Orientation Selectivity based Visual Pattern for Reduced-Reference Image Quality Assessment

Jinjian Wu[a], Weisi Lin[b,*], Guangming Shi[a], Leida Li[c], Yuming Fang[d]

[a]*School of Electronic Engineering, Xidian University, Xi'an, 710071, China*
[b]*School of Computer Engineering, Nanyang Technological University, 639798, Singapore*
[c]*School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, 221116, China*
[d]*School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, 330032, China*

Abstract

Image quality assessment (IQA) is in great demand for high quality image selection in the big data era. The challenge of reduced-reference (RR) IQA is how to use limited data to effectively represent the visual content of an image in the context of IQA. Research on neuroscience indicates that the human visual system (HVS) exhibits obvious orientation selectivity (OS) mechanism for visual content extraction. Inspired by this, an OS based visual pattern (OSVP) is proposed to extract visual content for RR IQA in this paper. The OS arises from the arrangement of the excitatory and inhibitory interactions among connected cortical neurons in a local receptive field. According to to the OS mechanism, the similarity of preferred orientations between two nearby pixels is first analyzed. Then, the orientation similarities of pixels in a local neighborhood are arranged, and the OSVP is built for visual information representation. With the help of OSVP, the visual content of an image is extracted and mapped into a histogram. By calculating the changes between the two histograms of reference and distorted images, a quality score is produced. Experimental results on five public databases demonstrate that the proposed RR IQA method has performance consistent with the human perception under a small amount of reference data (only 9 values).

Keywords: Orientation Selectivity Mechanism, Image Quality Assessment, Reduced-Reference, Excitatory and Inhibitory Interactions, Preferred Orientation, Visual Pattern

1. Introduction

In the digital era, multimedia signals are tremendously increased, and the visual media becomes increasingly important in the daily life [9]. Multimedia content processing during acquisition, compression, and transmission will cause distortions into the original signals [21, 26]. Extracting high-quality information efficiently from massive digital data is still an open problem. Generally, subjective assessment is the most straightforward and reliable way for high quality signal selection. However, it is time-consuming and cumbersome for intelligent systems [19, 34]. Therefore, objective quality assessment algorithms, which performs automatically and consistently with human perception, are in great demand in various signal processing applications.

During the past decades, a large number of image quality assessment (IQA) algorithms have been introduced. According to the amount of the reference information, the existing IQA methods can be divided into three categories [21]: 1) Full-Reference (FR) IQA, which requires the full reference image (the original image without distortion); 2) Reduced-Reference (RR) IQA, which requires only part of the reference image; and 3) No-Reference (NR) IQA, where the reference image is not required. With the full access to the reference image, FR IQA is easy to implement with high evaluation accuracy [12, 18, 42]. However, the full reference image is not always available in real applications, so an NR IQA method is required. Without the guidance of the reference image, it is quite difficult for NR IQA method to accurately evaluate image quality [11, 22, 24, 41]. As a compromise between the FR IQA and NR IQA methods, the RR IQA method aims to predict image quality accurately with a limited amount of reference information [35]. In this paper, we mainly focus on building an effective RR IQA method.

RR IQA method generally attempt to extract global features, which can effectively represent the distortion. For instance, in [36] and [40], a quality-aware image was extracted from the original

☆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), the Research Fund for the Doctoral Program of Higher Education (No. 20130203130001), the International cooperation project of Shaanxi science and technology R&D program (No. 2014KW01-02), and the Ministry of Education (MoE) AcRF Tire 1 Funding, Singapore (Grant 2014-T1-002-068).

*Corresponding author.

Email addresses: `jinjian.wu@mail.xidian.edu.cn` (Jinjian Wu), `wslin@ntu.edu.sg` (Weisi Lin), `gmshi@xidian.edu.cn` (Guangming Shi), `lileida@cumt.edu.cn` (Leida Li), `fa0001ng@e.ntu.edu.sg` (Yuming Fang)

image according to the statistical distribution of the image content to evaluate the quality. Based on the assumption that natural scenes often follow some stable statistical properties, the statistical distribution of the wavelet coefficients in an image was calculated in [37] to build a quality-aware image. Then, the quality was estimated as the Kullback-Leibler (KL) divergence between the reference and distorted quality-aware images [37]. In order to improve the performance, a divisive normalization transform procedure is adopted to normalize the statistical coefficients for quality prediction [20]. Moreover, the statistical distributions of wavelets, curvelets, and contourlets coefficients were throughly analyzed, and a multiscale geometric analysis based RR IQA method was introduced in [8]. However, these statistics-based RR IQA methods either fail to perform well across different distortion types or require a set of training parameters on databases.

Recently, some non-statistical RR IQA methods have been introduced [23, 31, 39]. Inspired by the brain cognition theory, Wu et al. [39] separately measured the information fidelities on the primary visual content and the disorderly uncertainty for quality evaluation. This algorithm performs well for individual distortion type but it cannot be generalized to behave well for datasets with many distortion types. In [31], the quality degradation was evaluated with the scaled entropies of wavelet coefficients on each divided block, and a large amount of reference information was required to achieve good performance. Therefore, these existing RR IQA methods either perform not good enough with limited reference data or require a large amount of reference data to achieve promising performance.

Since the objective IQA aims to achieve the performance at the level of human perception, we explore the visual information processing in the human visual system (HVS) and attempt to develop a perception-oriented RR IQA method in this study. It is well known that the HVS is an immensely complex organ, and how it processes the input visual information is still not well understood. With the effect of the researchers in neuroscience during the past decades, scientists claim that the HVS presents substantial orientation selectivity (OS) mechanism [1, 14, 15], based on which visual patterns of an input scene are extracted for perception and understanding. Inspired by the related research findings in neuroscience, we imitate the OS mechanism and extract visual patterns for IQA.

Orientation selectivity arises from the spatial arrangement of intracortical responses in a local receptive field of the primary visual cortex (PVC) [5, 32]. Moreover, there are two spatial opposite kinds of intracortical responses: excitatory and inhibitory interactions [4]. Cortical neurons with similar preferred orientations are more likely to present excitatory interactions, and vice versa [7, 10]. When a scene is perceived, input signals excite individual arrangements of excitatory/inhibitory interactions for different local receptive fields, which generate different kinds of OS based visual patterns (OSVP) for image understanding.

Inspired by the OS mechanism, an OSVP based RR IQA algorithm is introduced in this paper. Firstly, the gradient direction is computed as the preferred orientation of each pixel. Next, the spatial relationship between a central pixel and its circularly symmetric neighbors is analyzed. By mimicking the excitatory and inhibitory interactions in a local receptive field, pixels with similar gradient directions are regarded as excitatory interactions and vice versa. According to the arrangement of the excitatory and inhibitory interactions in a local neighborhood, the OSVP is built. Finally, the visual content of an image is represented by an OSVP histogram. The quality degradation is estimated by the histogram difference between the reference (the original image) and distorted (the test image which is contaminated by a certain type of distortion) images. Experimental results demonstrate that the OSVP based histogram can efficiently represent the visual content of an image, and the proposed RR IQA algorithm performs highly consistent with the subjective perception.

