International Journal of Engineering

A Novel Image Structural Similarity Index Considering Image Content Detectability Using Maximally Stable Extremal Region Descriptor

M. H. Khosravi*, H. Hassanpour

Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

P A P E R   I N F O

A B S T R A C T

The image content detectability and image structure preservation are closely related concepts with undeniable role in image quality assessment. However, the most attention of image quality studies has been paid to image structure evaluation, few of them focused on image content detectability. Examining the image structure was firstly introduced and assessed in Structural SIMilarity (SSIM) measure, in which, the definition of image structure is constrained to the intensity covariance between the reference and test images. Indeed, this measure discerns the luminance changes in the pixels of the reference and test images, by employing the low-level statistical features. But this minimal definition of image structure does not cover the issue of image content detectability. In this study, we found that the status of image region smoothness can reflect its structural content. So, we proposed a novel smoothness measure based on the maximally stable extremal regions (MSER) descriptor. Subsequently, we proposed a novel image structural similarity measure, in which the fidelity of image region smoothness is also taken into account. Experimental results on five popular benchmark image databases, including A57, LIVE, CSIQ, TID2008 and TID2013, are provided, which confirm that the proposed approach has a reasonable prediction performance compared to the state-of-the-art image quality metrics.

doi: *10.5829/idosi.ije.2017.30.02b.03*

1. INTRODUCTION

Developments in media devices, besides the growing networking technologies, have aroused extra demands for qualitative media products. Hence the media producers seek a way to automatically assess the quality of produced images and videos before delivery. In addition, in the image and video enhancement applications, existence of a validation stage, which assesses the quality of enhanced output is vital. Specifically, in image enhancement procedures, gauging the amount of image impairments, e.g., noisiness, blockiness, blurriness, and contrast altering is a fundamental necessary. The above two fields of applications make the image quality assessment (IQA) as one of the most active research areas in machine vision.

Intuitively human being is the final and sometimes the best judge, and his/her opinion, which is often referred to as the subjective quality score, determines the quality of image. But the subjective assessment is very time-consuming and inefficient, and cannot be used for real-time applications. Considerable efforts have been performed to develop objective quality metrics to predict the human judges automatically and provide scores as close as possible to the subjective ones [1].

In general, the existing objective IQA measures can be classified into three major categories according to dependence to the reference image. Full-reference (FR) models require the original distortion-free images known as the reference images. These models are widely used to evaluate the performance of image processing algorithms and are frequently employed in optimization procedures of image enhancement methods [2-5]. Reduced-reference (RR) IQA methods, which only need the quality aware features from reference images, are useful to assess the quality of received

Corresponding Author's Email: mohokhosravi@shahroodut.ac.ir (M. H. Khosravi)

images in a noisy transformation channel [6, 7]. The third class is No-reference (NR) or blind IQA methods which do not require the reference images or any information from them [8-10]. These models are more challenging due to their reference image dependency.

Among the FR IQA methods, the Structural SIMilarity (SSIM) measure proposed in the literature [11] can be considered as a mutation in IQA developments. This measure employs the statistical features of image to form its structural indicator. Let **x** and **y** be the reference and distorted image patches, each with M pixels, respectively. The SSIM employs the local comparison of three components: luminance, $l(x, y)$, contrast, $c(x, y)$ and structure, $s(x, y)$ with the following definitions:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} , \tag{1}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} , \tag{2}$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} , \tag{3}$$

in which, μ_x, σ_x and σ_{xy} are the image average, variance and covariance, respectively, as below:

$$\mu_x = \frac{1}{M}\sum_{i=1}^{M} x_i , \tag{4}$$

$$\sigma_x = \left(\frac{1}{M-1}\sum_{i=1}^{M}(x_i - \mu_x)^2\right)^{1/2} , \tag{5}$$

$$\sigma_{xy} = \frac{1}{M-1}\sum_{i=1}^{M}(x_i - \mu_x)(y_i - \mu_y), \tag{6}$$

and C_1, C_2 and C_3 are positive constants to avoid instability when the denominators are very close to zero. The general form of SSIM index is defined by pooling the above three statistics as below:

$$SSIM(x,y) = l(x,y)^\alpha .c(x,y)^\beta .s(x,y)^\gamma . \tag{7}$$

By setting $\alpha = \beta = \gamma = 1$, the resulting SSIM measure is given by:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} . \tag{8}$$

Although the SSIM is an outstanding measure for its interesting theory, low complexity and reasonable results, its structural descriptors, i.e. image mean, variance and covariance, are not sufficient to capture all aspects of image structures [12]. All of the three components in SSIM deal with image luminance and its

variations, while the HVS is also sensitive to the spatial distributions of intensities and their deformations. Several studies in the literature have been done to further promote the SSIM performance by considering more complex structural descriptors. MS-SSIM [13] is a multi-scale version of SSIM in which the three major components of SSIM are also applied to filtered and down-sampled versions of images. CW-SSIM as a complex version of SSIM in wavelet domain is more sensitive to phase than magnitude distortions [14]. Information content weighted SSIM (IW-SSIM) [15] incorporates the idea of weighted pooling, in which the weights derived via image information content. Gradient similarity (GSIM) [16], calculated the contrast and structural similarity of reference and distorted images based on their gradient maps. Riesz-transform based feature similarity (RFSIM) [17] employed the coeffients of the 1st-order and 2nd-order Riesz transform, and masked the resulted maps by the image's edge locations. Feature-based similarity (FSIM) used the image phase congruency and gradient magnitude to gauge the overall image quality [18]. The recently proposed IDSSIM [19] divides an image into edge and texture components, and then computes the luminance and contrast similarity in texture component, and the structural similarity in edge component. IDSSIMc is the extension of IDSSIM which computes the variations of chrominance components too.

In this paper we propose a new structural IQA measure by considering a novel region descriptor which takes into account the amount of image content detectability, as one of the most essential characteristics for IQA. The human perception of an image is very keen to the distinguishability of the image contents. There are also computational theories in object recognition science that support this hypothesis [20]. Here, to assess the amount of image content degradation, we introduce and employ a suitable measure of image smoothness. Image global and local smoothness are important descriptors which show the amount of degradation in image acuity. When an image with distinguishable contents is degraded, to the extent that the regions' smoothness of the original and degraded images is identical, one should expect that the image contents remain recognizable. Obviously, after any blur-like distortion, the image surface will be smooth, while additive-noise distortions have an opposite result and make the image regions harsh. In both cases the image content detectability will be decreased, and subsequently a poor quality score must be reported.

