A Novel Approach for Computing Quality Map of Visual Information Fidelity Index

Yu Shao, Fuchun Sun, Hongbo Li and Ying Liu

Abstract The visual information fidelity (VIF) index gained widespread popularity as a tool to assess the quality of images and to evaluate the performance of image processing algorithms and systems. But VIF is not a map-based quality metric if its quality map is calculated by traditional sliding window approach. This map-based property is owned by the other quality metrics such as structural similarity (SSIM) and mean-squared error (MSE). In this article, we first construct a novel VIF quality map in pixel domain, which makes VIF become a Minkowski norm of its quality map. Furthermore, we deduce the gradient of VIF by taking the derivative of VIF index with respect to the reference image. The gradient of VIF is easy to calculate and has many useful applications. Experimental results show that the proposed quality map can provide useful guidance on how local image quality is similar to reference image.

Keywords Visual information fidelity · Quality map · Structural similarity · Image quality assessment

1 Introduction

Image quality assessment (IQA) is a fundamental issue in many image processing applications. It can be applied to optimize and design the image processing systems and algorithms to improve the visual quality. For an IQA method, its quality map and gradient are two most important attributes. Quality map measures image quality locally and is able to capture local dissimilarities when compared to reference image. In quality map image, the brighter regions mean better quality.

Y. Shao (✉) · F. Sun · H. Li · Y. Liu
State Key Laboratory of Intelligence Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, People's Republic of China
e-mail: shaoyu2011@foxmail.com

F. Sun et al. (eds.), *Foundations and Applications of Intelligent Systems*, Advances in Intelligent Systems and Computing 213,
DOI: 10.1007/978-3-642-37829-4_14,

Traditionally, MSE and SSIM have been widely used not only for the evaluation of image quality, but also for the design and optimization of signal processing algorithms and systems [1]. That is because their quality map and gradient can be easily computed. For instance, the SSIM index [2, 3] is computed locally at each pixel of the image and can be visualized as an image, often referred to as a SSIM map, which provides useful information on the localization of distortions. The gradient of SSIM with respect to the image has been derived in [4–6], and this gradient is used as a fidelity term in iterative optimization procedures.

The VIF is the most accurate image quality metric according to the performance evaluation of major image quality assessment algorithms performed in [7]. In spite of its high level of accuracy, this index has not been given as much consideration as the SSIM index in a variety of applications, far behind its widespread usage as purely an assessment or comparison tool in other applications. There has been also a growing interest of using VIF as an objective function in optimization problems in a variety of image processing applications [8]. One major problem that could strongly impede the progress of further applications is the lack of understanding and desirable mathematical properties of VIF. For example, the VIF is a nonmap-based quality metric which gives a final score for a distortion image. Unlike the SSIM and MSE which compute a quality (or distortion) map between the reference and distorted images to depict the distribution of quality degradation at image pixels, the overall quality is usually computed as a mean over all the pixels in the distortion map. Second, SSIM or MSE methods can easily calculate its gradient. However, due to the high computational complexity of VIF (6.5 times the computation time of the SSIM index according to [9]) and its nonmap-based quality metric character, it is hard to get the gradient of VIF.

Some researchers are trying to study these two essential properties of VIF. For instance, Seshadrinathan [10] analyzed the properties of SSIM and VIF and established a relationship between SSIM and VIF. Li [11] proposed a spatial information theoretical weighting map for SSIM and VIF. Brighter regions in this weighting map indicate larger weights during error pooling process of IQA. But these studies are only qualitative research. According to the best of our knowledge, there is no complete formula about the quality map and gradient of VIF. In this article, we first propose a novel quality map for VIF in pixel domain. We convert the VIF from a nonmap-based quality metric to a map-based quality metric. Experimental results show that the proposed VIF quality map provides useful guidance on how local image quality is similar to reference image. Based on the formulation of VIF quality map, we then deduce the gradient of VIF by taking the derivative of VIF with respect to reference image. This gradient can be used to solve the minimization problems with VIF term in image processing field.

