An Attempt to Combine Structure and Color for Assessing Image Quality

Yunyu Shi, Youdong Ding, Ranran Zhang, Jun Li
School of Computer Engineering and Science
Shanghai University
Shanghai, China
newmelody, ydding, whitezrr, lijunlijun@shu.edu.cn

Abstract

The attempt to combine image structure and color information together for improving objective image quality assessment performance is introduced in this paper. Through analyzing the respective limitation of both SSIM, which focus on image structure, and CD-PSNR, which is based on S-CIELAB and works with image color only, the additive model (SAC) and the multiplicative model (SMC) are defined respectively. Our experimental results show that the proposed combined models all perform better than single method, and we could have different choices considering different purpose and the tradeoff between complexity and efficiency.

1. Introduction

Image quality assessment (IQA) has been developed rapidly in the last ten years. Generally speaking, subjective measures by human beings, such as Mean Opinion Score (MOS), are most accurate since human eyes ultimately receive processed digital images. But it's time-consuming and expensive for real applications. So objective metrics are designed to quantify image quality efficiently and automatically in agree with subjective judgment. Currently, there are three criteria to classify objective image quality measures[1]:

1.the availability of an original image: objective methods belong to three types, full reference (FR), reduced reference (RR), or no reference (NR).

2.the application scopes: calculation models are designed for general purpose or specific applications.

3.the philosophy used in constructing Human Visual System (HVS): down-up methods simulate almost all function modules of HVS and then combine them together as a unified model, while top-down methods are based on the conjecture that the purpose of HVS is to efficiently extract and make use of the information represented in natural scenes.

[2] evaluates the performance of several prominent full reference image quality assessment algorithms, among which Multi-scale Structure Similarity (SSIM-MS) method[3] and Visual Information Fidelity (VIF)[4] have much better performance by being tested on the subjective database[5], which includes 5 types of distortion, 779 distorted images and corresponding human judgment scores. Two of them are top-down assumptions about HVS. Comparatively, a lot of improvement models are based on SSIM, because of its almost the simplest formulations among all existing image quality metrics and high accuracy, although its performance is a little poorer than VIF, which is from the view point of information theory and accordingly complex.

SSIM woks with luminance only, but sometimes color could represent image contents and influence humans' judgments. So in this paper, we attempt to combine structure and color information together for evaluating image quality. Our experimental results indicate that this attempt is reasonable and efficient.

2. Image quality assessment methods

2.1. SSIM algorithm

SSIM[6], the more recent sophistication of Wang-Bovik Index[7], is under the assumption that natural image signals are highly "structured", human visual perception is highly adapted for extracting structural information from a scene, so a measurement of structural similarity (or distortion) should provide a good approximation to perceptual image quality[1, 6, 7].

Suppose x and y are two non-negative image signals, which have been aligned with each other. If we consider one of the signals to have perfect quality, then the similarity measure can serve as a quantitative measurement of the quality of the second signal. Three components are combined to yield an overall similarity measure: luminance,

978-0-7695-3565-4/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDIP.2009.56

contrast and structure comparison, which are defined as follows, respectively[6]:

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

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

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

where μ_x,σ_x,σ_{xy} are the mean of x, the variance of x, and the covariance of x and y, respectively. C_1,C_2 and C_3 are small constants.

The general form of SSIM index between signal x and y is defined as:

$$SSIM = [l(x,y)]^\alpha[c(x,y)]^\beta[s(x,y)]^\gamma, \tag{4}$$

usually,$\alpha = \beta = \gamma = 1$.

SSIM indexing algorithm is applied for image quality assessment using a sliding window approach. The window moves pixel-by-pixel across the whole image space. At each step, the SSIM index is calculated within the local window. In order to reduce "blocking" artifacts, [6] uses an 11×11 circular-symmetric Gaussian weighting function $\omega = \{\omega_i | i = 1, 2, \cdots, N\}$, with standard deviation of 1.5 samples, normalized to unit sum ($\sum_{i=1}^{N} \omega_i = 1$).

Finally, a mean SSIM (MSSIM) index is used to evaluate the overall image quality:

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j), \tag{5}$$

where X and Y are the reference and the distorted images, respectively; x_j and y_j are the image contents at the j-th local window; and M is the number of local windows in the image. Better image quality is indeicated by bigger MSSIM value, which satisfies three conditions[6]:

1.Symmetry:$f(x,y) = f(y,x)$;

2.Boundedness: $f(x,y) \leq 1$;

3.Unique maximum: $f(x,y) = 1$ if and only if $x = y$ (in discrete representations, $x_i = y_i$ for all $i = 1, 2, \cdots, N$).

