International Journal of Electronics and Communication Engineering (IJECE)
ISSN 2278-9901
Vol. 2, Issue 2, May 2013, 135-142

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

WHY SSIM? - A FULL REFERENCE IMAGE QUALITY ASSESSMENT

SWATI A. GANDHI[1] & C. V. KULKARNI[2]

Department of E&TC, MIT, COE, University of Pune, Pune, India

ABSTRACT

Digital image processing and image quality improvement techniques are involved during image transmission and restoration process. Image received is not perfect in quality. Hence, analysis of image is done using full reference IQA, where distorted and original image both are present. One of the method to extract structural information from distorted image using SSIM which is more practical approach over traditional IQA methods explain here.

KEYWORDS: Human Visual System Model (HVS), Image Quality Assessment (IQA), Mean Square Error (MSE), Structural Similarity Index (SSIM)

INTRODUCTION

Recent advances in digital imaging technology, speed and storage capacity and networking makes image suffer from different distortions like acquisition, processing , compression, enhancement and restoration, degrades the image quality. At receiver side human being are the observers. Hence, it is necessary to evaluate the quality of image through some mathematical calculation. Image quality is the degree to which an image satisfies the usefulness and naturalness requirement which determines quality of this image. Image quality is characteristics of an image that measures the perceived image degradation (typically, compared to an ideal or perfect image). IQA plays an important role in image processing applications. It can be used to evaluate performance of different image compression algorithm where number of bits are reduced in order to store as well as to increase the speed of transmission. It can also be used in acquisition and display system to monitor image quality[8].

IQA can be done widely using subjective and objective analysis[9]. In case of subjective analysis, human being is ultimate end user who uses visual perception to quantify image. Double stimuli continuous quality score (DSCQS) and Single stimuli continuous quality score (SSCQS) are generally used where a linear quality score is scored on a scale of 0 to 100. In DSCQS, number of users are asked to quantify both original and distorted image on the score scale, then mean opinion score (MOS) and differential mean opinion score (DMOS) is calculated. Higher the MOS better will be quality of image. Main disadvantage of this analysis is time consuming, inconvenient, and expensive. Objective IQA can be done using three methods namely Full Reference where distorted and original images are present, Reduced Reference where partial information is extracted from original image and on that basis IQ is assessed and No or Blind Reference where no information about original image. This paper explain the implementation of SSIM based full reference methods for IQA.

MEAN SQUARE ERROR BASED MATRIX

The simplest and most widely used traditional full reference image quality assessment method, MSE, is calculated by averaging the squared intensity differences of distorted and original image pixels. These are widely used due to simple and mathematically convenient[1].

If x and y are two non negative gray scale images, then Minkowski metric is calculated using following mathematical relation,

$$E\gamma = \left(\frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|^2 \right)^{\frac{1}{\gamma}} \tag{1}$$

Where x_i and y_i are the i^{th} sample of image x and y, N is number of image samples and γ is in the range of $\gamma \in [1\ \infty)$. For $\gamma = 2$, one obtains the well-known equation

$$E_2 = \left(\frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|^2 \right)^{\frac{1}{2}} = \sqrt{MSE} \tag{2}$$

Mean Square Error (MSE) is calculated using above equation(2) if ignoring the square root

$$\text{MSE}= \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|^2 \tag{3}$$

PSNR is calculated as

$$\text{PSNR}=10 \log_{10} \left(\frac{255^2}{MSE} \right) \tag{4}$$

Where 255 are maximum gray levels of 8 bits/pixel of image. If we consider image with MSE=0 and PSNR as infinite, then image is perfect image. As shown in Fig.4, we consider 6 images out which one is original and other 5 with different types of distortions such that all have set to equal MSE. The visual quality between any two images is different.

Two distorted image signals with same amount of error energy may have different structures of error and hence different perceptual quality. This is major disadvantage of MSE [5].