This paper is organized as follows. The principle of VIF is introduced in Sect. 1. Section 3 presents the proposed quality map of VIF, and its gradients are discussed in Sect. 4. The experiments are analyzed in Sect. 5, and conclusions are drawn in Sect. 6.

2 The VIF Metric

The VIF proposed by Sheikh [9] is an IQA method that consistently outperforms almost all other approaches. It treats the IQA as an information fidelity problem based on natural scene statistics (NSS) theory. There are two types of VIF: wavelet domain version and pixel domain version. Considering that wavelet domain version VIF is more complex and our proposed quality map is based on pixel domain, we only discuss the pixel domain version of VIF.

Support C and D denote the random fields (RFs) from the reference and distorted images, respectively. Let $C^N = (C_1, C_2, \cdots C_N)$ and denote N elements from C, and let $D^N = (D_1, D_2, \cdots D_N)$ be the corresponding N elements from D. C is a product of two stationary RFs that are independent of each other:

$$C = S \cdot U = \{S_k \cdot U_k : k \in I\}, \tag{1}$$

where I denotes the set of spatial indices for the RFs, S is an RFs of positive scalars, and U is a Gaussian scalar RFs with mean zero and variance σ_U^2. The image distortion model is a signal attenuation and additive Gaussian noise, defined as follows:

$$D = GC + V = \{g_k C_k + V_k : k \in I\}, \tag{2}$$

where G is a deterministic scalar attenuation field and V is a stationary additive zero-mean Gaussian noise RFs with variance σ_V^2.

The human visual system (HVS) model in VIF quantifies the impact of the image that flows through HVS:

$$\begin{aligned} E &= C + N, \\ F &= D + N, \end{aligned} \tag{3}$$

where E and F denote the cognitive output of the reference and test images extracted from the brain, respectively; N represents stationary, white Gaussian noise RFs with variance σ_n^2.

VIF utilizes mutual information $I(C_k, E_k)$ to measure the information that can be extracted from the output of HVS when the reference image is being viewed:

$$I(C_k, E_k) = \frac{1}{2}\log_2\left(\frac{\left|s_k^2 C_U + \sigma_N^2 I\right|}{\left|\sigma_N^2 I\right|}\right) = \frac{1}{2}\log_2\left(1 + \frac{\sigma_{C_k}^2}{\sigma_N^2}\right). \tag{4}$$

In addition, information $I(C_k, F_k)$ is measured in the same way when the test image is being viewed:

$$I(C_k, F_k) = \frac{1}{2}\log_2\left(\frac{\left|g_k^2 s_k^2 C_U + (\sigma_N^2 + \sigma_{V_k}^2) I\right|}{\left|(\sigma_N^2 + \sigma_{V_k}^2) I\right|}\right) = \frac{1}{2}\log_2\left(1 + \frac{g_k^2 \sigma_{C_k}^2}{\sigma_N^2 + \sigma_{V_k}^2}\right). \tag{5}$$

The above mutual information assumes that the distortion model parameters g and σ_V^2 are known a priori, but these would need to be estimated in practice. An estimated model replaces the theoretical model in practice. The value of the filed g over block k is denoted as g_k, and the variance of the RFs V over block k is denoted as σ_{V_k}, both are estimated from the local variance of pixels based on maximum likelihood (ML) criteria, which are easily estimated by

$$\hat{g} = \sigma_{CD}/\sigma_C^2, \tag{6}$$

$$\hat{\sigma}_V^2 = \sigma_D^2 - \hat{g}\sigma_{CD}, \tag{7}$$

where

$$\begin{aligned} \mu_C &= w * C, \\ \mu_D &= w * D, \\ \sigma_C^2 &= w * (C - \mu_C)^2 = w * C^2 - \mu_C^2, \\ \sigma_D^2 &= w * (D - \mu_D)^2 = w * D^2 - \mu_D^2, \\ \sigma_{CD} &= w * (C - \mu_C)(D - \mu_D) = w * (CD) - \mu_C\mu_D, \end{aligned} \tag{8}$$

in which w is a symmetric low-pass kernel (e.g., 11×11 normalized Gaussian kernel). '$*$' denotes convolution.