2.2. CD-PSNR based on S-CIELAB

Color is a powerful descriptor that often simplifies object identification and extraction from a scene, and humans can discern thousands of color shades and intensities, compared to about only two dozen shades of gray[8]. Color, as one of the most important features used in image analysis, is important for human beings to judge image quality to some extent.

S-CIELAB[9, 10], which measures how accurate the reproduction of a color is to the original when viewed by a human observer, is suitable for quality assessment application. Actually, S-CIELAB, an extension of the CIE L*a*b* DeltaE color difference formula, is a "perceptual color fidelity" metric. It incorporates the human spatial-color sensitivity, and is one of most advanced models for color difference calculation in the area of colourometry. S-CIELAB has been widely applied in real-world applications[9]. Huang etc. proposed color difference mean square error (CD-MSE) and color difference peak signal-to-noise ratio (CD-PSNR)[11], which are defined similarly as follows, respectively:

$$CD-MSE = \frac{1}{N}\sum_{i=1}^{N}[\Delta E_i]^2, \tag{6}$$

$$CD-PSNR = 10\log_{10}\frac{100^2}{CD-MSE}, \tag{7}$$

where N is the number of image samples(pixels), ΔE_i^2 is the color difference value of i-th samples between two images, and the value 100 in Eq. (7) is the existed biggest color difference. CD-PSNR values are all bigger than zero, and bigger value indicates better quality. The experimental results in [11] show good consistency between CD-PSNR and subjective evaluation.

2.3. Limitations

SSIM works with image luminance only and is not very sensitive to color changes, while CD-PSNR works with color only, and doesn't refer to structured relationship among neighboring signals. The example, demonstrated in Figure 1, reflects the limitations of the two methods respectively. (a) is the reference image (assumed to have perfect quality) from LIVE Subjective database, which is available at [5]. (b) and (d) are color-changed versions of (a), especially (b) is of gray scale. (c) and (e) are Gaussian noise contaminated images also from the above database.

SSIM index gives better quality of (b) than (c). Compared to (a), we generally judge (c) is of better quality than (b) because of almost the same details of (c) and severe chromatic difference of (b). CD-PSNR values indicate that (e) is better than (d), which is also not consistent with our perceptual judgment.

3. Combined models and experimental results

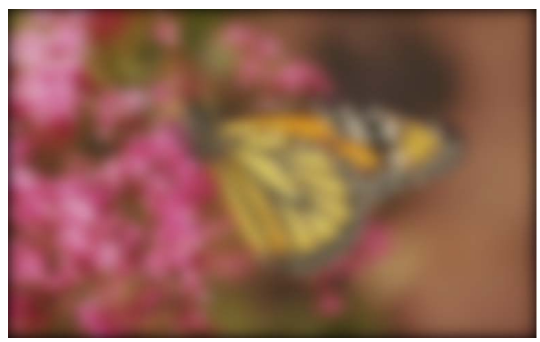
As mentioned above, SSIM, which focus on image structure and CD-PSNR, which works with color have limitations respectively. Because structure and color information

(a) monarch.bmp

(b) grayscale version of (a)
SSIM=0.97113

(c) Gaussian noise contamination
SSIM=0.87857

(d) color change version of (a)
CD−PSNR=11.965



(e) Gaussian noise contamination
CD−PSNR=12.832

Figure 1. Limitation examples of SSIM and CD-PSNR respectively.

are all important for human judgment, we attempt to combine them together to design a unified quality assessment model, which could be defined as follow:

$$SC(X,Y) = f(S(X,Y), C(X,Y)), \tag{8}$$

where $S(X,Y)$ is the structural similarity index drawn idea from SSIM method (Section 2.1), $C(X,Y)$ is the color difference calculation based on S-CIELAB.

Determining how to combine the two components is hard. Generally speaking, we could add or multiple them from the view point of mathematics because of their similar trend of values' growth, and could examine if the model is efficient through both theoretical analysis and experimental results.

