Objective Image Quality Assessment

1. Overview

After the color theme extraction, it is an indispensable step to evaluate the quality of those quantized images. Since human beings are the ultimate observers of the images and thus the judges of image quality, it is necessary to use image quality assessment (IQA) models with human subjective evaluation.

There are three classifications of IQA models (image quality, n.d.):

1. full-reference (FR) methods – access the quality of a test image by comparing with a reference image that has perfect quality
2. reduced-reference (RR) methods – access the quality of a test and reference based on the extracted features from both images
3. no-reference (NR) methods – access the quality of a test image without any information of the reference image

In this project, I focus on the FR-IQA models, which are used widely in the evaluation of image processing algorithms. Generally speaking, there are two main approaches to design FR-IQA metrics: bottom-top framework, which aims to model the overall function of human visual system (HVS) based, and top-bottom framework.

I use four different kinds of FR-IQA metrics to evaluate the quality of those quantized images.

2. Peak Signal-to-Noise Ratio (PSNR)

2.1. Introduction

On the basis of pixel information, PSNR computes the peak signal-to-noise ratio, in decibels, between two images. This ratio is used as a quality assessment between the reference image and quantized images. Definition

PSNR is defined via mean square error (MSE). Given a noise-free monochrome $m \times n$ image X and its noisy approximation Y, MSE is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[X(i,j) - Y(i,j)]^2$$

The $PSNR$(in dB) is defined as:

$$\begin{aligned} PSNR &= 10 \cdot log_{10}\left(\frac{{MAX_I}^2}{MSE}\right) \\ &= 20 \cdot log_{10}(MAX_I) - 10 \cdot log_{10}(MSE) \end{aligned}$$

In the equation above, MAX_I represents maximum fluctuation in the input image datatype. For example, in grayscale images, MAX_I is the maximum possible pixel value of the image, and the value of $MAX_1$is 255 when there are 8 bits in per pixel. To be more general, MAX_I is $2^B - 1$, where B represents B bits per pixel. (Peak signal-to-noise ratio, n.d.)

2.2. Application

PSNR is designed for grayscale images by default, but many approaches exist for computing PSNR of color images. since mimicking human vision is the goal of this experiment, PSNR should be as close to the assessment of human eyes as possible.

2.2.1. RGB color space

For color images in RGB color space (size $m \times n$), the range of three channels (red, green, and blue) is entirely same, so MSE for each channel is defined as:

$$MSE_{(RGB)} = \frac{1}{3mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[X(i,j) - Y(i,j)]^2$$

Based on this, the value of PSNR for each channel is same for a RGB image, and the relationship of PSNR in RGB and Grayscale color spaces is shown below:

$$PSNR_{(RGB_{/channel})} = \frac{1}{3} PSNR_{(Gray)}$$

2.2.2. L*a*b* color space

Human eyes are most sensitive to luma information which represents brightness in an image. With this consideration, in L*a*b* color space we can separate L* channel and only compute PSNR on it, so new *MSE* is defined as:

$$MSE_{L^*a^*b^*)} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I_{(L^*)}(i,j) - K_{(L^*)}(i,j)\right]^2$$

When we convert the color space of images from RGB to L*a*b*, the datatype of the image also experiences a transformation, to double-precision floating-point data type, so the value of MAX_I of PSNR becomes 1 in L*a*b* color space. We can define PSNR in this color space as follows:

$$PSNR_{(L^*a^*b^*)} = 10 \cdot log_{10}\left(\frac{1}{MSE_{L^*a^*b^*)}}\right)$$

2.3. Experiment

The table below shows the resulting score of comparing the original image and quantized images by using kmeans method. The range of the number of colors I chose for the color extraction is 1 to 20. The higher the PSNR, the better quality of the quantized image.

