The Role of Structure and Textural Information in Image Utility and Quality Assessment Tasks

Suiyi Ling▲, Patrick Le Callet▲, and Zitong Yu
Group IPI Lab LS2N, University of Nantes, Nantes, 44300, France
E-mail: suiyi.ling@univ-nantes.fr

Abstract. *The perceptual process of images is hierarchical. Human tends to first perceive global structural information such as shapes of objects and further focus on local regional details such as texture. Furthermore, it is widely believed that structure information plays the most important role in task of utility assessment and quality assessment, especially in new scenarios like free-viewpoint television, where the synthesized views contain geometric distortion around objects. We thus hypothesize that the degradation of structural information in an image is more annoying for human observers than the one of the textures in certain application scenarios. In order to confirm our hypothesis, a bilateral filtering based model (BF-M) is proposed referring to a recent subjective perceptual test. In the proposed model, bilateral filters are first utilized to separate structure from the texture information in images. Afterward, features that capture object properties and features that reflect texture information were extracted from the response and the residual of bilateral filtering separately. A contour, a shape related and a texture based estimator are then proposed with the corresponding extracted features. Finally, the model is designed by leveraging the three estimators according to target tasks. With the task-based model, one can then investigate the role of structure/texture information in certain task by checking the correspondence optimized weights assigned to the estimators. In this paper, the hypothesis and the performance of the BF-M is verified on CU-Nantes database as utility estimator and on SynTEX, IRCCyN/IVC-DIBR databases as quality estimator. Experimental results show that (1) structure information does play greater role in several tasks; (2) the performance of the BF-M is comparable to the state-of-the art utility metrics as well as the quality metrics designed for texture synthesis and views synthesis. It is thus validated that the proposed model can also be applied as a task-based parametric image metric.* © *2018 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.Percept.Imaging.2018.1.1.010501]

1. INTRODUCTION

Human visual system (HVS) tends to perceive global structure first and then fine-grained details of an image. The procedure of processing a scene proceeds from top of the hierarchy to the bottom (global to local) [1]. In another word, the global structure of a visual object within a human observer's effective global span will be comprehended before its local features. It has been pointed out in [1] that the global precedence accelerates several possible advantages including utilization of low-resolution information,

economy of processing resources, and disambiguation of indistinct details. Therefore, it is intuitively appealing to assume that structure information (i.e., edges, contours etc.) plays greater role in tasks like utility assessment where the objectiveness is to evaluate the usefulness instead of perceived quality of a distorted natural image. Because if structure information captured by an imaging system is useful, then the degradation will be tolerant as long as the underlying task is performed reliably. Examples of use cases include prevention of terrorist attack, fire control, emergency services and the military use imaging systems in real-time tactical scenarios for immediate decisions making on how best to respond to an incident [2–4], and so on.

"Visual texture" is usually defined as the portion of an image that is filled with repeated elements and often subject to some randomization in their location, size, orientation, and so on [5]. First, natural texture provides an important source of information of visible surfaces and details [6]. It is thus important for tasks like quality assessment, where texture descriptors were usually utilized as a proxy to quantify blurriness. Second, texture cues in images provide human observers with a potentially rich source of surface and shapes of objects [7]. In the field of quality assessment, distortions on both structure and texture regions affect how human observers judge the quality of an image. For instance, a three-component weighted SSIM (3-SSIM) has been proposed in [8] by assigning different weights to the SSIM scores according to the type of local regions: edge, texture, or smooth area. Recently, as immersive multimedia has developed in leaps and bounds, Free-viewpoint TV (FTV), Virtual Reality (VR), and so on have engaged a great amount of users and become the novel hot topic in the field. Taking FTV as an example, virtual views are commonly generated with Depth-Image-Based Rendering (DIBR) based algorithms. Quality assessment is important for selecting the appropriate view synthesis approach. Different from common images, synthesized views generated based on DIBR algorithms contain artifact mainly around disoccluded regions, including objects shifting, twisted shape of objects, blurriness along edges and even unfilled holes. It can be visually observed that structure related distortion (e.g., geometric distortion) and texture related distortion (e.g., blurriness) affect unequally the process of evaluating the quality of DIBR based synthesized images.

As discussed above, it is obvious that the effect of degradation on structural and texture regions differs with

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2018

010501-1

Jan.-June 2018
Human Vision and Electronic Imaging 2018

tasks. To the best of our knowledge, there is no related work that explores how different the roles of structure and texture information are playing in different tasks. If one knows which information plays larger or vital role in certain task, the task can be accomplished more accurately and efficiently toward a right direction. In this paper, we hence hypothesize that structure information and texture information play different roles in different tasks within different application scenarios.

To verify our hypothesis, a perceptually inspired BF-M is proposed. In the proposed scheme, a bilateral filter is first adopted to extract the structure and texture information separately based on a subjective study of human material perception in [9], i.e., structural features are extracted from the filter response while the texture features are extracted from the residual of the filter. Then, a "NICE" based edge estimator named bilateral Natural Image Contour Evaluation (BI-NICE), a shape related estimator named bilateral Histogram of Oriented Gradients estimator (BI-HOG) and a texture estimator named bilateral Local Radius Index estimator (BI-LRI) are introduced by calculating the dissimilarity between the original and distorted images with extracted features. Finally, the model is designed by leveraging the weights of the three proposed basic estimators to yield the best performance in different tasks. By doing so, one can determine to what extend the disruption of different information in an image affects the procedure of different tasks. The proposed hypothesis and the performance of the model are verified on CU-Nantes database as utility estimator and on SynTEX, IRCCyN/IVC-DIBR database as quality estimator.

Figure 1 is an example explaining the fundamental idea of the proposed bilateral filtering based model (BF-M): (a) By only observing the edge map of the response of bilateral filtering (the fourth column in Fig. 1), it is obvious that one can recognize the shape of the "teddy bear" easily from the first image (i.e., first row in the fourth column), while it is difficult to tell the second one (i.e., second row in the fourth column) is an image of "wood floor." (b) For the third image from the IRCCyN-DIBR database, one can observe not only the geometric distortion around objects but also the blurred regions. Obviously, the former disruption is more annoying considering the fake edges and changes of shape around the girl. (c) For the fourth image from the SynTEX database, one can see that the structure of the stones has been emphasized by comparing the edge maps of the original image (i.e., fourth row and second column) and one of the responses of bilateral filtering (i.e., fourth row and fourth column). Unrelated texture of the stones has been removed after bilateral filtering. It is thus more reasonable to extract structure related features from the response instead of the original image. (d) The last two images in Fig. 1 are from CU-Nantes with different quality. The previous one (i.e., the fifth row) is the reference of the other one (i.e., the sixth row). By checking the last column of these two images (i.e., the residual obtained by subtracting the response of the bilateral filter from the original image), one can see that there is more details/texture information maintained in the residual of the

reference image. An intuitive assumption on the basis of this observation could be that texture plays more important role in higher quality images in certain tasks.

The contribution of this paper is two-fold:

(1) This paper investigates the roles of structure and information in different tasks and presents a model to further explore and verify the weights of them in different tasks. With the proposed model, our hypothesis that structure information is more important in certain tasks has also been validated.

(2) The proposed model can serve as a task-based parametric image metric for different application scenarios. The performance of which has been tested and proven to be comparable to the state-of-the-art metrics in different tasks.

The remainder of the paper has the following organization. The second section introduces our hypothesis and related theoretical foundations of the proposed model. The third section describes the proposed basic estimators and how those estimators are combined into one according to specific application for the verification of our hypothesis. The experimental results are reported and analyzed in the fourth section. Finally, conclusions and future work are presented in the last section.

2. HYPOTHESIS AND THEORETICAL FOUNDATION

As discussed in [10], on one hand, structure information in visual scene provides HVS with more semantic information. Continuous edges/contours of an image could clearly reveal the visual objects inside the image, e.g., peoples. Those structural edges are important to the HVS and should be maintained as much as possible in digital image processing and tasks like object detection; On the other hand, the textures of a scene are usually the surface of objects which can also be the material of the targets, such as texture patterns of clothes on people, grass, sea and buildings' surface. Texture contains details of objects, could thus further augment the objects with more appealing properties, such as fine texture, smooth gray-scale transition and rich color, and make them vivid to human perception. In summary, structure and texture jointly render the users impressive perspective of visual scenes and structural contours provide HVS with most of the semantic information while details are provided by textures. Therefore, we hypothesize that features that focusing on structural properties and feature measuring details play different roles in different applications. To verify this, in the following sub-sections, we first explain the reason why local edge/contour can represent the structure and then further discuss how we separate texture from structure and extract different features separately as followed.

