine stereoscopic image $(I_k^L[0], I_k^R[0])$, $k \in \{1, \ldots, K\}$ and its distorted versions $(I_k^L[i], I_k^R[j])$, $i, j \in \{1, \ldots, N\}$, our aim is to find a prediction model S_θ such that

$$S_\theta((I_k^L[0], I_k^R[0]), (I_k^L[i], I_k^R[j])) \approx \mathrm{SUR}_k^{i,j}, \tag{1}$$

where $\mathrm{SUR}_k^{i,j}$ is the SUR value of the kth image for distortion level i of the left view and distortion level j of the right view. Note that in our experiments we set i=0 for H.265 All-Intra asymmetrically compressed images and i=1 for JPEG2000 asymmetrically compressed images.

Although the PJND is a discrete random variable, it is common practice [5], [9], [10] to model it as a Gaussian random variable. Consequently, the SUR is given by the function

$$\bar{\Phi}(x|\mu, \sigma^2) = 1 - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds, \tag{2}$$

where μ is the mean and σ^2 is the variance of the PJND samples. Fig. 2 shows an example of PJND samples, SUR curve, and 75% SUR.

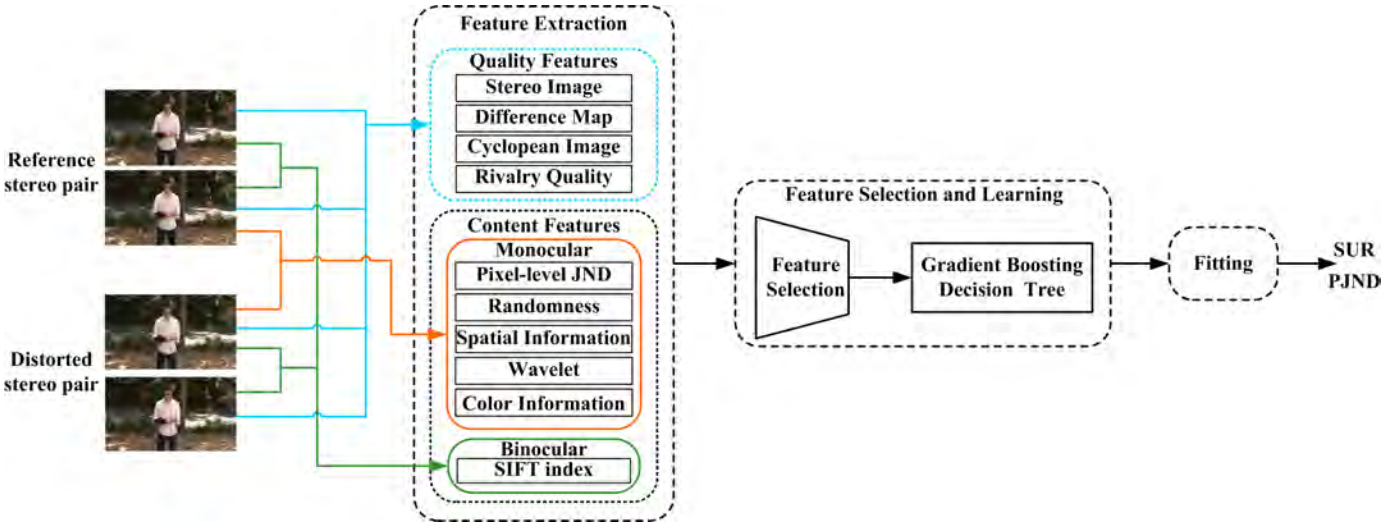


Fig. 1: Proposed architecture for SUR prediction.

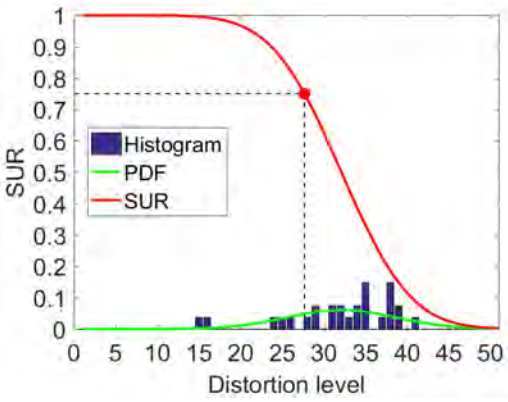


Fig. 2: Example of PJND samples (histogram in blue), approximation of the PJND by a Gaussian PDF, SUR curve corresponding to the Gaussian PDF, and 75% SUR (distortion level corresponding to the red dot). The data is from the first source image in the SIAT-JSSI dataset [10] and H.265 All-Intra symmetric compression.

3.2 Feature Extraction

In this section, we extract features from a pristine stereoscopic image-distorted version pair that enable our model S_θ to learn a mapping from the pair to the SUR values and accurately predict the SUR for a new input pair. For this reason, we target features that are correlated with the SUR of symmetrically and asymmetrically compressed stereoscopic images.

