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

Jinjian Wu*, Guangming Shi*, Weisi Lin†, and Xiaotian Wang*
*School of Electronic Engineering, Xidian University, Xi'an, Shannxi, China
†School of Computer Engineering, Nanyang Technological University, Singapore
Email: jinjian.wu@mail.xidian.edu.cn;gmshi@xidian.edu.cn; wslin@ntu.edu.sg

***Abstract*—Reduced-reference (RR) image quality assessment (IQA) method aims to accurately measure quality with part of the reference data. The challenge for RR IQA is how to effectively represent the visual content of an image with limited data for quality measurement. Inspired by the orientation selectivity (OS) mechanism in the primary visual cortex, we introduce an OS based visual pattern (OSVP) to extract visual content for RR IQA. The OS arises from the arrangement of the excitatory and inhibitory interactions among connected cortical neurons. Inspired by this, we investigate the correlation among neighbor pixels, and propose the OSVP to represent the visual content of an image. Then, the quality degradation is measured as the changes of OSVP, and a novel RR IQA model is proposed. Experimental results demonstrate that the OSVP based RR IQA model uses limited reference data (9 values) and performs highly consistent with the subjective perception.**

***Index Terms*—Reduced-Reference, Image Quality Assessment, Orientation Selectivity Mechanism, Preferred Orientation**

I. INTRODUCTION

With the development of the network and communication technology, digital signals are tremendously increased and becomes increasingly important in our daily life. During acquisition, processing, and transmission, various types of distortions are introduced into the digital signals. How to effectively acquire high quality digital signals becomes an open problem. Subjective quality assessment is the most reliable way to select high quality signals. However, this method is cumbersome and time-consuming for signal processing systems. Therefore, objective quality assessment algorithms are greatly demanded. In the past decade, a mass of image quality assessment (IQA) models have been proposed. According to the availability of the reference data, the existing IQA models can classified into three classes [1]: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR). In this paper, we focus on RR IQA modeling, where the part of the reference image is available for quality measurement.

RR IQA model aims to accurately measure quality with limited reference data. Generally, some kind of global feature is extracted to create a quality-aware map, which can effectively represent the visual content of an image with a small amount of values for RR IQA. For example, according to the assumption that natural scenes follow some stable statistical properties, a lot of natural scene statistic (NSS) based RR IQA models were proposed [2], [3], [4]. In [2], the statistic properties of images were analyzed in the wavelet domain, and the quality was measured by the changes of the statistical distribution of the wavelet coefficients between the reference and distorted images. In [3], Li and Wang [3] adopted a divisive normalization transform procedure to improve the RR IQA model. Furthermore, Gao et al. [4] throughly analyzed the statistical distribution of coefficients with wavelets, curvelets, and contourlets, and introduced a multiscale geometric based RR IQA method. Recently, Soundararajan and Bovik [5] proposed to measure the quality with the scaled entropies of wavelet coefficients on each divided block. Wu et al. [6] proposed a perception oriented RR IQA model, in which the information fidelities on the primary visual content and the disorderly uncertainty were separately measured. However these RR IQA models either required a large amount of reference information to achieve good performance, or performs not good enough with a small amount of reference data.

In this paper, we turn to investigate the perception on the human visual system (HVS), and introduce an orientation selectivity based visual pattern (OSVP) to extract visual information for RR IQA. Neuroscience researches indicate that the HVS presents obvious orientation selectivity (OS) mechanism for visual information extraction [7]. Moreover, OS arises from the spatial arrangement of excitations and inhibitions in the primary visual cortex (PVC) [8], [9]. By imitating the OS mechanism, we firstly analyze the spatial relations among pixels with the similarity on their preferred orientation. According to the correlation between a pixel and its circularly symmetric neighborhood, the OSVP feature is introduced. Then, the visual content of an input image is extracted and mapped into an OSVP based histogram. Finally, image quality is measured as the changes on the OSVP based histograms between the reference and distorted images. Experimental results demonstrate that the proposed OSVP based RR IQA model only requires a quite small amount of reference data and performs highly consistent with the human perception.