The rest of this paper is organized as follows. The mechanism of OS is firstly illustrated, and the OSVP feature is introduced in Section 2. Then, in Section 3, visual information is extracted with the OSVP feature, and a novel RR IQA method is proposed. Section 4 shows the experimental results. Finally, conclusions are given in Section 5.

2. Orientation selectivity based visual pattern

In this section, a brief introduction to the essential principle of the OS mechanism in the PVC is illustrated firstly. Then, by mimicking the excitatory/inhibitory interactions among cortical cells in a local receptive field, the OSVP feature is introduced for visual feature extraction.

2.1. *Mechanism of orientation selectivity*

It is well known that distortions degrade the visual structure of an image, which results in quality degradation [34, 38]. However, how to effectively represent visual structure is still an open problem. As a complex system, the HVS can effectively and efficiently extract visual structures for scene perception and understanding. Thus, exploring and imitating the underlying principle of the HVS on visual structure extraction gives a lot of inspirations on designing perception-oriented quality assessment systems. Research findings on brain cognition indicate that when perceiving an input scene, the HVS presents substantial OS mechanism [14]. Studies show that the HVS is extremely sensitive to the edge regions and highly adapted to extract their orientations for perception [2]. Therefore, the OS mechanism relates to the visual structure extraction [32].

During past decades, the origin of OS in the PVC has been thoroughly investigated, and OS becomes one of the standard models on visual signal representation in the PVC [7, 10, 28]. Many hypotheses and theory about the roots of OS are introduced [1–3, 14], among which the feedforward model is the most classical one. In the feedforward model, Hubel and Wiesel suggested that the OS lies on the intracortical responses to lateral geniculate nucleus (LGN) afferents [14]. As a thalathic station of visual projection pathway from the eye to the brain, the LGN transfers the stimuli from retina to the PVC [16]. Within each local receptive field, cortex cells present different responses according to the LGN afferents: cortex cells are more likely to present excitatory interactions when they receive stimuli similar with their preferred orientations, and present inhibitory interactions when they receive stimuli different from their preferred orientations. The orientation selectivity arises from the arrangement of the excitatory/inhibitory interactions in a small local receptive field [32]. A brief diagram for the origin of OS in the PVC is shown in Fig. 1.

2.2. *OSVP feature extraction*

The OS mechanism reveals the underlying principle of visual feature extraction in the HVS. Therefore, by imitating this mechanism, a novel OSVP feature is introduced for visual information representation. the related research on neuroscience indicates that the origin of OS is directly related to the arrangement of the interactions among cortical neurons in a local receptive field [7]. Inspired by this mechanism, the OSVP can be computed according to the arrangement of the cor-

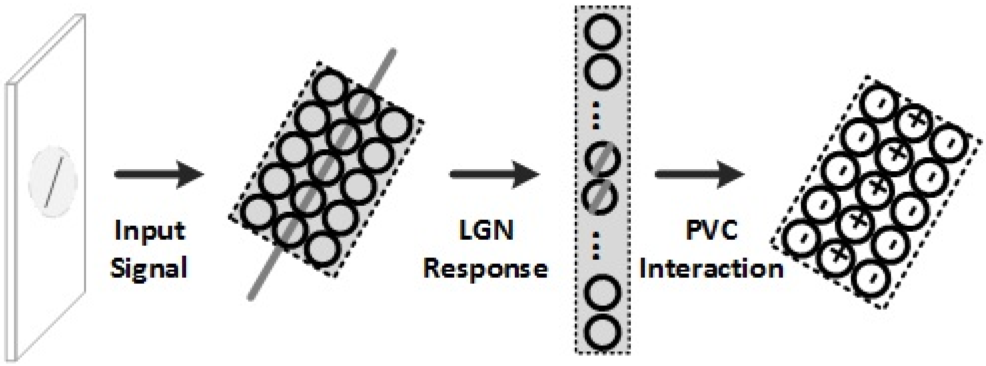


Figure 1: A brief diagram of the origin of OS.

relations among neighbor pixels for image analysis. For an image $\mathcal{F}$, the OSVP of the pixel $x \in \mathcal{F}$ is described as the arrangement of the spatial correlations with its circularly symmetric neighborhood $\mathcal{X}=\{x_1, x_2, \cdots, x_n\}$,

$$\mathcal{P}_v(x|\mathcal{X}) = \mathcal{A}(\mathcal{I}(x|\mathcal{X})) = \mathcal{A}(\mathcal{I}(x|x_1, x_2, \cdots, x_n)), \tag{1}$$

where $\mathcal{A}(\cdot)$ is the arrangement of spatial correlations, and $\mathcal{I}(x|\mathcal{X})$ is the spatial correlations between x and x_i in $\mathcal{X}$ (i.e., $\{x_i \mid x_i \in \mathcal{X}\}$).

Since each cortical neuron connects to a mass of cortical neurons through synapses and presents intricate relationships with other neurons, the correlations among neurons in a local receptive field are extremely complex. In order to effectively analyze the OS mechanism, Hubel and Wiesel only explored the dominant synapses between the central cell and its excited cells [14]. Inspired by this model, the correlation among pixels in the circularly symmetric neighborhood $\mathcal{X}$ is neglected, and only the correlations between the central pixel x and these neighbors x_i are considered. Thus, Eq. (1) can be reorganized as

$$\mathcal{P}_v(x|\mathcal{X}) \approx \mathcal{A}(\mathcal{I}(x|x_1), \mathcal{I}(x|x_2), \cdots, \mathcal{I}(x|x_n)), \tag{2}$$

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

The response of the HVS on visual pattern extraction depends on two types of interactions, namely, excitation and inhibition [6]. These two types of interaction play distinct roles in shaping OS: excitatory neurons connect to neurons that are well correlated in activity; and inhibitory neurons connect to neurons that are anti-correlated [4]. Moreover, the interaction type in the OS

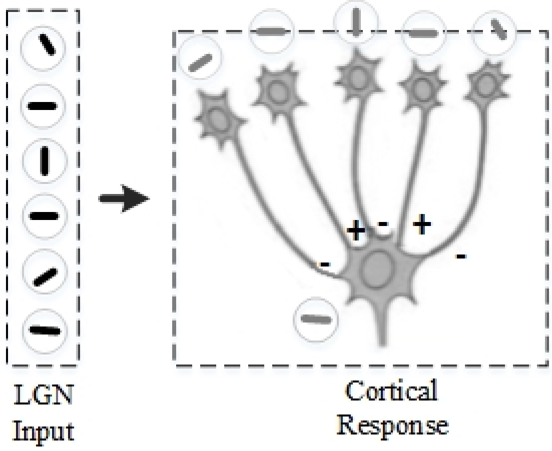


Figure 2: The excitatory ('+') and inhibitory ('-') interactions between neurons

mechanism is determined by the orientation similarity.

$$\mathcal{I}(x|x_i) = f(\theta(x), \theta(x_i)), \tag{3}$$

where $\theta(x)$ and $\theta(x_i)$ are the orientations of pixels x and x_i, respectively.

