

Visual Orientation Selectivity based Structure Description

Jinjian Wu, Weisi Lin, Guangming Shi, Yazhong Zhang, Weisheng Dong, and Zhibo Chen

Abstract—The human visual system is highly adaptive to extract structure information for scene perception, and structure character is widely used in perception-oriented image processing works. However, the existing structure descriptors mainly describe the luminance contrast of a local region, but cannot effectively represent the spatial correlation of structure. In this paper, we introduce a novel structure descriptor according to the orientation selectivity mechanism in the primary visual cortex. Researches on cognitive neuroscience indicate that the arrangement of excitatory and inhibitory cortex cells arise orientation selectivity in a local receptive field, within which the primary visual cortex performs visual information extraction for scene understanding. Inspired by the orientation selectivity mechanism, we compute the correlations among pixels in a local region based on the similarities of their preferred orientation. By imitating the arrangement of the excitatory/inhibitory cells, the correlations between a central pixel and its local neighbors are binarized, and the spatial correlation is represented with a set of binary values, which is named as the orientation selectivity based pattern. Then, taking both the gradient magnitude and the orientation selectivity based pattern into account, a rotation invariant structure descriptor is introduced. The proposed structure descriptor is applied in texture classification and reduced reference image quality assessment, as two different application domains to verify its generality and robustness. Experimental results demonstrate that the orientation selectivity based structure descriptor is robust to disturbance, and can effectively represent the structure degradation caused by different types of distortion.

Index Terms—Orientation Selectivity, Excitatory/Inhibitory Interaction, Structure Descriptor, Texture Classification, Image Quality Assessment

I. INTRODUCTION

Structure represents the main visual information of an image, and the human visual system (HVS) is highly adaptive to extract it for image understanding [1, 2]. Since the structural information plays an important role in visual perception, it is widely used in many perception-oriented image processing works, e.g., texture classification [3], image retrieval [4], face detection [5], quality assessment [6], and so on.

Jinjian Wu, Guangming Shi, Yazhong Zhang, and Weisheng Dong are with Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, School of Electronic Engineering, Xidian University. E-mail: jinjian.wu@mail.xidian.edu.cn.

Weisi Lin (corresponding author) is with the School of Computer Engineering, Nanyang Technological University, Nanyang 639798, Singapore. E-mail: wslin@ntu.edu.sg.

Zhibo Chen is with the Department of Electronic Engineering and Information Science, University of Science and Technology of China.

This work was supported by the Major State Basic Research Development Program of China (973 Program, No.2013CB329402), the National Natural Science Foundation of China (Nos. 61401325, 61472301, 61227004, 61401333, 61471281), the Research Fund for the Doctoral Program of Higher Education (No. 20130203130001), and the Fundamental Research Funds for The Central Universities (No. JB140227).

Though structure extraction is popular in many image processing tasks, how to effectively describe structural information is still an open problem. Generally speaking, some local image properties are used for structure analysis [7], such as the first-order statistic values (e.g., mean and variance) [2], the gray-level difference/luminance contrast [8], entropy [9], and so on. These structural descriptors are simple and easy to implement. However, they mainly represent the intensity character of the structure, and cannot effectively represent the spatial correlation [10]. By analyzing the co-occurring pixel values from neighborhood, Ojala et al. [11] introduced some new operators for structure description, within which the character of the spatial joint distribution can be effectively represented. Furthermore, in order to reduce the exponential growth of histogram size in these operators, Ojala et al. [3] adopted the signed gray-level differences to replace the absolute differences, and introduced the classical uniform local binary pattern (LBP) model. Though the LBP model succeeds in describing the spatial correlation of the structure, it is too sensitive to disturbance. Therefore, a more robust structure descriptor, which can effectively represent the intensity and spatial correlation, is required.

In this paper, we turn to investigate the subjective perception on image structure with the orientation characteristic. The human visual system (HVS) is highly sensitive to the orientation information when perceiving an input scene. Since orientation can effectively represent visual content, the local orientation information has been successfully used in many image processing works, such as in scale invariant feature transform (SIFT) based key point detection [12], histogram of oriented gradients (HOG) based human detection [5], and so on. Moreover, the HVS exhibits substantial orientation selectivity for scene perception [13, 14]. Neurophysiology research on visual cognition has found such a kind of neurons in the primary visual cortex, which tune specific orientations according to interactions among these correlated neurons [15]. Meanwhile, there are two spatial opponent types of interaction among these neurons, namely, excitation and inhibition [16]. These neurons with similar stimuli are more possible to tune in excitatory interaction, while those with dissimilar stimuli are more possible to act as inhibitory interaction [17]. The orientation selectivity arises from the arrangement of the excitatory and inhibitory interactions in a local receptive field, within which structure information is extracted in the primary visual cortex for scene perception [18, 19]. Therefore, orientation selectivity in the primary visual cortex reveals the inner mechanism for structure extraction.

Inspired by the orientation selectivity mechanism in the

B. Orientation Selectivity based Pattern

The orientation selectivity mechanism in the primary visual cortex reveals the general rule of feature extraction for visual stimuli recognition [26]. Inspired by the orientation selectivity mechanism, we imitate the excitatory and inhibitory interactions in the local receptive field and design a novel structure descriptor. The feedforward model posits that the orientation selectivity arises from the arrangement of cell responses in a local receptive field [13]; According to this, the orientation selectivity based pattern can be described as the organization of neighbor pixels. For a pixel x and its neighborhood $\mathcal{X}=\{x_1, x_2, \dots, x_n\}$, the pattern of its spatial correlation \mathcal{P} can be expressed as

$$\mathcal{P}(x|\mathcal{X}) = \mathcal{A}(\mathcal{I}(x|\mathcal{X})) = \mathcal{A}(\mathcal{I}(x|x_1, x_2, \dots, x_n)), \quad (1)$$

where $\mathcal{A}(\cdot)$ represents the arrangement of responses, namely, the pattern of correlations between the central pixel and its neighbor pixels; $\mathcal{I}(x|\mathcal{X})$ represents the interactions between x and \mathcal{X} , namely, the spatial correlations between x and pixels in \mathcal{X} .

Cortical cells in a local receptive field receive thousands of synapses altogether, and therefore, the correlations among $x_i \in \mathcal{X}$ are complex. For simplicity, in the feedforward model of orientation selectivity, only these synapses between the central cell and its excited cells are explored [13, 26]. According to the feedforward model, we neglect the correlations among $x_i \in \mathcal{X}$ and reorganize (1) as

$$\mathcal{P}(x|\mathcal{X}) \approx \mathcal{A}(\mathcal{I}(x|x_1), \mathcal{I}(x|x_2), \dots, \mathcal{I}(x|x_n)), \quad (2)$$

where $\mathcal{I}(x|x_i)$ is the interaction between x and x_i .