From this point of view, image distortions can be categorized in three classes. The first class includes distortions, which decrease the level of general and local image smoothness by adding artifacts to pristine image in its spatial representation. Additive noise distortions belong to this category. The distortions of

second class behave in opposite direction and increase image smoothness by attenuating its details. Gaussian blur and image denoising are examples of these distortions. In spectral domain, these two distortion classes are respectively called spurious high frequency additive (SHFA) and high frequency attenuating (HFA) distortions, due to their opposite effects on high-frequency components of image. There are also some kinds of distortions which have a mixed behavior. As examples, consider distortions like JPEG blocking, which introduce block regions with smooth body and non-smooth boundaries; and quantization noise which can divide originally smooth regions into some smooth sub-regions with sharp borders.

The main contributions of this work are as follows: First, we focus on the concept of image contents detectability and its relation to the image smoothness fidelity. Second, we define new image smoothness in term of image maximally stable extremal region (MSER), as an efficient image region descriptor. To the best of authors' knowledge, using of MSER descriptor for gauging the image smoothness is unprecedented. Third, based on the new proposed MSER-based image smoothness, we introduced a novel image structural similarity measure dubbed MSER-SSIM, which is thoroughly examined by extensive experiments conducted on five large scale databases. The results show that our proposed measure works relatively better than most of the other state-of-the-art IQA metrics, especially for SHFA distorted images. In addition, the computational complexity of our MSER-SSIM is quite promising.

The rest of this paper is organized as follows: In Section 2, we introduce the maximally stable extremal region (MSER), which will be used to propose a novel measure of image smoothness. In Section 3, the proposed measure is introduced. Experimental results and discussions are provided in Section 4. Section 5 concludes this paper.

2. MAXIMALLY STABLE EXTREMAL REGIONS

The concept of Maximally Stable Extremal Regions (MSERs) was first proposed by Matas et al. [21]. An MSER is a stable connected component of an image, obtained by thresholding the image at different gray levels. Suppose I_{bin}^t be a binary thresholded version of a gray image I in some threshold t, as below:

$$I_{bin}^t(x,y) = \begin{cases} 1 & \text{if } I(x,y) \geq t \\ 0 & \text{else} \end{cases}, \qquad (9)$$

in which the threshold $t \in [\min(I), \max(I)]$. The set of all connected components of all binary images I_{bin}^t, $(t = 0,...,255)$ is the set of all extremal regions. The

extremal regions with stable size over a large range of thresholds are maximally stable extremal regions. More formally definition of MSER is as follow: Let image I be a mapping $I : D \rightarrow \Phi$, in which $D \subset Z^2$ is the spatial domain of pixels and $\Phi = \{0,1,...,255\}$ include the pixel intensity values. A region R is a connected subset of D in terms of an adjacency (neighborhood) relation, with an outer boundary, depicted by B_R, which is the set of all pixels being adjacent to at least one pixel of R but not belonging to R. A region R is an extremal region if and only if $\forall p \in R, q \in B_R : I(p) > I(q)$ (maximum intensity region) or $I(p) < I(q)$ (minimum intensity region). Now, suppose $R_1,...,R_{i-1},R_i,...$ be a sequence of nested extremal regions, i.e. $R_i \subset R_{i+1}$. Extremal region R_{i*} is maximally stable if:

$$\Omega_i = \frac{|R_{i+\Delta} - R_{i-\Delta}|}{|R_i|}, \qquad (10)$$

has a local minimum at i^*, where $|\bullet|$ is the set cardinality (here the number of region points) and Δ is the stability parameter.

MSERs have some important properties, like invariance to affine transformation of image intensities, regions stability and low computational complexity, which make them suitable for various image processing applications like object correspondence [22], object recognition and matching [23, 24], object tracking [25, 26], face registration [27] and Google video grouping [28].

Obtaining the image MSERs is simple and fast. First, image pixels are sorted on intensity using a linear sorting algorithm. Since the pixel intensities are integer numbers in the range [0,255], the counting sort with the time cost $O(n)$ would be the best choice [29]. After pixel sorting, the connected components are formed using efficient union-find algorithm [29] with the computational complexity of $O(n\log\log n)$. The output is a set of region seed points, in which, for each seed point x, a set of neighbor points within the level set $I(x)$ are assigned to form a connected region.

2. 1. MSER-based Image Smoothness
The number of MSER seed points has a direct relationship with image smoothness, i.e. in an smooth region the number of seed points is less than the ones on a non-smooth region. This property of MSER in image smoothness description is an important characteristic, which persuades us to employ MSER as an effective and efficient approach for image smoothness measure. We define the MSER-based patch smoothness as below:

$$S_{MSER}(p) = 1 - \min\left(1, \frac{|MSER_{Seeds}(p)|}{c}\right), \qquad (11)$$

in which $MSER_{Seeds}(p)$ denotes the set of MSER seed points of image patch p, $C = 0.5|p|^2$ is an adaptive constant value greater than the number of patch seed points, and $|\bullet|$ indicates the set cardinality. Clearly the range of $S_{MSER}(p)$ is within [0,1], in which the values close to one, belong to more smooth patches. Based on Equation (11), we define the image smoothness measure in two globally and locally schemes. The globally smoothness of image I, indicated by $GS_{MSER}(I)$, can be considered as a particular case of Equation (11), in which the whole image I is the input of function S_{MSER}. To calculate the local-based smoothness measure, we generate a local smoothness map by partitioning the image I into non-overlapped 16×16 patches p, and calculating the smoothness measure of each patch using Equation (11). Finally, the image local-based smoothness measure is obtained as below:

$$LS_{MSER}(I) = \frac{1}{N_p} \sum_{p \in I} S_{MSER}(p), \qquad (12)$$

in which N_p is the number of image's patches.

2. 2. Analysis of MSER-based Image Smoothness
Here, we examine more closely on how well the proposed image smoothness measure works.