2.1 Local Edges/Contours Reveal Structure

According to [5], the perception of complex visual patterns and objects appears from neural activity as it is transformed through a cascade of areas in the cerebral cortex. Neurons
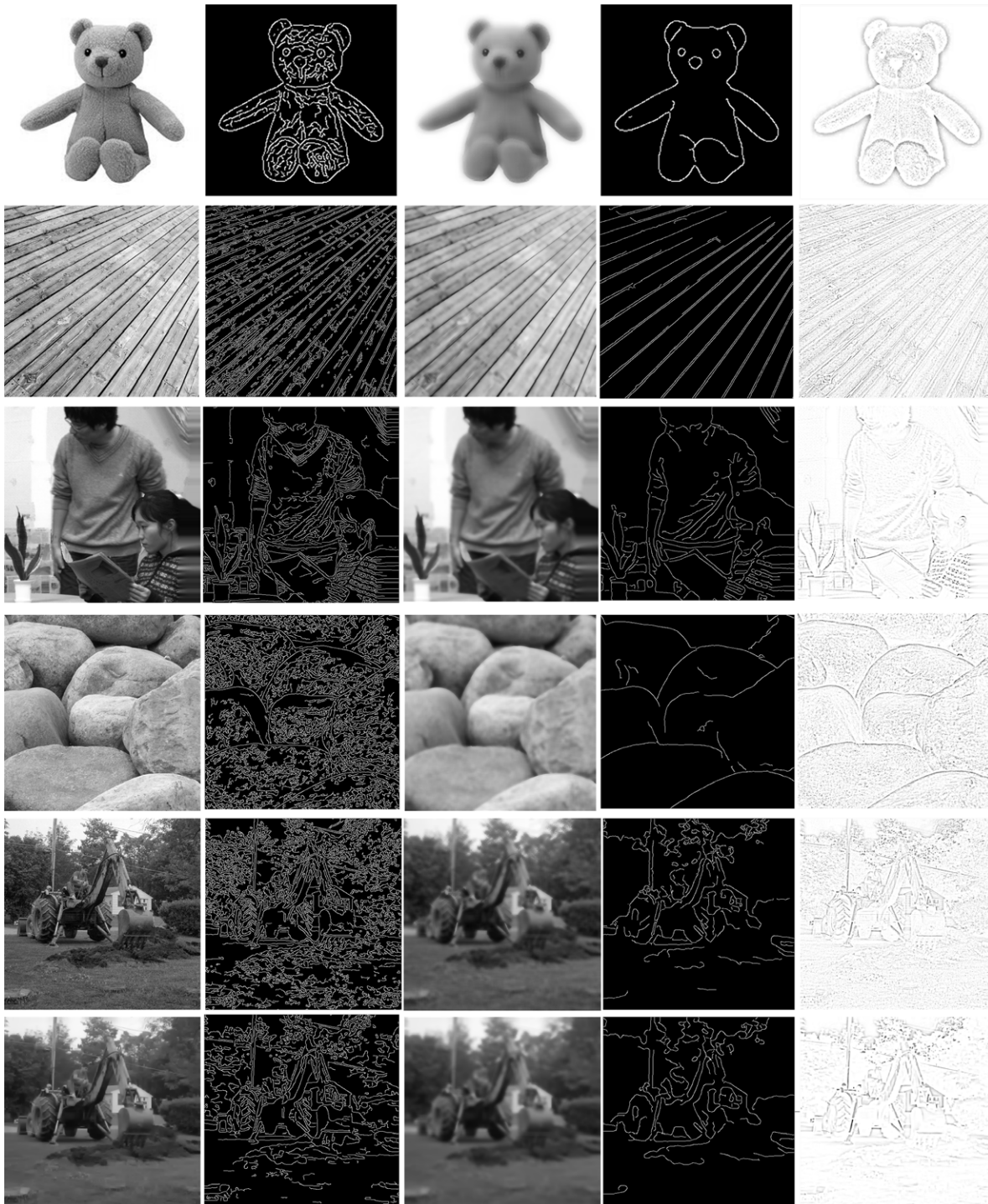
Figure 1. Example of separating structure information from texture information. First column: original image; Second column: edge map of the original image; Third column: response of the bilateral filter on the image; Fourth column: edge map of the response of the bilateral filter; Fifth column: residual of bilateral filtering obtained by subtracting the original image with the respond.

in the primary visual cortex (V1) are selective for local orientation and spatial scale of visual input [11–13]. Downstream regions contain neurons selective for more complex attributes, which is approximately achieved by assembling particular combinations of their upstream afferents. Considering the ubiquity of orientation selectivity in primary visual

cortex [14], it is intuitive to make the assumption that its computational purpose is to represent the local orientation of edges.

Furthermore, over the past decades, the mainstream view in both the biological and computational vision communities is that later stages of processing should

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2018

010501-3

Jan.-June 2018
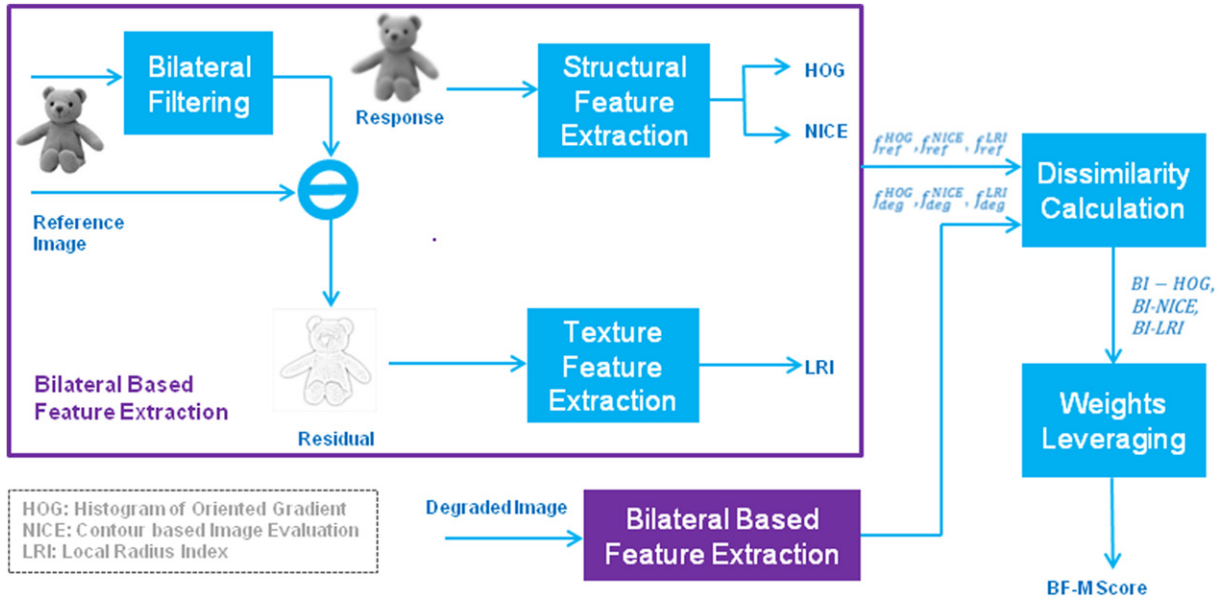Human Vision and Electronic Imaging 2018

Figure 2. Overall framework of the proposed model based on separating structure and texture information using bilateral filtering.

somehow combine these local edge elements to construct more extensive contours, eventually leading to shapes, forms, and objects [15]. Until recently, most researches on object recognition were built around this paradigm, as well as much of the study of mid-level pattern perception, and physiological measurements in areas V2 and V4 of the ventral stream.

It can thus be concluded that local edges and contours, which are local structural information in images, are vital foundation for the following process of higher level semantic structure understanding of images. Edges and contours features are important elements that reveal the structure information of an image. Therefore, in the proposed model, a contour based estimator as well as a histogram of oriented gradient based estimator are designed to quantify the amount of structural change due to disruption. Details of these estimators will be given in the following sections.

2.2 Separating Structure from Texture

Subjective test done in [9] about human material perception provides us with important clues about how to extract structural features and texture features separately. In [9], the subjective experiment is conducted in order to understand which features are useful for the recognition of material categories. In their experiments, images emphasizing local surface information and global structure information were generated separately with bilateral filtering, which is usually used as a non-linear, edge-preserving and noise-reducing smoothing filter for images.