The correlation-based rule of synaptic plasticity indicates that cortical neurons with similar preferred orientations are more likely to respond as excitation and have a higher probability of connection, and vice versa [10]. A schematic representation of excitatory and inhibitory interactions between neurons is shown in Fig. 2. The response between the central neuron and the third neuron is excitatory ('+'), since the two neurons have similar preferred orientations; while the responses between the central neuron and the first neuron is inhibitory ('-'), as they have different preferred orientations.

Since the interaction type between cortical neurons depends on their preferred orientations, $\mathcal{I}(x|x_i)$ is computed according to the orientation similarity between x and x_i. Firstly, the gradient direction θ of a pixel $x \in \mathcal{F}$ is regarded as its orientation,

$$\theta(x) = \arctan \frac{G_v(x)}{G_h(x)}, \tag{4}$$

where G_h and G_v are the gradient magnitudes along the horizontal and vertical directions, respec-

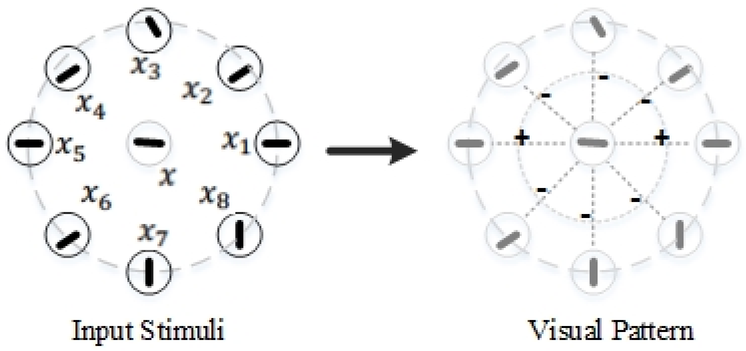


Figure 3: The 8-neighbor OSVP.

tively. In this work, G_h and G_v are calculated as,

$$G_h = \mathcal{F} * f_h, \quad G_v = \mathcal{F} * f_v, \tag{5}$$

$$f_h = \frac{1}{3}\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}, \; f_v = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \tag{6}$$

where f_h is the Prewitt filter along horizontal, f_v is the Prewitt filter along vertical, and $*$ denotes the convolutional operation.

According to the orientation similarity between two pixels, Eq. (3) is rewritten as

$$\mathcal{I}(x|x_i) = \begin{cases} + & \text{if } |\theta(x) - \theta(x_i)| < \mathcal{T} \\ - & \text{otherwise} \end{cases}, \tag{7}$$

where '+' represents excitatory interaction; '-' represents inhibitory interaction; and $\mathcal{T}$ is the similarity threshold, below which the two compared pixels are regarded to have similar orientations. The similarity threshold has been investigated with subjective viewing test on visual masking effect [3]. Subjective visual masking experiment demonstrates that the masking effects among nearby gratings are strong if they possess the same orientation. With the increase of orientation difference, the masking effect weakens. When the orientation difference is larger than the threshold (e.g., 12°), the masking effect becomes marginal. Therefore, considering the plus-minus of orientation difference, $\mathcal{T}$ is set as $\mathcal{T}$=6° in this paper.

With the orientation similarity based interaction type (Eq. (7)), the OSVP form of a pixel is represented by the arrangement of '+' and '-' within its circularly symmetric neighborhood. An intuitional example of 8-neighbor (i.e., n=8) OSVP is given in Fig. 3. According to the orientation similarities between x and its 8-neighbor x_i, their interaction types are computed and arranged for the OSVP of x, namely, $\mathcal{P}_v(x|\mathcal{X}) = \{+ - - - + - --\}$.

3. Quality degradation measurement

In this section, visual information of each pixel is firstly extracted by the OSVP feature, with which an image is mapped into a histogram. Next, by measuring the histogram difference between the reference and distorted OSVPs, a novel IQA algorithm is proposed.

3.1. Visual information extraction with OSVP

The RR IQA methods aim to obtain accurate visual quality prediction with a limited amount of reference data. Here we use OSVP, which can effectively represent the spatial correlation among pixels, to extract image feature for quality measurement.

With Eq. (2), the OSVP of each pixel is computed. Pixels possessing a same OSVP display the same visual content, while those with a different OSVP represent different visual information. Thus, pixels possessing the same OSVP are combined, and thus an input scene is mapped into an OSVP-based histogram,

$$\mathcal{B}(k) = \sum_{i=1}^{N} w(x_i)\ \delta(\mathcal{P}_v(x_i), \mathcal{P}_v^k) \tag{8}$$

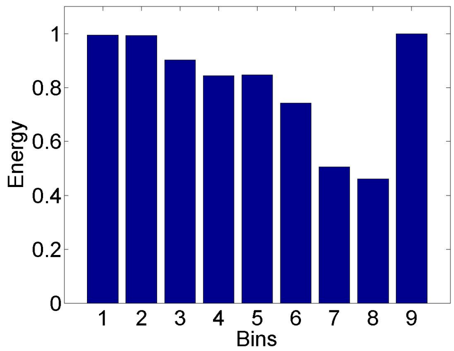
$$\delta(\mathcal{P}_v(x_i), \mathcal{P}_v^k) = \begin{cases} 1 & \text{if } \mathcal{P}_v(x_i) = \mathcal{P}_v^k \\ 0 & \text{else} \end{cases}, \tag{9}$$

where $\mathcal{B}(k)$ presents the histogram value for the $k-$bin, N presents the pixel number of the image, and $\mathcal{P}_v^k$ presents the index of the $k-$OSVP. $w(x_i)$ is the weighting factor. In general, pixels having large luminance changes are more attractive to human eyes. Thus, the weighting factor $w(x_i)$ for pixel x_i is directly related to its luminance changes, and in this work, $w(x_i)$ is set as

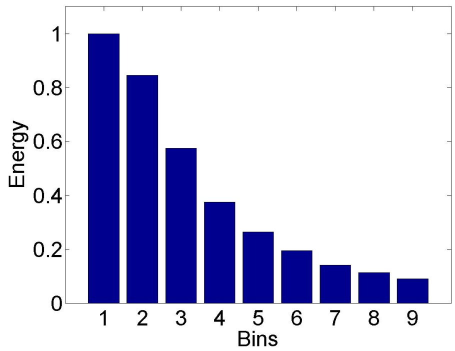
$$w(x_i) = \mathrm{var}(x_i), \tag{10}$$

(a) Hats

(b) Mountain



(c) OSVP for Hats



(d) OSVP for Mountain

Figure 4: Example for the OSVP feature extraction on images with different visual contents.

where $\text{var}(x_i)$ is the local variance of x_i.