The rest of this paper is organized as follows. The OSVP

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61401325, 61401333, 61472301, 61227004, 61301288), the Research Fund for the Doctoral Program of Higher Education (Nos. 20130203130001, 20130203120009), the 111 Project (No. B07048), the Fundamental Research Funds for The Central Universities (No. JB140227), and Shaanxi province natural science foundation of China (No. 2014JQ8296).

is designed to extract visual structure for RR IQA in Section II. In Section III, experimental results are given. Finally, conclusions are drawn in Section IV.

II. OSVP BASED RR IQA

In this section, the underlying principle for the OS mechanism in the PVC is firstly introduced. Then, by imitating the OS mechanism, the OSVP is introduced for feature extraction. Finally, with OSVP, a novel RR IQA model is proposed.

A. OSVP for Feature Extraction

The HVS is a senior visual signal processing system, which can effectively and efficiently process the input signals. Neuroscience researches indicate that during visual perception,the HVS presents substantial OS to extract visual structure for scene understanding [7]. Therefore, the OS mechanism in part reveals the underline principle of visual signal processing in the HVS.

During the past decades, many hypotheses about the roots of OS have been proposed. As the simplest and most classical one, the feedforward model suggests that the OS lies on the intracortical responses among cortical neurons [7]. When receiving different visual stimuli, cortical neurons present different responses: neurons which receive stimuli with similar preferred orientations are more likely to present excitatory interactions, while neurons which receive stimuli with dissimilar preferred orientations are more probability to present inhibitory interactions [9]. The OS arises from the arrangement of the excitatory and inhibitory interactions among neurons in a local receptive field [8]. A brief schematic representation of the interactions among neurons is shown in Fig. 1.

Inspired by the OS mechanism, we attempt to analyze the correlations among neighbor pixels and introduce a novel feature for visual information representation. Since the excitatory and inhibitory interactions relies on the similarity of preferred orientations between neurons, we firstly compute the preferred

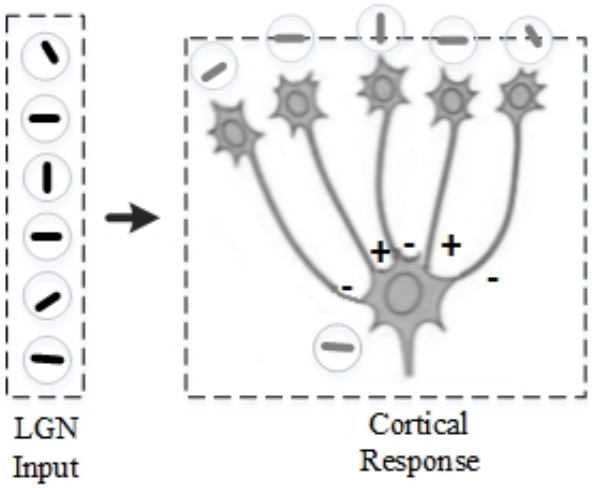


Fig. 1: The interactions between cortical neurons, where '+' means excitatory interaction and '-' inhibitory interaction.

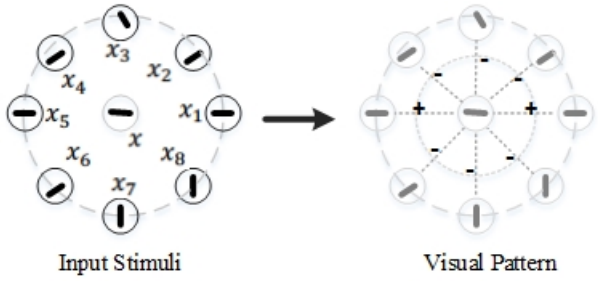


Fig. 2: The 8-neighbor OSVP.