3.2.1 Image Quality Features

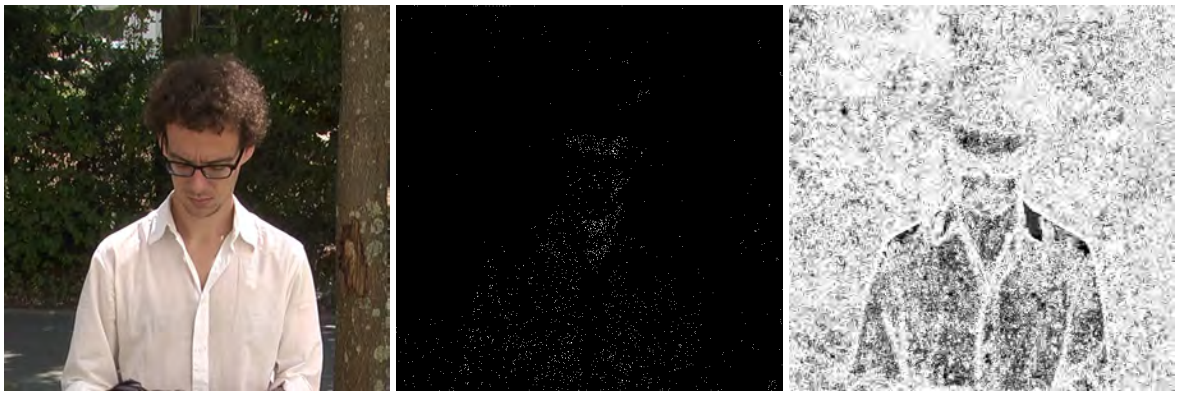
In H.265 intra coding, the quality of the compressed image is determined by the Quantization Parameter (QP). The work in [10] shows that there is a high correlation between the SUR and QP for compressed stereoscopic images. This motivates us to use the quality of the compressed stereoscopic image as a feature. We measured the quality of the compressed stereoscopic images using the Frequency Integrated Peak Signal-to-Noise Ratio (FIPSNR) [12] and denoted it by q_{stereo}. Binocular fusion and binocular rivalry are two fundamental properties of binocular vision. Binocular fusion means that when a left view and a right view with a certain disparity are represented in the two eyes, the brain generates a cyclopean image. To exploit binocular fusion, we generated two cyclopean images [13]: one from the two views of the reference image and one from the two views of the compressed image. Then we used the Peak Signal-to-Noise Ratio (PSNR) between the distorted cyclopean image and the reference cyclopean image as a feature q_{cyc} to represent the quality of the distorted cyclopean image. Binocular rivalry in the Human Visual System (HVS) means that perception alternates when two incompatible views are represented in the two eyes. To take binocular rivalry into account, we generated a difference map as the difference between the left view and the right view and used the PSNR of the difference map as a quality feature q_{diff}. As a fourth image quality feature, we used $q_{\text{rivalry}} = \max(q_l, q_r)$ where q_l is the PSNR between the distorted left view and the reference left view, and q_r is the PSNR between the distorted right view and the reference right view. Finally, we concatenated the four image quality features in a vector $\vec{f_Q} = (q_{\text{stereo}}, q_{\text{cyc}}, q_{\text{diff}}, q_{\text{rivalry}})$.

3.2.2 Image Content Features

In our previous work [10], we found that the SUR of stereoscopic images is highly correlated with image content. In this section, we extract the features that characterize monocular and binocular content.

Monocular visual features The visibility of a signal is reduced in the presence of another signal. This phenomenon is known as the visual masking effect. Because of visual masking, the HVS may not be able to perceive distortion in an image. Pixel-level JND models [14] exploit luminance adaptation and contrast masking effects. We first estimated the pixel-level JND threshold map T for the right view R of the reference stereoscopic image. Then, we obtained the ratio of the pixels in the distorted right view D that are outside the range defined by the threshold T as

$$M_{\text{JND}} = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \Psi(i,j), \tag{3}$$



(a) Distorted image with QP=20, M_{JND}=0.0043.

(b) Distorted image with QP=40, M_{JND}=0.1372.

Fig. 3: Comparison between pixel-level JND mask and randomness map for different distortion levels. First column: distorted image. The distorted image is obtained by encoding a reference image with H.265 All-Intra. Second column: pixel-level JND mask. Third column: randomness map.

where W and H are the width and height of the image, respectively, and

$$\Psi(i,j) = \begin{cases} 1, & \text{if } |R(i,j) - D(i,j)| > T(i,j) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

The second column in Fig. 3 shows an example of a pixel-level JND mask $\Psi(i,j)$. We found that when the distortion level increases, the ratio of the pixels that are outside the range defined by the JND threshold increases too. Therefore, we choose the ratio M_{JND} as a feature.

Intuitively, it is difficult to perceive the artefacts in an image with many random patterns. Spatial randomness is an index to measure such randomness. We first calculated the randomness map [15] for the right view of the reference and distorted images, respectively (see the third column in Fig. 3 for two examples of the randomness map). We found that with the increase of the distortion level, more areas of the randomness map become black, which means that the map contains less information. We extracted a 10-dimensional vector $(d_1, d_2, \ldots, d_{10})$ from a 10-bin histogram of the randomness map of the distorted image and a 10-dimensional vector $(r_1, r_2, \ldots, r_{10})$ from a 10-bin histogram of the randomness map of the reference image. Then, we built a randomness feature vector $\vec{M_{\text{Rand}}} = (d_1/r_1, \ldots, d_{10}/r_{10})$.

It was found [10] that there is a high correlation between the Spatial Information (SI) [16] and SUR for stereoscopic images. SI is an index of the edge energy in an image. We used the ratio of the SI from the right view of the distorted image to the SI from the right view of the reference image as a feature and denoted it by M_{SI}.

We also exploited the wavelet transform to extract features from the right view of the stereoscopic image. Specifically, we extracted the mean and standard deviation of the horizontal, vertical, and diagonal detail coefficients of a level 1 Haar wavelet decomposition. We concatenated these features in a feature vector and denoted it by $\vec{M_{\text{WT}}} = (M_{\text{meanh}}, M_{\text{stdh}}, M_{\text{meanv}}, M_{\text{stdv}}, M_{\text{meand}}, M_{\text{stdd}})$.

The above features were extracted from gray-scale images without considering color information. However, the HVS is also sensitive to color artifacts. To extract color features, we applied the logarithmic-scale transformation and the orthogonal transformation described in [17] to convert the RGB information into logarithmic-scale color information. The work in [17] shows that the logarithmic-scale color information follows a Gaussian distribution, and that the mean μ and variance σ^2 of the distribution are useful to predict the quality of the image. Using the mean and variance of each component, we built a six-dimensional feature vector $\vec{M_{\text{color}}} = (M_{\mu_1}, M_{\sigma_1^2}, M_{\mu_2}, M_{\sigma_2^2}, M_{\mu_3}, M_{\sigma_3^2})$. Finally, we concatenated all the monocular visual features into a single vector $\vec{f_M} = (M_{\text{JND}}, \vec{M_{\text{Rand}}}, M_{\text{SI}}, \vec{M_{\text{WT}}}, \vec{M_{\text{color}}})$.