According to Eq. (8), an input image can be represented by an OSVP-based histogram. Meanwhile, the size of the histogram (the number of the histogram bins) is determined by the size of the neighborhood $\mathcal{X}$, and the bin number increases exponentially with $\mathcal{X}$. As a result, even though with a small local neighborhood, the bin number is quite large (e.g., with an 8-neighborhood, there are 256 different kinds of OSVP). Generally, the RR IQA method is designed to use a limited amount of data to effectively represent the visual content of an image. Therefore, some OSVP types are combined to reduce the bin number.

Further analysis reveals those OSVP types having the same number of excitations, are more likely to present similar information. For example, both $\mathcal{P}_v = \{+ - - - - - - -\}$ and $\mathcal{P}_v = \{- - - - + - - -\}$ have only one excitatory interaction, and they are more probably to appear in disorderly regions (e.g., a forest region with massive uncertain structures); while $\mathcal{P}_v = \{- + + + + + + -\}$

and $\mathcal{P}_v = \{+ + + - + - ++\}$ have six excitatory interactions, and they are more likely to appear in orderly regions (e.g., a sky region with uniform structures). Therefore, these OSVP types with the same number of excitations are combined. As a result, all patterns from n-neighborhood can be classified into n+1 types, and the content of an image is represented by a histogram with (n+1) OSVP bins.

Fig. 4 shows an example of OSVP-based visual content extraction. The *hats* image has a large area of smooth structure (e.g., the sky and the wall), while the *mountain* image is mainly composed of disorderly structures (e.g., the woods and the stream). By comparing these two reference images, it can be seen that images with different visual contents present different OSVP-based histograms, as shown in Fig. 4 (c) and (d). Therefore, the proposed OSVP feature is useful for visual information extraction.

3.2. Image quality assessment

All kinds of distortion will degrade image structures to some degree, which results in quality drop. Due to the damage on spatial correlations among pixels, their corresponding OSVP types also change. For example, Gaussian white noise adds disturbances into images, which may change a uniform pattern (e.g., $\mathcal{P}_v = \{++++++++\}$) into a disorderly pattern (e.g., $\mathcal{P}_v = \{--------\}$); on the contrary, Gaussian Blur noise smoothens edge regions, which may change a disorderly pattern into a uniform pattern. Thus, quality degradation can be calculated based on the pattern change, and the quality of a distorted image is measured as the histogram similarity between the reference ($\mathcal{F}_r$) and distorted ($\mathcal{F}_d$) images,

$$Q(\mathcal{F}_d|\mathcal{F}_r) = \sum_{k=1}^{n+1} \frac{2\mathcal{B}_d(k)\,\mathcal{B}_r(k) + C}{\mathcal{B}_d^2(k) + \mathcal{B}_r^2(k) + C}, \tag{11}$$

where $\mathcal{B}_d$ ($\mathcal{B}_r$) is the OSVP-based histogram of $\mathcal{F}_d$ ($\mathcal{F}_r$), and C is a small constant to avoid the denominator being zero and is set as C=0.0001 in this work.

4. Experimental results

In this section, the effectiveness of the OSVP feature for distorted visual information extraction is firstly analyzed. Then, the performance of the proposed OSVP-based RR IQA method is illustrated. Finally, a comprehensive comparison between the proposed IQA method and the other IQA methods on publicly available databases is demonstrated.

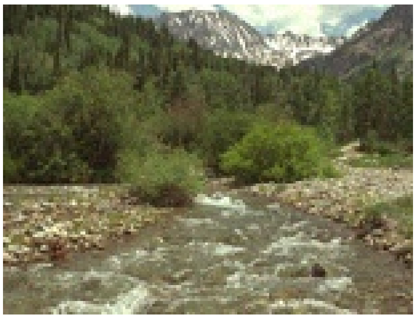
4.1. Effectiveness Analysis

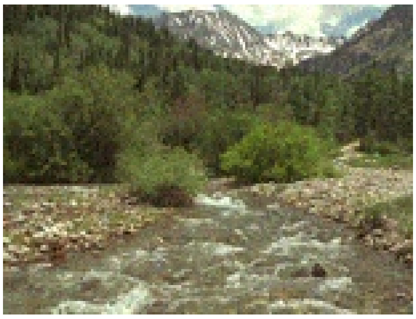
The proposed OSVP feature is sensitive to distortions in images and can effectively represent the visual information degradations. Fig. 5 shows the OSVP feature extraction for different images (the *hats* and the *mountain* images) under the same distortion type (Gaussian white noise, GWN for short). The corresponding reference images are shown in Fig. 4.

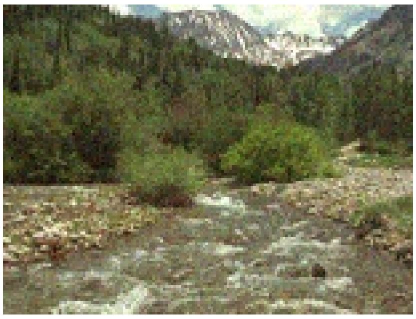
Though these two images are under the same distortion type, they present significantly different changes on OSVP histograms. In Fig. 5 (a)-(c), the energies of the GWN in *hats* images are 33.6dB, 30.5dB, and 27.5dB, respectively. Since the *hats* image contains a large flat area (e.g., the sky and the wall), it can conceal limited GWN. As can be seen, the noise in the three contaminated *hats* images is highly abrupt, especially in Fig. 5 (c). With the content degradations caused by the GWN on the set of *hats* images, their corresponding OSVP histograms also change. In Fig. 5 (i), since the GWN increases the disturbance in the *hats* image, the energies of most bins are obviously increased. However, the GWN has quite different impact on the *mountain* image. In Fig. 5 (d)-(f), the energies of the GWN on the set of *mountain* images are same to these on the set of *hats* images. Different from the *hats* image, the *mountain* image contains a large disorderly area (e.g., the woods and the stream), in which a mass of noise can be hidden. As a result, the GWN is less obvious in the *mountain* image than that in the *hats* image, as shown in Fig. 5 (d)-(f). Moreover, the impact of GWN on the OSVP histograms of the *mountain* images is limited , as shown in Fig. 5 (j). In summary, though distorted by the same noise, images with different visual contents present different degradations, and the proposed OSVP feature can effectively represent the degradations for different visual contents.