# of colors	RGB color space	L*a*b* color space	L* channel of L*a*b* color space
1	11.867	16.126	11.764
2	17.587	21.339	17.965
3	20.172	23.264	19.504
4	21.643	24.416	20.431
5	22.673	25.440	21.450
6	23.617	26.459	22.624
7	24.351	27.155	23.324
8	24.918	27.509	23.608
9	25.330	27.752	23.733
10	25.769	28.270	24.305
11	26.149	28.697	24.751
12	26.528	29.251	25.419
13	26.861	29.452	25.584
14	27.151	29.606	25.689
15	27.415	29.898	25.996

16	27.656	30.146	26.254
17	27.871	30.424	26.560
18	28.070	30.630	26.774
19	28.270	30.893	27.074
20	28.466	31.089	27.277

Table 1 The PSNR results of kmeans color extraction

3. Structural Similarity (SSIM)

3.1. Introduction

SSIM is perception-based model that considers image degradation as perceived change in structural information. As it shows above, PSNR just focuses on the absolute errors between images, so SSIM is designed to improve this traditional method.

While still considering luminance and contrast masking terms, the structural information is also aimed to find the inter-dependencies among pixels especially they are spatially close.

3.2. Definition

SSIM consists of 3 components L for luminance, C for contrast, S for structural information. (Wang, Bovik, Hamid , & Simoncelli, 2004)When we compute SSIM of the original image *x* and the quantized image *y*:

$$\mathrm{L}(x,y) = \frac{\left(2\boldsymbol{\mu_x}\boldsymbol{\mu_y} + \boldsymbol{C_1}\right)}{\left(\boldsymbol{\mu_x}^2 + \boldsymbol{\mu_y}^2 + \boldsymbol{C_1}\right)}$$

$$\mathrm{C}(x,y) = \frac{(2\boldsymbol{\sigma_{xy}} + \boldsymbol{C_2})}{(\boldsymbol{\sigma_x}^2 + \boldsymbol{\sigma_y}^2 + \boldsymbol{C_2})}$$

$$\mathrm{S}(x,y) = \frac{(\boldsymbol{\sigma_{xy}} + \boldsymbol{C_3})}{(\boldsymbol{\sigma_x}\boldsymbol{\sigma_y} + \boldsymbol{C_3})}$$

μ_x *is the average of* x, σ_x *is the variance of* x, σ_{xy} *is the covariance of* xy

$C_1 = k_1L^2,\ C_2 = k_2L^2, C_3 = \frac{C_2}{2},$

$k_1, k_2 \ll 1$ *are very small constant* $(k_1 = 0.01, k_2 = 0.03$ *by default*$)$,

L *is the dynamic range of the pixel value, typically this is* $2^{\#bits\ per\ pixel} - 1$

Then, SSIM can be defined as the combination of the three comparisons

$$SSIM(x,y) = L(x,y)^{\alpha} \cdot C(x,y)^{\beta} \cdot S(x,y)^{\gamma}$$

In order to simplify the formula, we set α = β = γ = 1, so the SSIM formula can be reduced to the form shown below when measuring between the reference image *x* and the test image *y* of size *N*x*N* is:

$$SSIM(\mathrm{x},\mathrm{y}) = \frac{\left(2\boldsymbol{\mu_x}\boldsymbol{\mu_y} + \boldsymbol{C_1}\right)(2\boldsymbol{\sigma_{xy}} + \boldsymbol{C_2})}{\left(\boldsymbol{\mu_x}^2 + \boldsymbol{\mu_y}^2 + \boldsymbol{C_1}\right)(\boldsymbol{\sigma_x}^2 + \boldsymbol{\sigma_y}^2 + \boldsymbol{C_2})}$$

Practically, one usually requires a single overall quality measure of the entire image. Mean SSIM (MSSIM) index is designed to evaluate the overall image quality:

$$MSSIM(\mathrm{x},\mathrm{y}) = \frac{1}{M}\sum_{j=1}^{M} SSIM(\mathrm{x}_j, \mathrm{y}_j)$$

x_j and y_j are the image contents at the jth local window (8x8 square local window by default), and M is the number of local windows of the image.

3.3. Application

3.3.1. RGB color space

As for RGB color space, MSSIM can be applied directly to the reference image and quantized images.

$$MSSIM_{(RGB)} = \mathrm{MSSIM}(\mathrm{x}, \mathrm{y})$$

3.3.2. L*a*b* color space

Since human eyes have high sensitivity to lightness, I think we can use these two methods shown below to refine SSIM in L*a*b* color space.