More specifically, Sharan et al. followed the idea of Bae et al. [16] to extract the micro-structure of the surface by smoothing an image with bilateral filtering. Afterward, they utilized the residual image for further texture analysis. The residual image was obtained by subtracting the bilateral filtered results from the gray-scale versions of the original

images to emphasize details of surface structure, which is an operation similar to high-pass filtering. In their subjective test, observers were asked to categorize those distorted images into ten material categories. Based on their results, they concluded that texture is the important attribute of material appearance, while information about surface micro-structure is often related to certain categories, i.e., higher level semantics.

Based on their conclusion, in this paper, bilateral filtering is used as a proxy to separate structure and texture information. In the proposed model, structure features related to shape are extracted from the response of bilateral filtering while texture features were extracted from the residual.

3. THE PROPOSED BF-M MODEL FOR VALIDATING THE PROPOSED HYPOTHESIS

In order to verify the different roles that structure and texture information are playing in different perceptual related tasks, a model on the basis of separating these two pieces of information is proposed and will be introduced in detail in this section. Figure 2 is the overall framework of our proposed model. First and foremost, structure and details related features are extracted separately with bilateral filtering from both the original and the degraded images. More specifically, images are first separated into the base image, i.e., bilateral responses, and the residuals after bilateral filtering [9]. In order to generate the response more efficiently, a faster approximation of bilateral filters [17] is used. The scale σ_s of the spatial kernel and the range value σ_r are set differently according to different tasks [18]. Then, structure related features including Histogram of Oriented Gradients estimator (HOG) and the Natural Image Contour Evaluation estimator (NICE) are calculated with the

base image, while the texture related feature Local Radius Index (LRI) is extracted from the residual image. With the extracted features set f^{HOG}, f^{NICE} and f^{LRI} from both the reference and degraded images, dissimilarity scores are then calculated. After normalization, the three estimators, $BI\text{-}NICE$, $BI\text{-}HOG$, and $BI\text{-}LRI$ are combined with different assigned weights according to different applications. Finally, the roles of different information can be investigated by checking the optimized weights.

3.1 *Bilateral Filtering Based Contour-Based Image Evaluation Estimator (BI-NICE)*

As discussed in the second section that local contours reveal the structure of an image, in this section, a contour based estimator is thus introduced. It has been confirmed that fragments of contours can be used to successfully understand semantics in images [19–21], which further showcases the importance of structure information in semantics related tasks. Since contours are important for global structure understanding, the NICE estimator is improved by using bilateral filter to emphasize important structural local elements. First of all, the edge maps are generated only on the responses of bilateral filtering using the Canny edge detector. For the reference and the degraded images, the obtained contour maps are then denoted as C^{BI} and $\hat{C}^{BI}$ correspondingly.

Before calculating the distance, to probe and expand the shapes contained in the image, the contour maps are subjected to morphological dilation with a 3×3 "plus-sign" shaped structuring element E. In line with the one-scale NICE estimator, the object score was computed by comparing the binary contours map of the reference and the test images. Afterward, the final contour error map is obtained by exerting point-wise exclusive-or (XOR) operation of the dilated binary images, since XOR is the commonly used operation for contours maps comparison. In the end, the overall $BI\text{-}NICE$ score for a test image is defined as

$$BI\text{-}NICE = \frac{d_H(C^{BI} \otimes E, \hat{C}^{BI} \otimes E)}{N_C^{BI}}, \quad (1)$$

where N_C^{BI} is the number of contours elements, $d_H(X, Y)$ denotes the Hamming distance between the X and Y, and $C^{BI} \otimes E$ denotes the dilation operation of the contour map C^{BI} with the morphological structuring element E.

3.2 *Bilateral Filtering Based Histogram of Oriented Gradients Estimator (BI-HOG)*

Considering the fact that HOG [22] is a powerful shape related descriptor used in computer vision and image processing for the purpose of object detection, action recognition and so on, we extract HOG features from each response of bilateral filtering as a higher level structure feature. First, each image is divided into 8×8 cells/blocks. After calculating the histogram of each cell, spatial pooling strategy based on visual importance proposed in [20] is utilized to pool the dissimilarity values. This pooling strategy

was presented based on the perception study that humans tend to perceive "poor" regions in an image more severely than the "good" ones.

Finally, the shape related estimator named bilateral HOG estimator ($BI\text{-}HOG$) is then defined as

$$BI\text{-}HOG = \frac{1}{|b_{ij} \in B_p|} \sum_{b_{ij} \in B_p} D_e(H\text{-}HOG_{ij}^R, H\text{-}HOG_{ij}^D), \quad (2)$$

where $H\text{-}HOG_{ij}^R$ and $H\text{-}HOG_{ij}^D$ denote the histograms correspond to the cells at the i_{th} row and j_{th} column of the bilateral response of both the reference and distorted images. B_p is the lowest 60% of cells ranked by the dissimilarity values. $D_e(X, Y)$ denotes the euclidean distance between the two vectors X and Y.

3.3 *Bilateral Filtering Based Local Radius Index Estimator (BI-LRI)*

To represent detail information in images, texture related features are considered in this section. Different from [9], instead of extracting micro-jet and micro-SIFT features, the LRI [23] texture descriptor is extracted in this paper with a size limit of $K = 4$ and a threshold T equaling to the standard deviation of the image divided by 2. Similar to $BI\text{-}HOG$, LRI texture descriptor is extracted based on 8×8 cells/blocks. After extracting the LRI descriptors from the residual of the bilateral filtering from both of the reference and degraded images, the texture based estimator named bilateral LRI estimator ($BI\text{-}LRI$) is then defined as

$$BI\text{-}LRI = \frac{1}{|b_{ij} \in B_p|} \sum_{b_{ij} \in B_p} D_e(H\text{-}LRI_{ij}^R, H\text{-}LRI_{ij}^D), \quad (3)$$

where $H\text{-}LRI_{ij}^R$ and $H\text{-}LRI_{ij}^D$ denote the LRI feature histograms correspond to the cell at the i_{th} row and j_{th} column of the bilateral residual of both the reference and distorted images.

3.4 *The Final Bilateral based Model*

As discussed in the second section of this paper, different information plays different roles in different application scenarios. Therefore, we combine the three proposed estimators so that the weights of which can be tuned as parameters according to certain applications. The output of each estimator, which is the dissimilarity value calculated based on different features, is normalized to a range of [0, 1]. Finally, the proposed $BF\text{-}M$ model, which can also be utilized as a tasks-based parametric image metric, is designed as

$$BF\text{-}M = 1 - (\alpha \cdot BI\text{-}NICE + \beta \cdot BI\text{-}HOG + \gamma \cdot BI\text{-}LRI)$$
$$s.t. \quad \alpha + \beta + \gamma = 1, \quad (4)$$

where α, β, γ are the aforementioned weights for fine-tuning the roles of the contour, shape, and texture based estimators, respectively and $\alpha + \beta + \gamma = 1$. The configurations of these three estimators are set differently according to the specific task in our experiments and will be further discussed for

Table I. Results summarizing the performance of the parametric metric with different parameters in different quality ranges.

SROCC	Quality Range			
α, β, γ	[1 , 2]	[2 , 3]	[3 , 4]	[4 , 5]
1 ,0 ,0	0.687	0.752	0.659	0.661
0.9,0.1,0	**0.696**	**0.756**	0.719	0.854
0.8, 0.1,0.1	0.681	0.743	**0.737**	0.831
0.7, 0.1,0.2	0.694	0.755	0.728	**0.888**

the purpose of investigating the functionality of different information in images in the following section.

4. RESULTS AND ANALYSIS

To verify the assumption that structure information like edges/contours does not play the same role as detail information like texture in different tasks, the proposed *BF-M* model described in the previous section is served as an utility estimator on CU-Nantes database [24] and as a quality estimator on both SynTex database [25–27] and IRCCyN-DIBR database [28, 29]. With the best-fit weights assigned to *BI-NICE*, *BI-HOG* and *BI-LRI*, the roles of both structure and texture information in the correspondence tasks can be uncovered.

Task descriptions and baselines for utility/quality prediction performance for each use case will be given at the beginning of each following relative sub-sections. Additionally, since the proposed model can also be applied for many tasks by tuning the weights, performance of the metric and related experimental results are also concluded and analyzed in each sub-section. The performance of the model used as task-based parametric metric is evaluated according to Pearson's Correlation Coefficient (PCC), Spearman Rank Order Correlation Coefficient (SROCC), Kendall Rank Order Correlation Coefficient (KROCC), and the Root Mean Squared Error (RMSE).