Binocular Visual Feature Because of the horizontal parallax, the two eyes view an object from slightly different directions, so there exists a positional difference between

(a) Distorted image with QP=20, and 4272 SIFT matching keypoints.

(b) Distorted image with QP=40, and 645 SIFT matching keypoints.

Fig. 4: SIFT matching keypoints for the left and right views of a stereoscopic image.

the two retinal projections. This is known as binocular disparity. The human brain has the ability to deduce depth perception from the binocular disparity. When the human eyes view a stereoscopic image, depth perception is obtained by matching the left and right views. When the quality of the stereoscopic image decreases, the number of matching keypoints will decrease, and the sense of depth will also decrease. The Scale-Invariant Feature Transform (SIFT) [18] is widely used in computer vision tasks. We first detected the SIFT keypoints in the left and right views of the stereoscopic image. Then we applied stereoscopic matching to get the number of matching keypoints in the reference and distorted images. Fig. 4 shows SIFT matching keypoints for different distortion levels. We found that the number of matching keypoints is related not only to the distortion level but also to image content. Specifically, for a given stereoscopic image, when the distortion level increases, the number of matching keypoints between the left and right views decreases. This is mainly due to the loss of details in the compressed images at lower bit rates. In addition, there are more matching keypoints in images with complex textures. To eliminate the influence of image content, we divided the number of matching keypoints in the distorted image by the number of matching keypoints in the reference image and used the ratio as a SIFT index M_{SIFT}.

3.2.3 Overall feature vector

Our overall feature vector $\vec{F}$ consists of the concatenation of the image quality feature vector $\vec{f_{\text{Q}}}$, the monocular feature vector $\vec{f_{\text{M}}}$, and the binocular visual feature M_{SIFT}. Fig. 1 shows which images were used to compute each feature.

3.3 Feature Selection and Learning

Feature selection can improve the prediction performance of a model and decrease its time complexity by removing redundant features.

Given m features, we aim to select the most effective n $(n < m)$ features from the feature set. Feature selection techniques can be roughly divided into three categories: filter methods, wrapper methods, and embedded methods. Filter methods first score each feature according to its divergence or correlation and then use a threshold to select the features. Wrapper methods select or delete features iteratively with the help of an objective function. However, if the feature space is high-dimensional, the search space is too large and the time complexity is too high. Embedded methods select the features by assessing their importance during the training of the model. In this paper, we used the embedded method GBDT [8] for feature selection and learning.

GBDT builds the prediction model as a number of decision trees, where each decision tree is used as a weak regressor. When constructing a decision tree, the best feature is selected each time as a decision node. In our experiments, features are sorted according to their weight defined as the average gain across all splits. We used the mean squared error as the loss function for each decision tree. Fig. 5 shows the average weight of features for H.265 and JPEG2000 in a 10-fold cross-validation experiment. We found that features M_{SIFT}, M_{SI}, q_{stereo}, M_{JND}, and M_{stdd} had larger weights than the others. The weights are learned from the training data with GBDT and these features are found more effective in a data-driven way. We also found that the relative importance of the features depends on the image contents and

compression types (symmetric H.265, asymmetric H.265, symmetric JPEG2000, and asymmetric JPEG2000). For each compression type, we used a threshold to select the most important features. Our method can adaptively learn the weight for each feature according to the image contents and compression types.

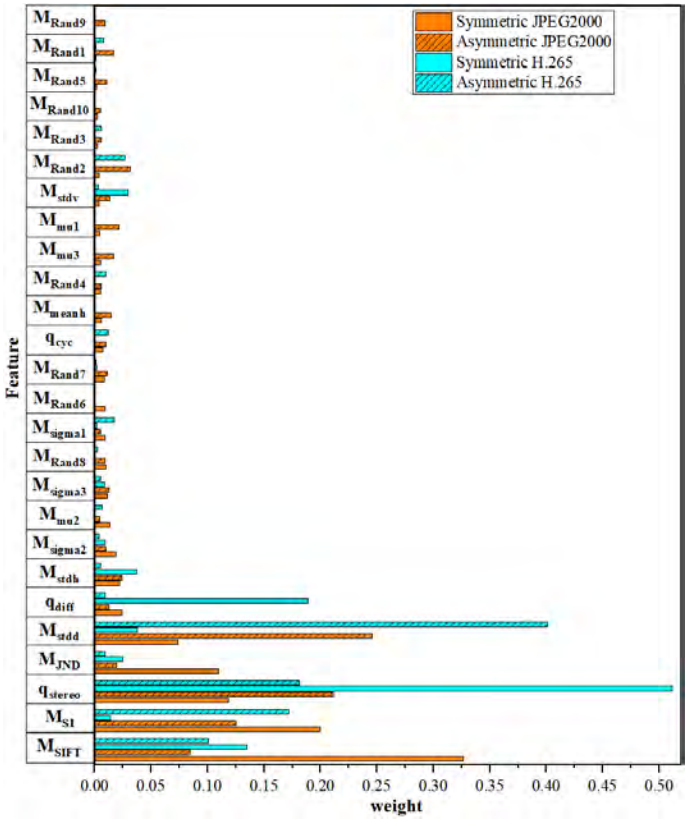


Fig. 5: Average weight of features for H.265 and JPEG2000 in a 10-fold cross-validation experiment.

We aim to learn a mapping function from the feature space to the SUR values. We consider a training set consisting of K stereoscopic images $(I_k^L[0], I_k^R[0])$, $k \in \{1, \ldots, K\}$ and their distorted versions $(I_k^L[i], I_k^R[j])$, $k \in \{1, \ldots, K\}$, $i, j \in \{1, \ldots, N\}$, where N is the number of distortion levels. We assume that for compressed stereoscopic images the PJND samples are normally distributed and use the training data to estimate the mean and standard deviation of the distribution. Then, we derive the ground truth SUR as CCDF of this distribution as given in Eq. (2). Next, using the ground truth SUR values, we use the GBDT model to learn the mapping function from the feature space to the SUR values. Here each decision tree is a regression tree. The final score can be calculated as

$$f_M(x) = \sum_{m=1}^{M} T(x; \Theta(m)), \tag{5}$$

where $T(x; \Theta(m))$ is the m-th decision tree, x is the feature vector, and M is the number of decision trees.