3.3.2.1. L* channel computation

Inspired from computing PSNR in L*a*b* color space, we can only compute SSIM for the L* channel of images. Since luminance plane consists of significant information in the image the evaluation score obtained is closer to human vision judgement.

$$MSSIM_{(L^*a^*b^*)} = \mathrm{MSSIM}\left(x_{(L^*)}, y_{(L^*)}\right) = \frac{1}{M}\sum_{j=1}^{M} SSIM\left({x_{(L^*)}}_j, {y_{(L^*)}}_j\right)$$

$$= \frac{1}{M}\sum_{j=1}^{M} \frac{\left(2\boldsymbol{\mu}_{x_{(L^*)}}\boldsymbol{\mu}_{y_{(L^*)}} + \boldsymbol{C_1}\right)\left(2\boldsymbol{\sigma}_{x_{(L^*)}y_{(L^*)}} + \boldsymbol{C_2}\right)}{\left(\boldsymbol{\mu}_{x_{(L^*)}}{}^2 + \boldsymbol{\mu}_{y_{(L^*)}}{}^2 + \boldsymbol{C_1}\right)\left(\boldsymbol{\sigma}_{x_{(L^*)}}{}^2 + \boldsymbol{\sigma}_{y_{(L^*)}}{}^2 + \boldsymbol{C_2}\right)}$$

3.3.2.2. Weighted computation

L component in SSIM represents the luminance measurement, thus we can change the relative importance of the three components by setting higher value of α than β, γ in SSIM format shown above.

For instance, we can set $\alpha = 1$ than $\beta = \gamma = 0.1$, so

$$SSIM_{(L^*a^*b^*)} = \mathrm{L}(\mathrm{x}, \mathrm{y})^1 \cdot \mathrm{C}(\mathrm{x}, \mathrm{y})^{0.1} \cdot \mathrm{S}(\mathrm{x}, \mathrm{y})^{0.1}$$

$$MSSIM_{(L^*a^*b^*)} = \frac{1}{M}\sum_{j=1}^{M} \mathrm{L}\left(\mathrm{x}_j, \mathrm{y}_j\right)^1 \cdot \mathrm{C}\left(\mathrm{x}_j, \mathrm{y}_j\right)^{0.1} \cdot \mathrm{S}\left(\mathrm{x}_j, \mathrm{y}_j\right)^{0.1}$$

3.4. Experiment

In the experiment, MSSIM assessment is applied to compare the reference image and test images after kmeans quantization in RGB, L*a*b* color spaces, and L* channel as well. Weighted computation hasn't been implemented yet. The higher score represents higher similarity of the reference and test images.

# of colors	RGB color space	L*a*b* color space	L* channel
1	0.1849	0.6958	0.3098
2	0.4717	0.7744	0.5642
3	0.6541	0.7957	0.6003
4	0.7396	0.8244	0.6501
5	0.7700	0.8510	0.7063
6	0.7913	0.8725	0.7609
7	0.8283	0.8845	0.7854
8	0.8569	0.8907	0.7936

9	0.8718	0.8950	0.7937
10	0.8734	0.9049	0.8212
11	0.8735	0.9123	0.8357
12	0.8894	0.9192	0.8553
13	0.8935	0.9228	0.8597
14	0.8969	0.9258	0.8621
15	0.9054	0.9295	0.8696
16	0.9092	0.9312	0.8724
17	0.9158	0.9344	0.8790
18	0.9191	0.9363	0.8832
19	0.9170	0.9394	0.8899
20	0.9247	0.9415	0.8939

Table 2 The MSSIM results of kmeans color extraction

4. Visual Information Fidelity (VIF)

4.1. Introduction

VIF index introduces natural scene statistical model in conjunction with a distortion(channel) model to quantify the information shared between the reference and test images, so this measurement does not rely on any Human Visual System (HVS) or viewing geometry parameter. It treats HVS as a communication channel and predicts the subjective image quality by computing how much the information within the perceived reference image is preserved in the distorted one.

4.2. Definition

In VIF, an approach called "information-theoretic setting" is included. The reference image and the test image are in two different processes through the measurement.