4.1 *Results: Objective Estimates of Perceived Utility*

In utility assessment task, human observers estimate the usefulness of a natural image as a substitute for a reference. In such a task, structure information is important since the main purpose is to quantify the amount of useful information from an image instead of evaluating its quality. For example, as long as the license plate numbers of vehicles are captured by the surveillance camera, the image is useful for tracking them. More interestingly, according to what have been analyzed in [24] based on the results obtained on the CU-Nantes database, there is a linear relationship between perceived quality score and perceived utility score for images with quality scores under 30, while the one is non-linear for those whose quality scores are higher. It was concluded that observers evaluate very low quality images in terms of the ability to interpret the content. About why the relationship between them is non-linear for higher quality images, one possible explanation could be that texture information plays

different role in task of utility and quality assessment for higher quality images. Because the higher the quality, the more details will be maintained. Disruption of texture, e.g., blurriness, is annoying for human observers while judging the quality of the image. For example, in Fig. 1, the image in the last row is one degraded image while the one in the second last row is its reference image. It can be observed from the last columns of the two rows (i.e., the residual of the correspondence images) that there are more texture information in the residual of the reference image than the one of the degraded images. For the task of quality assessment of high quality images, details are important but may not be the same case in utility assessment. Therefore, we also hypothesize that the roles of structure and texture information in the task of utility assessment vary with the quality.

To affirm the assumption that (1) structure information plays the main role in this task, (2) the role of texture and structure differs in different quality ranges, the proposed *BF-M* model is utilized as the utility estimator and is tested on the CU-Nantes database [24]. The CU-Nantes database consists of 9 reference gray-scale images and 235 distorted images. Each image was degraded by one of the five processes including JPEG compression, blocking, JPEG2000 with dynamic contrast-based quantization, texture smoothing (TS) and texture smoothing with high-pass filtering. To further check how the weights of different information vary with quality, one best configuration will be selected for each sub-interval divided according to the perceived quality score, i.e., the mean opinion score (MOS). To confirm the feasibility of using the proposed model as estimator, ReDLOG [30], most apparent distortion (MAD) [31] metric, multi-scale SSIM (MS-SSIM) [32], the visual information fidelity criterion (VIF) [33], the contours based image evaluation (NICE) [2] metric, the multi-scale version of NICE (MS-NICE) and the multi-scale difference of Gaussian utility (MS-DGU) [34] metrics are chosen to be the compared metrics for utility prediction performance evaluation.

In the experiment, since each sample in the database is labeled not only with the utility score but also with the quality score ranging from 1 to 5, we divide the whole range into quarters and optimize one configuration for each sub-range. Table I illustrates the correlation between objective and subjective scores in different quality interval along with the relative weights configuration. As it can be observed from Table I, for images locate in the quality range of [1,3], the proposed model performs the best with a configuration of $\alpha = 0.9$, $\beta = 0.1$, $\gamma = 0$, while for higher quality images which locate in the range of [3,5], the model performs better with a higher weight for the texture estimator. Overall speaking, it can be concluded that structure plays a vital role in utility assessment, especially for lower quality images. Furthermore, it is also obvious that texture also plays certain role in evaluating the utility of higher quality images. It has been verified that the role of structure and texture information is different among different quality ranges in utility evaluation task.
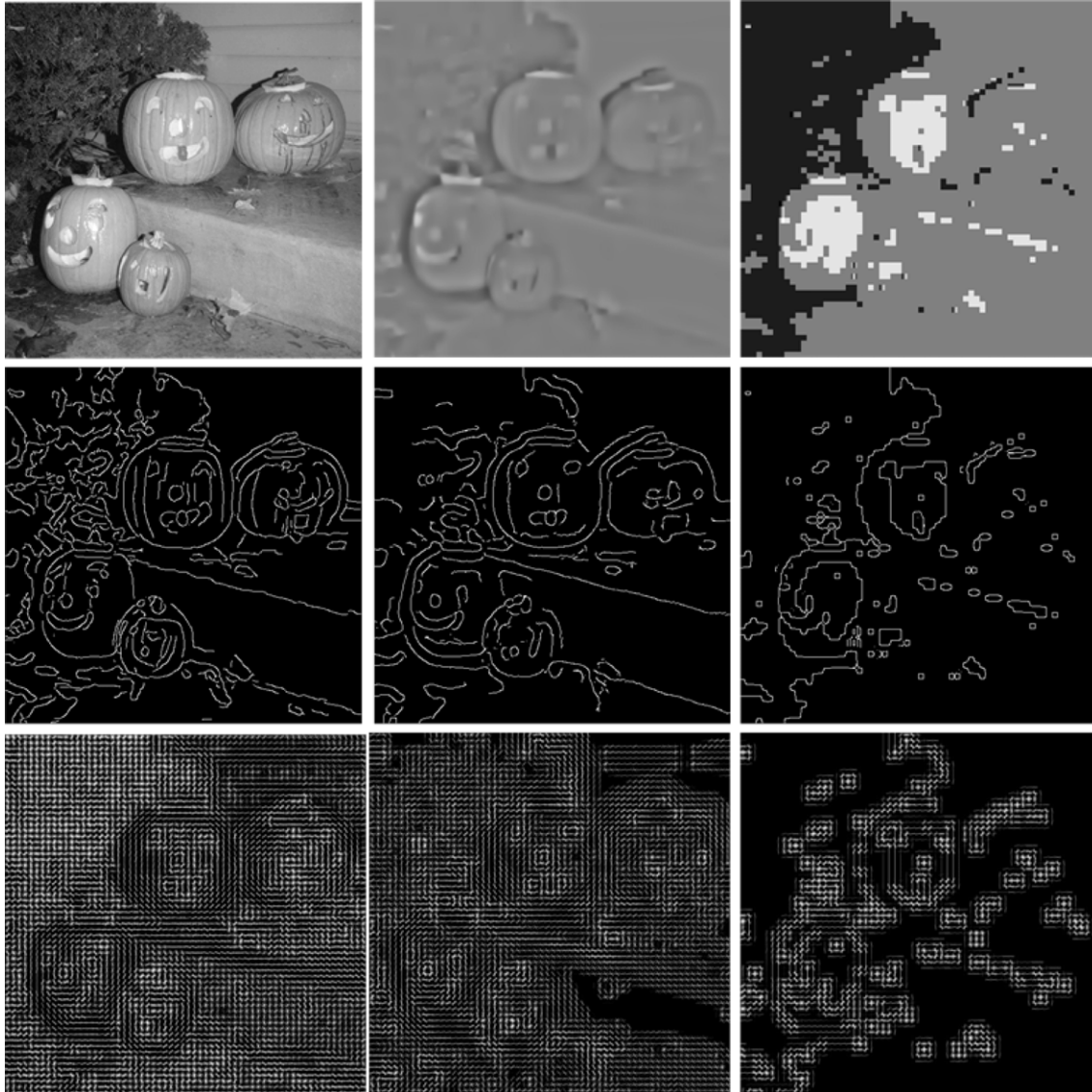
J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2018

010501-6

Jan.-June 2018
Human Vision and Electronic Imaging 2018

Figure 3. Examples explaining why structure related information plays greater role in the task of utility assessment.

For performance evaluation, best weights are selected for the three basic estimators for images with different quality according to Table I. The overall performance of the metrics is concluded in Table II. Among the compared metrics, the proposed *BF-M* performs the best. It is proven that the proposed model is qualified for the task of utility assessment.

To better understand why structure related information is more vital in the case of utility assessment, the edge maps and the extracted HOG descriptors are visualized in the second and third rows of Figure 3. In the figure, the first column is the reference image while the other two are the degraded image and the one in the second column has higher utility score than the third one (10.528 > −47.638). By only observing the edges and HOG maps, one can observe that the shapes of the "pumpkin lanterns" on the floor in the first and second columns are recognizable while the ones in the third column are not. It can be thus concluded that, for low quality

Table II. Results summarizing the performance of various estimators as utility estimator.

