Evaluation of video quality by CWSSIM method

Yu-Ta Chen[a], Shau-Wei Hsu[b], Bao-Jen Pong[b], Mang Ou-Yang[c]
[a]Department of Optics and Photonics, National Central University, Taiwan
[b]Center for Measurement Standards, Industrial Technology Research Institute, Taiwan
[c]Institute of Electrical and Control Engineering, National Chiao Tung University, Taiwan

ABSTRACT

Several estimative factors of image quality have been developed for approaching the human perception objectively[1-3]. We propose to take systematically distorted videos into the estimative factors and analyze the relationship between them. Several types of noise and noise weight were took into COSME standard video and verified the image quality estimative factors which were MSE (Mean Square Error)、SSIM (Structural SIMilarity)、CWSSIM (Complex Wavelet SSIM)、PQR (Picture Quality Ratings) and DVQ (Digital Video Quality). The noise includes white noise, blur and luminance…etc. In the results, CWSSIM index has higher sensitivity at image structure and it could estimate the distorted videos which have the same noise type at the different levels. PQR is similar to CWSSIM, but the ratings of distribution were banded together; SSIM index divides the noise types into two groups and DVQ has linear relationship with MSE in the logarithmic scale.

Keyword: Evaluation factors, image quality, objective metrics, CWSSIM.

1. INTRODUCTION

Displays grow very fast in recent years. The images on the TVs have different performance with human vision, high or low image quality is the subjective judgments of human. We do not have a standard objective metric to connect with the subjective judgments of human. MSE is a common estimation in the signal processing field over 50 years[4]. That is a simple estimation method to obtain the signal quality, but it is not effective to estimate in the image quality. MSE method is like most of the image evaluation factors that were calculated point by point such as RMS, PQR [5] and DVQ[6, 7], another type of method such as SSIM[4, 8] or CWSSIM[8] is calculated by a window cross the image after filtering process. There are many objective metrics could quantified the subjective of human vision, but the performance of accuracy and relationship is a very important point.

In this paper, we select several parts of video from Color Space Management Evaluation Material (CoSME) and add some noise into the video. The noises have JPG compression (JPG), Gaussian blur (Blur), white noise (WN), luminance gain (Lm), gamma contrast (Ct), and speckle noise (SN). Then we use the objective metrics to estimate the distorted video, and observe the relationship between them.

Mathematics of Data/Image Coding, Compression, and Encryption with Applications XII,
edited by Mark S. Schmalz, Gerhard X. Ritter, Junior Barrera, Jaakko T. Astola, Proc. of SPIE
Vol. 7799, 77990T · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.860391

2. OBJECTIVE METRICS

MSE is a simple algorithm and generally used in the signal processing fields, but it was not represent that MSE was accurate at human perception. Therefore many objective metrics were developed.

DVQ metric is an estimating algorithm that purposed to combine much information about human visual[7]. The information included SCSF (spatial contrast sensitivity function), local contrast, color and temporal effect, so the estimation is in the Discrete Cosine Transform (DCT) domain. The figure 1 is the processing of DVQ algorithm.

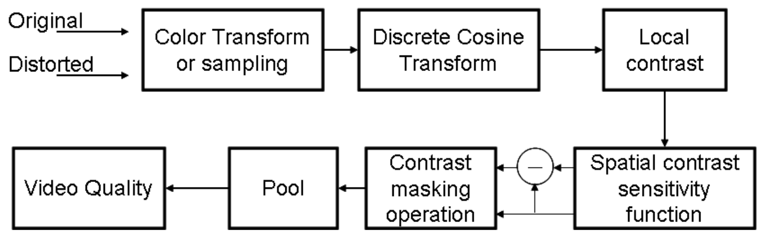

Figure 1. The processing of DVQ algorithm[6, 7].

PQR metric was estimated by Tektronix-PQA500 which could relate with differential mean opinion score (DMOS). We set the display condition as default LCD-FPD, and the viewer type with typical spatial acuity and excellent temporal response before comparison processing.

SSIM metric was developed by Wang. Z et al[4] and separate the image into three dimensional space which were luminance, contrast and structural. The parameters were default and image quality value was mean SSIM.

CWSSIM metric was similar to SSIM. It transferred the image into complex-wavelet domain, and calculated the value of image quality.

3. EXPERIMENTS

Figure 2. Parts of CoSME video

We selected the four sections video from CoSME shown in Fig. 2. The contents have sport, landscape, toy train and indoor. Each video has 96 frames and six types of noise. The noise was shown in the Fig. 3.