Figure 1 shows the image 'I23.bmp' from TID2013 image database [30], in which some patches with different smoothness status are selected and their MSER-based smoothness values are sticked beside them. It can be observed that the corresponding smoothness values increase with increasing the patch homogeneity and decrease by adding any primitive. The overall smoothness of images are evaluated in Figure 2 in which the reference images of TID2013 database are arranged in a descending order of MSER-based global smoothness (GS_{MSER}).
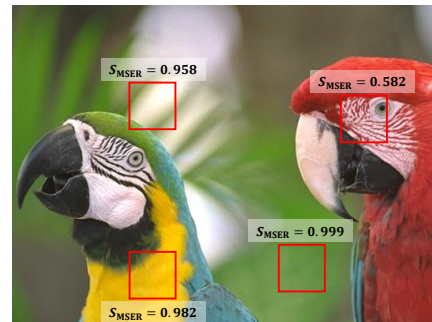

Figure 1. The reference image 'I23.bmp' from TID2013 database, in which the MSER-based smoothness for four different patches are calculated and depicted

Figure 2. (Top-left to bottom-right) The reference images of TID2013 database, arranged in descending order of GS_{MSER}.

It can be seen that this sorted list of images is intuitively consistent with our perception of image smoothness, i.e. the images in the beginning of this list have large smooth regions, while the ones in the ending can be considered as crowd images with small smooth regions.

Any degradation of image primitives, which alters the image content detectability, can be monitored by tracking the image global and local smoothnesses. When a pristine image is polutted by an SHFA distortion (like additive Gaussian noise (AGN), or high-frequency noise (HFN)), the number of MSER seeds increases. In opposite, a degraded image with HFA distortion (like blur or jpeg2000 ringing artifact), has less MSER seed points with respect to its pristine image. We showed this relationship through Figure (3), in which Figure 3(a), 3(c) and 3(e) show the reference image 'I15.bmp' from the TID2013 database, and its corresponding blurred and HFN distorted images, respectively. Obviously, the detectability of image contents is altered in both of the two distortions. Figure 3(b), 3(d) and 3(f) indicate the corresponding MSER map, besides the number of region seed points. Here, the MSER map shows the overlapping MSERs, in which the value of each pixel is equal to the number of overlapping extremal regions, rescaled to [0-255] for presentation purpose. It can be seen that the HFN distortion increases the number of region seed points drastically, while the blur distortion, decreases it. This emphasizes the fact that the number of image regions obtained by MSER detector is sensitive to the image distortions and thus can serve as a suitable distinguishing property.

3. MSER-BASED STRUCTURAL SIMILARITY

Inspired from the adequacy evidences of the MSER-based smoothness measure in image region description, we define a novel image structural similarity measure, which evaluates the image content detectability through assessing its region smoothness provided by MSER detector. We found that in SHFA distortions (e.g. additive noise) the deviation of general smoothness between reference and distorted image are discriminative enough to cover the images differences, but in HFA distortions (e.g. image blurring) this deviation is not adequate and we must resort to the image average smoothness by considering the smoothness map. The good news is that we can distinguish these two distortion types by image global smoothness (i.e. for SHFA distortions, the global smoothness of reference image is higher than the global smoothness of degraded image, while for HFA distortions this relation is opposite). Let x and y be the pristine reference and distorted images. We define the

MSER-based image Content Detectability (MSER_CD) as below:

$$MSER\_CD(x,y) = \begin{cases} 1 - \left| GS_x - GS_y \right| & GS_x > GS_y \\ \dfrac{2LS_x LS_y + C_4}{LS_x^2 + LS_y^2 + C_4} & else \end{cases} \quad (13)$$

in which, for simplicity, the acronyms GS_x and LS_x stand for $GS_{MSER}(x)$ and $LS_{MSER}(x)$, respectively. Clearly, the MSER-CD of two images, provides a scalar value in the range of [0,1], in which the higher value indicates more fidelity of content detectability in two images.

Having the MSER-CD, we provide a new quality measure that encompasses all three major components of SSIM measure besides our novel quality factor MSER-CD, to compensate the lack of content detectability evaluation of SSIM. The overall proposed quality metric is as below,

$$MSER\_SSIM(x,y) = \\ w_1 SSIM(x,y) + w_2 MSER\_CD(x,y), \quad (14)$$

in which the coefficients w_1 and w_2 are used to adjust the relative importance of different components and must be chosen to satisfy the condition $w_1 + w_2 = 1$.

(a) Reference image   (b) #Region seeds = 910

(c) Blurred image   (d) #Region seeds = 44
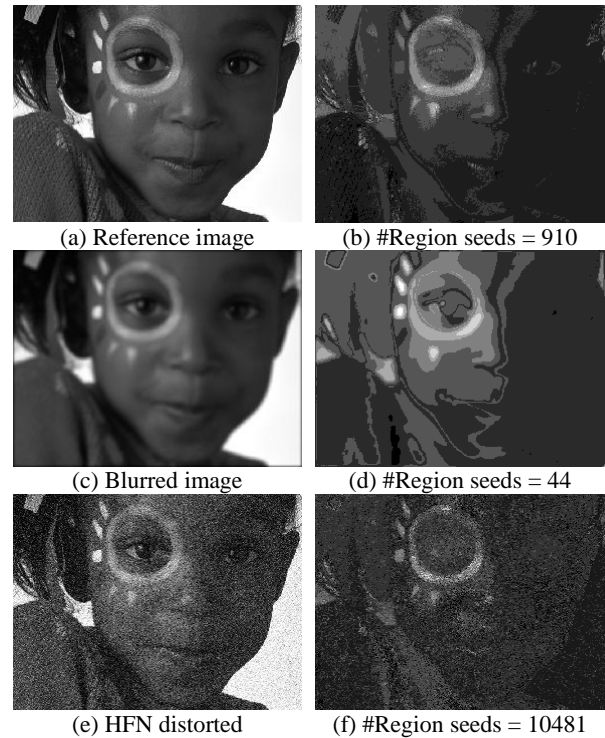
(e) HFN distorted   (f) #Region seeds = 10481

Figure 3. (from top to bottom) Left: An image and its corresponding blurred and high-frequency noise distorted images. Right: The corresssponding MSER regions.