	SROCC	KROCC	PCC	RMSE
ReDLOG [30]	0.7757	0.5847	0.7575	39.89
MAD [31]	0.7303	0.5736	0.7241	42.1
MS-SSIM [32]	0.8510	0.6769	0.833	33.8
VIF [33]	0.959	0.821	0.943	12.4
NICE$_{canny}$ [2]	0.937	0.785	0.935	13.3
MS-NICE [2]	0.959	0.821	0.911	15.4
MS-DGU [34]	**0.960**	**0.825**	**0.961**	**10.3**
BF-M	**0.961**	**0.829**	**0.961**	**10.2**

distorted images, where most of the texture information is lost, structure is the most important information for judging its utility.
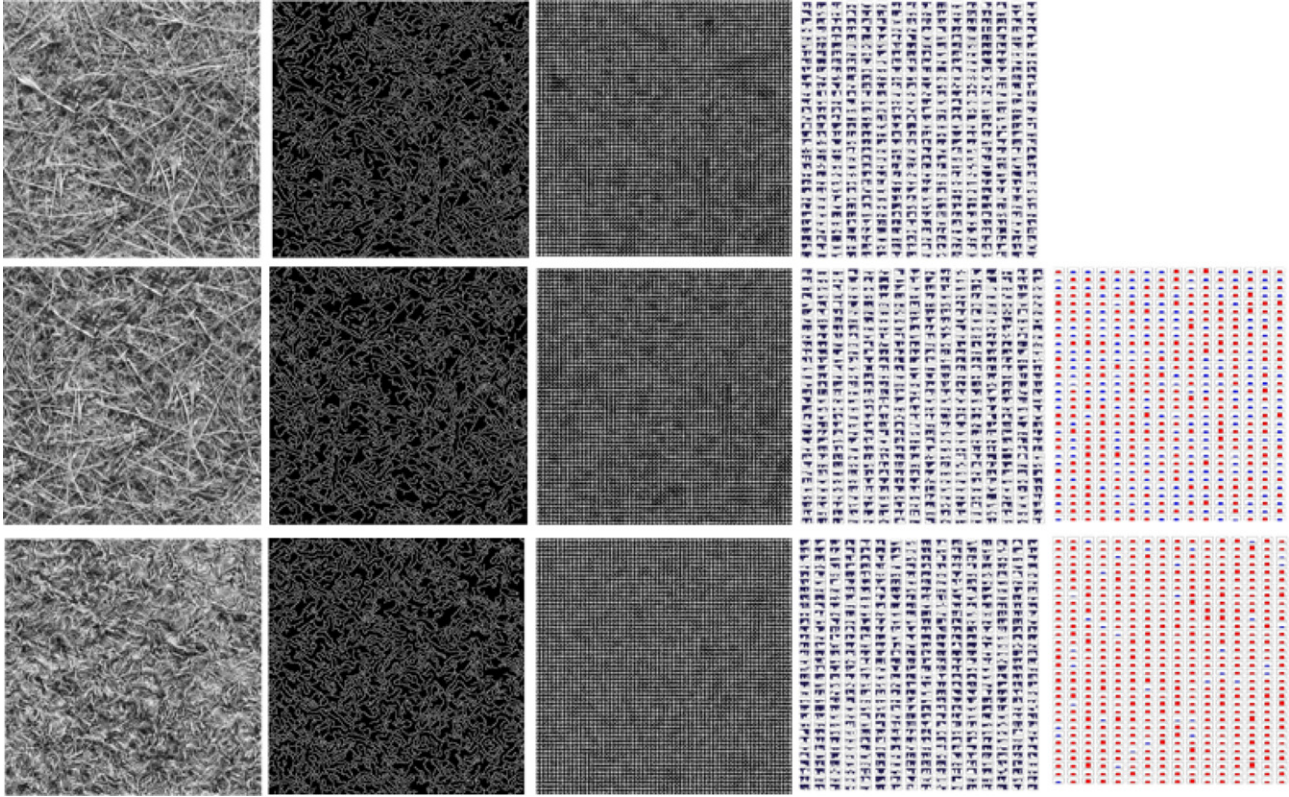
J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2018

010501-7

Jan.-June 2018
Human Vision and Electronic Imaging 2018

Figure 4. Examples explaining why texture related information plays greater role in the task of quality assessment for synthesis texture.

4.2 *Results: Objective Estimates of Perceived Quality for Synthesized Texture Images*

Texture synthesis is a broadly and commonly used technique for bit-rate saving in image, video compression, in-painting (e.g., used for error concealment or disoccluded regions filling for view synthesis in FTV system) and so on. The purpose of quality assessment for texture synthesized images is to estimate the perceived quality of the synthesized texture referring to the original texture in images. Therefore, the role of texture information is definitely more important than one of the structure information in such case.

In verifying what has been discussed above, we test the proposed *BF-M* model on the SynTEX Granularity database [25–27]. This database contains 21 reference textures and 105 synthesized texture images generated with five different texture synthesis algorithms. For *BF-M*, relatives parameters of the model are set as described in the previous section to obtain the most correlated objective scores with the MOS. According to [35], CWSSIM [36], WCWSSIM [37], parametric metric that proposed in [38] and STQA [35] are the four most promising metrics on the SynTEX Granularity database for evaluating the quality of synthesized texture. Therefore, the performance of the proposed model used as an estimator of perceived quality for texture synthesized images is tested on the same database and compared to these four methods.

During the experiment, when $\alpha = 0.2$, $\beta = 0.2$, $\gamma = 0.6$ the proposed model gets the most consistent objective quality

Table III. Results summarizing the performance of various estimators as quality estimator for synthesized texture.

	SROCC	PCC	RMSE
WCWSSIM [37]	0.497	0.546	0.170
CWSSIM [36]	0.644	0.663	0.198
Parametric [38]	0.481	0.412	0.253
STQA [35]	**0.755**	**0.766**	**0.799**
BF-M	0.719	0.708	0.162

score with the subjective one. Since the weights for the texture estimator (i.e., BI-LRI) account for the greatest proportion, we can then draw the conclusion that texture is more important than structure in the task of quality assessment for texture synthesis. In addition, the overall performance of the model applied as quality estimator for texture synthesized images is concluded in Table III. Although *BF-M* does not outperform STQA, the performance is still comparable to the others. This result proves the feasibility of using the proposed model as a quality estimator for texture synthesized images.

For further interpreting why texture information is most important for quality assessment of synthesis texture, we visualized the edges, HOG, LRI maps and the error map between the LRI maps of the reference and the synthesis texture images in the second to fourth columns in Figure 4. In the figure, the first row corresponds to the reference image

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2018

010501-8

Jan.-June 2018
Human Vision and Electronic Imaging 2018

while the second and third rows correspond to synthesis texture images and the second one has higher perceived quality score (4.647 > 1.235). For better observation, when generating the visualized LRI maps, we select a slightly larger block size 16×16 and crop only the top left part of the image. In the visualized LRI map, each sub-figure is an LRI histogram representing the texture information of the local block. LRI is a statistical texture feature that considers inter-edge distance distribution along different angles, i.e., eight directions by comparing the current pixel value to the closest edge pixel value along each direction. The magnitude of each bin in the histogram is decided by the pixel number between the current pixel and the closest edge pixel along the direction and the sign of the bin is decided by comparing the two pixels' value. Therefore, the more saturated the histogram, the smoother the block. By comparing the edges, HOG maps of the synthesis texture images with the ones of the original image, it is almost impossible for human observers to tell the difference between them, let alone asking them to judge which one is better synthesized. On the contrary, the LRI map can provide more clues about the statistical difference between the texture in images. By comparing the error maps calculated using Euclidean distance between the LRI histograms of the original image and the synthesized one, it can be observed that the overall error in the third row is larger than the one in the second row, which is consistent to the perceived quality score (bins in the error map larger than 1.2 are labeled with red color). It can be concluded from this sub-section that texture information is more important when task involves mainly fine-granular texture in the images. Since there is no clear main structure (e.g., boundaries of objects) in these images, details of these images (i.e., the texture) are the dominant factor in the task.

4.3 *Results: Objective Estimates of Perceived Quality for DIBR based Synthesized Views*

Depth-Image-Based-Rendering (DIBR) techniques are indispensable for three-dimensional (3D) video applications including 3D Television (3DTV) and free-viewpoint video. Views that synthesized with DIBR based techniques contain specific distortions like object shifting, incorrect rendering, flickering, blurriness and geometry distortion around disoccluded regions. Since HVS is more sensitive to local severe disruptions than the global consistent ones [39], we hypothesize that structure information plays greater role than texture information during the process of assessing the quality of synthesized views.

Targeting at evaluating the synthesized images' quality properly, several metrics have been proposed to improve common used metrics. In [40], VSQA was proposed to improve SSIM with three visibility maps which help in characterizing complexity of the images. Federica et al. [41] presented the 3DswIM on the basis of statistical features of wavelet sub-bands. Considering the fact that multi-resolution image quality assessment approaches perform better than the single-resolution ones, Dragana [42] first deployed morphological wavelet decomposition for quality

assessment of synthesized images named Morphological Wavelet Peak Signal-to-Noise Ratio metric (MW-PSNR). Later, they devised PSNR with morphological pyramids decomposition (MP-PSNR) instead of morphological wavelet decomposition to obtain better performances [43]. However, none of the aforementioned researches has verified the importance of structure information.