Moreover, different distortion types will result in different quality drops, and the proposed OSVP feature can also effectively capture the degradations from different distortion types. As shown in Fig. 6, the *hats* image is contaminated by the GWN, the JPEG compression noise (JPG), and the Gaussian blur noise (GBN), respectively. Though the noise levels (PSNR=24.5dB) on the three contaminated images are same, their quality degradations are obviously different: the GWN

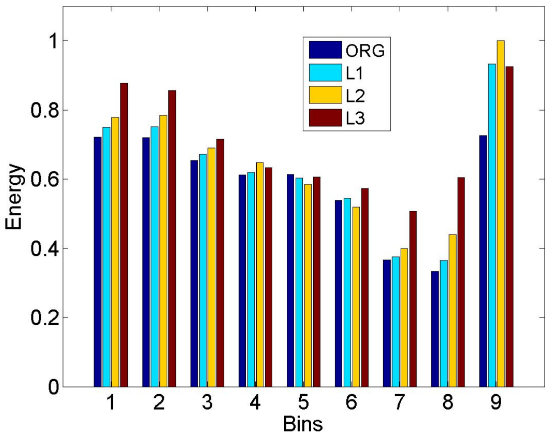
(a) PSNR=33.6dB (b) PSNR=30.5dB (c) PSNR=27.5dB



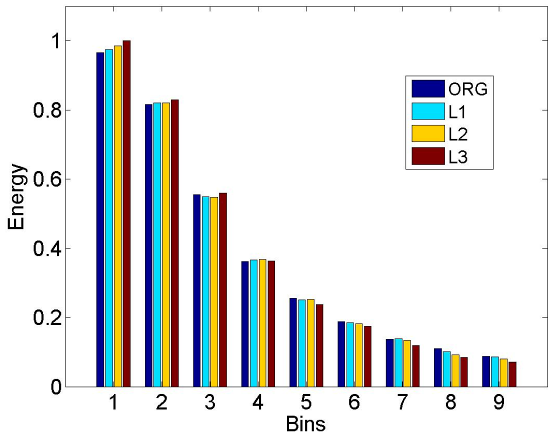




(d) PSNR=33.6dB (e) PSNR=30.5dB (f) PSNR=37.5dB



(g) OSVP histograms for hats images

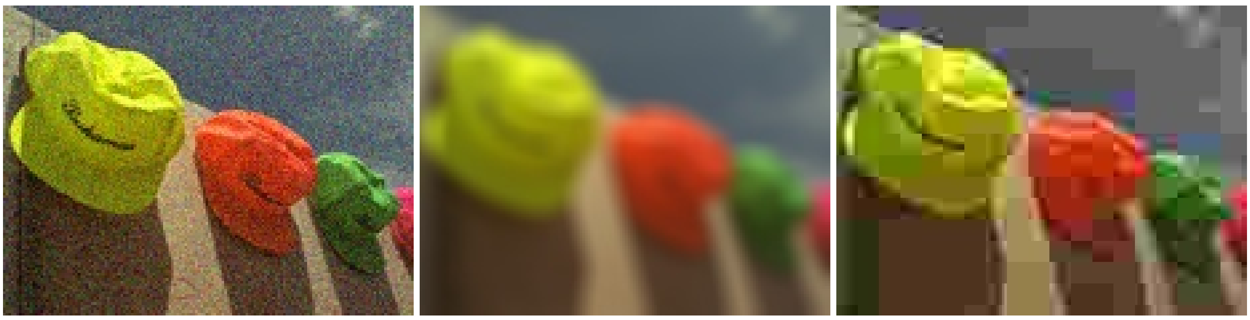


(h) OSVP histograms for mountain images

Figure 5: Examples for the OSVP feature extraction on two different images (hats and mountain) contaminated by the same type of distortion (Gaussian White noise, under three different noise levels, L_1, L_2, and L_3).

mainly increases the random disturbance for the whole image (in Fig. 6 (a)); the GBN smoothens the high frequency contents (e.g., the edge in Fig. 6 (b)); and the JPG not only damages the original edge regions, but also creates some new edges on the smooth regions due to the blockiness artifacts (in Fig. 6 (c)).

With visual information degradations from three types of distortions, the corresponding OSVP histograms also change. As the light blue bars shown in Fig. 7, due to the random disturbance, the GWN increases the energies of most bins. That is because when the GWN is large enough (with a large PSNR value), a flat region is probably distorted into a texture region; as a result, pixels with the 9−OSVP ('++++++++', which always appears in the flat region) are changed into different OSVP types. The GBN decreases most bins' energies except for the last one, as shown by the yellow bars in Fig. 7. Because the GBN turns some unsmooth regions into flat ones, the number of the 9−OSVP greatly increases and its corresponding energy also increases. The change of OSVP histogram caused by JPG is quite different (the wine red bars in Fig. 7). Under the blockiness, the changes on most bins are limited except for the last bin, because the blockiness not only damages some real edges but also creates some fake edges on the smooth regions. As a result, the energy values for most bins remained. Meanwhile, the blockiness averages out small disturbance; as a result, the energy for these pixels which belong to the 9−th OSVP type are decreased. Thus, we can conclude that the proposed OSVP can effectively capture structure degradations from different type of distortions.



(a) GWN (b) GBN (c) JPG

Figure 6: The *hats* image distorted by three types (GWN, GBN, and JPG) of noise. The noise levels of them are the same (PSNR=24.5dB)

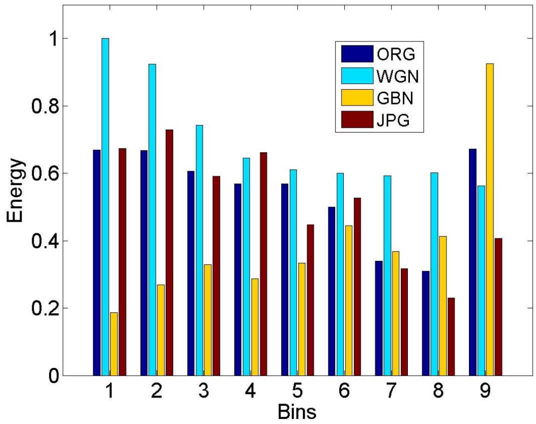


Figure 7: The changes of the OSVP histograms for the three *hats* images as shown in Fig. 6.