Sheikh [9] uses a more sophisticated vector GSM model for VIF. Rezazadeh [12] uses scalar GSM instead of vector GSM in modeling the images for VIF computation. Rezazadeh [13] shows that the first-level approximation subband of decomposed images plays an important role in improving quality assessment performance and also in complexity reduction. So we restrict our analysis to a scalar version of the VIF metric, where the natural scene model is identical to that used in the scalar IFC (information fidelity criterion) [14] index.

Considering a single subband, we obtain the sample VIF as follows:

$$\mathrm{VIF}(C,D) = \frac{I(C^N,F^N)}{I(C^N,E^N)} = \frac{\sum_k I(C_k,F_k)}{\sum_k I(C_k,E_k)} = \frac{\sum_{k=1}^{N} \log_2\left(1 + \frac{g_k^2 \sigma_{C_k}^2}{\sigma_N^2 + \sigma_V^2}\right)}{\sum_{k=1}^{N} \log_2\left(1 + \frac{\sigma_{C_k}^2}{\sigma_N^2}\right)}. \tag{9}$$

The denominator of the above equation represents the amount of information that the HVS can extract from the original image. The numerator represents the amount of information that the HVS can extract from the distorted image. The ratio of these two quantities hence is a measure of the amount of information in the distorted image relative to the reference image and has been shown to correlate very well with visual quality. The one extra parameter in this model namely the variance of the neural noise σ_N^2 is hand-optimized in [9] and chosen to be 2.

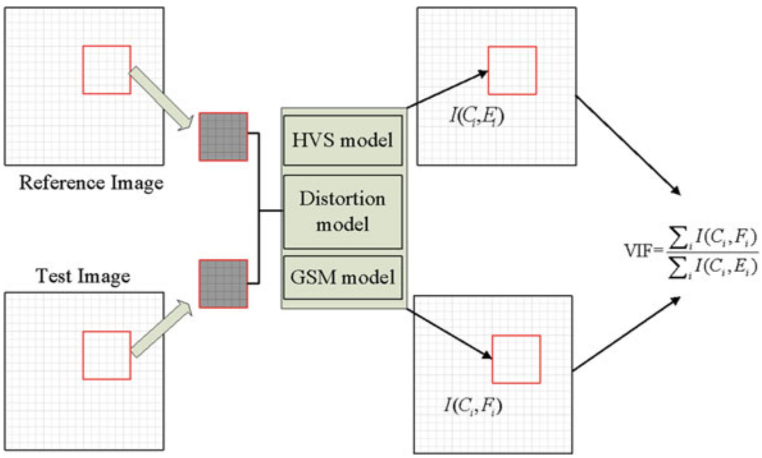

Fig. 1 A schematic of VIF

As depicted in Fig. 1, VIF first decomposes the image into several blocks. Then, VIF measures the visual information by computing mutual information in the different models in each block by Eqs. (4) and (5). Finally, the image quality value is measured by integrating visual information for all the blocks by Eq. (9).

3 Quality Map of VIF

There is no doubt that quality maps are important to the IQA method. Different image quality maps can provide a substantially distinct prediction of local image quality. Although a lot of research effort has been put into investigating the perceptual error map, such as the absolute difference map and the SSIM map, much less has been done for studying the perceptual error map or quality map of VIF. In the former case[15], the VIF is only one number that quantifies the information fidelity for the entire image, whereas in the latter case, a sliding window approach could be used to compute a quality map that could visually illustrate how the visual quality of the test image varies over space. However, in contrast to most previous quality assessment methodologies, the VIF is not a Minkowski norm of the quality map. In other words, a norm of the VIF quality map is not the appropriate measure of image quality.

On the right side of the Eq. (9), the numerator is basically IFC and the denominator can be thought as a content-dependent adjustment. For reference image $\mathbf{X}$ and test image $\mathbf{Y}$, we define the numerator $R(\mathbf{X}, \mathbf{Y}, i, j)$ and denominator $P(\mathbf{X})$ as follows:

$$R(\mathbf{X}, \mathbf{Y}, i, j) = \log_2\left(1 + \frac{g_k^2 \sigma_{X_k}^2}{\sigma_N^2 + \sigma_V^2}\right), \tag{10}$$

$$P(\mathbf{X}) = \sum_{k=1}^{N} \log_2\left(1 + \frac{\sigma_{X_k}^2}{\sigma_N^2}\right) \tag{11}$$

here, N is the number of pixels in either of the input images.