There are two spatial opponent interactions, namely, excitation ('+') and inhibition ('-'). The interaction type between cortical cell pair is determined by the correlation of their received stimuli. According to the correlation-based rule of synaptic plasticity, cortical cells with similar preferred orientations have a higher probability of connection and respond as excitatory interactions, and vice versa [25, 30]. As shown in Fig. 1, the third stimulus has the same orientation as the last stimulus, and the synapse between them responds as excitation. While the other three stimuli have different orientations with the last stimulus, and their synapses respond as inhibition. Therefore, we try to determine the interaction $\mathcal{I}(x|x_i)$ according to the preferred orientations of x and x_i . For an input image \mathcal{F} , the preferred orientation θ of each pixel $x \in \mathcal{F}$ is computed as

$$\theta(x) = \arctan \frac{G_v(x)}{G_h(x)}, \quad (3)$$

where G_h and G_v are the gradient magnitudes along the horizontal and vertical directions, which are computed with the Prewitt filters,

$$G_h = \mathcal{F} * f_h, \quad G_v = \mathcal{F} * f_v, \quad (4)$$

$$f_h = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \quad f_v = \frac{1}{3} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}, \quad (5)$$

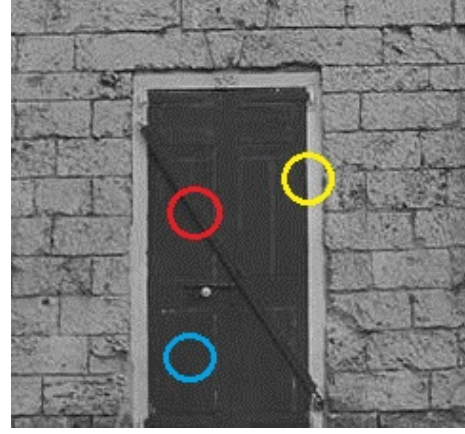


Fig. 2: An example of preferred orientation demonstration. The preferred orientation for the central pixel locating at the yellow circle is $\theta=89.87^\circ$, for the red one is $\theta=122.47^\circ$. For the pixel in the plain region, as shown in the blue circle, we set its preferred orientation $\theta=360^\circ$.

where $*$ denotes the convolution operation. With (3), (4), and (5), the preferred orientation θ of each pixel is acquired, and its range is $\theta \in [-180, 180]$. An example is shown in Fig. 2. For the pixel locating at the vertical edge, as the yellow circle shown in Fig. 2, its gradient magnitudes are $G_h=0.33$ and $G_v=150.67$, and its preferred orientation $\theta=89.87^\circ$. For the pixel locating at the oblique batten, as the red circle shown in Fig. 2, its gradient magnitudes are $G_h=-9.33$ and $G_v=19.33$, and its preferred orientation $\theta=122.47^\circ$. While for the pixel locating at the plain region, as the blue circle shown in Fig. 2, its gradient magnitudes along the horizontal and vertical directions are equal to zero, and there is no solution with (3). Since plain regions present different visual contents from the other types of region (e.g., edge and texture region, which have a certain gradient magnitudes), pixels in plain regions have distinct orientation characteristic. Therefore, we set the preferred orientation $\theta=360^\circ$ (be different from the preferred orientations of other pixels) for pixels in plain regions in this work.

The interaction $\mathcal{I}(x|x_i)$ can be estimated according to the similarity of the preferred orientations,

$$\mathcal{I}(x|x_i) = \begin{cases} 1 & \text{if } |\theta(x) - \theta(x_i)| < \mathcal{T} \\ 0 & \text{else} \end{cases}, \quad (6)$$

where '1' means excitatory interaction, '0' means inhibitory interaction, and \mathcal{T} is the judging threshold. Subjective viewing test on visual masking demonstrates that nearby gratings with similar orientations cause strong masking effect, and the masking effect is obviously reduced when the orientation difference is larger than a certain threshold (e.g., 12°) [31]. Considering the orientation difference on the two sides of x (i.e., plus or minus), we set $\mathcal{T}=6^\circ$ in this paper. With Eqs. (2) and (6), the pattern of the spatial correlation can be expressed by a set of binary values. In order to visualize how the orientation selectivity based pattern varies as a function of orientation correlations, we set the local receptive field as a 8-neighborhood region (i.e., we set $n=8$ for the neighborhood

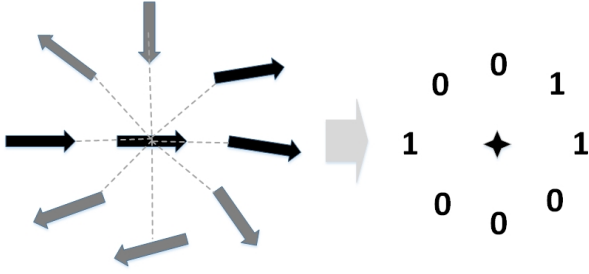


Fig. 3: An example of orientation selectivity based pattern. Pixels with similar preferred orientations to the central one respond as excitation ('1'), and these dissimilar ones respond as inhibition ('0').

\mathcal{X}), and give an example of the spatial arrangement of excitatory and inhibitory interactions and its corresponding pattern, which is shown in Fig. 3. As can be seen, the interactions between the central stimulus and its neighbor stimuli can be simplified as a 8-binary-value pattern.

C. Fundamental Patterns of Orientation Selectivity

By analyzing the response arrangement in a local receptive field, the orientation selectivity based pattern is created in the above subsection. However, the number of pattern increases exponentially with the neighbor number. For example, a local receptive field with 8 neighbors will present 256 different types of patterns.

In order to reduce the number of patterns, we explore the relationships between these orientation selectivity based patterns and try to combine the similar ones. During experiment we found that these patterns with same excitatory subfield (where the excitatory interaction respond) are more correlated and represent much similar response. For example, the patterns [00000011] and [00001100] have a same size of excitatory subfield, as shown in the third column of Fig. 4, the excitatory interactions of the two patterns locating at a 90° sector domain. Moreover, the excitatory interactions of the patterns [00000111] and [00000101] locate in a 135° sector domain, as shown in the fourth column of Fig. 4. Therefore, we combine these patterns with a same size of excitatory subfield. For a 8-neighborhood local receptive field, the excitatory subfield can be divided into 8 sector domains, i.e., $\{0^\circ, 45^\circ, \dots, 315^\circ\}$. As a result, all of the 256 orientation selectivity based patterns can be located into 8 types of fundamental patterns (\mathcal{P}_f) according to their corresponding sector domains, as shown in Fig. 4. Moreover, the fundamental patterns present rotation invariant, e.g., the three patterns [00000001], [00000010], and [00000100] locate at the same sector domain (with 45°), as shown in the second column of Fig. 4, which will be described as a same fundamental pattern (e.g., \mathcal{P}_f^1).