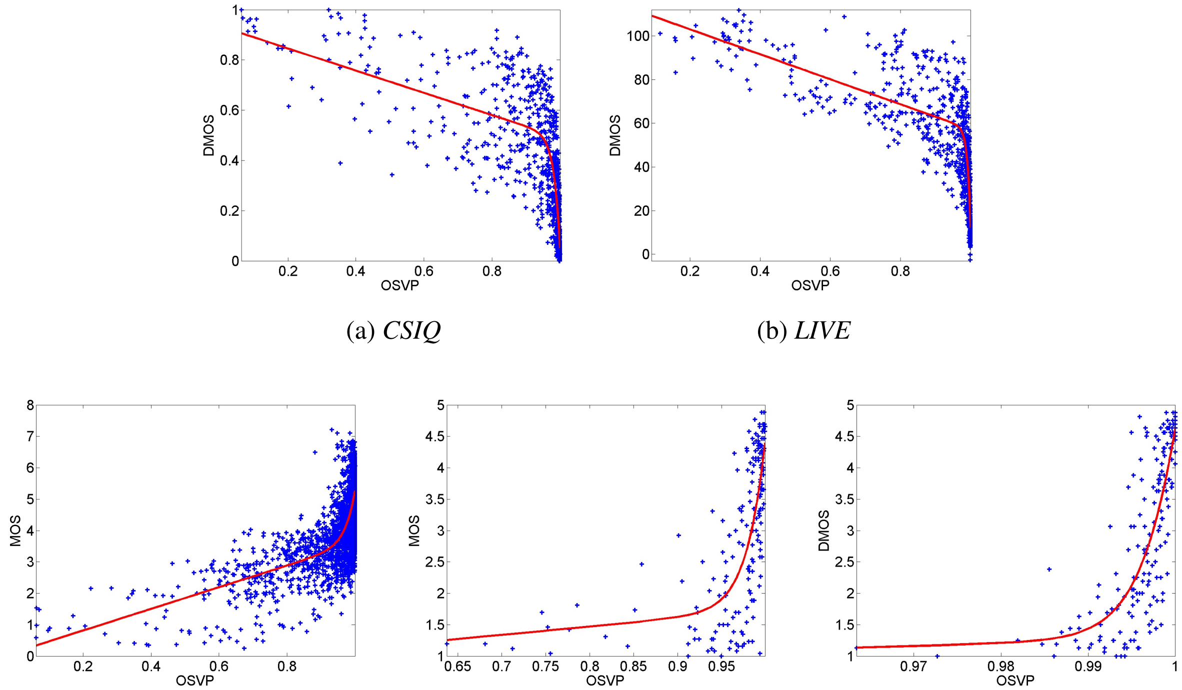
4.2. IQA Performance Analysis

In order to verify the effectiveness of the proposed OSVP feature for IQA, five publicly available IQA databases, *CSIQ* [17], *LIVE* [29], *TID2013* [27], *IVC* [25], and *TOYAMA* [13], are used in the comparison experiment. For the five databases, *LIVE* contains 29 reference images and 779 distorted images (with 5 distortion types); *CSIQ* contains 30 reference images and 866 distorted images (with 6 distortion types); *TID2013* contains 25 reference images and 3000 distorted images (with 24 distortion types); *IVC* contains 10 reference images and 185 distorted images (with 4 distortion types); and *TOYAMA* contains 14 reference images and 168 distorted images (with 2 distortion types).

Meanwhile, three widely used performance criteria, namely, the Pearson linear correlation coefficient (PLCC), the Spearman rank-order correlation coefficient (SRCC), and the root mean squared error (RMSE), are employed to measure the correlation between the predicted quality score and the ground truth quality (i.e., the mean opinion score (MOS) or the difference MOS (DMOS) provided by these databases). A better IQA method will obtain higher PLCC and SRCC values and a lower RMSE value. Before measuring the correlation, a five parameter mapping function [33] is generally adopted to nonlinearly regress the predicted quality score,

$$Q' = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2 (Q - \beta_3))} \right) + \beta_4\, Q + \beta_5, \tag{12}$$

where Q' is the regressed quality score, and $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ are the parameters to be fitted.



(a) *CSIQ* (b) *LIVE*

(c) *TID2013* (d) *IVC* (e) *TOYAMA*

Figure 8: Logistic fittings of the OSVP outputs against the ground truth qualities (i.e., MOS or DMOS) on the five public IQA databases.

Logistic regression is used to fit lines (Fig. 8) to the predicted quality scores and ground truth values of the five databases (i.e., *CSIQ*, *LIVE*, *TID2013*, *IVC*, and *TOYAMA*). As can be seen, the predicted quality scores (the blue crosses) are always converged around the fitted lines (the red lines), which confirms that the performance of the proposed OSVP-based IQA method is consistent with the human perception.

In order to further demonstrate the performance of the proposed OSVP based RR IQA method, three recently-developed RR IQA methods (i.e., RRED [31], RRVIF [39], and WNISM [36]) and three classic FR IQA metrics (i.e., PSNR, SSIM [34], and MS-SSIM [35]) are adopted for comparison. The performances of these IQA methods on *LIVE* database are listed in Tab. 1. There are five different types of distortions in *LIVE* database, namely, GWN, GBN, JPG, J2K, and Fastfading noise (FF). For a RR IQA method, it is desired to use less reference data for quality assessment. Hence, the reference data used by each IQA method is firstly given in the second row

Table 1: IQA Performance Comparison on *LIVE* Database

Dist.	Algo. / Crit.	RR				FR		
		OSVP	RRED	RRVIF	WNISM	PSNR	SSIM	MS-SSIM
No. of scalars		9	9	2	18	*N*	*N*	*N*
GWN	PLCC	*0.977*	0.937	0.958	0.892	0.988	0.969	0.975
	SRCC	*0.975*	0.935	0.972	0.872	0.985	0.963	0.973
	RMSE	*5.96*	9.77	8.03	12.67	4.33	6.90	6.21
GBN	PLCC	*0.941*	0.946	0.741	0.896	0.784	0.874	0.959
	SRCC	*0.928*	0.940	0.753	0.920	0.782	0.894	0.959
	RMSE	*6.26*	6.00	12.40	8.21	11.47	8.96	5.21
JPG	PLCC	*0.962*	0.953	0.886	0.921	0.890	0.951	0.981
	SRCC	*0.964*	0.949	0.877	0.919	0.881	0.944	0.980
	RMSE	*8.67*	9.67	14.79	12.39	14.55	9.89	6.13
J2K	PLCC	*0.892*	0.949	0.960	0.945	0.900	0.941	0.970
	SRCC	*0.896*	0.941	0.955	0.939	0.895	0.936	0.965
	RMSE	*11.42*	7.98	7.03	8.27	11.01	8.51	6.15
FF	PLCC	*0.940*	0.891	0.901	0.929	0.890	0.945	0.929
	SRCC	*0.940*	0.915	0.901	0.939	0.891	0.941	0.932
	RMSE	*9.74*	12.91	12.33	10.53	13.01	9.35	10.57
Overall	PLCC	*0.862*	0.815	0.722	0.743	0.872	0.904	0.941
	SRCC	*0.867*	0.818	0.738	0.755	0.876	0.910	0.945
	RMSE	*13.84*	15.83	18.90	18.28	13.36	11.67	9.27

of Tab. 1. As can be seen, a very limited amount of reference data is used for all of the four RR IQA methods (9, 9, 2, and 18 values for OSVP, RRED, RRVIF, and WNISM, respectively); while the whole reference data (the size (N) of the image) is required for the three FR IQA metrics.

The proposed OSVP method performs effectively on *LIVE* database. As shown in Tab. 1, OSVP achieves larger PLCC and SRCC values than the other RR IQA methods on GWN, JPG, and FF. OSVP also yields the smallest RMSE values on the three types of distortion among the four RR IQA methods. In other words, OSVP performs the best on three of the five distortion types among these RR IQA methods. Meanwhile, OSVP performs as well as the best one on GBN, and slightly worse than the other methods on J2K. Furthermore, as the overall performances shown in the last row of Tab. 1, OSVP performs the best (has the largest PLCC and SRCC values, and smallest RMSE value) through the whole *LIVE* database among the four RR IQA methods. Therefore, both the experimental results on the individual distortion type and the whole database verify that the proposed OSVP outperforms other existing RR IQA methods on *LIVE* database.