3.4 Fitting

Given a test stereoscopic image $(I_k^L[0], I_k^R[0])$ and its N distorted versions $(I_k^L[i], I_k^R[j])$, $j \in \{1, \ldots, N\}$, i=j for symmetric compression, i=0 for H.265 asymmetric compression, and i=1 for JPEG2000 asymmetric compression, we use the learned GBDT model to predict the N SUR values $\text{SUR}_k^{i,j}$ given by Eq. (1). We then use the method of least squares to fit a Gaussian CCDF to the predicted SUR values. That is, we compute the mean $\hat{\mu}$ and the variance $\hat{\sigma^2}$ of the distribution as

$$(\hat{\mu}, \hat{\sigma}^2) = \arg\min_{\mu, \sigma^2} \sum_{j=1}^{N} \left| \bar{\Phi}(j|\mu, \sigma^2) - \text{SUR}_k^{i,j} \right|^2. \tag{6}$$

The fitted SUR curve is given by $\bar{\Phi}(x|\hat{\mu}, \hat{\sigma}^2)$ as in Eq. (2).

4 Experimental Results and Analysis

To evaluate the performance of the proposed method, we conducted experiments on symmetrically compressed images from the SIAT-JSSI dataset [10] and asymmetrically compressed images from the SIAT-JASI dataset [10]. SIAT-JSSI and SIAT-JASI contain 10 reference stereoscopic images and their compressed versions with H.265 All-Intra and JPEG2000. For H.265 All-Intra, the QP ranges from 1 to 51. For JPEG2000, the Compression Ratio (CR) ranges from 1 to 300. Thus, SIAT-JSSI contains 3510 symmetrically compressed stereoscopic images. In SIAT-JASI, the left view is the reference image and the right view is distorted by varying QP from 1 to 51 for H.265 All-Intra and CR from 1 to 300 for JPEG2000. Thus, the dataset contains 7520 asymmetrically compressed stereo images.

We conducted k-fold (k=10) cross-validation and predicted the SUR curve and the 75% SUR for each reference stereo image in the datasets. Specifically, each dataset was divided into 10 non-overlapping subsets, each of which contains a reference stereoscopic image and its distorted versions. Each time, nine subsets were used for training and validation, and the remaining one was used for testing. The reported result is the average for the ten tests.

To analyze the results, we used the following measures: ΔPJND, ΔSUR, ΔFIPSNR, and the Bhattacharyya distance between the predicted and ground truth PJND (Gaussian) distributions. Here, given a reference stereoscopic image and its N distorted versions,

$$\Delta\text{PJND} = \mid \text{PJND}^P - \text{PJND}^{gt} \mid, \tag{7}$$

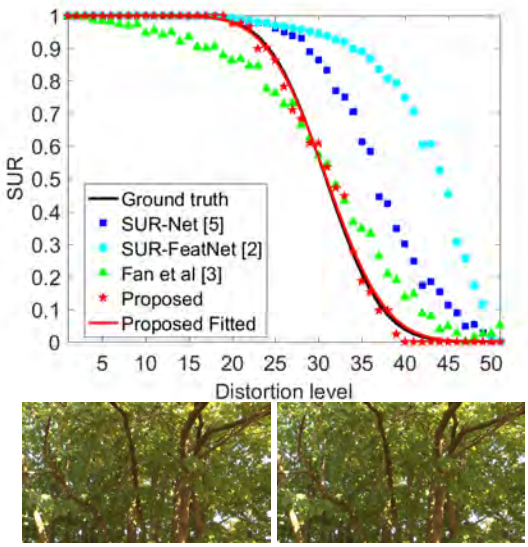
where PJND^P is the predicted 75% SUR and PJND^{gt} is the ground truth 75% SUR.

$$\Delta\text{SUR} = \frac{1}{N} \sum_{i=1}^{N} \mid \text{SUR}_i^P - \text{SUR}_i^{gt} \mid, \tag{8}$$

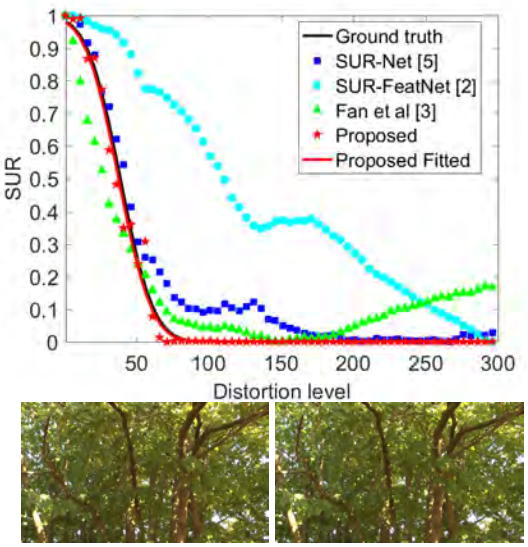
where SUR_i^P is the predicted SUR value at distortion level i and SUR_i^{gt} is the ground truth SUR at distortion level i.

$$\Delta\text{FIPSNR} = \mid \text{FIPSNR}^P - \text{FIPSNR}^{gt} \mid, \tag{9}$$

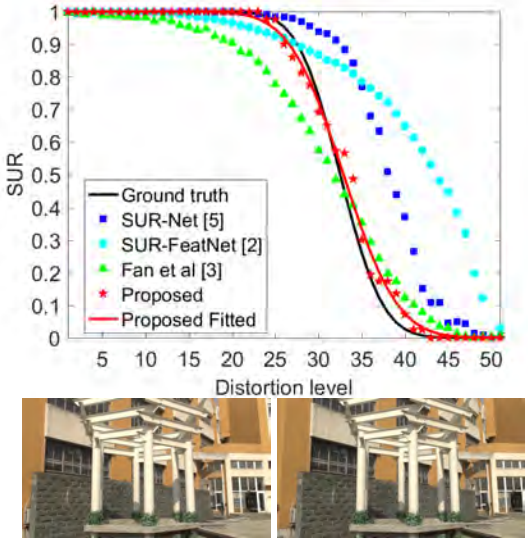
where FIPSNR^P is the stereoscopic image quality corresponding to the predicted 75% SUR and FIPSNR^{gt} is the stereoscopic image quality corresponding to the ground truth 75% SUR.