orientation for each pixel in an image. In this paper, the gradient direction θ is adopted to represent the orientation. For a given image $\mathcal{F}$, the orientation of a pixel $x \in \mathcal{F}$ is computed as,

$$\theta(x) = \arctan \frac{G_v(x)}{G_h(x)}, \tag{1}$$

where G_h and G_v represent the gradient magnitudes along the horizontal and vertical directions, respectively. In this work, G_h and G_v are calculated with the Prewitt filters,

$$G_h = \mathcal{F} * f_h, \quad G_v = \mathcal{F} * f_v, \tag{2}$$

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

where f_h/f_v is the horizontal/vertical Prewitt filter, and $*$ is the convolution operation.

The interaction between two pixels is determined by the similarity of their preferred orientation, which can be computed as

$$\mathcal{I}(x|x_i) = \begin{cases} + & \text{if } |\theta(x) - \theta(x_i)| < \mathcal{T} \\ - & \text{else} \end{cases}, \tag{4}$$

where '+' represents excitatory interaction, and '-' represents inhibitory interaction. The similarity threshold $\mathcal{T}$ decides the interaction type between two pixels. According to the subjective visual masking test [10], we set $\mathcal{T}$=6° in this work.

With the interaction types from (4), the OSVP type of a pixel x can be represented by the arrangement of '+' and '-' between pixel x and its circularly symmetric neighborhood. Fig. 2 shows an example of 8-neighbor (namely, n=8) OSVP, whose pattern type is OSVP=$\{+---+--\!-\}$.

B. Quality Measurement

According to the analysis above, the visual pattern (i.e., OSVP) for each pixel is acquired. The OSVP type presents the visual structure of a pixel. For a given image, pixels with a same OSVP type display a same visual content, and pixels with different OSVP types display different visual contents. Therefore, we combine pixels with the same OSVP type, and a given image can be mapped into a histogram for visual content

representation,

$$\mathcal{B}(k) = \sum_{i=1}^{N} w(x_i)\ \delta(\mathcal{P}_v(x_i), \mathcal{P}_v^k), \tag{5}$$

$$\delta(\mathcal{P}_v(x_i), \mathcal{P}_v^k) = \begin{cases} 1 & \text{if } \mathcal{P}_v(x_i) = \mathcal{P}_v^k \\ 0 & \text{else} \end{cases}, \tag{6}$$

where $\mathcal{B}(k)$ is the value of the $k-$th bin, N is the total pixel number of the image, and $\mathcal{P}_v^k$ is the index of the $k-$th OSVP type. The parameter $w(x_i)$ is the weighting factor of pixel x_i. Pixels with large luminance changes always carry much information. Therefore, we adopt the luminance change of each pixel as its corresponding weighting factor $w(x_i)$, which is calculated as

$$w(x_i) = \text{var}(x_i), \tag{7}$$

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

With (5), the visual content of an image is extracted and represented by an OSVP based histogram. For an 8-neighbor OSVP, there are 2^8 different types, which means the visual content of an image is represented by 2^8 values. In order to reduce the number of reference data for RR IQA, we analyze all of the OSVP types and try to combine these types which present similar visual structure. With subjective viewing test, we have found that these OSVP types possess a same number of excitatory interactions present similar visual structure. Thus, we combine these OSVP types with a same number of excitatory interactions. As a result, the visual content of an image can be mapped into a $(n+1)$-bin OSVP based histogram.

Since noise will distort image structures, we measure the quality as the changes on OSVP types. For a given distorted image, the quality is measured as the distance of its OSVP based histogram from that of the reference image,