3. 1. Visual Masking Property of MSER-SSIM   An interesting property of the proposed MSER-CD, which will be discussed here, is its visual masking property. In MSER domain, the increasing or decreasing of seed points depends on the intrinsic smoothness nature of the original image. To be more concretely, increasing the number of MSER seed points due to an SHFA distortion in an intrinsicaly smooth region, is more than the ones in an innately non-smooth region. In a similar way, decreasing the number of MSER seed points resulted from a HFA distortion in a non-smooth region is more than the ones happened for a smooth region. This is completely consistent with the visual masking property of HVS, which claims that the visibility of an image additive distortion is reduced in crowd regions. We show this in Figure 4, in which two natural images with different intrinsic smoothness are chosen from TID2013 database for comparison. Figure 4(a) and 4(b) show these two images contaminted by HFN, while Figure 4(c) and 4(d) indicate their blurred versions, besides the corresponding SSIM and MSER_CD values. It can be seen that, although the subjective quality scores (MOS) of Figure 4(a) and 4(b) (and also 4(c) and 4(d)) are very close, the corresponding SSIM values are different. Indeed, according to SSIM, the image in Figure 4(b) is far worse than 4(a), while MSER_CD measure coincides with MOS, for these two distorted images. For the blurred image pairs, the SSIM prediction is more accurate than the previous case, but still the MSER_CD prediction is better.

4. EXPERIMENTAL RESULTS

4. 1. Image Databases   To evaluate the performance of the proposed method we employed five frequently used public image quality databases,

(a) MOS=5.39, SSIM=0.93 MSER_CD= 0.99

(b) MOS=5.33, SSIM=0.59 MSER_CD=0.90

(c) MOS=5.40, SSIM=0.98 MSER_CD=0.99
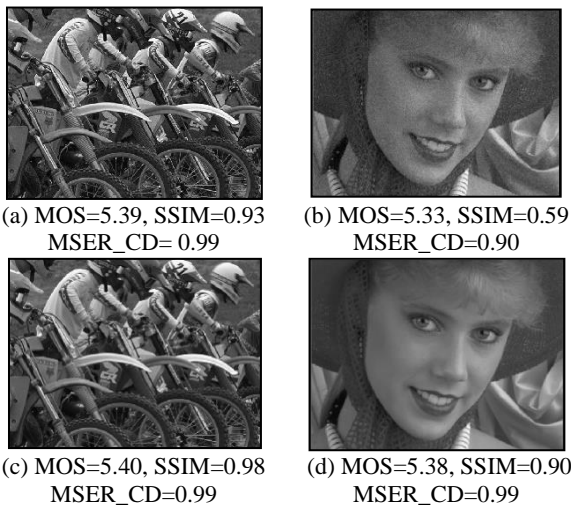
(d) MOS=5.38, SSIM=0.90 MSER_CD=0.99

Figure 4. Comparison of SSIM and MSER_CD on two natural images with different intrinsic smoothnesses (a) and (b) HFN contaminated images, (c) and (d) blurred images

including A57 [31], CSIQ [32], LIVE [33], TID2008 [34] and TID2013. The subjective scores of the images in TID2008 and TID2013 are in the form of MOS, while the images in A57, LIVE and CSIQ have subjective scores in the form of DMOS. Table 1 shows the major characteristics of these databases.

4. 2. Performance Metrics   We used four criteria to assess the performance of the proposed algorithm following the literature [35]: (1) The Pearson correlation coefficient (PCC), which measures the amount of predictions correlation with the subjective scores, (2) the Spearman rank order correlation (SROCC) and (3) the Kendall rank order correlation (KROCC), which measure both the relative monotonicity between the predictions and subjective scores, and (4) the root mean square error (RMSE) which validates the predictions accuracy, like PCC. Before evaluating the performance criteria, we applied the logistic transform suggested by [35], to the values obtained from our proposed measure to bring them on the same scales as the LIVE, TID2008 and TID2013's MOS, and A57 and CSIQ's DMOS values.

4. 3. Implementation Details   Here, the proposed MSER-SSIM measure was compared with SSIM and SSIM-inspired IQAs including: MSSIM, IWSSIM, RFSIM, GSIM, FSIM, CWSSIM and IDSSIM. In addition some SSIM-irrelevant well-performed state-of-the-art IQMs such as PSNR, visual information fidelity (VIF) [36], internal generative mechanism-based quality metric (IGM) [37] and visual signal to noise ratio (VSNR) [31], are also incorporated in our comparison. These IQMs were applied using their default implementations.

To obtain the MSER smoothness measures, we extract the MSER descriptors of non-overlapped patches using VLFeat open source Matlab codes [38]. The number of MSER seed points and their neighbors can be controlled by parameter Δ in Equation (10), which defines how the stability is calculated. Other MSER parameteres are: 1) MinArea and MaxArea, which control the minimum and maximum area of the regions relative to the image domain area, 2) MaxVariation, which can make a tradeoff between the region stability and the number of regions, and 3) MinDiversity, a threshold indicates when two nested stable regions can be merged together.

TABLE 1. Benchmark image datasets for IQA

1	Reference Images	Distorted Images	Distortion Types	Image Dimensions	Score Ranging
A57	3	54	6	512×512	[0-1]
LIVE	29	779	5	Various	[0-100]
CSIQ	30	866	6	512×512	[0-1]
TID2008	25	1700	17	512×384	[0-9]
TID2013	25	3000	24	512×384	[0-9]

The best values for these MSER parameters, which lead to the best SROCC performance, are as follows: $\Delta = 5$, MinArea=0.05, MaxArea=0.75, MaxVariation=0.25 and MinDiversity=0.5. In addition, we used the default values of constants c_1 to c_3, as they set in the original SSIM, and choose the values $C_4 = 6.5$ and $w_1 = 0.2$, employed in Equation (13) and Equation (14), respectively.

4. 4. Performance Evaluation   We compared the SROCC performance criterion of the proposed method with the ones obtained by other methods on each distortion types in TID2008 and TID2013 databases. Among all of the existing image quality databases, the TID2008 and TID2013 cover most of the distortion types, i.e. 17 and 24 different types, respectively. The distortions of TID2008 are additive Gaussian noise (AGN), additive noise in color components (ANC),

spatially correlated noise (SCN), masked noise (MN), high frequency noise (HFN), impulse noise (IN), quantization noise (QN), Gaussian blur (GB), image denoising (DEN), JPEG compression (JPEG), JPEG2000 compression (JP2K), JPEG transmission errors (JPTE), JPEG2000 transmission errors (J2TE) , non-eccentricity pattern noise (NEPN), local block-wise distortions of different intensity (Block), mean shift (MS), and contrast change (CTC). The first 17 distortions of the TID2013 are the same as the TID2008. The rest are as follow: change of color saturation (CCS), multiplicative Gaussian noise (MGN), comfort noise (CS), lossy compression of noisy images (LCNI), image color quantization with dither (ICQD), chromatic aberrations (CHA) and finally, sparse sampling and reconstruction (SSR). Table 2 shows the results of SROCC comparison, in which for each distortion, we bolded the three best performed algorithms.