The noise strength was increased with frames in each video, except Lm and Ct. When the frames of Lm and Ct were increased or decreased, the noise strength was decreased. Therefore the noise strength of Lm and Ct was least while the frame was in the middle of Lm and Ct noise videos. Some objective metrics could accept video format input but the other could not, hence the video should be divided into several pictures and was estimated after comparison processing.

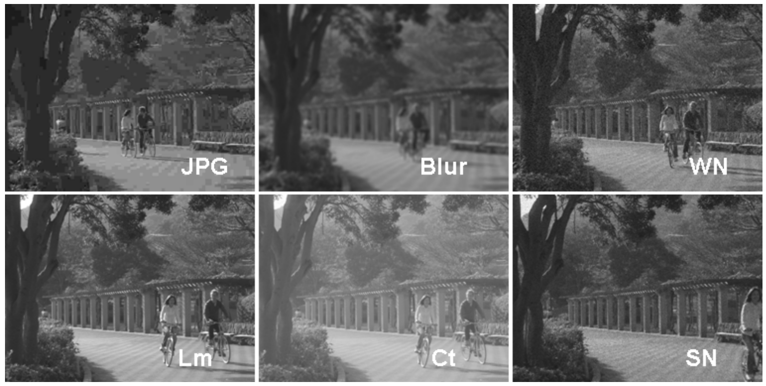

Figure 3. Example frames of distorted video V1.

After transferring the file format, if the objective metrics could not accept the color pictures, we processed the color information of frames by mean the digital level of three colors.

4. RESULTS AND DISCUSSIONS

We can obtain the values after the objective metrics, and have comparison between MSE and the others. Figure 4 to figure 7 are objective metrics comparison with MSE in logarithmic scale.

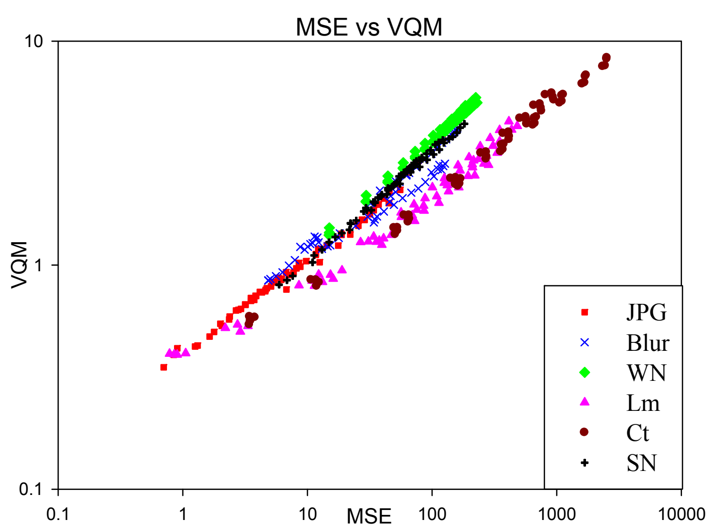

Figure 4. Relationship between VQM and MSE in logarithmic scale.

The figure 4 is relationship between VQM and MSE in the logarithmic scale. We fitted the curves by quadratic polynomial[10] and shown it in the table 1. The parameters a, b, c were coefficient of quadratic polynomial and shown at (1). Almost coefficients of videos are similar to each other in the same noise type, in the Fig. 4, most distribution of several types of noise was like several indistinct lines. We were difficult to separate the lines into two groups, and it's meaning that the content information is invisible. Therefore the objective metric depended on noise types and ignored the content effect on the human vision.

$$y = ax^2 + bx + c \quad (1)$$

Table 1. Parameters of fitting curve after VQM estimation.