HVS BASED MODEL

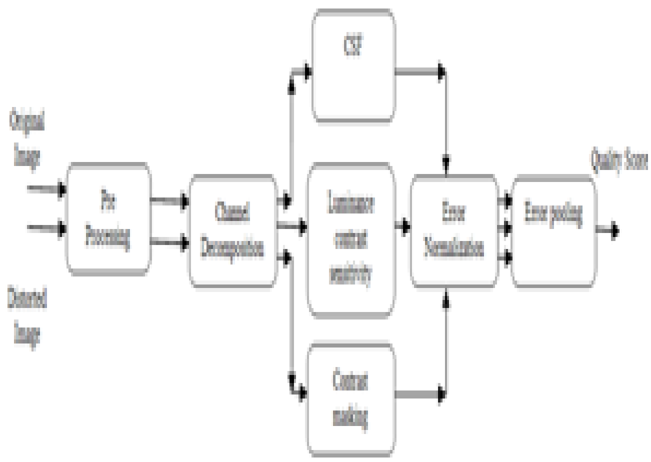

Figure 1: Human Visual System Based Model

Human Visual system (HVS) based approach is based on conjecture[1]. The purpose of entire visual observation process is to efficiently extract and make use of information represented in natural scenes, whose statistical properties are playing important role in evaluation, development and adaptation of HVS. HVS based methods used to solve the quality assessment problem by measuring the threshold of visibility of signals. Then these thresholds are used to normalize the error between reference and distorted images to obtain a error metric which is perceptually perfect. Visibility of threshold is measured by considering different aspects of visual processing like response to average brightness, contrast, spatial frequencies and orientations[6].

As shown in Figure 1[6], first stage of HVS model is preprocessing where registration and calibration are done. In the second stage, channel decomposition (Frequency Analysis) stage, the reference and test images into two different channels or sub bands with different spatial frequencies and orientations using a set of linear filters. CSF provides characterization of its frequency response. It is nothing but a band pass filter. Contrast sensitivity determines amount of energy in each sub band that is required to detect target in flat mid gray image referred to just noticeable difference (JND). According to Webbers Law, human eyes are sensitive to luminance contrast rather than absolute luminance value. The ratio $\Delta I/I$ of just noticeable luminance difference ΔI and luminance I is constant for wide range of luminance.

$$\Delta I/I \;=K \text{ constant} \;\approx 0.33 \tag{5}$$

Reduction in visibility of one image component by the presence of another with similar spatial location and frequency defines contrast masking. Hence the presence of multitude of frequency components in varying intensity field image hides or mask the presence of noise field. The last stage of model is error pooling which provide metric to combine the errors that have been commutated for each spatial frequency and orientation band, each spatial location into a single number for each pixel of the image or single number for whole image.

Due to Some Disadvantages Like

- Tested metrics are to be oriented on some average conditions of image visualization. Then experiments intended on obtaining observers opinion (MOS) in creating and exploiting test image databases have to be carried out with reasonable variations of image visualization conditions.
- An image database cannot be used for design and verification of HVS models since creation of HVS database would require stringent control of visualization and observation conditions which may not be feasible.
- HVS based metric give a better correlation with the human perception but faces some problems like complexity in design and difficulty in setting threshold for visibility.

Hence HVS is complex and difficult to implement.

STRUCTURAL SIMILARITY INDEX

All natural images are highly structured which means the signal samples exhibit strong dependencies amongst themselves. The principle premise of structural similarity approach is that the major goal of visual observation is to extract such information, for which the HVS is highly adapted [4].