To verify our hypothesis, the proposed model is applied as a quality estimator and is tested on the IRCCyN/IVC-DIBR images database [28, 29]. Images from this database were generated from three multi-view video plus depth sequences including Book Arrival (1024×768, 16 cameras with 6.5 cm spacing), Lovebird1 (1024×768, 12 cameras with 3.5 cm spacing) and Newspaper (1024×768, 9 cameras with 5 cm spacing). Seven DIBR algorithms labelled as A1–A7 [44–49] processed the three sequences to generate four new virtual views for each of them. The database is composed of 84 synthesized views and 12 original frames extracted from the corresponding sequence along with subjective score in the form of MOS. The difference mean opinion score (DMOS) is then calculated measuring the subjective difference between the reference and synthesized images. In [42, 43], images synthesized with A1 are excluded from the experiment due to the significant shifting artifacts compared to the others. However, according to the MOS, images synthesized with A1 have better quality compared to the others, and thus are more similar to advanced synthesized algorithms. Since the main purpose of developing a quality metric is to evaluate the performance of synthesis algorithms, the tested database should be in line with the images/videos synthesized with the state-of-the-art synthesis algorithms to follow the trend. Based on the previous discussion, in our experiments, we include the image set generated by A1 and check the performance on the full IRCCyn/IVC DIBR database. As claimed in [42, 43, 50], MP-PSNR MW-PSNR performed the best among the state-of-the-art metrics designed for synthesized views. According to Dragana et al. [50], PSNR is more consistent with human judgment when calculated at higher morphological decomposition scales. They thus proposed a reduced version of the morphological multi-scale measures, which are reduced MP-PSNR, and reduced MW-PSNR correspondingly, by using only detail images from higher decomposition scales. The reduced versions outperform the full ones. Therefore, in this section, we mainly compare our proposed model with MW-PSNR$_{full}$, MP-PSNR$_{full}$ MW-PSNR$_{reduced}$ and MP-PSNR$_{reduced}$. To obtain the best performance of them, a 5×5 size of SE is used for MP-PSNR and a min Haar wavelet decomposition is used for MW-PSNR as reported in [50].

The overall performance of the metrics is concluded in Table IV. In the experiment, by setting α, β, γ to 0.5, 0.2 and 0.3, the performance of the our model peaks. This configuration indicates the fact that both structure and texture information play roles in evaluating the quality of synthesized views, and one of the structures is greater than the other. In another word, artifact that interferes the structure of the view is more annoying for HVS, which

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2018

010501-9
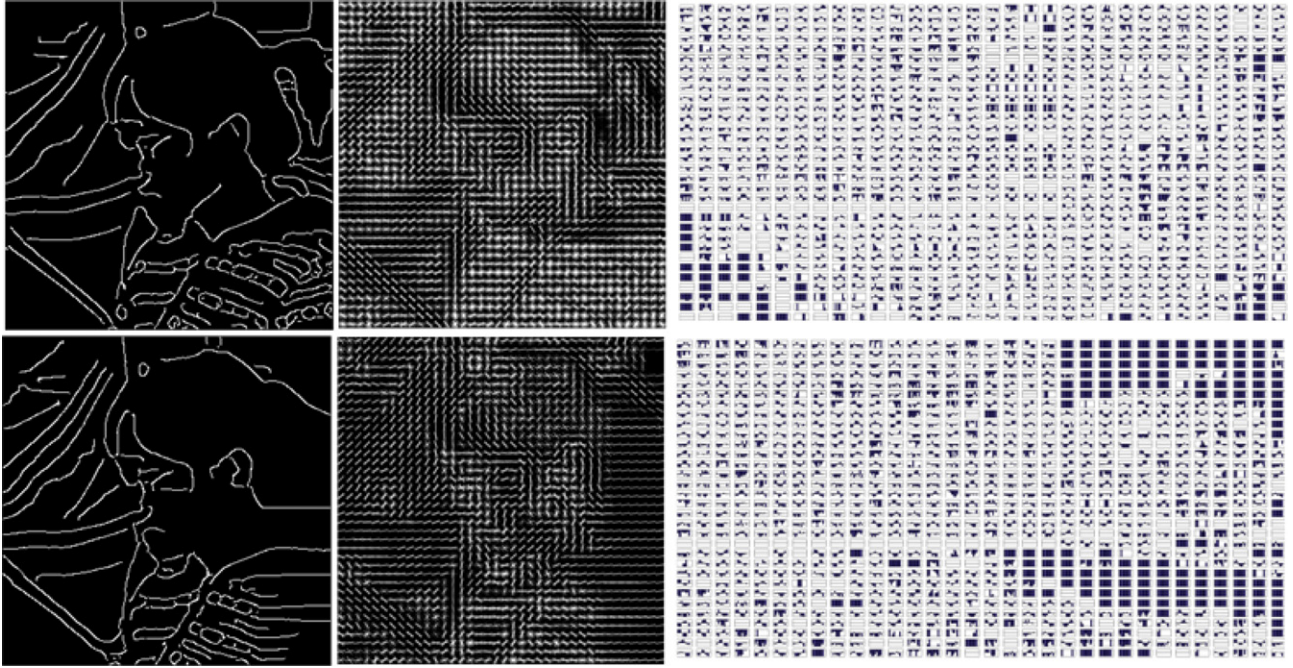
Jan.-June 2018
Human Vision and Electronic Imaging 2018

Figure 5. Examples explaining why both structure and texture information plays considerable role in the task of quality assessment synthesized views.

Table IV. Performance comparison of the proposed metric with state-of-the-art metrics for synthesized views.

	PCC	SROCC	RMSE
ReDLOG [30]	0.1400	0.3361	0.6271
MP-PSNR$_{full}$	0.6553	0.6239	0.5029
MP-PSNR$_{reduced}$	0.6733	**0.66**	0.4923
MW-PSNR$_{full}$	0.6089	0.5738	0.4348
MW-PSNR$_{reduced}$	0.6444	0.6218	0.5091
BF-M	**0.6980**	0.5885	**0.4768**

verifies our previous assumption. Moreover, according to Table IV, the proposed *BF-M* achieves 0.6980 value of PCC, which outperforms all of the compared metrics designed for synthesis images. Compared to the second best performing MP-PSNR$_{redeuce}$ our proposed model obtains a gain of 0.0247 in PCC, which verifies its capability of assessing the perceived quality of synthesized views.

In order to understand how different information loss affects the perceived quality in the scenario of quality evaluation for synthesized views, the edges, HOG and LRI maps of the synthesized image in the third row of Fig. 1 and one of its original images are made visible in Figure 5. By only comparing the edges and HOG map, one can easily notice the geometric distortion around the face of the girl, specially at the right part where the entire regions are blurred. Therefore, it is obvious that structure related information is more important in this task since the deformation of objects'

shapes caused by synthesized algorithms is more eye catchy and can be well captured by structure related descriptors. More interestingly, by comparing the right part of the two LRI maps, one can easily notice the big differences of the histograms in that part. Due to the blurriness introduced by the DIBR algorithms, texture information has been modified, and start become annoying. That is why BI-LRI accounts for 20% in this task.

4.4 *Discussion and Failure Cases*
In summary, the optimized configurations of *BI-NICE*, *BI-HOG*, and *BI-LRI* in tasks of utility assessment, quality assessment for synthesized texture and views are concluded in Figure 6. The setting of the weights is selected according to the performance of the proposed model tested on the Cu-Nantes, SynTEX, and IRCCyN-DIBR database correspondingly. According to the optimized setting, two main conclusions can be made:

(1) Our hypothesis has been verified: It is obvious that structure information does play greater role than texture information in tasks like utility assessment and quality assessment for synthesized views. Nevertheless, in the context of texture synthesis quality assessment, information of details is more important.