When comparing with the three classic FR IQA metrics on *LIVE* database, OSVP also shows some advantages. According to the experimental results between OSVP and PSNR, we can see that OSVP performs better on three of the five distortion types (i.e., GBN, JPG, and FF), and the overall performance of OSVP is comparable to PSNR. When compared with SSIM, OSVP also performs better on three of the five distortion types (i.e., GWN, GBN, and JPG), OSVP performs a little worse than MS-SSIM on both individual distortion type and the whole database. In summary, though OSVP uses limited reference data (9 values), it is comparable with PSNR and SSIM, and a little worse than MS-SSIM.

4.3. Comprehensive Performance Comparison on IQA Databases

In order to present a comprehensive analysis on these IQA methods, the performances (i.e., PLCC, SRCC, and RMSE) of these models on the other four IQA databases are listed in Tab. 2. As can be seen, OSVP performs quite well on the four databases. On *CSIQ* database, OSVP has much larger PLCC and SRCC values and smaller RMSE value than the other three RR IQA methods, which means OSVP performs significantly better than the existing RR IQA methods on this database. Moreover, OSVP outperforms PSNR and SSIM on this database, while worse than

Table 2: IQA Performance Comparison on Four Databases

Dist.	Algo. / Crit.	RR				FR		
		OSVP	RRED	RRVIF	WNISM	PSNR	SSIM	MS-SSIM
CSIQ	PLCC	*0.843*	0.677	0.598	0.735	0.800	0.815	0.900
	SRCC	*0.849*	0.758	0.633	0.771	0.800	0.838	0.914
	RMSE	*0.141*	0.193	0.210	0.178	0.158	0.152	0.115
TID2013	PLCC	*0.724*	0.729	0.577	0.629	0.702	0.686	0.831
	SRCC	*0.654*	0.671	0.451	0.523	0.703	0.627	0.785
	RMSE	*0.856*	0.849	1.013	0.964	0.883	0.902	0.689
IVC	PLCC	*0.757*	0.633	0.489	0.534	0.720	0.792	0.893
	SRCC	*0.747*	0.590	0.453	0.446	0.688	0.779	0.884
	RMSE	*0.795*	0.943	1.063	1.030	0.846	0.743	0.548
TOYAMA	PLCC	*0.827*	0.727	0.783	0.779	0.731	0.847	0.916
	SRCC	*0.825*	0.726	0.779	0.774	0.722	0.840	0.905
	RMSE	*0.743*	0.906	0.821	0.828	0.901	0.701	0.531

MS-SSIM. On *TID2013* database, OSVP performs as well as RRED, and much better than the other two RR IQA methods. Meanwhile, it also outperforms PSNR and SSIM on this database, while worse than MS-SSIM. The performances on *IVC* database is much similar as that on *CSIQ* database, where OSVP performs the best among these RR IQA methods, also better than PSNR, slightly worse than SSIM, while worse than MS-SSIM. On *TOYAMA* database, OSVP outperforms the other RR IQA methods, also outperforms PSNR, slightly worse than SSIM, and worse than MS-SSIM.

According to the performance comparison on these public IQA databases (as listed in Tab. 1 and Tab. 2), OSVP performs the best among RR IQA methods on four of the five databases (i.e., *LIVE*, *CSIQ*, *IVC*, and *TOYAMA*), and it performs as well as the best RR IQA method on the left database (i.e., *TID2013*). Meanwhile, by comparing with these FR IQA metrics, OSVP performs better than PSNR on four of the five databases (i.e., *CSIQ*, *TID2013*, *IVC*, and *TOYAMA*), and

Table 3: The average IQA Performance Comparison

Crit. \ Algo.	RR				FR		
	OSVP	RRED	RRVIF	WNISM	PSNR	SSIM	MS-SSIM
PLCC	*0.803*	0.716	0.634	0.684	0.765	0.809	0.896
SRCC	*0.788*	0.712	0.611	0.654	0.760	0.799	0.887

Table 4: Performance Comparison with F-test (Statistical Significance with 95% confidence)

DB \ Algo.	RR			FR		
	RRED	RRVIF	WNISM	PSNR	SSIM	MS-SSIM
LIVE	1	1	1	0	-1	-1
CSIQ	1	1	1	1	1	-1
TID2013	0	1	1	0	1	-1
IVC	1	1	1	0	0	-1
TOYAMA	1	0	0	1	0	-1

also better than SSIM on two of the five databases (i.e., *CSIQ* and *TID2013*). Moreover, the average PLCC and SRCC values (the average RMSE is not computed since the ranges of RMSE values on these databases are different) on the five databases are listed in Tab. 3. As can be seen, OSVP has much larger PLCC and SRCC values than RRED, RRVIF, WNISM, and PSNR; besides, it has quite similar values with SSIM, and has smaller values than MS-SSIM. In summary, the performance comparisons on the five databases confirm that OSVP is highly effective for IQA.

Besides these correlation criteria, the statistical significance is computed to further demonstrate the effectiveness of the proposed OSVP method. In this experiment, the F-test [30] criterion is adopted to evaluate the statistical significance of OSVP against the other IQA methods. For F-test, the residual between the ground truth and the predicted quality score (after nonlinear mapping) is firstly calculated. Then, the residual variance is computed, and the ratio F between two residual variances from two IQA methods on one database is acquired. According to the relationship between F and its corresponding judgement threshold $F_{critical}$ (determined by the number of samples and the confidence level), the statistical significance is finally required: if $F > F_{critical}$, the method on the denominator is significantly better than the method on the numerator; if $F < \frac{1}{F_{critical}}$, the method on the denominator is significantly worse than the method on the numerator; for other conditions, the performance of the two methods are statistical indistinguishable.

In this experiment, the confidence is set to 95%. The statistical significance between OSVP and the other IQA methods are listed in Tab. 4, in which '1', '0', and '-1' means OSVP is statistically better, indistinguishable, and worse than the compared method on the corresponding database, respectively. As listed in Tab. 4, when compared with RRED, OSVP performs statistically better on *LIVE*, *CSIQ*, *IVC*, and *TOYAMA* databases, and performs similarly on *TID2013* database. When compared with both RRVIF and WNISM, OSVP performs statistically better on *LIVE*, *CSIQ*, *IVC*, and *TID2013* databases, and performs similarly on *TOYAMA*. Moreover, when compared with the three classic FR IQA metrics, OSVP performs statistically better than PSNR on *CSIQ* and *TOYAMA* databases, and similarly on the other three databases, performs better than SSIM on *CSIQ* and *TID2013* databases, and worse on *LIVE* database, performs worse than MS-SSIM on all of the five databases. In summary, the performance comparison results from the F-test are much similar to those from these correlation criteria. So far, we can conclude that OSVP outperforms the existing RR IQA methods, and is comparable to the classical FR IQA metrics.