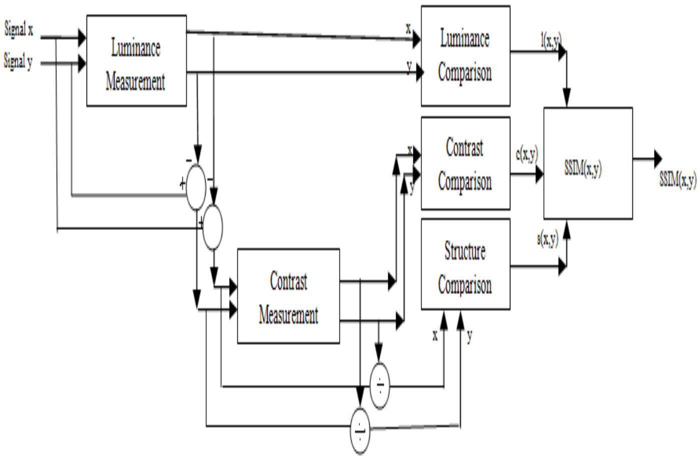

Figure 2: Structural Similarity (SSIM) Measurement System

Luminance of the surface of an object being observed is the product of illumination and reflectance, but the structures of objects in the scene are independent of illumination. Hence system will separate out the influence of illumination from information. Illumination change is the variation of the average luminance and contrast in the image. Since luminance and contrast can vary across a scene, they are preferably measured locally. This leads to localized image similarity measure that separates influence of luminance and contrast variations from the remaining attributes of local image region.

x and y are two non negative discrete image signals, which have been aligned with each other. Similarity measure can serve as a quantitative measure of quality of one of the signals if other has perfect quality.

Let x and y be represented as

$x=\{x_i \mid i=1,2,3,....N\}$ and $y=\{y_i \mid i=1,2,3....N\}$

where i=sample index, N=No. of signal samples (pixels).

As shown in Figure 2 above, the system separates the image into three components namely luminance, contrast and structure.

Luminance of each signal is estimated as mean intensity

$$\mu_x=\bar{x}=\frac{1}{N}\sum_{i=1}^{N} x_i \qquad (6)$$

The luminance comparison function l(x, y) is a function of μ_x and μ_y equals to

$$l(x, y)=l(\mu_x , \mu_y) \qquad (7)$$

Removing mean intensity from the signal, the resulting signal $(X-\mu_x)$ corresponds to projection of luminance vector onto the hyper plane of

$$\sum_{i=1}^{N} x_i=0 \qquad (8)$$

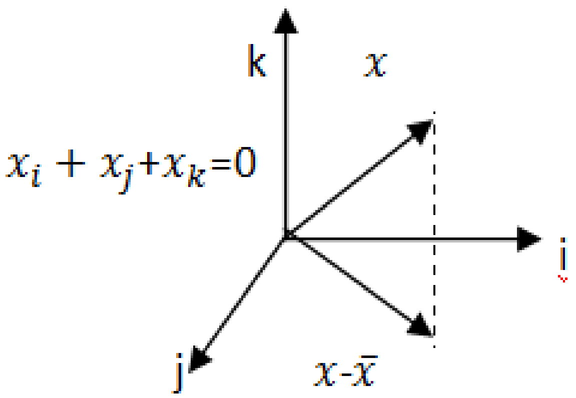

Figure 3: Three Dimensional Representation of Changes in Luminance, Contrast and Structural Information of Reference Image

Using the standard deviation (the square root of variance) as an estimate of signal contrast.

$$\sigma_x=[\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_x)^2]^{1/2} \qquad (9)$$

Contrast comparison c(x,y) is then comparison of σ_x and σ_y and is given by

$$c(x,y)=c(\sigma_x, \sigma_y) \qquad (10)$$

The signal is normalized (divided) by its own standard deviation, so that the two signals being compared have unit deviation. The structure comparison s(x, y) is conducted on these normalized signals.

$$s(x, y)=s\left(\frac{x-\mu_x}{\sigma x}, \frac{y-\mu_y}{\sigma y}\right) \quad (11)$$

The three components are combined to yield an overall similarity measure

$$SSIM(x,y)=f(l(x,y), c(x,y), s(x,y)) \quad (12)$$

All three points are independent on each other.

Similarity measure satisfies the following conditions:

- **Symmetry**: S (x, y) = S (y, x). Exchanging the order of input signals should not affect the resulting similarity measure.
- **Boundedness** S(x, y) <= 1.Upper bound can serve as an indication of how to close the two signals are to being perfectly identical.
- **Unique Maximum:** S (x, y) =1. If and only if x=y.