(2) In the task of utility assessment, interesting result can be found: the roles of structure and texture information change as the quality of images varies. The fact that texture starts to play greater important role with increasing quality may also be true in the case of normal image quality assessment and will be verified in the future work.
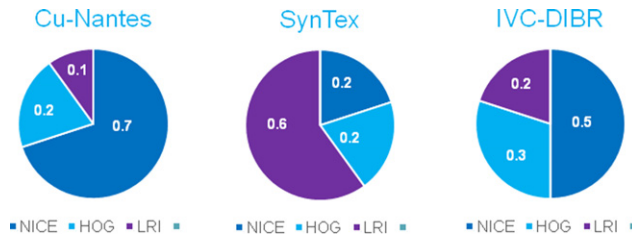
J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2018

010501-10

Jan.-June 2018
Human Vision and Electronic Imaging 2018

Figure 6. Optimized configurations of BI–NICE, BI–HOG, and BI–LRI in tasks of utility assessment, quality assessment for synthesized texture and views.
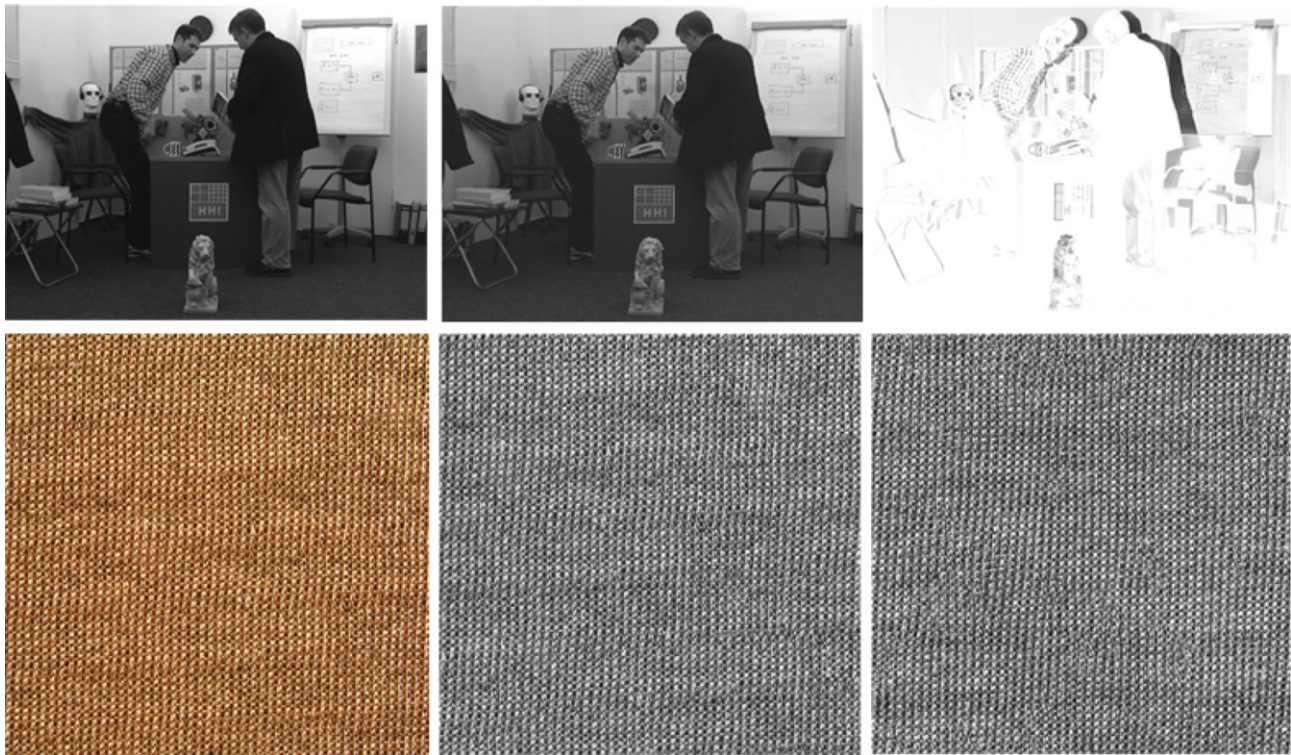

Figure 7. Examples of failure cases that BF–M fails to measure.

Although our model manages to quantify the structural/textural errors for certain cases in different tasks, there are still some cases that our model fails to handle. Examples are shown in Fig. 7: (1) For quality assessment of synthesized views, since BI-LRI and BI-HOG are calculated at block level, the global shifting artifact has been over-penalized to some extend. For example, in the fist row of Fig. 7, the three images are the reference, synthesized image and the error map (the darker the color the more errors there are) between them correspondingly. There is an obvious shift of the object but is not noticeable by only checking the synthesized image. It is obvious that this shifting artifact is durable for human observers compared to the severe local geometric distortions, but it is over-penalized by our model. This can be improved by block matching before dissimilarity calculation; (2) For quality assessment of synthesis texture image, our proposed BI-LRI estimator is not rotation invariant enough, meaning that it is not invariant enough for acceptable rotation of

textures. For example, in the second row of Fig. 7, where the first column is the original image and the rest are two synthesized texture images with slightly different subjective quality score. The characteristic of this "goldWeave" is that it consists of a set of regular small geometric figures. Slight rotation of the geometric figure does not affect significantly the overall perceived quality, but will be over-penalized by our model. In this paper, the LRI descriptor is chosen mainly for its capability of being easily visualized, which makes it easier for analysis. It can be improved significantly by using more advanced texture descriptor, e.g., code-book based model.

4.5 BF-M with other Features or Measures
As the main goal of this paper is to explore the role of structure and texture information in different tasks, simple structure and texture descriptors, which are easier, clearer for visualization, are selected. However, with more

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2018

010501-11

Jan.-June 2018
Human Vision and Electronic Imaging 2018

Table V. Performance of the modified BF-M with more advanced features/measures.

	PCC	SROCC	RMSE
BI-ST	0.8013	0.7556	0.3983
BI-CSF	0.7279	0.6409	0.4565
BF-M	0.6980	0.5885	0.4768
BF-M$_{new}$	0.8105	0.7672	0.3752

powerful structural and texture relative features or measures, the proposed model has the capability to achieve better performance. In order to check how far BF-M can be improved as well as to verify whether the weights will change with different features, in this sub-section, we take quality assessment of synthesized view as an example and replace the simple descriptors with more complicated ones.

More specifically, the mid-level contour descriptor based measure ST-IQA proposed in [51] (other structure related measures like [52, 53] could be utilized too) is used to replace the structure related estimators *BI-NICE* and *BI-HOG* as *BI-ST*, while the spatial contrast sensitivity (CSF) based texture descriptor proposed in [54] is used to replace the texture related estimator *BI-LRI* as *BI-CSF*. Thus, equation (4) can be modified as

$$BF\text{-}M_{new} = 1 - [(\alpha + \beta) \cdot BI\text{-}ST + \gamma \cdot BI\text{-}CSF]$$
$$s.t. \quad \alpha + \beta + \gamma = 1, \tag{5}$$

and the performance of the $BF\text{-}M_{new}$ is summarized in Table V. The optimized performance of $BF\text{-}M_{new}$ is obtained while $(\alpha + \beta)$ is set as 0.8 and γ as 0.2. This best-fit configuration is similar to the optimum configuration of $BF\text{-}M$ as described in the previous section. The conclusion that structural information plays the major role in the task of synthesized views' quality assessment is still solid in this case. As it can be seen from the table, the performance of $BF\text{-}M$ is much better than the one of $BF\text{-}M$ with more advanced structural/texture estimators. It can be thus concluded that, (1) the weights of structural and texture estimators do not fluctuate with different features/measures; (2) the performance of the proposed model can be improved significantly with more advanced structure/textures measures.

5. FUTURE WORK

In the future, to obtain more robust model, subjective tests will be conducted to generate larger database for parameters training. Furthermore, we will try more powerful structure/texture descriptors by incorporating more perceptual factors and explore more use cases with the proposed model including: (1) Select appropriate patches for patch-based training tasks, e.g., patch-based deep learning. (2) Select appropriate images for material recognition training. (3) Separate image/sequence into texture and contours regions as a pre-process stage for object detection. (4) Select reliable

samples of subjective tests from crowd sourcing data. (5) Quality evaluation for higher quality images/videos.

6. CONCLUSION

Human observers tend to perceive global structure first then finer granularity details like texture. Based on which, we hypothesize that structure and texture information plays different roles in different tasks considering the characteristics of the tasks. To validate this assumption, a contour, a coarse-grained structure related, and a texture estimator are first introduced using bilateral filtering. A BF-M is then designed to combine the three estimators differently according to different applications. Experiments are conducted on three different databases for different tasks. The optimized configurations of the model serve as a proxy for checking the roles of structure and texture information in those tasks. According to the experimental results, our hypothesis has been verified and the performance of the proposed model applied as a tasks-based parametric metric is proven to be comparable to the state-of-the-art utility/quality metrics. In the future, more use cases will be explored and tested with the proposed model.

REFERENCES

[1] D. Navon, "Forest before trees: The precedence of global features in visual perception," Cogn. Psychol. **9**, 353–383 (1977).

[2] D. M. Rouse and S. S. Hemami, "Natural image utility assessment using image contours," *16th IEEE Int'l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2009), pp. 2217–2220.

[3] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment (Synthesis Lectures on Image, Video, and Multimedia Processing)* (Morgan Claypool, San Rafael, CA, 2006).

[4] M. I. Leszczuk, I. Stange, and C. Ford, "Determining image quality requirements for recognition tasks in generalized public safety video applications: Definitions, testing, standardization, and current trends," *IEEE Int'l. Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)* (IEEE, Piscataway, NJ, 2011), pp. 1–5.