The reference image is regarded as natural image source that goes through HSV channel before being processed to the brain. The information comes from the reference image is quantized as being mutual information between input and output of HSV channel. This kind of information is that the brain could ideally extract from the reference image.

The same measurement is also applied for the test image. There is a distortion channel that distorts the output of the natural image source before it comes to HSV channel, thereby we can get the information that brain can extract from the test image.

Finally, we can combine those two results from each phase to get the VIF evaluation. The Fig. 1 below shows the entire measure process, where VIF system model consists of three sub-models: source model, distortion model and HVS model. (Visual Information Fidelity, n.d.)

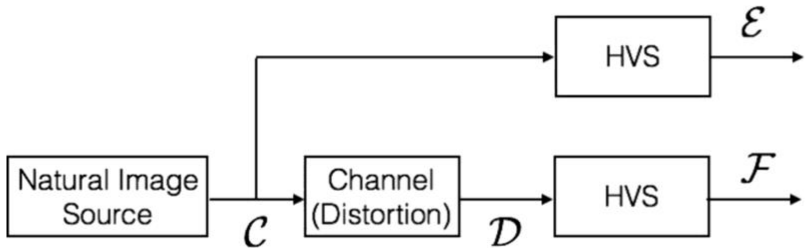

Fig. 1 VIF measurement process

4.3. Application

Based on the definition of VIF, the same VIF assessment model can be applied for both RGB color space and L*a*b* color space.

4.4. Experiment

VIF computes the similarity of the original image and quantized images in both RGB and L*a*b* color spaces. Table 3 shows the results of this experiment.

# of colors	RGB color space	L*a*b* color space
1	0.0586	0.0259
2	0.1081	0.5756
3	0.1439	0.6955
4	0.1774	0.7440
5	0.2084	0.7904
6	0.2408	0.8324
7	0.2586	0.8543
8	0.2686	0.8627
9	0.2727	0.8703
10	0.2940	0.8840
11	0.3053	0.8936
12	0.3232	0.9056
13	0.3337	0.9108
14	0.3394	0.9133
15	0.3485	0.9187
16	0.3524	0.9222
17	0.3589	0.9261

18	0.3668	0.9295
19	0.3780	0.9339
20	0.3841	0.9362

Table 3 The VIF results of kmeans color extraction

5. Gradient Magnitude Similarity Deviation (GMSD)

5.1. Introduction

Practically, a good FR-IQA method should be not only effective, but also efficient. But unfortunately, it is very hard to achieve both aspects simultaneously, and those FR-IQAs mentioned before can only perform well in one of these two parts.

GMSD is aimed to fill this need. It computes a local quality map (LQM) by comparing the gradient magnitude maps of the test and reference images and uses standard deviation as the pooling strategy to compute the final quality score. The Fig. 2 shown below illustrates this two-step framework. The gradient feature of images can capture image local structures efficiently, to which the HVS is highly sensitive. (Xue, Zhang, Mou, & Bovik, 2014)

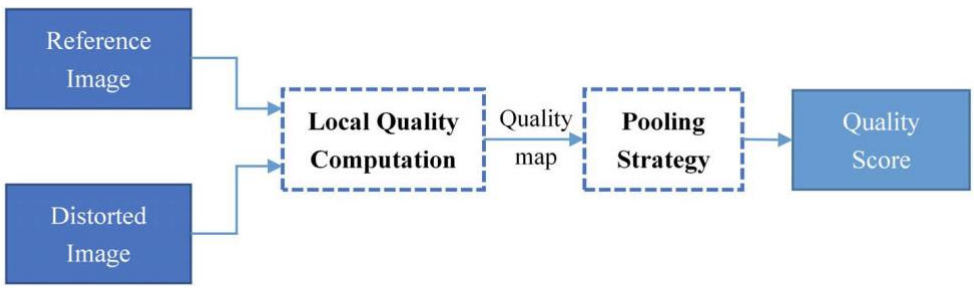

Fig. 2. The flowchart of a class of two-step FR-IQA models.