5. Conclusion

In this paper, a novel RR IQA method has been introduced with the inspiration of the OS mechanism in the HVS. Since the aim of objective IQA is to perform consistently with the human perception, the processing of visual information in the HVS has been investigated. Neuroscience research indicates that the HVS exhibits obvious OS mechanism, within which visual information is extracted for scene perception and understanding. Thus, we have attempted to imitate the OS mechanism for feature extraction from an image.

The OS arises from the arrangement of the excitatory and inhibitory interactions among cortical neurons in a local receptive field. According to this mechanism, we have analyzed the correlations among nearby pixels, and extract the OSVP feature for visual information description. Next, the visual content of an image has been extracted with OSVP, and then mapped into an OSVP-based histogram. Finally, according to the changes on the two histograms from the reference and distorted images, an OSVP-based RR IQA method has been proposed. Experimental results on five publicly available databases have demonstrated that the proposed RR IQA method performs highly consistent to the HVS perception with limited reference data (9 values).

[1] Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of neurophysiology*, 52(6):1106–1130.

[2] Ben Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92(9):3844–3848.

[3] Campbell, F. W. and Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *The Journal of Physiology*, 187(2):437–445.

[4] Cardin, J. A., Palmer, L. A., and Contreras, D. (2007). Stimulus feature selectivity in excitatory and inhibitory neurons in primary visual cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 27(39):333–344.

[5] Chapman, B., Zahs, K. R., and Stryker, M. P. (1991). 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*, 11(5):1347–1358.

[6] Ferster, D. (1988). Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *The Journal of Neuroscience*, 8(4):1172–1180.

[7] Ferster, D. and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual review of neuroscience*, 23:441–471.

[8] Gao, X., Lu, W., Tao, D., and Li, X. (2009). Image quality assessment based on multiscale geometric analysis. *IEEE Transactions on Image Processing*, 18(7):1409–1423.

[9] Gasper, K. and Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, 13(1):34–40.

[10] Hansel, D. and Vreeswijk, C. (2012). The mechanism of orientation selectivity in primary visual cortex without a functional map. *The Journal of Neuroscience*, 32(12):4049–4064.

[11] He, L., Tao, D., Li, X., and Gao, X. (2012). Sparse representation for blind image quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1146 –1153.

[12] Hong, R., Pan, J., Hao, S., Wang, M., Xue, F., and Wu, X. (2014). Image quality assessment based on matching pursuit. *Information Sciences*, 273:196–211.

[13] Horita, Y., Shibata, K., Kawayoke, Y., and Sazzad, Z. M. P. Mict image quality evaluation database.

[14] Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the

cat's visual cortex. *The Journal of Physiology*, 160(1):106–154.

[15] Hubel, D. H. and Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology*, 28:229–289.

[16] Kastner, S., O'Connor, D. H., Fukui, M. M., Fehd, H. M., Herwig, U., and Pinsk, M. A. (2004). Functional imaging of the human lateral geniculate nucleus and pulvinar. *Journal of Neurophysiology*, 91(1):438–448.

[17] Larson, E. C. and Chandler, D. M. (2004). Categorical image quality (csiq) database.

[18] Larson, E. C. and Chandler, D. M. (2010). Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006.

[19] Li, L., Zhu, H., Gao, Y., and Qian, J. (2014). Referenceless measure of blocking artifacts by tchebichef kernel analysis. *IEEE Signal Processing Letters*, 21(1):122–125.

[20] Li, Q. and Wang, Z. (2009). Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):202–211.

[21] Lin, W. and Kuo, C. C. J. (2011). Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312.

[22] Manap, R. A. and Shao, L. (2015). Non-distortion-specific no-reference image quality assessment: A survey. *Information Sciences*, 301:141–160.

[23] Masry, M., Hemami, S. S., and Sermadevi, Y. (2006). A scalable wavelet-based video distortion metric and applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(2):260–273.

[24] Moorthy, A. K. and Bovik, A. C. (2010). A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 17(5):513–516.

[25] Ninassi, A., Callet, P. L., and Autrusseau, F. Subjective quality assessment-ivc database.

[26] Pan, Z., Zhang, Y., and Kwong, S. (2015). Efficient motion and disparity estimation optimization for low complexity multiview video coding. *IEEE Transactions on Broadcasting*, 61(2):166–176.

[27] Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., and Kuo, C.-C. (2013). Color image database TID2013: Peculiarities and preliminary results. In *2013 4th European Workshop on Visual Information Processing (EUVIP)*, pages 106–111.

[28] Shapley, R., Hawken, M., and Xing, D. (2007). The dynamics of visual responses in the primary visual cortex. *Progress in brain research*, 165:21–32.

[29] Sheikh, H. R., Seshadrinathan, K., Moorthy, A. K., Wang, Z., Bovik, A. C., and Cormack, L. K. (2006). Image and video quality assessment research at live.

[30] Sheskin, D. (2011). *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC, Boca Raton.

[31] Soundararajan, R. and Bovik, A. (2012). RRED indices: Reduced reference entropic differencing for image quality assessment. *IEEE Transactions on Image Processing*, 21(2):517 –526.

[32] Troyer, T. W., Krukowski, A. E., Priebe, N. J., and Miller, K. D. (1998). Contrast-invariant orientation tuning in cat visual cortex: Thalamocortical input tuning and correlation-based intracortical connectivity. *The Journal of Neuroscience*, 18(15):5908–5927.

[33] VQEG (2003). Final report from the video quality experts group on the validation of objective models of video quality assessment ii. Video Quality Expert Group.

[34] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

[35] Wang, Z., Simoncelli, E., and Bovik, A. (2003). Multiscale structural similarity for image quality assessment. In *Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2003. Conference*, volume 2, pages 1398–1402.

[36] Wang, Z. and Simoncelli, E. P. (2005). Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *SPIE*, volume 5666, pages 149–159.

[37] Wang, Z., Wu, G., Sheikh, H., Simoncelli, E., Yang, E., and Bovik, A. C. (2006). Quality-aware images. *IEEE Transactions on Image Processing*, 15(6):1680–1689.

[38] Wu, J., Lin, W., Shi, G., and Liu, A. (2013a). Perceptual quality metric with internal generative mechanism. *IEEE Transactions on Image Processing*, 22(1):43 –54.

[39] Wu, J., Lin, W., Shi, G., and Liu, A. (2013b). Reduced-reference image quality assessment with visual information fidelity. *IEEE Transactions on Multimedia*, 15(7):1700–1705.

[40] Wu, J., Shi, G., Lin, W., and Wang, X. (2015). Reduced-reference image quality assessment with orientation selectivity based visual pattern. In *2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pages 663–666.

[41] Ye, P. and Doermann, D. (2012). No-reference image quality assessment using visual codebooks. *IEEE Transactions on Image Processing*, 21(7):3129 –3138.

[42] Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2011). FSIM: a feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386.