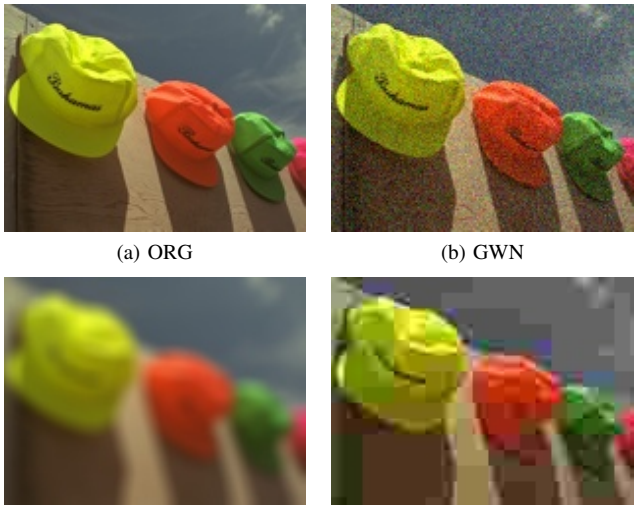
$$\mathcal{Q}(\mathcal{F}_d|\mathcal{F}_r) = \sum_{k=1}^{n+1} \frac{2\mathcal{B}_d(k)\,\mathcal{B}_r(k)}{\mathcal{B}_d^2(k) + \mathcal{B}_r^2(k)}, \tag{8}$$

where $\mathcal{B}_d(k)$/$\mathcal{B}_r(k)$ is the k-th bin of the distorted/reference histogram.

III. Experimental Results

In the experimental section, the effectiveness of the proposed OSVP on visual structure extraction is firstly illustrated. And then, the performance of the proposed RR IQA model is demonstrated on publicly available databases.

Different types of distortion cause different visual structure degradations. The proposed OSVP feature can effectively capture the degradations caused by different types of distortion. Fig. 3 shows a *hats* image contaminated by three different types of distortions: the Gaussian white noise (GWN), the Gaussian blur noise (GBN), and the JPEG compression noise (JPG). Though the distortion energies on the three contaminated images (i.e., Fig. 3 (b)-(d)) are the same (with PSNR=24.5dB), their quality are clearly different. It is obvious that Fig. 3 (b) has a much better quality than the other two contaminated images (i.e., Fig. 3 (c) and (d)). With further



(a) ORG (b) GWN

(c) GBN (d) JPG

Fig. 3: The *hats* image contaminated by three types of distortions, namely, Gaussian white noise, Gaussian blur noise, and JPEG noise. The noise energies on the three distorted images are same, with PSNR=24.5dB.

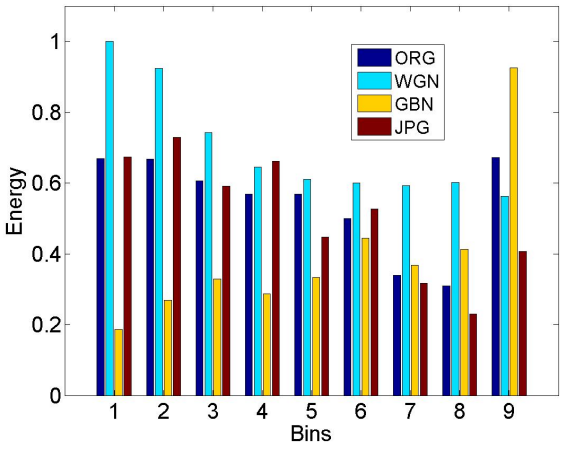


Fig. 4: The OSVP histograms for the *hat* image and its corresponding three contaminated images (as shown in 3).

analysis on these images in Fig. 3, we can see that: the GWN randomly adds the disturbance to the whole image, which has limited effect on visual structure, as shown in Fig. 3 (a); the GBN smooths out most of the high frequency contents, which obviously degrades image structure, as shown in Fig. 3 (b); with the blockiness artifacts, the JPG not only distorts the original edges but also adds some new structures on the smooth regions, as shown in Fig. 3 (c).