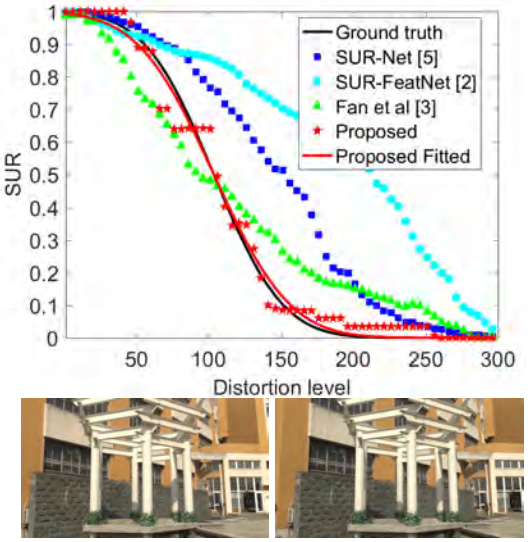
(a) Tree. Symmetric H.265 intra coding.



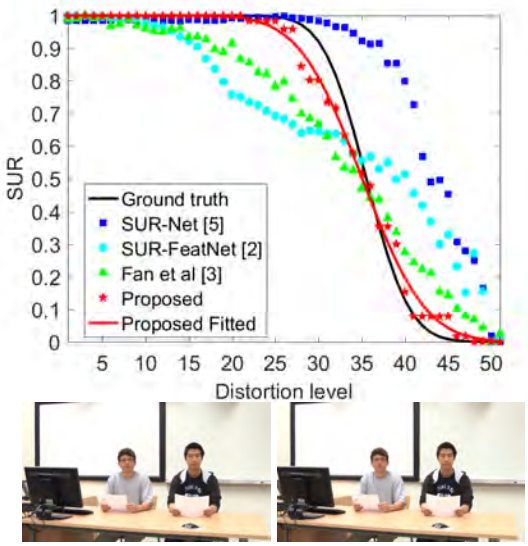
(b) Tree. Symmetric JPEG2000 compression.



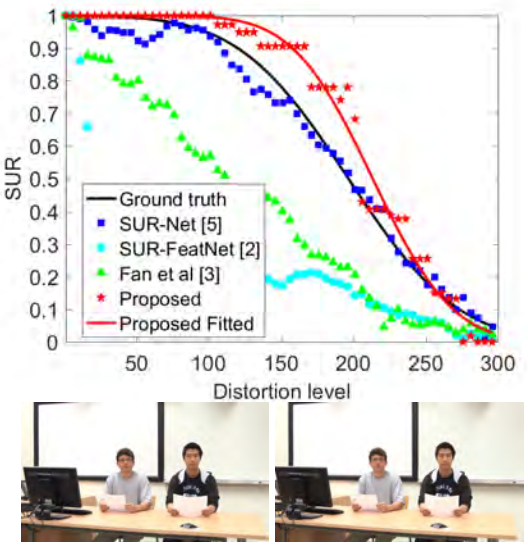
(c) NBuild. Symmetric H.265 intra coding.



(d) NBuild. Symmetric JPEG2000 compression.



(e) News. Symmetric H.265 intra coding.



(f) News. Symmetric JPEG2000 compression.

Fig. 6: Examples of the predicted SUR curve.

TABLE 1: Prediction for H.265 All-Intra coding on SIAT-JSSI and SIAT-JASI.

Comp.		Ground truth				Fitted Prediction				ΔPJND	ΔSUR	ΔFIPSNR	Bhattacharyya
Type	Image	QP	FIPSNR	μ	σ	QP	FIPSNR	μ	σ				distance
	People	29	53.95	32.76	6.01	27	55.43	30.24	4.76	2	0.05	1.47	0.04
	Basket.	29	56.18	34.29	7.74	32	53.52	34.67	4.44	3	0.06	2.66	0.07
	News.	33	54.35	35.73	4.22	32	55.15	36.08	5.93	1	0.03	0.80	0.03
	Tree.	27	54.90	30.87	5.08	27	54.90	31.08	5.95	0	0.02	0.00	0.01
Sym	Flower	32	52.65	35.55	5.28	29	55.28	32.53	5.19	3	0.06	2.63	0.04
	Comp.	29	54.84	34.47	7.45	31	53.16	33.95	4.55	2	0.06	1.69	0.06
	Volley.	31	53.57	35.27	6.85	30	54.57	33.69	4.75	1	0.04	1.01	0.04
	Pavi.	26	56.55	29.13	4.27	27	55.72	32.97	8.19	1	0.14	0.83	0.14
	NBuild.	30	55.27	32.90	3.66	31	54.28	33.77	4.41	1	0.02	0.99	0.01
	Fruit.	26	58.46	30.53	6.82	30	55.01	33.68	4.79	4	0.07	3.45	0.07
									Average	**1.8**	**0.05**	**1.55**	**0.05**
	People	29	57.00	34.29	7.43	28	57.73	32.21	6.14	1	0.04	0.73	0.02
	Basket.	28	60.10	34.36	9.47	35	53.83	37.50	3.78	7	0.10	6.27	0.21
	News.	25	63.25	32.57	10.75	33	57.44	37.10	5.82	8	0.10	5.81	0.12
	Tree.	31	54.89	34.04	4.48	28	57.18	32.62	6.84	3	0.04	2.29	0.05
Asym	Flower	36	52.34	39.19	4.02	30	57.84	34.41	6.40	6	0.09	5.50	1.92
	Soccer	33	54.16	38.36	7.42	30	56.79	33.73	5.65	3	0.09	2.64	0.08
	KBuild.	31	56.57	35.29	6.24	29	58.52	32.12	4.44	2	0.06	1.95	0.07
	Pavi.	26	59.76	30.89	7.42	34	53.07	37.04	4.07	8	0.12	6.69	0.22
	NBuild.	30	58.40	34.18	6.04	31	57.41	35.35	6.05	1	0.02	0.99	0.00
	Fruit.	22	64.79	30.82	12.53	26	61.56	30.73	7.15	4	0.08	3.23	0.07
									Average	**4.3**	**0.07**	**3.61**	**0.28**