[5] J. A. Movshon and E. P. Simoncelli, "Representation of naturalistic image structure in the primate visual cortex," *Cold Spring Harbor Symposia on Quantitative Biology* (Cold Spring Harbor, NY, USA, 2014), Vol. 79, pp. 115–122.

[6] J. Aloimonos, "Shape from texture," Biol. Cybern. **58**, 345–360 (1988).

[7] J. R. Kender, "Shape from texture: An aggregation transform that maps a class of textures into surface orientation," *Proc. 6th Int'l. Joint Conf. on Artificial Intelligence* (Morgan Kaufmann Publishers Inc., Burlington, MA, USA, 1979), Vol. 1, pp. 475–480.

[8] C. Li and A. C. Bovik, "Three-component weighted structural similarity index," Proc. SPIE **7242**, 1–9 (2009).

[9] L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adelson, "Recognizing materials using perceptually inspired features," Int. J. Comput. Vis. **103**, 348–371 (2013).

[10] L. Xu, W. Lin, L. Ma, Y. Zhang, Y. Fang, K. N. Ngan, S. Li, and Y. Yan, "Free-energy principle inspired video quality metric and its use in video coding," IEEE Trans. Multimedia **18**, 590–602 (2016).

[11] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," J. Physiol. **160**, 106–154 (1962).

[12] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," J. Physiol. **195**, 215–243 (1968).

[13] S. L. Brincat and C. E. Connor, "Underlying principles of visual shape selectivity in posterior inferotemporal cortex," Nature Neurosci. **7**, 880 (2004).

[14] N. J. Priebe and D. Ferster, "Mechanisms of neuronal computation in mammalian visual cortex," Neuron **75**, 194–208 (2012).

15 M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," Nature Neurosci. **2** (1999).

16 S. Bae, S. Paris, and F. Durand, "Two-scale tone management for photographic look," *ACM Transactions on Graphics (TOG)* (ACM, New York, NY, USA, 2006), Vol. 25, pp. 637–645.

17 S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," Int. J. Comput. Vis. **81**, 24–52 (2009).

18 F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Transactions on Graphics (TOG)* (ACM, New York, NY, USA, 2002), Vol. 21, pp. 257–266.

19 J. Shotton, A. Blake, and R. Cipolla, "Multiscale categorical object recognition using contour fragments," IEEE Trans. Pattern Anal. Mach. Intell. **30**, 1270–1281 (2008).

20 I. Biederman and G. Ju, "Surface versus edge-based determinants of visual recognition," Cogn. Psychol. **20**, 38–64 (1988).

21 J. D. Winter and J. Wagemans, "Contour-based object identification and segmentation: Stimuli, norms and data, and software tools," Behav. Res. Methods Instrum. Comput. **36**, 604–624 (2004).

22 N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2005. CVPR 2005* (IEEE, Piscataway, NJ, 2005), Vol. 1, pp. 886–893.

23 Y. Zhai, D. L. Neuhoff, and T. N. Pappas, "Local radius index-a new texture similarity feature," *IEEE Int'l. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, Piscataway, NJ, 2013), pp. 1434–1438.

24 D. M. Rouse, R. Pepion, S. S. Hemami, and P. Le Callet, "Image utility assessment and a relationship with image quality assessment," Proc. SPIE **7240**, 724010 (2009).

25 S. Varadarajan and L. J. Karam, "A reduced-reference perceptual quality metric for texture synthesis," *IEEE Int'l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2014), pp. 531–535.

26 S. A. Golestaneh, M. M. Subedar, and L. J. Karam, "The effect of texture granularity on texture synthesis quality," Proc. SPIE **9599**, 959912 (2015).

27 S. Varadarajan and L. J. Karam, "A no-reference perceptual texture regularity metric," *IEEE Int'l. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, Piscataway, NJ, 2013), pp. 1894–1898.

28 E. Bosc, R. Pepion, P. Le Callet, M. Koppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, "Towards a new quality metric for 3-d synthesized view assessment," *IEEE Journal of Selected Topics in Signal Processing* (IEEE, Piscataway, NJ, 2011), 5, pp. 1332–1343.

29 'Irccyn ivc dibr database website," http://ftp.ivc.polytech.univnantes.fr/IRCCyN_IVC_DIBR_Images/.

30 S. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of dwt coefficients of locally weighted gradient magnitudes," IEEE Trans. Image Process. **25**, 5293–5303 (2016).

31 E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," J. Electronic Imaging **19**, 011006 (2010).

32 Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," *Conf. Record of the Thirty-Seventh Asilomar Conf. on Signals, Systems and Computers, 2004* (IEEE, Piscataway, NJ, 2003), Vol. 2, pp. 1398–1402.

33 H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Trans. Image Process. **15**, 430–444 (2006).

34 E. T. Scott and S. S. Hemami, "Image utility estimation using difference-of-gaussian scale space," *IEEE Int'l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2016), pp. 101–105.

35 S. A. Golestaneh and L. J. Karam, "Reduced-reference synthesized-texture quality assessment based on multi-scale spatial and statistical texture attributes," *IEEE Int'l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2016), pp. 3783–3786.

36 Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," *Proc. IEEE Int'l. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'05)* (IEEE, Piscataway, NJ, 2005), Vol. 2, pp. ii–573.

37 A. C. Brooks, X. Zhao, and T. N. Pappas, "Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions," IEEE Trans. Image Process. **17**, 1261–1273 (2008).

38 D. Siddalinga Swamy, "*Quality Assessment of Synthesized Textures*," Ph.D. dissertation, (Oklahoma State University, 2011).

39 A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," IEEE J. Sel. Top. Signal Process. **3**, 193–201 (2009).

40 P. H. Conze, "Objective view synthesis quality assessment," Proc. SPIE **8288**, 53 (2012).

41 F. Battisti, E. Bosc, M. Carli, P. L. Callet, and S. Perugia, "Objective image quality assessment of 3d synthesized views," Signal Process, Image Commun. **30**, 78–88 (2015).

42 D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "Dibr synthesized image quality assessment based on morphological wavelets," *Seventh Int'l. Workshop on Quality of Multimedia Experience (QoMEX)* (IEEE, Piscataway, NJ, 2015), pp. 1–6.

43 D. Sandic-Stankovic, D. Kukolj, and P. Le Callet, "Dibr synthesized image quality assessment based on morphological pyramids," *3DTV-Conf.: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)* (IEEE, Piscataway, NJ, 2015), pp. 1–4.

44 C. Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," Proc. SPIE **5291**, 93–104 (2004).

45 A. Telea, "An image inpainting technique based on the fast marching method," J. Graphics Tools **9**, 23–34 (2004).

46 Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3d warping using depth information for ftv," Signal Process. Image Commun. **24**, 65–72 (2009).

47 K. Mueller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View synthesis for advanced 3d video systems," EURASIP J. Image Video Process. **2008**, 1–11 (2009).

48 P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3-d video," IEEE Trans. Multimedia **13**, 453–465 (2011).

49 M. Köppel, P. Ndjiki-Nya, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand, "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," *IEEE 17th Int'l. Conf. on Image Processing* (IEEE, Piscataway, NJ, 2010), pp. 1809–1812.

50 D. Sandić-Stanković, D. Kukolj, and P. L. Callet, "Dibr-synthesized image quality assessment based on morphological multi-scale approach," EURASIP J. Image Video Process. **2017**, 4 (2016).

51 S. Ling and P. Le Callet, "Image quality assessment for free viewpoint video based on mid-level contours feature," *IEEE Int'l. Conf. on Multimedia and Expo (ICME)* (IEEE, Piscataway, NJ, 2017), pp. 79–84.

52 S. Ling, P. Le Callet, and G. Cheung, "Quality assessment for synthesized view based on variable-length context tree," *IEEE 19th Int'l. Workshop on Multimedia Signal Processing (MMSP)* (IEEE, Piscataway, NJ, 2017), pp. 1–6.

53 S. Ling and P. Le Callet, "Image quality assessment for dibr synthesized views using elastic metric," *Proc. ACM on Multimedia Conf.* (ACM, New York, NY, USA, 2017), pp. 1157–1163.

54 Y. Rai, A. Aldahdooh, S. Ling, M. Barkowsky, and P. Le Callet, "Effect of content features on short-term video quality in the visual periphery," *IEEE 18th Int'l. Workshop on Multimedia Signal Processing (MMSP)* (IEEE, Piscataway, NJ, 2016), pp. 1–6.

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2018

010501-13

Jan.-June 2018
Human Vision and Electronic Imaging 2018