Both SSIM and CD-PSNR methods are respectively tested on the LIVE Subjective database, of which the five types of distortion are jpeg2000 compression (JPEG2K for short), jpeg compression (JPEG), white noise (WN), Gaussian blur (GBLUR) and Rayleigh fast fading channel (FF). The performance metric is the linear correlation coefficient (CC) between DMOS (Different MOS) and algorithm scores after non-linear regression. The nonlinearity chosen for regression is a 5-parameter logistic function[2], as defined in Eq. (9) and Eq. (10).

$$Quality(x) = \beta_1 logistic(\beta_2, (x - \beta_3)) + \beta_4 x + \beta_5, \quad (9)$$

$$logistic(\tau, x) = \frac{1}{2} - \frac{1}{1 + \exp(\tau x)}. \quad (10)$$

The results, shown in Table 1, are reported for different methods on individual datasets as well as on the entire database. From the table, we see that on the entire database, SSIM and CD-PSNR have similar accordance with subjective score. For every individual dataset, SSIM has higher (JPEG2K, FF), or similar (JPGE, WN, GBLUR) accordance. The scatter plots are shown in Figure 2 (a) and (b).

3.1. Additive model(SAC)

SSIM and CD-PSNR values belong to different ranges. We preprocessed CD-PSNR values and restricted them in the rang $(0, 1]$ in order to satisfy three conditions above as SSIM, and the additive model is defined as:

$$\begin{aligned} SAC(X,Y) &= f(S(X,Y), C(X,Y)) \\ &= \frac{\alpha S(X,Y) + \beta C(X,Y)}{\alpha + \beta}, \end{aligned} \quad (11)$$

where α, β are the relative importance of the two components. We found that the performance is best when $a = 0.7$,$b = 1.0$, by applying least square method and computing CC between DMOS and quality score after non-linear regression. The results and scatter plots are shown in Table 1, Figure 2 (c) respectively.

3.2. Multiplicative model(SMC)

The multiplicative model (SMC) is defined as below, based on the fact that both two components have similar trend of values' growth:

$$\begin{aligned} SMC(X,Y) &= g(S(X,Y), C(X,Y)) \\ &= [S(X,Y)]^{\mu}[C(X,Y)]^{\nu}, \end{aligned} \quad (12)$$

where μ, ν are the weighted values of the two components. Applying the same method mentioned in Additive model, we determine $\mu = 1.0$,$\nu = 0.4$ to get the best performance. The results and scatter plots are shown in Table 1, Figure 2 (d) respectively.

3.3. Discussions

From Table 1, we see that two combined models have almost the same accordance with subjective score (SMC is a little better). We must notice that SMC model has surprising accuracy on white noise dataset, which belongs to pointwise distortion. Except that, SAC model performs a little better than SMC for JPEG2K, JPEG, GBLUR and FF distortion. Two combined models all perform better than single SSIM and CD-PSNR methods on the entire database. For general purpose, we prefer SAC model because it ranges over much more distortion types. SSIM may be chosen for FF and SMC for WN distortion during specific applications.

Table 1. Performance comparison of different models on LIVE Subjective database.

	SSIM	CD-PSNR	SAC	SMC
JPEG2K	0.97074	0.94712	**0.97170**	0.96771
JPEG	0.97511	0.97183	**0.98313**	0.97916
WN	0.98259	0.98707	0.98915	**0.99317**
GBLUR	0.93713	0.93403	**0.94681**	0.94443
FF	**0.97261**	0.92885	0.96595	0.96657
All data	0.95202	0.95073	0.96133	**0.96283**

Actually, the performance improvement of combined models may be not very obvious tested on the above database, of which the images with 5 types of distortion don't contain too much color changes.

4. Conclusion

In this paper, we analysis SSIM and CD-PSNR limitations for color image quality assessment, and attempt to combine image structure and color information together in order to improve the performance of objective quality model. Our experiments on LIVE Subjective database indicate that the attempt is reasonable and efficient. The proposed additive model and multiplicative model all perform better than SSIM, which focus on structure, and CD-PSNR, which works with image color only. For different purpose, we could have different choices of objective quality assessment approaches, such as SSIM for Rayleigh fast fading channel, multiplicative model for white noise, additive model for general purpose, considering the tradeoff between complexity and efficiency. The development either of structure similarity or of color difference approaches could promote further improvement.

There are a number of issues that are worth investigation. For example, the combination way of two components, the determination of their relative importance. More extensive experiments are needed to fully rationalize the attempt.