At each point k with its coordinate (i, j), $\mathrm{VIF}_{\mathrm{map}}$ is an indication of the local similarity between reference image $\mathbf{X}$ and test image $\mathbf{Y}$

$$\mathrm{VIF}_{\mathrm{map}}(\mathbf{X}, \mathbf{Y}, i, j) = \frac{R(\mathbf{X}, \mathbf{Y}, i, j)}{P(\mathbf{X})}. \tag{12}$$

The quality map $\mathrm{VIF}_{\mathrm{map}}(\mathbf{X}, \mathbf{Y}; i, j)$ is then added up to obtain a single quality score for the entire image. The VIF for an image $\mathbf{Y}$, with respect to the reference image $\mathbf{X}$, is given by the following equation:

$$\mathrm{VIF}(\mathbf{X}, \mathbf{Y}) = \sum_{\forall i,j} \mathrm{VIF}_{\mathrm{map}}(\mathbf{X}, \mathbf{Y}, i, j). \tag{13}$$

This VIF is a Minkowski norm of its quality map. So we convert the VIF from a nonmap-based quality metric to a map-based quality metric. Though our proposed approach for computing VIF quality map has a form that is more complicated than that of SSIM, it still remains analytically tractable as discussed in subsequent sections. Specifically, we can deduce the deviation of VIF. This makes VIF being able to produce a spatially varying quality map in which quality varies across the image. So the final output of the VIF is either a spatial map showing the image quality at different spatial locations or a single number describing the overall quality of the image.

4 Gradient of VIF

Based on our formulation of VIF and its quality map given above, we first compute ∂ VIF/ $\partial Y(a, b)$ and then $\nabla_{\mathbf{Y}}\mathrm{VIF}(\mathbf{X}, \mathbf{Y})$.

$$\frac{\partial}{\partial Y(a,b)}\mathrm{VIF} = \sum_{\forall i,j} \frac{\partial}{\partial Y(a,b)}\mathrm{VIF}_{\mathrm{map}} \tag{14}$$

in which

$$\frac{\partial}{\partial Y(a,b)}\mathrm{VIF}_{\mathrm{map}} = \frac{\partial \sigma_{XY}}{\partial Y(a,b)} \frac{\partial \mathrm{VIF}_{\mathrm{map}}}{\partial \sigma_{XY}} + \frac{\partial \sigma_Y^2}{\partial Y(a,b)} \frac{\partial \mathrm{VIF}_{\mathrm{map}}}{\partial \sigma_Y^2}. \tag{15}$$

By calculating partial derivatives of the parameters defined in Eq. (8) with respect to $Y(a, b)$, we have

$$\frac{\partial \sigma_{XY}}{\partial Y(a,b)} = \omega(i-a, j-b)(X(a,b) - \mu_X), \tag{16}$$

$$\frac{\partial \sigma_Y^2}{\partial Y(a,b)} = 2\omega(i-a, j-b)(Y(a,b) - \mu_Y). \tag{17}$$

By substituting the partial derivates in Eq. (15) and collecting $\omega(i-a, j-b)$, the summation in Eq. (14) turns into a weighted sum of three convolutions:

$$\nabla_{\mathbf{Y}} \mathrm{VIF}(\mathbf{X}, \mathbf{Y}) = w * \mathbf{M_1} + \left(w * \frac{\partial \mathrm{VIF_{map}}}{\partial \sigma_{XY}} \right) \mathbf{X} + \left(w * \frac{\partial \mathrm{VIF_{map}}}{\partial \sigma_Y^2} \right) \mathbf{Y} \tag{18}$$

where the simplified auxiliary variable $\mathbf{M_1}$ is

$$\mathbf{M_1} = -\mu_X \frac{\partial \mathrm{VIF_{map}}}{\partial \sigma_{XY}} - 2\mu_Y \frac{\partial \mathrm{VIF_{map}}}{\partial \sigma_Y^2}. \tag{19}$$