TABLE 2: Prediction for JPEG2000 compression on SIAT-JSSI and SIAT-JASI.

Comp.		Ground truth				Fitted Prediction				ΔPJND	ΔSUR	ΔFIPSNR	Bhattacharyya
Type	Image	CR	FIPSNR	μ	σ	CR	FIPSNR	μ	σ				distance
	People	54	50.85	77.06	33.90	35	54.22	52.20	26.21	19	0.08	3.37	0.10
	Basket.	114	51.41	152.50	56.96	91	52.85	128.72	55.72	23	0.08	1.44	0.02
	News.	156	56.15	196.56	59.68	184	54.83	213.36	43.29	28	0.06	1.31	0.04
	Tree.	27	52.81	38.91	17.96	26	53.16	37.59	17.91	1	0.00	0.35	0.00
Sym	Flower	45	52.47	68.06	34.27	39	53.32	52.58	19.40	6	0.06	0.85	0.12
	Comp.	123	53.17	160.44	54.86	122	53.17	166.06	65.45	1	0.03	0.00	0.01
	Volley.	170	54.03	211.45	61.85	123	56.83	176.86	79.34	47	0.11	2.79	0.97
	Pavi.	20	56.20	28.81	12.66	46	50.23	117.02	104.90	26	0.32	5.97	0.89
	NBuild.	78	55.32	103.48	37.53	76	55.97	103.56	41.33	2	0.01	0.65	0.00
	Fruit.	56	56.53	77.10	30.62	70	53.83	93.90	35.90	14	0.06	2.70	0.04
									Average	**16.7**	**0.08**	**1.94**	**0.22**
	People	72	53.34	112.97	60.73	43	55.93	73.44	45.77	29	0.13	2.59	0.09
	Basket.	137	53.30	180.65	64.64	131	53.40	168.52	55.88	6	0.04	0.10	0.01
	News.	153	58.70	209.03	83.46	170	57.93	221.73	77.36	17	0.04	0.77	0.00
	Tree.	44	52.59	82.88	57.98	43	52.66	65.25	33.08	1	0.08	0.07	0.09
Asym	Flower	78	52.93	132.35	79.87	70	54.16	125.54	81.88	8	0.02	1.23	0.00
	Soccer	78	49.83	137.06	87.29	58	53.17	81.17	34.20	20	0.20	3.34	0.28
	KBuild.	63	54.31	116.24	78.76	40	57.53	88.78	71.97	23	0.09	3.22	0.02
	Pavi.	31	57.03	78.79	70.33	54	52.97	84.35	45.67	23	0.05	4.06	0.05
	NBuild.	114	56.05	164.50	75.06	102	56.12	170.38	101.31	12	0.05	0.08	0.02
	Fruit.	89	55.54	136.15	69.22	112	54.45	181.73	103.42	23	0.14	1.09	0.07
									Average	**16.2**	**0.08**	**1.66**	**0.06**

4.1 SUR Prediction for Symmetric and Asymmetric H.265 All-Intra Coding

In this section, we evaluate the SUR prediction for symmetric and asymmetric H.265 All-Intra coding for SIAT-JSSI and SIAT-JASI. Table 1 presents the prediction results for the 10 source images in each dataset. The table shows the ground truth and the fitted prediction, the mean μ and the standard deviation σ of the corresponding normal distribution, the QP parameter corresponding to the 75% SUR, and the FIPSNR for this QP. It also shows ΔPJND, ΔSUR, ΔFIPSNR, and the Bhattacharyya distance between the ground truth and the fitted prediction.

We find that: 1) For symmetric compression, ΔPJND was very low, with an average value of 1.8 and a maximum value of 4. For asymmetric compression, ΔPJND was slightly higher, with an average value of 4.3 and a maximum value of 8. 2) ΔSUR was very low. The maximum value was 0.14, and the average value was 0.05 and 0.07, for symmetric and asymmetric compression, respectively. 3) ΔFIPSNR was also low. For symmetric compression, the average prediction error was 1.55 dB, and the maximum prediction error was 3.45 dB. For asymmetric compression, the average and maximum errors were higher (3.61 dB and 6.69 dB, respectively). 4) For both symmetric and asymmetric compression, the Bhattacharyya distance was small, reaching zero for some images.

4.2 SUR Prediction for Symmetric and Asymmetric JPEG2000 compression

In this section, we evaluate the SUR prediction for symmetric and asymmetric JPEG2000 compression for SIAT-JSSI and SIAT-JASI. Table 2 presents the results for the 10 source images in each dataset. 1) For both symmetric and asymmetric compression, ΔPJND was higher than for H.265

All-Intra and varied widely across the images. For example, for symmetric compression, the minimum value was 1, while the maximum value was 47. 2) For both symmetric and asymmetric compression, ΔSUR was very low, with an average value of 0.08 and a maximum value of 0.32 and 0.20, respectively. 3) ΔFIPSNR was also low. For symmetric compression, the average prediction error was 1.94 dB, and the maximum prediction error was 5.97 dB. The average and maximum errors for asymmetric compression were lower (1.66 dB and 4.06 dB, respectively). 4) The Bhattacharyya distance was smaller for asymmetric compression than for symmetric compression.