D. Orientation Selectivity based Structure Extraction

A successful image structure descriptor should represent both the spatial correlation and intensity of image content. In the above subsection, the spatial correlation of structure is described with the orientation selectivity based pattern. The

intensity of image structure can be computed as the gradient magnitude,

$$\mathcal{M}(x) = \sqrt{(G_h(x))^2 + (G_v(x))^2}. \quad (7)$$

So far, a novel orientation selectivity based structure descriptor $\{\mathcal{P}_f/\mathcal{M}\}$ is created for image structure extraction. The structural characteristic of an image is usually represented by its structural histogram. Therefore, we calculate the structural information (i.e., $\{\mathcal{P}_f/\mathcal{M}\}$) of all pixels to build a structural histogram. There are two kinds of mapping methods for structural histogram construction: 1) directly calculate the number of structural patterns; 2) considering the contribution of gradient magnitude and calculate the weighted distribution of structural patterns. For the first condition, the structural histogram, which is named as *Orientation Selectivity based Structural histogram (OSS)*, is mapped as,

$$H(k) = \sum_{x=1}^N \delta(\mathcal{P}_f(x), \mathcal{P}_f^k), \quad (8)$$

$$\delta(\mathcal{P}_f(x), \mathcal{P}_f^k) = \begin{cases} 1 & \text{if } \mathcal{P}_f(x) = \mathcal{P}_f^k \\ 0 & \text{else} \end{cases}, \quad (9)$$

where N is the dimension of the image \mathcal{F} , and \mathcal{P}_f^k is the k -th fundamental pattern.

For the second condition, the contribution of gradient magnitude of structure is considered, and the weighted structural histogram, which is named as *Weighted Orientation Selectivity based Structural histogram (WOSS)*, is mapped as,

$$H_w(k) = \sum_{x=1}^N w(x) \delta(\mathcal{P}_f(x), \mathcal{P}_f^k), \quad (10)$$

where $w(x)$ is the weight factor due to the gradient magnitude, and we simply set $w(x) = \mathcal{M}(x)$ in this paper.

In (8) and (10), the bin number of the structural histogram is determined by the number of the fundamental patterns. For a 8-neighborhood local receptive field, it has 8 different types of fundamental patterns, and therefore, an image can be mapped into a 8-bin histogram.

III. EXPERIMENTAL RESULTS

The orientation selectivity based structure is quite consistent with subjective perception. In order to demonstrate the effectiveness, we apply the proposed structure descriptor for texture classification. Moreover, it is used to extract structural information from distorted image for quality assessment.

A. Texture Classification

In this subsection, the rotation invariance property of the orientation selectivity based structure is firstly illustrated. Three images with the same texture but different rotation angles (with 0° , 45° , and 90° , respectively) are shown in Fig. 5, and their corresponding OSS based histograms are shown in Fig. 6 (in which the 1th bin corresponds to patterns with angles 0° in Fig 4, the 2th bin corresponds to 45° , \dots , the 8th bin corresponds to 315°). As can be seen, the OSS histograms of the three images are quite similar, especially

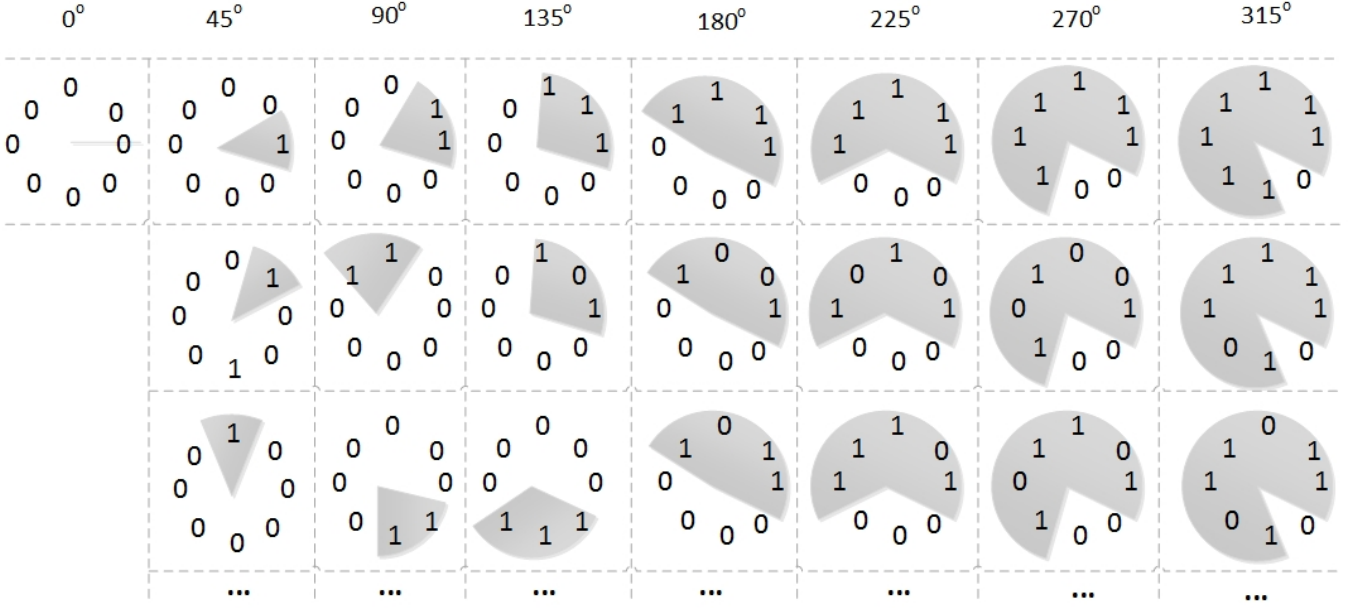


Fig. 4: The combination of orientation selectivity based pattern based on their excitation subfield.

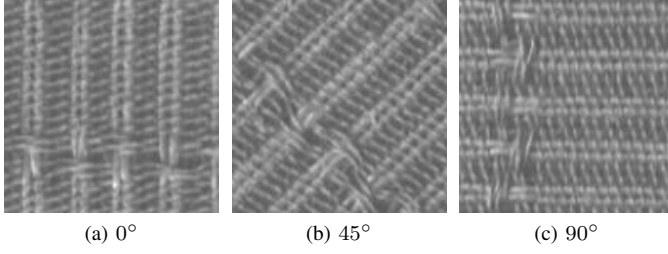


Fig. 5: Images with a same type of texture while under three different rotation angles.

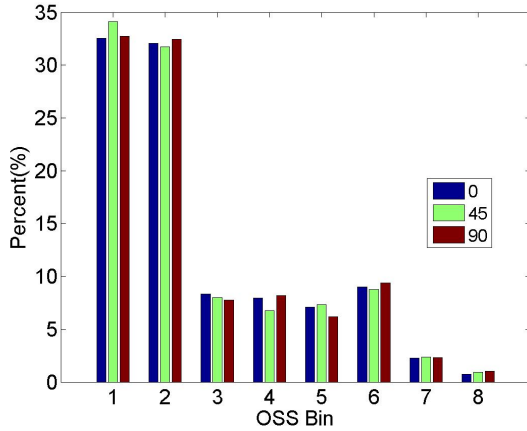


Fig. 6: The OSS based histogram for the three texture images in Fig. 5.

for the two texture images with angles 0° and 90° (as the blue and red bars shown in Fig. 6). The OSS histogram of Fig. 6 (b) (with 45° angle) is a little different from that of the other two images, and this is because the content of this image is a little different from the others. In summary, OSS presents rotation invariant for texture description.