TABLE 2. SROCC values of IQA indices for each type of distortions in TID2013 and TID2008. The best three performed metrics have been bolded for quick access. In addition, the MSER-SSIM measures which occupy the first rank, indicated by an '*' sign

DB.	Dist.	SSIM-irrelevant measures					MSSIM	IWSSIM	SSIM-related IQA measures					
		VIF	IGM	PSNR	VSNR	SSIM			CWSSIM	RFSIM	FSIM	GSIM	IDSSIM	MSER-SSIM
TID2013	AGN	0.8996	**0.9371**	**0.9338**	0.8270	0.8675	0.8663	0.8448	0.7920	0.8877	0.8984	0.9063	0.8981	**0.9522***
	ANC	0.8428	**0.8792**	**0.8667**	0.7266	0.8010	0.7729	0.7514	0.7228	0.8476	0.8207	0.8175	0.8316	**0.8673***
	SCN	0.8835	**0.9244**	**0.9245**	0.8024	0.8756	0.8543	0.8166	0.8340	0.8821	0.8749	0.9158	0.8822	**0.9555***
	MN	**0.8449**	0.8053	0.8355	0.7118	0.7766	0.8074	0.8019	0.6848	**0.8366**	0.7943	0.7292	0.8233	**0.8377**
	HFN	0.8972	**0.9260**	**0.9182**	0.8566	0.8633	0.8648	0.8589	0.8750	0.9145	0.8991	0.8869	0.9000	**0.9194**
	IN	0.8536	0.8591	**0.9000**	0.7343	0.7503	0.7628	0.7281	0.8802	**0.9062**	0.8072	0.7964	0.8595	**0.8648**
	QN	0.8161	**0.8913**	0.8754	0.8356	0.8657	0.8705	0.8467	0.8224	**0.8968**	0.8719	**0.8841**	0.8462	0.4627
	GB	0.9649	**0.9772**	0.9102	0.9469	0.9667	0.9672	**0.9701**	0.8855	**0.9697**	0.9550	0.9689	0.9692	0.9666
	DEN	0.9064	**0.9493**	**0.9503**	0.9104	0.9254	0.9267	0.9152	0.9001	0.9359	0.9301	0.9432	0.9405	**0.9453**
	JPEG	0.9191	**0.9491**	0.9217	0.9007	0.9200	0.9265	0.9197	0.9080	**0.9398**	0.9378	0.9284	**0.9414**	0.9226
	JP2K	.9516	**0.9680**	0.8858	0.9273	0.9063	0.9504	0.9506	0.9326	0.9518	0.9576	**0.9601**	**0.9589**	0.9064
	JGTE	0.8441	0.8452	0.8060	0.8181	0.8493	0.8475	0.8387	0.7203	**0.8786**	0.8463	0.8512	**0.8655**	**0.8919***
	J2TE	0.8760	**0.9183**	0.8905	0.8407	0.8828	0.8888	0.8656	0.6725	**0.9102**	0.8912	**0.9181**	0.9045	0.8787
	NEPN	0.7719	**0.8029**	0.6791	0.6652	0.7821	0.7968	0.8010	0.7882	0.7704	0.7917	**0.8130**	0.7470	**0.8104**
	Block	0.5306	0.5272	0.3297	0.1771	**0.5720**	0.4800	0.3716	0.3463	0.0338	0.5489	**0.6418**	0.5292	**0.6331**
	MS	0.6275	0.6091	0.7571	0.3632	0.7751	**0.7906**	**0.7833**	0.5485	0.5547	0.7530	**0.7874**	0.6626	0.7776
	CTC	**0.8523**	0.4601	0.4466	0.3319	0.4314	0.4633	0.4592	**0.7323**	**0.5591**	0.4686	0.4856	0.4858	0.4275
	CCS	0.3099	0.3225	**0.6388**	0.3676	0.4141	0.4099	0.4196	0.1425	0.0204	0.2748	0.3573	**0.6690**	**0.4567**
	MGN	0.8466	**0.8832**	**0.8831**	0.7644	0.7803	0.7785	0.7727	0.7481	0.8487	0.8469	0.8347	0.8788	**0.9085***
	CN	0.8948	**0.9201**	0.8413	0.8690	0.8565	0.8527	0.8761	0.8946	0.8917	**0.9120**	**0.9124**	0.9051	0.8652
	LCNI	0.9229	**0.9492**	0.9155	0.8821	0.9057	0.9067	0.9037	0.9283	0.9009	**0.9466**	**0.9562**	0.8988	0.8915
	ICQD	0.8463	**0.9071**	**0.9201**	0.8695	0.8542	0.8554	0.8401	**0.9061**	0.8959	0.8759	0.8972	0.8890	0.4966
	CHA	0.8848	**0.9142**	0.8797	0.8644	0.8774	0.8784	0.8681	0.7700	**0.8990**	0.8714	0.8822	0.8927	**0.8888**
	SSR	0.9371	**0.9672**	0.9108	0.9364	0.9460	0.9482	0.9474	0.8724	0.9325	**0.9564**	**0.9667**	0.9536	0.9102
TID2008	AGN	0.8838	**0.9069**	**0.9070**	0.7727	0.8106	0.8085	0.7869	0.7559	0.8415	0.8566	0.8606	0.8501	**0.9336***
	ANC	0.8750	**0.8947**	**0.8994**	0.7793	0.8029	0.8053	0.7920	0.6436	0.8621	0.8527	0.8090	0.8595	**0.9036***
	SCN	0.8709	**0.9152**	**0.9169**	0.7664	0.8143	0.8209	0.7713	0.7838	0.8475	0.8486	0.8941	0.8653	**0.9388***
	MN	**0.8683**	0.7968	0.8515	0.7294	0.7794	0.8106	0.8088	0.7097	**0.8533**	0.8021	0.7452	0.8523	**0.8751***
	HFN	0.9074	**0.9223**	**0.9270**	0.8800	0.8773	0.8733	0.8702	0.8797	0.9181	0.9152	0.8945	0.9091	**0.9281***
	IN	0.8464	0.8160	**0.8723**	0.6471	0.6732	0.6907	0.6464	**0.8589**	**0.8805**	0.7452	0.7234	0.8074	0.8445
	QN	**0.8816**	0.8788	0.8696	0.8261	0.8530	0.8588	0.8176	0.7762	**0.8950**	0.8564	**0.8799**	0.8481	0.4147
	GB	0.9540	**0.9682**	0.8684	0.9330	0.9544	0.9563	**0.9636**	0.8413	0.9408	0.9471	**0.9599**	0.9574	0.9435
	DEN	0.9182	**0.9704**	0.9416	0.9299	0.9529	0.9582	0.9473	0.9050	0.9399	**0.9602**	**0.9724**	0.9539	0.9478
	JPEG	0.9167	**0.9484**	0.8717	0.9174	0.9251	0.9321	0.9208	0.9127	0.9385	0.9369	**0.9393**	**0.9438**	0.9023
	JP2K	0.9709	**0.9845**	0.8131	0.9515	0.9629	0.9699	0.9738	0.9452	0.9487	**0.9773**	**0.9761**	0.9694	0.8783
	JGTE	0.8585	0.8635	0.7565	0.8113	0.8677	0.8680	0.8588	0.7046	0.8534	**0.8707**	**0.8790**	0.8701	0.8704
	J2TE	0.8500	**0.8893**	0.8308	0.7909	0.8576	**0.8606**	0.8202	0.5997	0.8591	0.8543	**0.8935**	0.8602	0.8433
	NEPN	0.7619	0.7295	0.5814	0.5715	0.7107	0.7376	**0.7724**	**0.7686**	0.7274	0.7491	0.7386	0.6824	**0.7934***
	Block	0.8320	0.7902	0.6192	0.1926	0.8462	0.7557	0.7623	0.2788	0.6258	**0.8493**	**0.8862**	0.7520	**0.8934***
	MS	0.5095	0.4887	0.7107	0.3714	**0.7230**	**0.7336**	0.7066	0.4221	0.4335	0.6720	0.7190	0.5514	**0.7213**
	CTC	**0.8403**	0.6411	0.6042	0.4746	0.4411	0.6380	0.6301	**0.8648**	0.5431	0.6481	**0.6691**	0.6459	0.4121