Using Eq. (18), we can compute $\nabla_{\mathbf{Y}} \mathrm{VIF}(\mathbf{X}, \mathbf{Y})$ for all pixels with just three convolutions and some element-wise multiplications and additions. The partial derivatives required in Eq. (19) are given below:

$$\frac{\partial \mathrm{VIF_{map}}}{\partial \sigma_{XY}} = \frac{1}{P} \left(1 + \frac{g^2 \sigma_X^2}{\sigma_V^2 + \sigma_n^2} \right)^{-1} \frac{2g(\sigma_V^2 + \sigma_n^2)\sigma_X^2 + 2g^2 \sigma_{XY}}{(\sigma_V^2 + \sigma_n^2)^2} \tag{20}$$

$$\frac{\partial \mathrm{VIF_{map}}}{\partial \sigma_Y^2} = \frac{1}{P} \left(1 + \frac{g^2 \sigma_X^2}{\sigma_V^2 + \sigma_n^2} \right)^{-1} \frac{-g^2 \sigma_X^2}{(\sigma_V^2 + \sigma_n^2)^2} \tag{21}$$

in which P is defined in Eq. (11).

Since VIF can be employed as the data-fidelity term of optimization function in some image processing field, if having the gradient of VIF, the gradient descent approach can be used with an iterative procedure to solve the optimization problem.

5 Experimental Results

Like SSIM map, our proposed VIF map is computed locally at each pixel of the distorted image and can be visualized as an image, which provides useful information on the location of distortions. To validate the proposed VIF map model, we

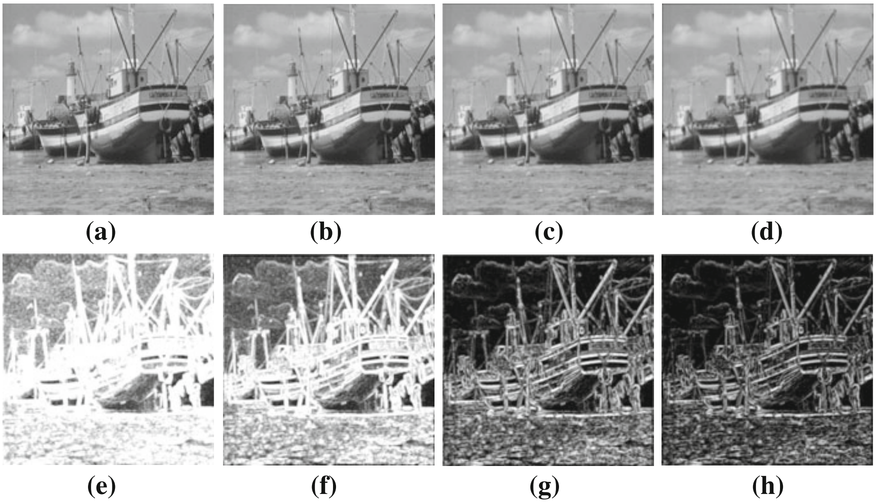

Fig. 2 Illustration the VIF map of distorted images with different blur levels. **a** Original image(VIF = 1). **b** Blured image,Gaussian blur kernelsize = 7 × 7 and $\sigma = 0.5$ (VIF = 0.674). **c** Blured image,Gaussian blur kernelsize = 7 × 7 and $\sigma = 1$ (VIF = 0.293). **d** Blured image,Gaussian blur kernelsize = 7 × 7 and $\sigma = 1.5$ (VIF = 0.195). **e** VIF map of (**a**). **f** VIF map of (**b**). **g** VIF map of (**c**). **h** VIF map of (**d**)

first test its performance using two types of image distortion: Gaussian blur and additive white Gaussian noise.