Then, the orientation selectivity based structure is adopted for texture classification. In order to compute the structural dissimilarity between two images, the chi-square distance is employed, which could be regarded as the weighted L^2 -norm between their structural histogram H_1 and H_2 ,

$$D(H_1, H_2) = \sum_{k=1}^n \frac{(H_1(k) - H_2(k))^2}{H_1(k) + H_2(k)}. \quad (11)$$

A well accepted texture database, Outex [32], is selected for texture classification experiment. Outex database is composed of 24 classes of textures with three illuminations ('horizon', 'inca', and 't184') and nine angles (0° , 5° , 10° , 15° , 30° , 45° , 60° , 75° , and 90°). The performance of the proposed method is evaluated according to the classification rate by using the chi-square distance and the nearest neighborhood classifier. Moreover, the classic texture classification method, i.e., local binary pattern (LBP) [3], its improved method LBPV [20] (LBP with variance) and CLBP (considering the local gray level, the sign and magnitude features of local difference) [33], and the local phase quantization method (LPQ) [34] are adopted for comparison. For fair comparison, the size of the neighborhood for all methods is set as $n=8$.

The experiments are carried on two test suites of Outex, namely, Outex_TC_00010 (TC10) and Outex_TC_00012 (TC12), which include the same 24 classes of textures. For the training and testing datasets choosing, we follow the default

TABLE I: Classification rate (%) by different descriptors.

DB \ AI	LBP	OSS	LBPV	WOSS	LPQ	CLBP
TC10	55.16	84.53	91.41	88.83	51.6	96.56
TC12	49.00	71.81	76.41	78.80	54.6	90.30

setting of Outex. Thus, as in [20, 33], the experimental setups for the two suites of Outex are listed as follows,

- For TC10, these images under ‘inca’ illumination and ‘0°’ angle (with a number of 480 images) are chosen for classifier training, and the other images (with a number of 480×8 images) for test.
- For TC12, the classifier is trained under the same training condition as TC10, and tested with all images under ‘t184’ and ‘horizon’ illuminations and nine angles (with a number of $480 \times 2 \times 9$ images).

The classification results on TC10 and TC12 are listed in Table I. As can be seen, the proposed OSS achieves 20% – 30% increase against LBP and LPQ on both TC10 and TC12, the proposed WOSS performs similar with LBPV on the two databases, and worse than CLBP which considers more local information than the other methods. Therefore, the proposed structural descriptor is comparable with these improved LBP-based structural descriptors for texture classification.

Furthermore, the proposed structural descriptor is very robust to disturb, while the LBP based structural descriptor is quite sensitive to small disturb. In order to give a clear view about this, the function of white noise on LBP/OSS based histogram is analyzed (Here, we choose LBP/OSS based structural histogram, rather than LBPV/WOSS based histogram. That is because the former is directly based on the number of LBP/OSS, which can better represent the change on structural pattern).

Fig. 7 shows a texture image contaminated by weak white noise (with $\text{PSNR}=30\text{dB}$) and strong white noise (with $\text{PSNR}=23\text{dB}$) respectively. As shown in Fig. 7 (b), when the image is contaminated by the weak white noise, we can hardly sense the distortion. In other words, the weak white noise in Fig. 7 (b) has limited damage on the structural information. However, the LBP based structural histogram is obviously changed under the weak white noise, as the green bars shown in Fig. 8 (a). Since the weak white noise has limited affection on image orientation, the OSS based histogram is almost invariant under it, as the green bars shown in Fig. 8 (b). When the image is contaminated by strong white noise, its visual structure is obviously distorted, as shown in Fig. 7 (c). Under the affection of strong noise, both of the LBP and OSS based structural histograms are changed, as the red bars shown in Fig. 8. Therefore, OSS is robust for structural information extraction under different levels of noise, which performs quite consistently with the human perception.

In order to present a comprehensive analysis on the robustness of the texture classification algorithms against noise, we carried out the texture classification experiments on white noise contaminated TC10 and TC12 databases. Three different levels of white noise (with PSNR of 30 dB, 27 dB, and

TABLE II: Classification rate (%) on contaminated images by different descriptors.

DB \ AI	PSNR	LPQ	LBP	OSS	LBPV	WOSS	CLBP
TC10	30	44.84	44.19	80.78	86.88	88.73	94.87
	27	38.49	37.68	73.20	79.84	87.27	93.36
	23	22.99	24.74	61.59	62.27	77.16	84.87
TC12	30	46.71	44.61	69.49	73.84	78.73	86.00
	27	38.31	38.59	66.27	67.64	77.64	84.03
	23	25.05	26.81	52.73	53.84	68.94	75.32

23 dB, under which images represent little, obvious, and severe distortion, respectively) are injected into the two texture databases.

The classification results are listed in Table II, and the decreases of the classification rates against the original results are shown in Fig. 9. As the black curves shown in Fig. 9 (a) and (b), the classification rates based on LBP structure are greatly decreased with the disturbance of the white noise. The LPQ performs much similar with LBP on these contaminated textures. The decreasing rate of OSS is much smaller than that of LBP and LPQ, as shown in Table II, OSS achieves about 40% increase on TC10 against LBP and LPQ under white noise (the increase is about 30% for the original TC10 database), and achieves about 25% increase on TC12 (the original increase is about 20%). WOSS performs quite steady under white noise. As shown in Fig 9, the classification rates are almost unchanged on both databases under weak white noise ($\text{PSNR}=30\text{dB}$), decreased a little bit when the noise becomes stronger ($\text{PSNR}=27\text{dB}$), and obviously decreased under sever noise ($\text{PSNR}=23\text{dB}$). While the classification rate of LBPV is obviously effect by white noise. As a result, WOSS performs much better than LBPV under white noise, as shown in Fig 9, which achieves about 15% increase under sever noise ($\text{PSNR}=23\text{dB}$) for both TC10 and TC12 databases. Moreover, though CLBP considers much more local information than other descriptors, WOSS performs as robust as CLBP (as the purple curve shown in Fig 9) on both TC10 and TC12 databases.

According to the analysis above, we can conclude that the proposed orientation selectivity based structure descriptor is robust to noise, and performs comparable with the state-of-the-art algorithms for texture classification.

B. Image Quality Assessment

In this subsection, the orientation selectivity based structure is used for reduced-reference (RR) image quality assessment (IQA). The effectiveness of the orientation selectivity based structure for representing visual information degradation is firstly demonstrated. Then, we propose a novel RR IQA method, and compare it with three state-of-the-art RR IQA algorithms and three classic full reference (FR) IQA metrics.