4.3 Performance Comparison

To further assess the proposed method, we compared it to our previous method [3], and to two methods designed for 2D images: SUR-Net [5] and SUR-FeatNet [2]. We did not retrain SUR-Net and SUR-FeatNet for the SIAT-JSSI and SIAT-JASI datasets and used the trained versions in [5] and [2]. A prediction was made for the left view and the right view separately, and the result was averaged. Tables 3 and 4 compare the predictions for H.265 All-Intra and JPEG2000 compression in SIAT-JSSI and SIAT-JASI. For SUR-Net, SUR-FeatNet, and our previous method, we obtained the predicted PJND from the predicted SUR values as $\arg\min_{n=1,\ldots,N} |\mathrm{SUR}_k^n - 0.75|$. In terms of average ΔPJND, the proposed method outperformed the three other methods for symmetric and asymmetric H.265 All-Intra and for symmetric JPEG2000. For asymmetric JPEG2000, it was second best behind SUR-Net [5]. One major limitation of the methods designed for 2D images is that they did not consider binocular vision.

Fig. 6 compares the predicted SUR for the images "Tree", "Pavi", and "News". "Tree" has the highest texture complexity, followed by "Pavi" and "News". As shown in Fig. 6 (a) to (d), the predicted SUR curves of the proposed method are very close to the ground truth. We note that SUR-Net performs better than SUR-FeatNet. This may be because the SUR-Net model was trained on PJND samples that are assumed to be normally distributed, while the SUR-FeatNet model was trained on PJND samples that are assumed to follow a generalized extreme value distribution. Compared to our previous method [3], we obtained a higher prediction accuracy because we extracted more features, applied feature selection, and used a more powerful machine learning technique. Fig. 6 (e) and (f) show an example where the four methods did not make an accurate prediction. This may be because the size and diversity of the training set are small. In future work, the performance can be improved by using a larger dataset.

5 Conclusions

We proposed a learning-based model to predict the SUR for symmetrically and asymmetrically compressed stereoscopic images. We considered various monocular and binocular features that reflect the quality and content of the original and compressed stereoscopic images and applied feature selection to reduce the dimension of the feature space. The proposed method shows high prediction accuracy for symmetric and asymmetric H.265 All-Intra coding and JPEG2000 compression of stereoscopic images.

Acknowledgments

This work was supported in part by the Key-Area Research and Development Program of Guangdong Province Grant 2019B010137002 and 2018A0303130126, Shenzhen Science and Technology Program under Grant JCYJ20180507183823045, JCYJ20200109110410133, and JCYJ20170818163403748, the NSFC under Grant 61871372, 61902389 and 61901459, Guangdong R&D Grants 2019B010137002 and 2018A030313943, Guangdong International Science and Technology Cooperative Research Project under Grant 2018A050506063, in part by Membership of Youth Innovation Promotion Association, Chinese Academy of Sciences under Grant 2018392, in part by CAS President's International Fellowship Initiative (PIFI) under Grant 2020VTA008, in part by China Postdoctoral Science Foundation under Grant 2019M653127.

References

[1] Jin L, Lin J Y, Hu S, et al. Statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis. Electronic Imaging, Image Quality and System Performance XIII, 2016, 9:1-9.
[2] Lin H, Hosu V, Fan C, et al. SUR-FeatNet: predicting the satisfied user ratio curve for image compression with deep feature learning. Quality and User Experience, 2020, 5:1-5.
[3] Fan C, Zhang Y, Hamzaoui R, et al. Satisfied user ratio prediction with support vector regression for compressed stereo images. IEEE International Conference on Multimedia and Expo Workshops (ICMEW). London, July 2020.
[4] Liu H, Zhang Y, Zhang H, et al. Deep learning-based picture-wise just noticeable distortion prediction model for image compression. IEEE Transactions on Image Processing, 2019, 29: 641-656.
[5] Fan C, Lin H, Hosu V, et al. SUR-Net: predicting the satisfied user ratio curve for image compression with deep learning. IEEE Eleventh International Conference on Quality of Multimedia Experience (QoMEX). Berlin, June 2019.
[6] Huang Q, Wang H, Lim S C, et al. Measure and prediction of HEVC perceptually lossy/lossless boundary QP values. IEEE Data Compression Conference (DCC). Utah, Apr. 2017.
[7] Jiang Q, Shao F, Gao W, et al. Unified no-reference quality assessment of singly and multiply distorted stereoscopic images. IEEE Transactions on Image Processing, 2019, 28(4): 1866-1881.
[8] Friedman J. Greedy function approximation: a gradient boosting machine. Annals of Statistics, 2001: 1189-1232.
[9] Wang H, Katsavounidis I, Zhou J, et al. VideoSet: A large-scale compressed video quality dataset based on JND measurement. Journal of Visual Communication and Image Representation, 2017, 46: 292-302.
[10] Fan C, Zhang Y, Zhang H, et al. Picture-level just noticeable difference for symmetrically and asymmetrically compressed stereoscopic images: Subjective quality assessment study and datasets. Journal of Visual Communication and Image Representation, 2019, 62: 140-151.
[11] Hadizadeh H, Heravi A R, Bajić I V, et al. A perceptual distinguishability predictor for JND-noise-contaminated images. IEEE Transactions on Image Processing, 2019, 28(5): 2242-2256.
[12] Lin Y-H, Wu J-L, Quality assessment of stereoscopic 3D image compression by binocular integration behaviors, IEEE Transactions on Image Processing, 2014, 23(4): 1527-1542.
[13] Chen M, Su C, Kwon D, et al. Full-reference quality assessment of stereopairs accounting for rivalry. Signal Processing: Image Communication, 2013, 28(9): 1143-1155.
[14] Wu J, Li L, Dong W, et al. Enhanced just noticeable difference model for images with pattern complexity. IEEE Transactions on Image Processing, 2017, 26(6): 2682-2693.
[15] Hu S, Jin L, Wang H, et al. Compressed image quality metric based on perceptually weighted distortion. IEEE Transactions on Image Processing, 2015, 24(12): 5594-5608.