Figure 2 illustrates the VIF map of distorted images with different Gaussian blur levels. First row shows the reference image (a) and distorted images (b–d) obtained from the reference using Gaussian blur with an incremental variance. Second row shows the corresponding VIF maps at each pixel displayed as an image. Figure 3 illustrates the VIF map of distorted images with different noise levels. First row shows the reference image (a) and distorted images (b–d) obtained from the reference using additive, white Gaussian noise with an incremental noise level. Second row shows the corresponding VIF maps. In the VIF map image, bright regions correspond to better quality and dark regions correspond to worst quality. From Figs. 2 and 3, we obviously see that with noise or blur level increase, the visual quality of distortion image deteriorates gradually, and its quality map brightness also reduces gradually. The proposed VIF map clearly displays the regions of the distorted image that are visually annoying to the human observer.

Figure 4 shows a reference image that has been distorted with three different types of distortion and its corresponding VIF map. The distortion types illustrated are contrast stretch, Gaussian blur, and JPEG compression. VIF map of Fig. 4e–h shows the spread of structural information. In flat image regions such as the sky area, the information content of the image is low, whereas in textured regions and regions containing strong edges such as the outline of buildings, the image quality is high. The contrast-enhanced image Fig. 4b has a brighter quality map than

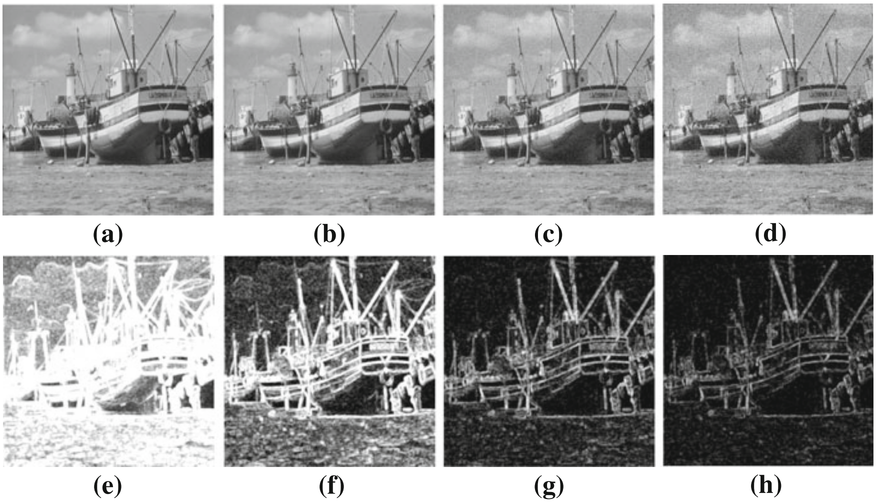

Fig. 3 Illustration the VIF map of distorted images with different noise levels. **a** Original image(VIF = 1). **b** Noised image,Gaussian whitenoise with $\sigma = 5$ (VIF = 0.442). **c** Noised image,Gaussian whitenoise with $\sigma = 15$ (VIF = 0.156). **d** Noised image,Gaussian whitenoise with $\sigma = 25$ (VIF = 0.083). **e** VIF map of (**a**). **f** VIF map of (**b**). **g** VIF map of (**c**). **h** VIF map of (**d**)