It can be seen that the proposed measure can faithfully assess image quality. Particularly, for the SHFA distortions (include AGN, ANC, SCN, JGTE, and MGN) our algorithm is superior to the entire competing IQA methods on TID2013 and TID2008. In addition, we see that the results of the proposed measure on HFA distortions (like GB, DEN and CHA) are also comparable with the ones obtained using the state-of-the-art metrics although not the best.

In these cases, the sensitivity of MSER descriptor decreased in high-severity distortions and similar seed points are reported for different severities. In applications which only the HFA distortions are targeted, one can increase the performance of the proposed measure by tunning the constant C in Equation (11), and obtain better results. However, we didn't do it here to preserve generality of the method. For exotic distortions (defined in [30]) like JGTE, NEPN and Block the performance of proposed method is still acceptable and stand among the best. It must be mentioned that for some distortions (like QN and ICQD) the performance of MSER-SSIM is not promising, because of their mixed nature, as mentioned in Section 1. Indeed these distortions exhibit the behaviour of SHFA and HFA distortions, simultaniously, and therefore have unpredictable smoothness status in different severities. Table 3 lists the average of PLCC, SROCC, KROCC and RMSE results of the proposed MSER-SSIM and the other IQA measures on A57, LIVE, CSIQ, TID2008 and TID2013 databases. For each performance criterion, the three IQA indices producing the best results were highlighted in boldface for each database. It can be seen that the results of the proposed method for A57, LIVE and TID2013 databases, stand among the three best IQA metrics. For CSIQ and TID2008 databases the proposed method's results are only slightly worse than the best and the differences are not significant. As we mentioned in the previous paragraphs, although the performance of our proposed measure for SHFA distortions is substantial, its performance for HFA distortions is not as well as the ones for SHFA distortions, and by this, the overall performance of MSER-SSIM is lower than metrics like IGM, FSIM and IDSSIM. It is worthy to note that the IGM metric, as one of the best performers on TID2008, TID2013 and CSIQ database, is a very time-consuming metric, with limited applicability in real-time situations.

TABLE 3. Comparison of average performance criteria of IQA indices on four image databases. The best three performed metrics have been bolded for quick access. The MSER-SSIM measures which occupy the first rank, indicated by an '*' sign