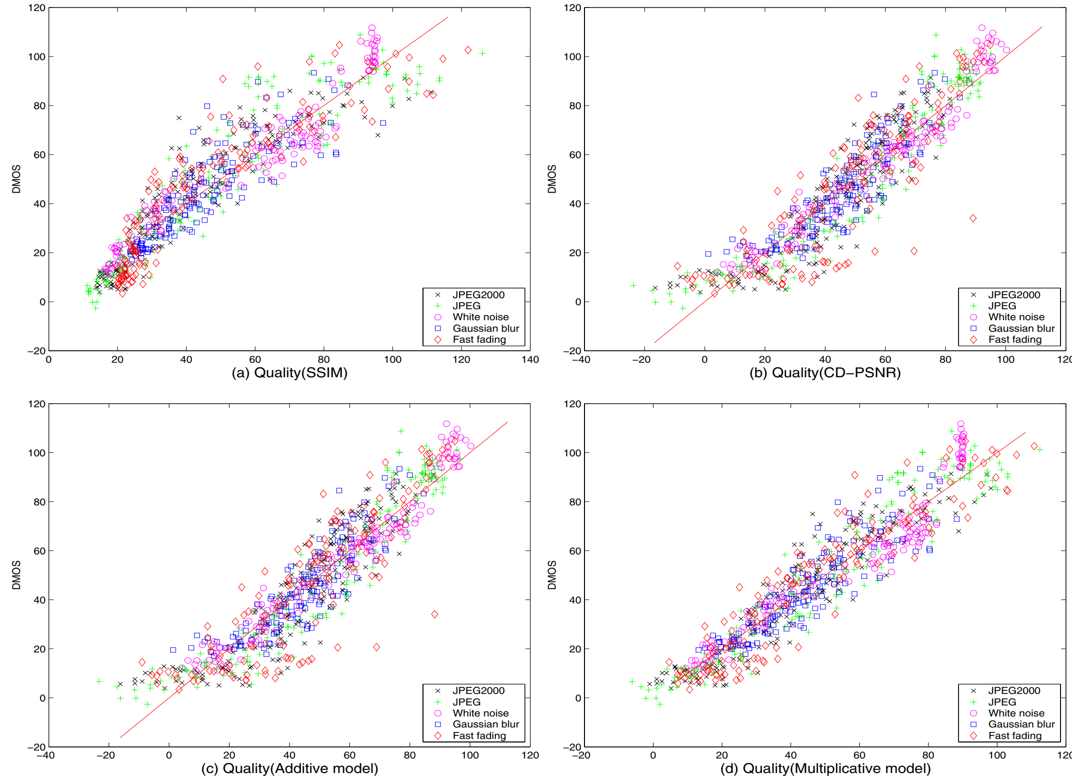


Figure 2. Scatter plots of DMOS versus different model predictions. Each sample point represents one test image in the LIVE Subjective database.

References

[1] Z.Wang and A.C.Bovik, *Modern Image Quality Assessment*. Morgan & Claypool Publisher, 2006.

[2] H.R.Sheikh, M.F.Sabir, and A.C.Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Transactions on Image Processing*, vol. 15, pp. 3440–3451, Nov. 2006.

[3] Z.Wang, E.P.Simoncelli, and A.C.Bovik, “Multi-scale structural similarity for image quality assessment,” in *the 37th IEEE Asilomar Conference on Signals,Systems and Computers*, (Pacific Grove, CA), 2003.

[4] H.R.Sheikh, A.C.Bovik, and G. Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” *IEEE Transactions on Image Processing*, vol. 14, pp. 2117–2128, Dec. 2005.

[5] H.R.Sheikh, Z.Wang, L.Cormack, and A.C.Bovik, “LIVE image quality assessment database release 2.” http://live.ece.utexas.edu/research/quality, 2005.

[6] Z.Wang, A.C.Bovik, H.R.Sheikh, and E.P.Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, Apr. 2004.

[7] Z.Wang and A.C.Bovik, “A universal image quality index,” *IEEE Signal Processing Letters*, vol. 9, pp. 81–84, Mar. 2002.

[8] R.C.Gonzalez and R.E.Woods, *Digital Image Processing*. Beijing: Publishing House of Electronics Industry, second ed., 2007.

[9] X.M.Zhang and B.A.Wandell, “Color image fidelity metrics evaluated using image distortion maps,” *Signal Processing*, vol. 70, pp. 201–214, Nov. 1998.

[10] X.M.Zhang, “S-CIELAB:a spatial extension to the CIE L*a*b* DeltaE color difference metric.” http://white.stanford.edu/˜brian/scielab/scielab.html, 1998.

[11] X.Q.Huang, J.S.Shi, J.Yang, and J.C.Yao, “Study on color image quality evaluation by mse and psnr based on color difference,” *Acta Photonica Sinica(in Chinese)*, vol. 36, pp. 295–298, 2007.