TABLE 3: Comparison of prediction accuracy for symmetric and asymmetric H.265 All-Intra coding on SIAT-JSSI and SIAT-JASI.

Comp. Type	Schemes	Image (ΔPJND/ΔSUR) People	Basket.	News.	Tree.	Flower	Compu.	Volley.	Pavi.	NBuild.	Fruit.	Average
Sym	SUR-Net [5]	10/0.15	12/0.16	15/0.20	1/0.05	2/0.04	7/0.08	4/0.07	12/0.20	5/0.08	15/0.18	8.3/0.12
	SUR-FeatNet [2]	7/0.13	3/0.14	12/0.17	13/0.23	10/0.18	1/0.09	5/0.07	18/0.33	7/0.18	16/0.27	9.2/0.18
	Fan et al. [3]	3/0.07	2/0.04	5/0.09	3/0.07	5/0.08	0/0.05	1/0.04	8/0.17	3/0.06	3/0.12	3.3/0.08
	Proposed	1/0.05	3/0.06	2/0.03	0/0.00	5/0.09	3/0.07	1/0.05	4/0.12	0/0.02	5/0.07	1.8/0.06
	Schemes	People	Basket.	News.	Tree.	Flower	Soccer.	KBuild.	Pavi.	NBuild.	Fruit.	Overall
Asym	SUR-Net [5]	11/0.19	12/0.17	8/0.16	6/0.12	4/0.08	12/0.17	8/0.12	12/0.24	5/0.11	11/0.16	8.9/0.15
	SUR-FeatNet [2]	6/0.10	3/0.11	3/0.09	9/0.18	8/0.13	11/0.14	11/0.17	19/0.30	7/0.14	19/0.26	9.6/0.16
	Fan et al. [3]	3/0.04	5/0.07	0/0.05	4/0.09	7/0.11	12/0.17	2/0.03	6/0.11	5/0.07	6/0.11	5.0/0.09
	Proposed	1/0.04	7/0.10	8/0.10	3/0.04	6/0.09	3/0.09	2/0.06	8/0.12	1/0.02	4/0.08	4.3/0.07

TABLE 4: Comparison of prediction accuracy for symmetric and asymmetric JPEG2000 compression on SIAT-JSSI and SIAT-JASI.

Comp. Type	Schemes	Image (ΔPJND/ΔSUR) People	Basket.	News.	Tree.	Flower	Compu.	Volley.	Pavi.	NBuild.	Fruit.	Average
Sym	SUR-Net [5]	47/0.24	21/0.08	11/0.03	2/0.05	13/0.11	32/0.14	21/0.07	33/0.29	30/0.14	30/0.15	24.0/0.13
	SUR-FeatNet [2]	3/0.10	68/0.17	144/0.44	41/0.31	89/0.39	30/0.13	42/0.13	137/0.61	65/0.31	101/0.41	71.9/0.30
	Fan et al. [3]	34/0.20	95/0.14	72/0.13	15/0.08	30/0.17	60/0.10	108/0.35	5/0.04	10/0.05	10/0.08	43.9/0.13
	Proposed	19/0.08	23/0.08	28/0.06	1/0.00	6/0.06	1/0.03	47/0.11	26/0.32	2/0.01	14/0.06	16.7/0.08
	Schemes	People	Basket.	News.	Tree.	Flower	Soccer.	KBuild.	Pavi.	NBuild.	Fruit.	Overall
Asym	SUR-Net [5]	21/0.11	14/0.03	9/0.07	18/0.12	8/0.08	21/0.14	11/0.09	26/0.11	4/0.04	9/0.04	14.1/0.08
	SUR-FeatNet [2]	16/0.04	83/0.16	147/0.44	29/0.16	68/0.22	44/0.11	38/0.09	134/0.44	35/0.12	57/0.32	65.0/0.21
	Fan et al. [3]	25/0.08	61/0.11	50/0.14	14/0.04	14/0.06	35/0.15	1/0.03	28/0.07	51/0.15	23/0.09	30.0/0.09
	Proposed	29/0.13	6/0.04	17/0.04	1/0.08	8/0.02	20/0.20	23/0.09	23/0.05	12/0.05	23/0.14	16.2/0.08

[16] Winkler S. Analysis of public image and video databases for quality assessment. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(6): 616-625.
[17] Liu Y, Yan W, Zheng Z, et al. Blind stereoscopic image quality assessment accounting for human monocular visual properties and binocular interactions. IEEE Access, 2020, 8: 33666-33678.
[18] Lowe D G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60: 91-110.

Chunling Fan received her Ph.D. degree from the University of Chinese Academy of Sciences in 2020. She is currently with Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. Her research interests are image and video quality assessment.

Yun Zhang received his Ph.D. degree at Institute of Computing Technology, CAS, Beijing. From 2009 to 2014, He was a visiting scholar in Department of Computer Science, City University of Hong Kong. He is currently a Professor at High Performance Computing Center, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. His research interests are video compression, 3D video processing, and visual perception.

Raouf Hamzaoui received the Dr.rer.nat. degree from the University of Freiburg, Germany, in 1997 and the Habilitation degree in computer science from the University of Konstanz, Germany, in 2004. He is currently a Professor in the Institute of Engineering Sciences, De Montfort University.

Djemel Ziou received the Ph.D. Degree from Lorraine INP. He served as a Lecturer at several French universities. Now, he is a Full Professor at the Université de Sherbrooke, Canada. He was the recipient of the prestigious NSERC research chair. His research interests include image processing, computer vision, and pattern recognition.

Qingshan Jiang received the Ph.D. degree in mathematics from the Chiba Institute of Technology, Japan, in 1996, and the Ph.D. degree in computer science from the University of Sherbrooke, Canada, in 2002. He is currently a Professor with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. His research interests include data mining, information security, and database technology.