DB.	Criterion	SSIM-irrelevant measures					SSIM-related IQA measures							
		VIF	IGM	PSNR	VSNR	SSIM	MSSSIM	IWSSIM	CWSSIM	RFSIM	FSIM	GSIM	IDSSIM	MSER-SSIM
TID2013	**PLCC**	0.7720	**0.8561**	0.6902	0.7402	0.7895	0.8329	0.8319	0.7908	0.8333	**0.8589**	0.8464	**0.8584**	0.8327
	SROCC	0.6769	**0.8097**	0.6862	0.6812	0.7417	0.7859	0.7779	0.7628	0.7744	0.8015	0.7946	**0.8304**	**0.8088**
	KROCC	0.5147	-	0.5043	0.5084	0.5588	0.6047	0.5977	0.5833	0.5951	**0.6289**	0.6255	**0.6451**	**0.6410**
	RMSE	0.7880	-	0.8976	0.8392	0.7608	0.6861	0.6880	**0.5319**	0.6852	0.6349	0.6603	0.6358	**0.4935***
TID2008	**PLCC**	**0.8084**	**0.8857**	0.5309	0.6820	0.7732	0.8451	0.8579	0.7650	0.8645	**0.8738**	0.8422	0.8646	0.8453
	SROCC	0.7491	**0.8902**	0.5245	0.7046	0.7749	0.8542	0.8559	0.7442	0.8680	**0.8805**	0.8504	**0.8736**	0.8261
	KROCC	0.5860	**0.7104**	0.3696	0.5340	0.5768	0.6568	0.6636	0.5605	0.6780	**0.6946**	0.6596	**0.6827**	0.6456
	RMSE	0.7899	**0.6228**	1.1372	0.9815	0.8511	0.7173	0.6895	**0.5491**	0.6746	0.6525	0.7235	0.6742	**0.4833***
CSIQ	**PLCC**	**0.9277**	**0.9280**	0.8001	0.8002	0.8613	0.8991	0.9144	0.9150	0.9179	0.9120	0.8964	**0.9317**	0.9096
	SROCC	0.9195	**0.9401**	0.8057	0.8106	0.8756	0.9133	0.9213	0.9202	**0.9295**	0.9242	0.9108	**0.9451**	0.9071
	KROCC	0.7537	**0.7872**	0.6080	0.6247	0.6907	0.7393	0.7529	0.7570	**0.7645**	0.7567	0.7374	**0.7947**	0.7396
	RMSE	0.0980	**0.0978**	0.1575	0.1575	0.1334	0.1449	0.1063	0.0985	0.1042	0.1022	0.1164	**0.0953**	0.0975
LIVE	**PLCC**	**0.9604**	0.9578	0.8721	0.9231	0.9449	0.9489	0.9522	0.8725	0.9354	**0.9597**	0.9512	0.9473	**0.9657***
	SROCC	**0.9636**	0.9580	0.8755	0.9274	0.9479	0.9513	0.9567	0.9027	0.9401	**0.9634**	0.9561	0.9516	0.9616
	KROCC	0.8282	**0.8319**	0.6864	0.7616	0.7963	0.8045	0.8175	0.7319	0.7816	**0.8337**	0.8150	0.8063	**0.8399***
	RMSE	**7.6137**	0.9248	13.368	10.506	8.9455	8.6188	8.3473	11.2960	9.6642	**7.6780**	8.4327	8.7514	**5.8673***
A57	**PLCC**	0.6158	0.9230	0.6587	**0.9472**	0.8017	0.8504	0.9035	0.7444	0.8475	0.9252	0.7231	**0.9282**	0.9349
	SROCC	0.6223	0.8984	0.6189	**0.9355**	0.8066	0.8394	0.8706	0.6557	0.8215	**0.9181**	0.9018	**0.9285**	0.8556
	KROCC	0.4589	0.7359	0.4309	**0.8031**	0.6058	0.6478	0.6848	0.5093	0.6324	0.7639	**0.8724**	**0.7741**	0.7407
	RMSE	0.1936	0.0945	0.1849	**0.0781**	0.1469	0.1293	0.1052	0.1641	0.1305	0.0933	0.1206	**0.0900**	**0.0781***

5. CONCLUSIONS

In this paper we emphasize the value of image content detectability in the context of image quality assessment, and accentuate the role of region smoothness state in image content clarity. Inspired by this idea, a new MSER-based image smoothness measure is proposed and benefited from the suitable distinguishing feature of this measure; we introduced a novel structural similarity measure. We evaluated the performance of the proposed method in terms of correlation with human perceptual opinion scores and found it highly competitive against most of the state-of-the-art image quality methods.

6. REFERENCES

1. Chandler, D.M., "Seven challenges in image quality assessment: Past, present, and future research", *ISRN Signal Processing*, Vol. 2013, (2013), 53-60.

2. Wang, Z., "Applications of objective image quality assessment methods [applications corner]", *Signal Processing Magazine, IEEE*, Vol. 28, No. 6, (2011), 137-142.

3. Hassanpour, H. and Asadi, S., "Image quality enhancement using pixel wise gamma correction", *International Journal of Engineering-Transactions B: Applications*, Vol. 24, No. 4, (2011), 301-311.

4. Hassanpour, H., Azari, F. and Asadi, S., "Improving dark channel prior for single image dehazing", *International Journal of Engineering-Transactions C: Aspects*, Vol. 28, No. 6, (2015), 880-889.

5. Khosravi, M. and Hassanpour, H., "Image denoising using anisotropic diffusion equations on reflection and illumination components of image", *International Journal of Engineering-Transactions C: Aspects*, Vol. 27, No. 9, (2014), 1339-1348.

6. Soundararajan, R. and Bovik, A.C., "Rred indices: Reduced reference entropic differencing for image quality assessment", *Image Processing, IEEE Transactions on*, Vol. 21, No. 2, (2012), 517-526.

7. Rehman, A. and Wang, Z., "Reduced-reference image quality assessment by structural similarity estimation", *Image Processing, IEEE Transactions on*, Vol. 21, No. 8, (2012), 3378-3389.

8. Kamble, V. and Bhurchandi, K., "No-reference image quality assessment algorithms: A survey", *Optik-International Journal for Light and Electron Optics*, Vol. 126, No. 11, (2015), 1090-1097.

9. Manap, R.A. and Shao, L., "Non-distortion-specific no-reference image quality assessment: A survey", *Information Sciences*, (2015), 203-210.

10. Li, L., Lin, W., Wang, X., Yang, G., Bahrami, K. and Kot, A.C., "No-reference image blur assessment based on discrete orthogonal moments", *IEEE Transaction on Cybernetics*, Vol. 46, No. 1, (2016), 39-50.

11. Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P., "Image quality assessment: From error visibility to structural similarity", *Image Processing, IEEE Transactions on*, Vol. 13, No. 4, (2004), 600-612.

12. Wu, J., Qi, F. and Shi, G., "Image quality assessment based on improved structural similarity, in Advances in multimedia information processing–PCM", (2012), Springer.153-163.

13. Wang, Z., Simoncelli, E.P. and Bovik, A.C., "Multiscale structural similarity for image quality assessment", in Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers, . Vol. 2, (2003), 1398-1402.

14. Sampat, M.P., Wang, Z., Gupta, S., Bovik, A.C. and Markey, M.K., "Complex wavelet structural similarity: A new image similarity index", *Image Processing, IEEE Transactions on*, Vol. 18, No. 11, (2009), 2385-2401.

15. Wang, Z. and Li, Q., "Information content weighting for perceptual image quality assessment", *Image Processing, IEEE Transactions on*, Vol. 20, No. 5, (2011), 1185-1198.

16. Chen, G.-H., Yang, C.-L. and Xie, S.-L., "Gradient-based structural similarity for image quality assessment", in Image Processing, IEEE International Conference on. (2006), 2929-2932.

17. Zhang, L., Zhang, D. and Mou, X., "Rfsim: A feature based image quality assessment metric using riesz transforms", in Image Processing (ICIP), 17th IEEE International Conference on. (2010), 321-324.