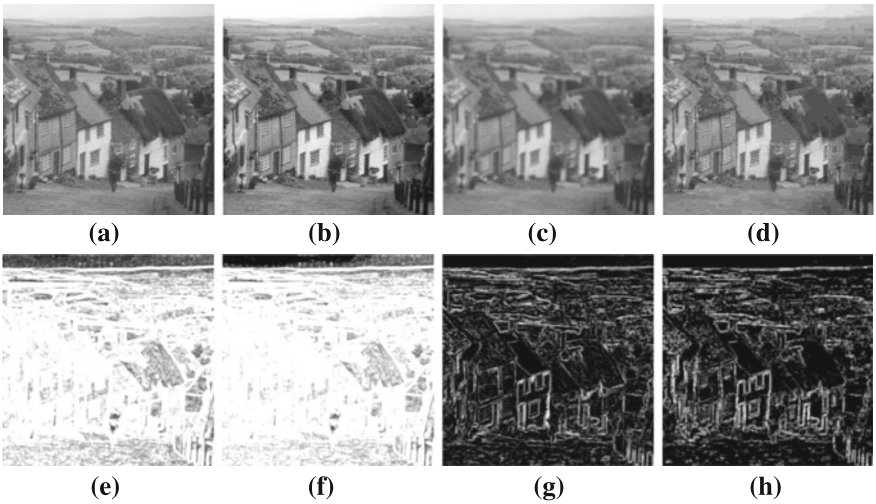

Fig. 4 The VIF map can capture the quality loss or improvement in the distorted image. **a** reference image (VIF = 1), **b** distorted image by contrast stretch (VIF = 1.1), **c** distorted image by Gaussian Blur (VIF = 0.07), (**d**) distorted image by JPEG compression (VIF = 0.10), **e**–**h** are corresponding VIF map of image **a**–**d**.

reference image Fig. 4a, and its VIF value is larger than unity. In contrast, both the blurred image Fig. 4c and the JPEG-compressed image Fig. 4d have a darker quality map compared with the reference image, and its VIF value is smaller than unity. It is interesting to see that the brightness of VIF map reflects its VIF value. It indicates the relative image information that is present in the distorted image. So this proposed VIF map captures the improvement or loss of the distorted image in visual quality and appears to be a good indicator of image quality in the pixel domain.

6 Conclusion

The goal of this paper was to analyze the properties of recently most widely used IQA paradigm VIF based on information theoretical framework. We first showed that the numerator term in VIF, which is the mutual information between the test image and the reference image, can be interpreted as a quality map for VIF. The experimental results indicate that the proposed VIF map provides useful guidance on how local image quality is similar to reference image. Additionally, VIF map can also predict improvement in quality over space. Based on the formulation of VIF map, we then deduce the gradient of VIF by taking the derivative of VIF with respect to reference image. We pointed out that this gradient can be used to solve optimization problem where there exists VIF term. This VIF value is the sum of VIF map. In the future, we would like to use other pool strategies to compute the overall VIF value and extend our analysis to the muti-subband/vector VIF model.

Acknowledgments This study was jointly funded by the National Basic Research Program of China (Grant No:2012CB821206) and the Tsinghua Self-innovation Project (Grant No:20111081111).

References

1. Wang Z, Bovik AC (2009) Mean squared error: love it or leave it? a new look at signal fidelity measures. IEEE Sign Process Mag 26:98–117
2. Wang Z, Simoncelli EP (2004) Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics. Human vision and electronic imaging IX. Proc SPIE 5292:99–108
3. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Proc 13:600–612
4. Avanaki AN (2009) Exact global histogram specification optimize for structural similarity. Opt Rev 16:613–C621
5. Rehman A, Rostami M, Wang Z, Brunet D, Vrscay ER (2012) SSIM-inspired image restoration using sparse representation. EURASIP J Adv Signal Proc 2012:16
6. Wang Z, Simoncelli EP (2008) Maximum differentiation (MAD) competition a methodology for comparing computational models of perceptual quantities. J Vision 8 8:1–13

7. Sheikh HR, Sabir MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans Image Proc 15:3440–3451
8. Seshadrinathan K, Bovik AC (2007) New vistas in image and video quality assessment. Proc SPIE 6492:649202
9. Sheikh HR, Bovik AC (2006) Image information and visual quality. IEEE Trans Image Proc 15:430–444
10. Seshadrinathan K, Bovik AC (2008) Unifying analysis of full reference image quality assessment. In: 15th IEEE international conference on image processing, San Diego, CA, United states pp 1200–1203
11. Li Q (2009) Objective image and video quality assessment with applications. PhD thesis, The University of Texas at Arlington
12. Rezazadeh S, Coulombe S (2011) A novel discrete wavelet transform framework for full reference image quality assessment. Sign Image Video Proc pp 1–15
13. Rezazadeh S, Coulombe S (2010) Low-complexity computation of visual information fidelity in the discrete wavelet domain. In: IEEE international conference on acoustics speech and signal processing (ICASSP) 2010:2438–2441
14. Sheikh HR, Bovik AC, de Veciana G (2005) An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Trans Image Proc 14:2117–2128
15. Sheikh HR (2004) Image quality assessment using natural scene statistics. PhD thesis, The University of Texas at Austin