It is well known that different types of distortion result in different quality degradations [35]. An example is shown in Fig. 10, where the reference image is distorted by three different types of noise, i.e., additive white Gaussian

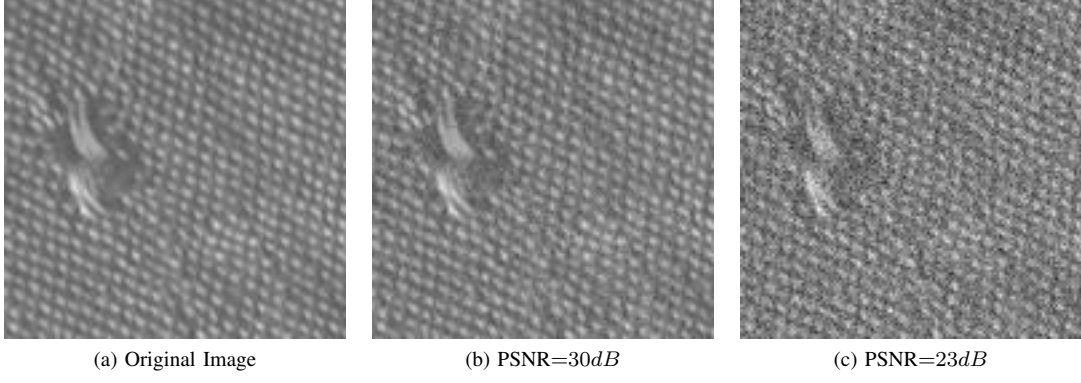
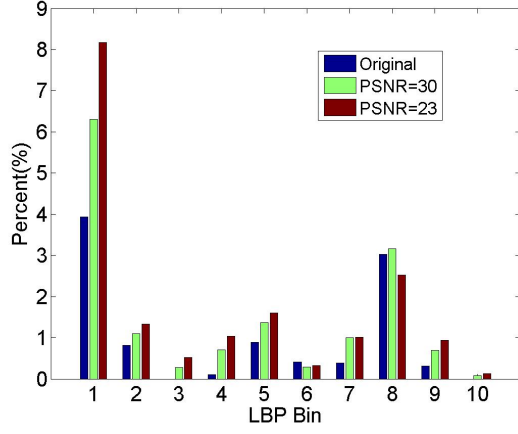
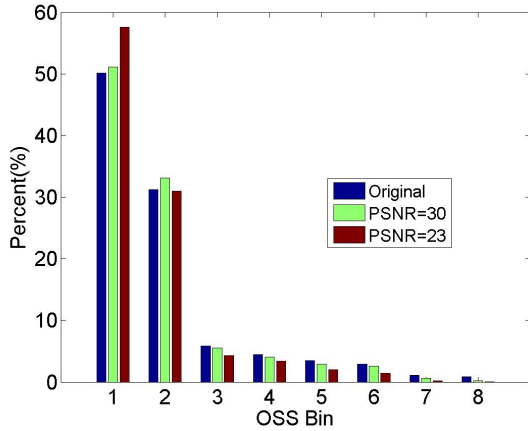


Fig. 7: Texture image distorted by different levels of white noise.

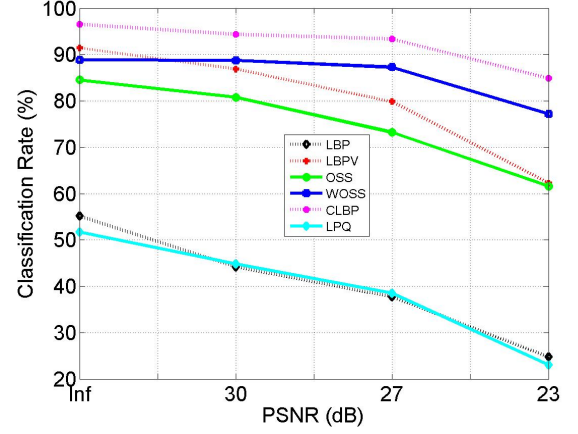


(a) LBP histogram

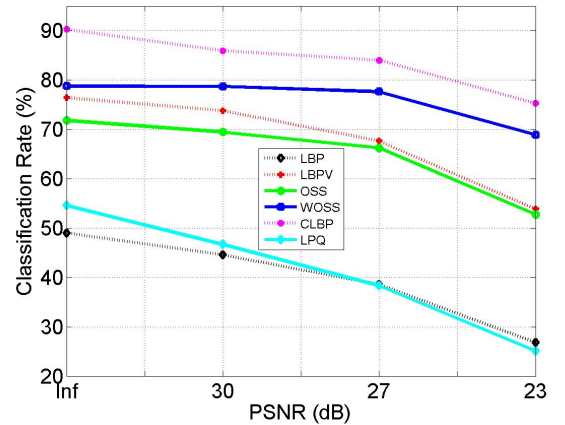


(b) OSS histogram

Fig. 8: The robustness of the LBP/OSS based structural histograms.



(a) TC10



(b) TC12

Fig. 9: Decrease of classification rates under different levels of white noise.