Due to the degradations on visual structures of the three contaminated images, their corresponding OSVP histograms are also changed, as shown in Fig. 4. By adding disturbance into the image, the GWN increases the energies of all OSVP types except for the last type, as the light blue bars shown in Fig. 4. That is because with the disturbance from GWN, a flat

TABLE I: IQA Performance Comparison on Three Databases

Dist.	Algo. / Crit.	RR				FR	
		OSVP	RRED	RRVIF	WNISM	PSNR	SSIM
No. of scalars		9	9	2	18	N	N
CSIQ	PLCC	**0.843**	0.677	0.598	0.735	0.800	0.815
	SRCC	**0.849**	0.758	0.633	0.771	0.800	0.838
	RMSE	**0.141**	0.193	0.210	0.178	0.158	0.152
LIVE	PLCC	**0.862**	0.815	0.722	0.743	0.872	0.904
	SRCC	**0.867**	0.818	0.738	0.755	0.876	0.910
	RMSE	**13.84**	15.83	18.90	18.28	13.36	11.67
TID2013	PLCC	**0.724**	0.729	0.577	0.629	0.702	0.686
	SRCC	**0.654**	0.671	0.451	0.523	0.703	0.627
	RMSE	**0.856**	0.849	1.013	0.964	0.883	0.902

region may be distorted into an unsmooth region, where pixels with the 9-th OSVP type (i.e., '++++++++') are distorted into other OSVP types. On the contrary, the effect of GBN is shown as the yellow bars in Fig. 4, where the energies for most OSVP bins are decreased except for the last one. That is because the GBN degrades some unsmooth regions into flat regions, and as a result, the number of the last OSVP type is greatly increased. The JPG causes different changes on OSVP histogram from that of GWN and GBN, as the wine red bars shown in Fig. 4. Since the blockiness artifacts not only degrades original edges but also adds new edges, the energies for most OSVP bins remain similar. Meanwhile, the JPG averages out small disturbance on the flat block, which decrease the variance values for the last OSVP type. In summary, the proposed OSVP feature can effectively represent structure degradations from different distortion types.

In order to give a comprehensive analysis on the proposed OSVP based IQA (OSVP for short), three publicly available IQA databases, *CSIQ*, *LIVE*, and *TID2013*, are chosen in this experiment. Meanwhile, three latest RR IQA models (i.e., RRED [5], RRVIF [6], and WNISM [2]) and two classic FR IQA metrics (i.e., PSNR, SSIM [11]) are adopted for comparison. The performance of these IQA models are measured with 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). A better IQA model will return higher PLCC and SRCC values and a lower RMSE value [1].

The comparison result is listed in Tab. I. As the second row shown in Tab. I, the four RR IQA models uses a very limited amount of reference data (i.e., OSVP, RRED, RRVIF, and WNISM with 9, 9, 2, and 18 values, respectively); and all of the reference data (the size (N) of the image) is required for the two FR IQA metrics. By comparing the performance we can see that, OSVP outperforms the other RR IQA models on both CSIQ and LIVE databases (have larger PLCC and SRCC values and smaller RMSE values). Meanwhile, on TID2013 database, OSVP performs a slight worse than RRED, while much better than RRVIF and WNISM. Moreover, OSVP is comparable with the two FR IQA metrics: OSVP outperforms both PSNR and SSIM on CSIQ and TID2013, while worse than them on LIVE. According to the above comparison result, we can conclude that OSVP outperforms the existing RR IQA models, and is comparable with the classical FR IQA metrics.

IV. Conclusion

In this paper, we have designed a novel OSVP feature to extract visual structure for RR IQA. Inspired by the OS mechanism in the PVC, we firstly analyzed the correlations among nearby pixels with the similarities of their preferred orientation. Then, by arranging the correlations between a pixel and its neighbor pixels, an OSVP is designed to represent its visual information. Finally, according to the changes on OSVP between the reference and distorted images, a novel RR IQA model is introduced. Experimental results on three publicly databases demonstrated that the OSVP based RR IQA model uses limited reference data (9 values) and performs highly consistent to the subjective perception.

References

[1] 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.

[2] 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.

[3] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[4] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.

[5] 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, Feb. 2012.

[6] 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, Nov. 2013.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.