For luminance comparison

$$l(x,y) = \frac{(2\,\mu_x\,\mu_y + c1)}{(\mu_x{}^2 + \mu_y{}^2 + c1)} \quad (13)$$

c1is used to avoid instability when ($\mu_x^2 + \mu_y^2$) is close to zero.

$$c1=(k_1L)^2 \quad (14)$$

Where L = dynamic range of pixel values (255 for 8 bit gray scale image) $k \leq 1$ as small constant.

According to Webber law, R represents the ratio of luminance change relative to background luminance. Luminance of distorted signal can be written as

$$\mu_y=(1+R)\,\mu_x \quad (15)$$

Therefore,

$$l(x,y)=\frac{2(1+R)}{1+(1+R)^2+\frac{c1}{\mu_x{}^2}} \quad (16)$$

So l(x,y) is a function only of R instead of $\Delta I=\mu_y - \mu_x$

$$c(x,y)=\frac{2\sigma_x\sigma_y+c2}{\sigma_x{}^2+\sigma_y{}^2+c2} \quad (17)$$

where $c2=(k_2L)^2$ and $k_2 \leq 1$ as small constant.

With the same amount of contrast change $\Delta\sigma = \sigma_y - \sigma_x$, this measure is less sensitive to high base contrast σ_x than low base contrast. This explains contrast masking features [1].

Structure is calculated after luminance subtraction and contrast normalization. Structure of image is calculated by the direction of two unit vectors $\frac{x-\mu_x}{\sigma x}$ and $\frac{y-\mu_y}{\sigma y}$. Correlation between $\frac{x-\mu_x}{\sigma x}$ and $\frac{y-\mu_y}{\sigma y}$ is equivalent to correlation coefficient between x and y. Correlation means closest match between two prototypes.

$$s(x, y)=\frac{\sigma_{xy}+c3}{\sigma_x\,\sigma_y+c3} \quad \text{where } c3=\frac{c2}{2} \quad (18)$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y) \quad (19)$$

$$\text{SSIM}(x,y)=\frac{(2\,\mu_x\mu_y+c1)(2\sigma_x\sigma_y+c2)}{(\mu_x{}^2+\mu_y{}^2+c1)(\sigma_x{}^2+\sigma_y{}^2+c2)} \quad (20)$$

IQA BASED ON SSIM INDEX

SSIM index can be applied locally rather than globally for that image conditions required are [4]:

- Image statistical features are usually highly spatially non stationary.
- Image distortions which may or may not dependent on image statistics, may also be space varient.
- Typical viewing distance, only a local area in the image can be perceived with resolution by human observer at one time instant.
- Localized quality measurement can provide a spatially varying quality map of image which delivers more information about the quality degradation.

A circular symmetric Gaussian weighting function

$w=\{w_i|i=1,2,3\ldots N\}$ with unit sum $\sum_{i=1}^{N} w_i$=1 is applied, then local statistic modified are features given by

$$\mu_x=\sum_{i=1}^{N} w_i\, x_i \quad (21)$$

$$\sigma_x=\left(\sum_{i=1}^{N} w_i(x_i-\mu_x)^2\right)^{\frac{1}{2}} \quad (22)$$

$$\sigma_{xy}=\sum_{i=1}^{N} w_i\,(x_i-\mu_x)(y_i-\mu_y) \quad (23)$$

Mean SSIM index to evaluate overall image quality

$$\text{MSSIM}=\frac{1}{M}\sum_{j=1}^{M} w_j\ SSIM_j\,. \quad (24)$$

RESULTS

An experiment is done by taking reference images with different types of distortions. Here tries to assess images based on cross distortions. MSSIM is calculated using Gaussian weighing mechanism with 11*11 window size and $\sigma=1.5$, proves best as compare to MSE and PSNR traditional methods as shown in Fig.4.