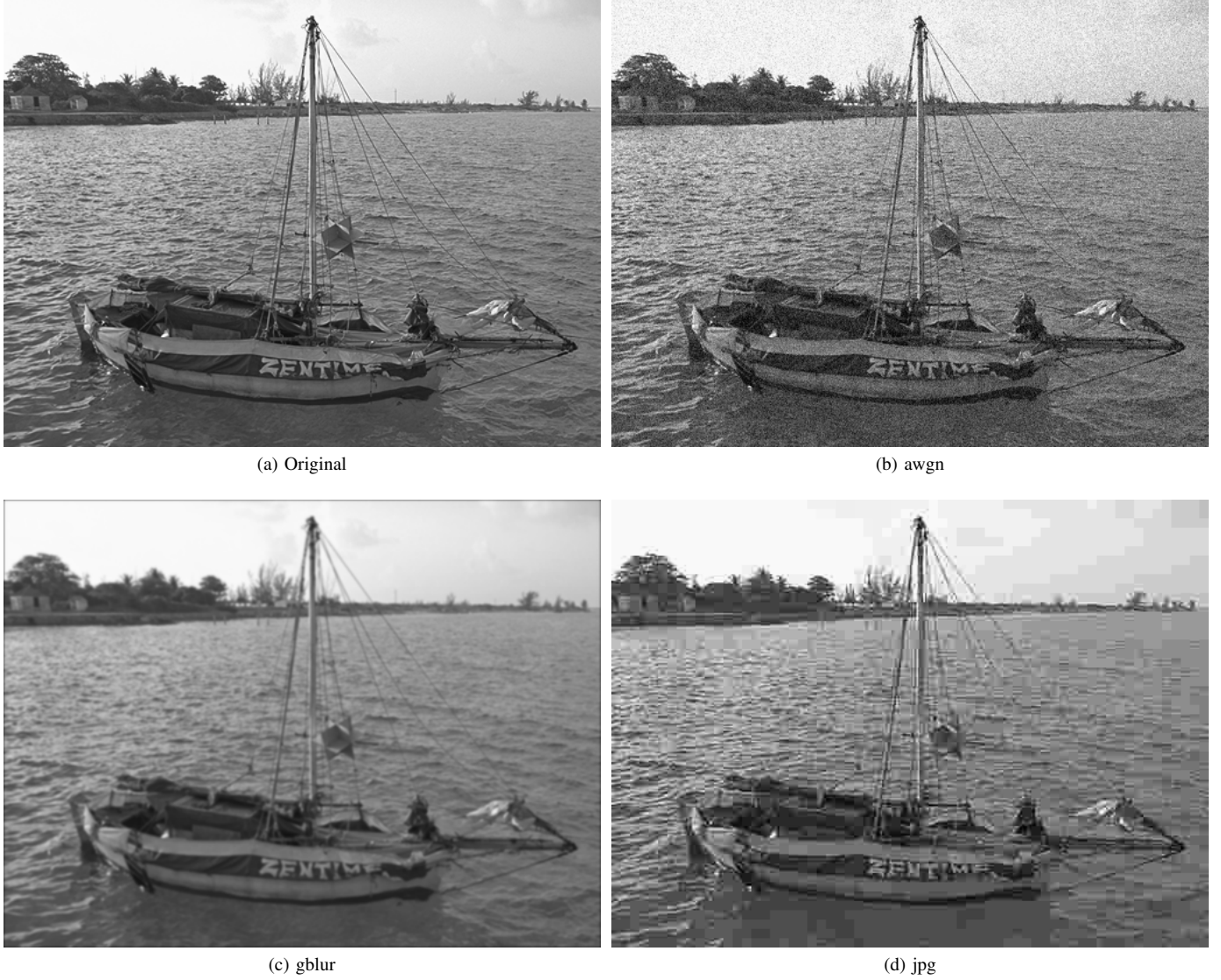


Fig. 10: The boat image distorted by three different types of noise, namely, additive white Gaussian noise (with PSNR=24.7 dB), Gaussian blur (with PSNR=25.1 dB), and JPEG compression noise (with PSNR=24.7 dB).

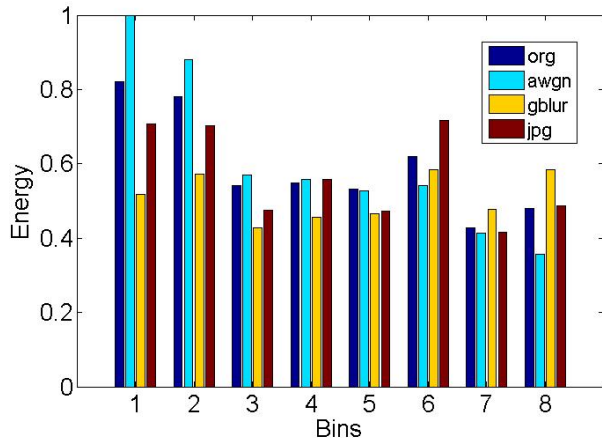


Fig. 11: Structure degradation under different types of distortion.

noise (awgn), Gaussian blur noise (gblur), and JPEG compression noise (jpg), respectively. Meanwhile, the noise levels in the three distorted images are much similar, where Fig. 10 (b) with PSNR=24.7 dB, Fig. 10 (c) with PSNR=25.1 dB, and Fig. 10 (d) with PSNR=24.7 dB. With (10), the structural information of an image is extracted, and the visual content of the image is represented by a structural histogram.

Different types of distortion will cause various structure decays. Intuitively, as shown in Fig. 10, the awgn mainly adds random disturbance into the whole image, which causes much more obvious distortion on the plain region (e.g., the sky) than that on the other regions; On the contrary, the gblur mainly weakens the luminance contrast of texture and edge regions, while has limited affection on the plain region; The jpg smooths out some high frequency information, meanwhile, it brings new false edges due to the blockiness artifact.

As a result, the three types of noise generate different

changes on the structural histograms. As the babe blue bars shown in Fig. 11, with the distortion of the awgn, the energies of the first and second bars (corresponding to structures which are more likely to appear in the disorderly regions) are obviously increased, while that for the last bar (corresponding to the structure which is more likely to appear in the plain regions) is decreased. That is because the awgn increases the variance values for most of the image regions, and some plain regions are tuned into disorderly (e.g., the sky). On the contrary, the gblur decreases the variance values for most of the image regions, and tunes some texture regions into smooth ones. As a result, the energies of most bins are decreased, especially for the first and second bars, while the energies of the last bar is increased, as the yellow bars shown in Fig. 11. Since the jpg creates new edge due to the blockiness artifact, the energy of the sixth red bar which represents edge structure is increased. Meanwhile, the jpg removes some high frequency information, the energies for the other bars are decreased, especially for the first and second bars. In summary, different types of noise cause different degradations on image structures, and the orientation selectivity based structural histogram can effectively represent the degradations from different types of noise.

According to the analysis hereinabove, we measure image quality according to the change on structural histogram, and introduce a novel RR IQA method. For a test image $\hat{\mathcal{F}}$ (and its corresponding reference image \mathcal{F}), the structural information can be extracted by Eq. (10), and its visual content can be represented by a 8-bin structural histogram. Furthermore, images are composed with many sizes. In order to offer a certain uniformity in scene representation over multiple scales, the eye will automatically zoom in and out according to the viewing condition [36], which results in different projection size of the local receptive field for an object [37]. In order to extract more visual information from an image, a multiscale framework is employed. For simplicity, we down sample the original image and create a L -level pyramid. In this experiment, we set $L=2$, and therefore, the structural information of a test image $\hat{\mathcal{F}}$ (and its reference image \mathcal{F}) can be represented by a 16-bin structural histogram \hat{H}_w (H_w). The visual content degradation is then measured as the similarity between \hat{H}_w and H_w ,

$$\mathcal{S}(k) = \frac{2\hat{H}_w(k)H_w(k)}{\hat{H}_w^2(k) + H_w^2(k)}, \quad (12)$$

where $k \in \{1, 2, \dots, 16\}$. As a result, the content degradations on all of the 16 bins are acquired.

Since each bin represents an individual structural pattern, its change results in different quality degradation. Therefore, we employ the support vector machine regressor (SVR) rather than equally accumulation for feature pooling. Here, the LIB-SVM package [23] is adopted to implement the SVR with a radial basis function kernel, and the quality score of an image is predicted as,

$$\mathcal{Q}_s(\hat{\mathcal{F}}) = \text{SVR}(\mathcal{S}, \text{MOD}), \quad (13)$$

where \mathcal{Q}_s is the quality score, and MOD is a trained model for regression.

We compare the proposed RR IQA method (which we called WOSS) with three state-of-the-art RR IQA methods (i.e., RRED [22], WNISM [21], and RRVIF [38]) and four classic full-reference (FR) IQA (i.e., SSIM [2], MSSIM [39], FSIM [40], and PSNR [41]). Meanwhile, three popular large databases in the quality assessment society are selected for the comparison experiments: LIVE database [42], including 29 reference images and 799 distorted images cross 5 different types of noise; CSIQ database [43], including 30 reference images and 866 distorted images cross 6 different types of noise; and TID database [44], including 25 reference images and 1700 distorted images cross 17 types of noise.