18. Zhang, L., Zhang, D. and Mou, X., "Fsim: A feature similarity index for image quality assessment", *Image Processing, IEEE Transactions on*, Vol. 20, No. 8, (2011), 2378-2386.

19. Yang, J., Lin, Y., Ou, B. and Zhao, X., "Image decomposition-based structural similarity index for image quality assessment", *EURASIP Journal on Image and Video Processing*, Vol. 2016, No. 1, (2016), 31-40.

20. Biederman, I., "Recognition-by-components: A theory of human image understanding", *Psychological review*, Vol. 94, No. 2, (1987), 115-122.

21. Matas, J., Chum, O., Urban, M. and Pajdla, T., "Robust wide-baseline stereo from maximally stable extremal regions", *Image and Vision Computing*, Vol. 22, No. 10, (2004), 761-767.

22. Perd'och, M., "Maximally stable extremal regions and local geometry for visual correspondences", Citeseer, (2011),

23. Forssen, P.-E., "Maximally stable colour regions for recognition and matching", in IEEE Conference on Computer Vision and Pattern Recognition,. (2007), 1-8.

24. Obdrzalek, S. and Matas, J., Object recognition using local affine frames on maximally stable extremal regions, in Toward category-level object recognition. (2006), Springer.83-104.

25. Donoser, M. and Bischof, H., "Efficient maximally stable extremal region (MSER) tracking", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06),. Vol. 1, (2006), 553-560.

26. Roth, P.M., Donoser, M. and Bischof, H., "Tracking for learning an object representation from unlabeled data", in Proceedings of the Computer Vision Winter Workshop (CVWW). (2006), 46-51.

27. Murphy-Chutorian, E. and Trivedi, M.M., "N-tree disjoint-set forests for maximally stable extremal regions", in BMVC., (2006), 739-748.

28. Sivic, J., Schaffalitzky, F. and Zisserman, A., "Object level grouping for video shots", *International Journal of Computer Vision*, Vol. 67, No. 2, (2006), 189-210.

29. Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C., "Introduction to algorithms, MIT press Cambridge, Vol. 6, (2001).

30. Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M. and Battisti, F., "Image database tid2013: Peculiarities, results and perspectives", *Signal Processing: Image Communication*, Vol. 30, (2015), 57-77.

31. Chandler, D.M. and Hemami, S.S., "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images", *Image*

Processing, IEEE Transactions on, Vol. 16, No. 9, (2007), 2284-2298.

32. Larson, E.C. and Chandler, D.M., "Most apparent distortion: Full-reference image quality assessment and the role of strategy", *Journal of Electronic Imaging*, Vol. 19, No. 1, (2010), 011006-011021.

33. Sheikh, H.R., Wang, Z., Cormack, L. and Bovik, A.C., "Live image quality assessment database release 2" (2005).

34. Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M. and Battisti, F., "TID2008-a database for evaluation of full-reference visual quality assessment metrics", *Advances of Modern Radioelectronics*, Vol. 10, No. 4, (2009), 30-45.

35. Group, V.Q.E. "Final report from the video quality experts

group on the validation of objective models of video quality assessment, phase II (FR_TV2)" (2003); Available from: ftp://ftp.its.bldrdoc.gov/dist/ituvidq/Boulder_VQEG_jan_04/VQEG_PhaseII_FRTV_Final_Report_SG9060E.doc.

36. Sheikh, H.R. and Bovik, A.C., "Image information and visual quality", *Image Processing, IEEE Transactions on*, Vol. 15, No. 2, (2006), 430-444.

37. Wu, J., Lin, W., Shi, G. and Liu, A., "Perceptual quality metric with internal generative mechanism", *Image Processing, IEEE Transactions on*, Vol. 22, No. 1, (2013), 43-54.

38. A. Vedaldi, B.F. {VLFEAT}: "An open and portable library of computer vision algorithms" (2008); Available from: \url{http://www.vlfeat.org/}.

A Novel Image Structural Similarity Index Considering Image Content Detectability Using Maximally Stable Extremal Region Descriptor

M. H. Khosravi, H. Hassanpour

Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

P A P E R I N F O	چکیده
	قابل تشخیص بودن محتوای تصویر و حفظ ساختار آن دو مفهوم کاملاً در هم تنیده و معادل در ارزیابی کیفیت تصویر هستند که اهمیت سنجش آنها بر کسی پوشیده نیست. اما بررسی مطالعات این حوزه نشان می‌دهد بیشترین اهمیت به ارزیابی ساختار تصویر داده شده و پژوهش‌های اندکی به بررسی میزان قابل تشخیص بودن محتوای تصویر پرداخته‌اند. ساختار تصویر برای اولین بار در معیار شباهت ساختاری (SSIM) مورد سنجش قرار گرفت، که در آن ساختار تصویر از طریق بررسی واریانس و کوواریانس مقادیر شدت روشنایی در دو تصویر مرجع و تست، ارزیابی می‌شود. به عبارت دیگر این معیار با بکارگیری آماره‌های سطح پایین، تأکید بر همانندی تغییرات شدت روشنایی در دو تصویر تحت آزمون دارد. اما این تعریف حداقلی از ساختار تصویر، در خصوص میزان قابل تشخیص بودن محتوای تصویر بی‌توجه است. در این پژوهش، ما دریافتیم که از طریق ارزیابی میزان وفاداری وضعیت همواری یا عدم همواری نواحی تصویر، می‌توان به درک خوبی از حفظ یا بهم‌ریختگی محتوای تصویر رسید. بنابراین با استفاده از توصیف‌گرهای نواحی حداکثری (MSER) یک معیار همواری کارآمد پیشنهاد و بعنوان یک عامل سنجش میزان وفاداری محتوای تصویر برای تعریف یک معیار جدید ارزیابی کیفیت مورد استفاده قرار گرفت. نتایج حاصل از آزمایشات بر روی پنج بانک تصویر رایج (A57، LIVE، CSIQ، TID2008 و TID2013) نشان می‌دهند که معیار پیشنهادی در ارزیابی کیفیت تصویر عملکرد قابل قبولی دارد و کارآمدی آن بالاتر از بسیاری از معیارهای نوین قرار می‌گیرد. *doi*: 10.5829/idosi.ije.2017.30.02b.03