CONCLUSIONS

Image quality assessment is important in image processing techniques.Traditional methods which are easy but having lack of matching with human perception model.HVS algorithm are complex and difficult to implement .Structuural similarity index which having maximum value=1, indicates perfect image in which luminace, contrast and structer are relatively independent.

ACKNOWLEDGMENTS

The authors would like to thanks Prof. A.C.Bovik, Prof.Z.Wang and Anish Mittal for giving the solutions and allowing to use the database for objective evaluation.

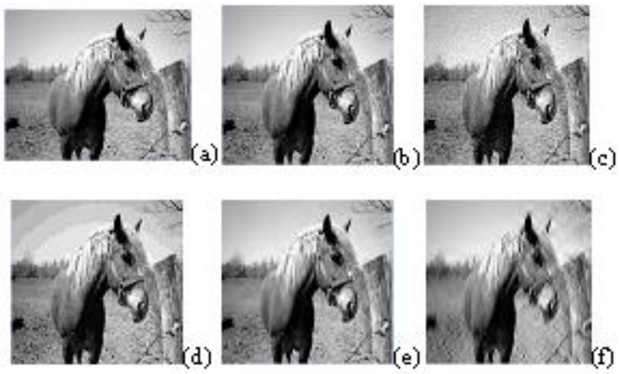

Figure 4: Comparision of "Horse"Images with Different Types of Distortions a)Original Image, MSE=0; PSNR=INF;MSSIM=1 b)Blurr Image , MSE=275.78; PSNR=23.69; MSSIM=0.6989 c)Noisy Image, MSE=304; PSNR=23.30; MSSIM=0.5760 d) JPEG Image, MSE=282.43; PSNR=23.62; MSSIM=0.6726 e)JPEG2 Image, MSE=309.06; PSNR=23.33; MSSIM=0.6323 f)Filter Image, MSE=272.53;PSNR=23.77; MSSIM=0.8211

REFERENCES

1. K. Moorthy, Z. Wang and A. C. Bovik, "Visual perception and quality assessment," in *Optical and Digital Image Processing* (Cristobal, Schelkens and Thienpont, eds.), Wiley publisher, 2010.
2. K. Seshadrinathan, T. N. Pappas, R. J. Safranek, J. Chen, Z. Wang, H. R. Sheikh and A. C. Bovik, “Image quality assessment,” in *Essential Guide to Image Processing*, Elsevier, 2009.
3. Z. Wang, A. C. Bovik and H. R. Sheikh, "Structural similarity based image quality assessment," in *Digital Video Image Quality and Perceptual Coding* (H. R. Wu, and K. R. Rao, eds.), Marcel Dekker Series in Signal Processing and Communications, Nov. 2005.
4. Z. Wang, A. C. Bovik and E. P. Simoncelli, "Structural approaches to image quality assessment," in *Handbook of Image and Video Processing* (Al Bovik, eds.), second edition, Academic Press, May 2005.
5. Z. Wang and A. C. Bovik, “Mean squared error: love it or leave it? - A new look at signal fidelity measures,” *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98-117, Jan. 2009.
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,no.4,Apr.
7. Z.Wang,A.C.Bovik and L.Lu,"Why is image quality assessment so difficult?"IEEE International 2004Conference on Acoustics,speech and Signal Processing,May 2002.
8. Z.Wang, "Application of Objective Image Quality Assessment Methods,"IEEE SIGNAL PROCESSING MAGAZINE,pp-137-142,Nov. 2011.
9. Kim-Han Thung,Paramesran Raveendran,"A Survey Of Image Quality Measures,"TECHPOS,International Conference,pp-1-4,Dec.2009
10. LIVE, http://live.ece.utexas.edu
11. UT Austin, http://www.utexas.edu
12. "Z.Wang Plublication Home Page", https://ece.uwaterloo.ca/~z70wang/publications.htm#journal_papers
13. "Video lectures,"http://iitk.ac.in/on line lectuer series on Digital Image Processing, P.K.Biswas, IIT.