Since the predicted quality scores from different IQA algorithms always have different ranges, a five-parameter mapping function is adopted to nonlinearly regress the scores into a common space [35],

$$\mathcal{Q} = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(\mathcal{Q}_s - \beta_3))} \right) + \beta_4 \mathcal{Q}_s + \beta_5, \quad (14)$$

where $\beta_1, \beta_2, \beta_3, \beta_4$, and β_5 are five parameters to be fitted.

Then, three well accepted performance criteria, namely, linear correlation coefficient (CC), Spearman rank-order correlation coefficient (SRCC), and root mean squared error (RMSE), are employed to judge the performances of different IQA algorithms, which compute the correlations between the regressive scores \mathcal{Q} and their corresponding subjective quality scores (i.e., the mean opinion score (MOS) or the different MOS (DMOS)). A good IQA algorithm returns high CC/SRCC values and a low RMSE value. Meanwhile, a better RR IQA algorithm should use less data from reference images and achieve higher prediction accuracy. In this experiment, the quantities of reference data used in WOSS, RRED, WNISM, and RRVIF are 16, 16, 18, and 2, respectively. Meanwhile, the FR IQA metrics use the whole information of the reference image, so the quantities of reference data they used are N (the size of the reference image).

Firstly, the LIVE database is chosen for performance comparison, which includes five types of noise: JPEG2000 compression noise (j2k), JPEG compression noise (jpg), additive white Gaussian noise (awgn), Gaussian blur (gblur), and a Rayleigh fast-fading channel simulation noise (ff). Since the proposed IQA method adopts SVR for feature pooling, a train-test procedure is required. In this experiment, 80% of the reference images and their corresponding distorted images in the LIVE database are randomly chosen for training, and the rest for testing (which ensures that no overlap occurs between the training and testing images). Moreover, in order to eliminate performance bias, the random train-test procedure is repeated for 100 times, and the average performance is calculated to return the final result. The performances of the IQA algorithms on the LIVE database are listed in Table III. As can be seen, WOSS performs better than the other RR IQA algorithms on awgn and gblur, performs almost same with the best RR IQA algorithm on jpg, and performs quite similar with the best RR IQA algorithm on j2k and ff. Meanwhile, the overall performance of WOSS is better than the other RR IQA algorithms. Furthermore, WOSS performs better than the

TABLE III: PERFORMANCE OF IQA INDICES ON LIVE DATABASE.

Dist.	Algo. Crit.	RR				FR			
		OSS	RRED [22]	WNISM [21]	RRVIF [38]	PSNR	SSIM [2]	MSSIM [39]	FSIM [40]
No. of scalars		16	16	18	2	N	N	N	N
j2k	CC	0.932	0.952	0.924	0.932	0.896	0.941	0.957	0.978
	SRCC	0.921	0.946	0.920	0.950	0.889	0.936	0.953	0.971
	RMSE	8.91	7.74	6.18	5.88	11.0	8.51	4.69	5.22
jpg	CC	0.958	0.959	0.876	0.895	0.860	0.951	0.943	0.984
	SRCC	0.944	0.953	0.851	0.885	0.841	0.944	0.942	0.983
	RMSE	8.91	9.08	7.71	7.15	14.5	9.89	5.33	5.72
awgn	CC	0.979	0.946	0.890	0.957	0.982	0.969	0.974	0.965
	SRCC	0.972	0.946	0.870	0.946	0.985	0.963	0.973	0.965
	RMSE	5.58	9.08	7.29	4.66	4.334	6.901	3.65	7.34
gblur	CC	0.972	0.956	0.888	0.955	0.784	0.874	0.955	0.969
	SRCC	0.962	0.952	0.915	0.961	0.782	0.894	0.954	0.971
	RMSE	4.19	5.42	7.22	4.66	11.5	8.96	4.69	4.57
ff	CC	0.924	0.895	0.925	0.944	0.890	0.945	0.947	0.946
	SRCC	0.907	0.918	0.923	0.941	0.890	0.941	0.947	0.950
	RMSE	10.4	12.7	6.25	5.42	13.0	9.36	5.30	9.21
overall	CC	0.863	0.827	0.710	0.725	0.872	0.904	0.943	0.960
	SRCC	0.865	0.830	0.703	0.732	0.876	0.910	0.945	0.963
	RMSE	12.4	15.3	18.4	17.6	13.4	11.7	9.09	7.68

FR IQA metrics on gblur, and a slightly worse than the best one on jpg and awgn.

In addition, another two databases, namely, CSIQ and TID, are chosen for comparison. The same train-test procedure as that for LIVE database is adopted, and the performances on the CSIQ and TID databases are listed in Table IV. As can be seen, WOSS performs much better (>20% improvement) than the other RR IQA algorithms on the CSIQ database, and also outperforms the other three RR IQA algorithms on the TID database. Meanwhile, WOSS even performs better than two FR IQA metrics (i.e., PSNR and SSIM) on both CSIQ and TID databases, and is quite comparable with MSSIM and FSIM (FR metrics) on the two databases.

According to the performances on the three large databases, we can conclude that as a RR IQA algorithm, WOSS outperforms the existing RR IQA algorithms, and is quite comparable with the classic FR IQA metrics.

IV. CONCLUSION

In this paper, we have introduced a novel structure descriptor based on the orientation selectivity mechanism of the primary visual cortex. The HVS is highly adaptive to extract structure information for visual perception, and structure character is widely used in perception-oriented image processing tasks, such as image retrieval, face detection, texture classification, quality assessment, and so on. However, existing structural descriptors mainly describe the intensity character, but can not effectively represent spatial correlation of structure. Inspired by the orientation selectivity mechanism, we explore the excitatory/inhibitory interactions among cortex cells for visual structure extraction. By imitating the arrangement of the excitatory/inhibitory cells in a local receptive field, the orientation selectivity based pattern was introduced to represent the spatial correlation of structure. Then, according to

the gradient magnitude and the orientation selectivity based pattern, a novel structure descriptor was proposed.

With the novel structure descriptor, image structure was extracted and an orientation selectivity based structural histogram was created to represent the image content. Then, the novel structure descriptor was applied to two different domains of applications, namely, texture classification and reduced reference image quality assessment, to demonstrate its generality and robustness. Experimental results on texture classification demonstrated that the proposed structure descriptor is rotation invariant, and it is robust to disturbance. Experimental results on reduced reference image quality assessment demonstrated that the proposed structure descriptor can effectively represent degradation on visual content, and it performs quite consistently with human perception.

REFERENCES

- [1] D. Marr, "Visual information processing: the structure and creation of visual representations," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 290, pp. 199–218, 1980.
- [2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [3] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [4] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 1, pp. 39–62, 1999.