5.2. Definition

5.2.1. Gradient Magnitude Similarity

In GMSD, Similarity function called gradient SSIM (G-SSIM) is derivate from SSIM.
For digital images, the gradient magnitude is defined as the root mean square of image directional gradients along two orthogonal directions. The gradient is computed by convolving an image with a linear filter -Prewitt filter.

The horizontal (x) and vertical (y) directions of Prewitt filters are defined as:

$$h_x = \begin{bmatrix} \frac{1}{3} & 0 & -\frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \end{bmatrix},\ h_y = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

The horizontal and vertical gradient images of r and d are the results of convolving h_x and h_y with the reference and quantized images, and then we can compute the gradient magnitude of r and d at location i, denoted by $m_r(i)$ and $m_d(i)$ as follows:

$$m_r(i) = \sqrt{(r \otimes h_x)^2(i) + \left(r \otimes h_y\right)^2(i)}$$

$$m_d(i) = \sqrt{(d \otimes h_x)^2(i) + \left(d \otimes h_y\right)^2(i)}$$

where "⊗" denotes the convolution operation.

Based on the gradient magnitude images m_r and m_d we get, the gradient magnitude similarity (GMS) map then can be computed as follows:

$$\text{GMS}(i) = \frac{2m_r(i)m_d(i) + c}{m_r(i)^2 + m_d(i)^2 + c}$$

where c is a positive constant to supply numerical stability.

The GMS map serves as the local quality map (LQM) of the quantized image d. In the GMS map, the brighter the gray level, the higher the similarity, and thus the higher the predicted local quality.

5.2.2. Pooling with Standard Deviation

Average pooling is the most common used pooling strategy to compute the final quality score, so it is applied to the GMS map as Gradient Magnitude Similarity Mean (GMSM):

$$\text{GMSM} = \frac{1}{N}\sum_{i=1}^{N} GMS_{(i)}$$

In estimating the overall image quality, the main issue about average pooling is that the importance of each pixel is same, which means that it ignores the fact that the global variation of image local quality degradation can reflect its overall quality.

In order to solve this problem, the standard deviation of GMS map is computed and taken as the final IQA index, namely Gradient Magnitude Similarity Deviation (GMSD):

$$\text{GMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (GMS(i) - GMSM)^2}$$

With the value of GMSD, the range of distortion severities shows the reflection in an image.

5.3. Application

GMSD can be fit into both RGB and L*a*b* color spaces.

5.4. Experiment

As for GMSD, the lower the score, the better quality of the quantized image. Table 4 below shows the GMSD evaluation results of images after color extraction by kmeans in RGB and L*a*b* color spaces.

# of colors	RGB color space	L*a*b* color space
1	0.2170	0.2293
2	0.1760	0.1268
3	0.1675	0.1281
4	0.1514	0.0955
5	0.1271	0.0964
6	0.1220	0.0809
7	0.1051	0.0624
8	0.0934	0.0602
9	0.0800	0.0555
10	0.0790	0.0478
11	0.0734	0.0417
12	0.0632	0.0382

13	0.0598	0.0367
14	0.0576	0.0359
15	0.0552	0.0336
16	0.0532	0.0316
17	0.0511	0.0302
18	0.0497	0.0293
19	0.0499	0.0278
20	0.0462	0.0272

Table 4 The GMSD results of kmeans color extraction

Bibliography

Visual Information Fidelity. (n.d.). Retrieved from wikipedia: https://en.wikipedia.org/wiki/Visual_Information_Fidelity

Wang, Z., Bovik, A. C., Hamid , S. R., & Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE*.

image quality. (n.d.). Retrieved from wikipedia: https://en.wikipedia.org/wiki/Image_quality

image quality. (n.d.). Retrieved from wikipedia: https://en.wikipedia.org/wiki/Image_quality

Peak signal-to-noise ratio. (n.d.). Retrieved from wikipedia: https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio

structural similarity. (n.d.). Retrieved from wilkipedia: https://en.wikipedia.org/wiki/Structural_similarity

Visual Information Fidelity. (n.d.). Retrieved from wikipedia: https://en.wikipedia.org/wiki/Visual_Information_Fidelity

Xue, W., Zhang, L., Mou, X., & Bovik, A. C. (2014). Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index. *IEEE*.