TABLE IV: OVERALL PERFORMANCE OF IQA INDICES ON CSIQ and TID DATABASES.

DB	Algo. Crit.	RR				FR			
		WOSS	RRED [22]	WNISM [21]	RRVIF [38]	PSNR	SSIM [2]	MSSIM [39]	FSIM [40]
CSIQ	CC	0.901	0.695	0.696	0.698	0.800	0.815	0.900	0.928
	SRCC	0.895	0.773	0.705	0.733	0.806	0.838	0.914	0.940
	RMSE	0.113	0.189	0.189	0.182	0.158	0.152	0.115	0.098
TID	CC	0.782	0.712	0.572	0.535	0.573	0.641	0.842	0.886
	SRCC	0.759	0.702	0.495	0.500	0.579	0.627	0.853	0.890
	RMSE	0.829	0.943	1.101	1.134	1.100	1.03	0.723	0.623

- [5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," vol. 1. IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.
- [6] W. Lin and C. C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.
- [7] T. Ojala, K. Valkealahti, E. Oja, and M. Pietikinen, "Texture discrimination with multidimensional distributions of signed gray-level differences," *Pattern Recognition*, vol. 34, no. 3, pp. 727–739, 2001.
- [8] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment base on gradient similarity," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500–1512, 2012.
- [9] J. Wu, F. Qi, G. Shi, and Y. Lu, "Non-local spatial redundancy reduction for bottom-up saliency estimation," *Journal of Visual Communication and Image Representation*, vol. 23, no. 7, pp. 1158–1166, 2012.
- [10] J. Wu, W. Lin, and G. Shi, "Image quality assessment with degradation on spatial structure," *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 437–440, 2014.
- [11] T. Ojala, M. Pietikinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [13] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, no. 1, pp. 106–154, 1962.
- [14] T. D. Albright, "Direction and orientation selectivity of neurons in visual area MT of the macaque," *Journal of neurophysiology*, vol. 52, no. 6, pp. 1106–1130, 1984.
- [15] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat," *Journal of neurophysiology*, vol. 28, pp. 229–289, 1965.
- [16] E. L. Bienenstock, L. N. Cooper, and P. W. Munro, "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex," *The Journal of neuroscience: the official journal of the Society for Neuroscience*, vol. 2, no. 1, pp. 32–48, 1982.
- [17] N. J. Priebe and D. Ferster, "Inhibition, spike threshold, and stimulus selectivity in primary visual cortex," *Neuron*, vol. 57, no. 4, pp. 482–497, 2008.
- [18] J. A. Cardin, L. A. Palmer, and D. Contreras, "Stimulus feature selectivity in excitatory and inhibitory neurons in primary visual cortex," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 27, no. 39, pp. 333–344, 2007.
- [19] O. Yizhar, L. E. Fenno, M. Prigge, F. Schneider, T. J. Davidson, D. J. OShea, V. S. Sohal, I. Goshen, J. Finkelstein, J. T. Paz, K. Stehfest, R. Fudim, C. Ramakrishnan, J. R. Huguenard, P. Hegemann, and K. Deisseroth, "Neocortical excitation/inhibition balance in information processing and social dysfunction," *Nature*, vol. 477, no. 7363, pp. 171–178, 2011.
- [20] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognition*, vol. 43, no. 3, pp. 706–719, 2010.
- [21] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *SPIE*, vol. 5666, 2005, pp. 149–159.
- [22] R. Soundararajan and A. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517–526, 2012.
- [23] C. C. Chang and C. J. Lin. (2001) Libsvm: a library for support vector machines. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [24] R. Ben Yishai, R. L. Bar-Or, and H. Sompolinsky, "Theory of orientation tuning in visual cortex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 9, pp. 3844–3848, 1995.
- [25] D. Hansel and C. Vreeswijk, "The mechanism of orientation selectivity in primary visual cortex without a functional map," *The Journal of Neuroscience*, vol. 32, no. 12, pp. 4049–4064, 2012.
- [26] D. Ferster and K. D. Miller, "Neural mechanisms of orientation selectivity in the visual cortex," *Annual review of neuroscience*, vol. 23, pp. 441–471, 2000.
- [27] R. Shapley, M. Hawken, and D. Xing, "The dynamics of visual responses in the primary visual cortex," *Progress in brain research*, vol. 165, pp. 21–32, 2007.
- [28] B. Chapman, K. R. Zahs, and M. P. Stryker, "Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborize within a single orientation column in ferret visual cortex," *The Journal of neuroscience: the official journal of the Society for Neuroscience*, vol. 11, no. 5, pp. 1347–1358,

- 1991.
- [29] D. Ferster, "Spatially opponent excitation and inhibition in simple cells of the cat visual cortex," *The Journal of Neuroscience*, vol. 8, no. 4, pp. 1172–1180, 1988.
 - [30] T. W. Troyer, A. E. Krukowski, N. J. Priebe, and K. D. Miller, "Contrast-invariant orientation tuning in cat visual cortex: Thalamocortical input tuning and correlation-based intracortical connectivity," *The Journal of Neuroscience*, vol. 18, no. 15, pp. 5908–5927, 1998.
 - [31] F. W. Campbell and J. J. Kulikowski, "Orientational selectivity of the human visual system," *The Journal of Physiology*, vol. 187, no. 2, pp. 437–445, 1966.
 - [32] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen, "Outex - new framework for empirical evaluation of texture analysis algorithms," in *16th International Conference on Pattern Recognition*, vol. 1, 2002, pp. 701–706.
 - [33] Z. Guo, D. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.
 - [34] E. Rahtu, J. Heikkil, V. Ojansivu, and T. Ahonen, "Local phase quantization for blur-insensitive image analysis," *Image and Vision Computing*, vol. 30, no. 8, pp. 501–512, Aug. 2012.
 - [35] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 43–54, 2013.
 - [36] E. H. Adelson, E. P. Simoncelli, and W. T. Freeman, "Pyramids and multiscale representations," *Pyramids and multiscale representations*, pp. 3–16, 1991.
 - [37] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," in *IEEE Symposium on Visual Languages*, 1996, pp. 336–343.
 - [38] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1700–1705, 2013.
 - [39] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, 2003, pp. 1398–1402.
 - [40] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *Image Processing, IEEE Transactions on*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
 - [41] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it?" *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.
 - [42] H. R. Sheikh, K. Seshadrinathan, A. K. Moorthy, Z. Wang, A. C. Bovik, and L. K. Cormack. (2006) Image and video quality assessment research at live. [Online]. Available: <http://live.ece.utexas.edu/research/quality/>.
 - [43] E. C. Larson and D. M. Chandler. (2004) Categorical image quality (csiq) database. [Online]. Available: <http://vision.okstate.edu/csiq>.
 - [44] N. Ponomarenko and K. Egiazarian. (2009) Tampere image database 2008 tid2008. [Online]. Available: <http://www.ponomarenko.info/tid2008.htm>.