	JPG			Blur			WN		
	a	b	c	a	b	c	a	b	c
V3	-6.45E-03	4.24E-01	-3.85E-01	1.38E-02	4.67E-01	-4.01E-01	-1.50E-03	5.00E-01	-4.45E-01
V5	-2.32E-02	4.46E-01	-3.98E-01	1.13E-02	4.49E-01	-4.51E-01	1.90E-02	4.25E-01	-3.82E-01
V11	2.93E-02	3.52E-01	-3.62E-01	7.69E-02	1.59E-01	-2.24E-01	5.70E-03	4.75E-01	-4.00E-01
V13	9.43E-02	2.75E-01	-3.97E-01	1.12E-01	6.68E-02	-7.79E-02	-8.15E-03	5.26E-01	-4.73E-01
	Lm			Ct			SN		
	a	b	c	a	b	c	a	b	c
V3	5.94E-02	2.15E-01	-3.96E-01	2.52E-02	3.14E-01	-4.42E-01	6.17E-02	3.41E-01	-3.94E-01
V5	5.45E-02	2.31E-01	-4.03E-01	2.85E-02	3.05E-01	-4.48E-01	8.46E-03	4.75E-01	-4.91E-01
V11	5.19E-02	2.47E-01	-3.70E-01	1.91E-02	3.39E-01	-4.32E-01	1.70E-02	4.49E-01	-4.47E-01
V13	6.06E-02	2.24E-01	-3.79E-01	2.41E-02	3.25E-01	-4.40E-01	-4.08E-03	5.21E-01	-5.04E-01

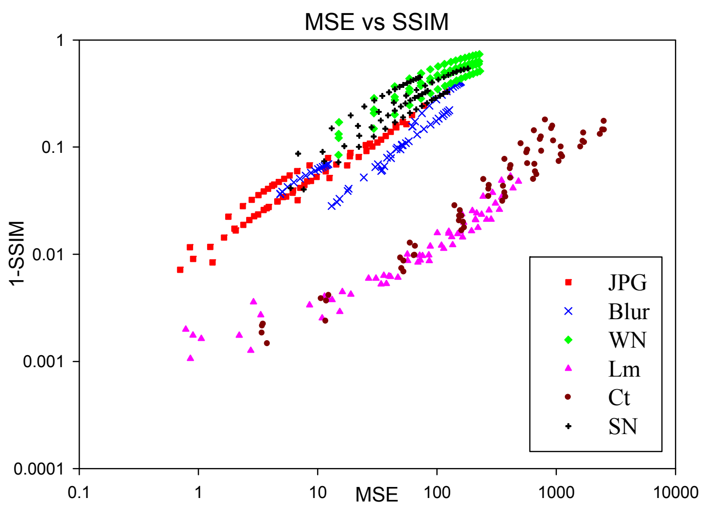

Figure 5. Relationship between 1-SSIM and MSE in logarithmic scale.

In the Fig. 5, we could see the group clearer than figure 4, and fitting result was shown in the table 2. The distribution was separated into two groups clearly that brought much content information in the Fig. 5, but it is not split enough. The coefficients of WN and SN were similarity, but the coefficients of JPG and Blur were different. The degree of image quality depended on noise types at human vision.

PQR did not separate the distribution into groups clearly if we do not use several symbols in the Fig. 6. It's meaning that the different degrees of noise types on the different contents were possible to have the same evaluation value at human vision. When the value of MSE was small, the value of PQR metric would be possibly bigger than neighbor points, like an upward parabolic curve. In the table 3, the coefficients were almost different, but in the Ct noise, we couldn't fit the curve.

Table 2. Parameters of fitting curve after SSIM estimation.

	JPG			Blur			WN		
	a	b	c	a	b	c	a	b	c
V3	1.57E-02	7.85E-01	-2.17E+00	-1.06E+00	5.25E+00	-6.79E+00	-2.09E-01	1.40E+00	-2.43E+00
V5	-2.35E-01	9.67E-01	-1.87E+00	-6.49E-01	1.82E+00	-2.38E+00	-2.55E-01	1.44E+00	-2.11E+00
V11	-3.88E-02	8.03E-01	-2.02E+00	-1.98E-01	1.52E+00	-3.01E+00	-2.40E-01	1.42E+00	-2.22E+00
V13	-4.06E-02	7.71E-01	-2.00E+00	-5.07E-01	2.81E+00	-4.32E+00	-2.40E-01	1.44E+00	-2.27E+00
	Lm			Ct			SN		
	a	b	c	a	b	c	a	b	c
V3	1.81E-01	1.65E-01	-2.98E+00	-1.76E-03	7.96E-01	-3.40E+00	-1.36E-01	1.16E+00	-2.31E+00
V5	1.42E-01	7.75E-02	-2.66E+00	2.69E-02	6.02E-01	-3.06E+00	-1.97E-01	1.23E+00	-1.95E+00
V11	1.70E-01	3.91E-02	-2.72E+00	4.60E-02	5.22E-01	-3.06E+00	-2.04E-01	1.32E+00	-2.20E+00
V13	1.90E-01	6.10E-02	-2.73E+00	-4.94E-03	7.59E-01	-3.16E+00	-1.12E-01	1.10E+00	-2.16E+00

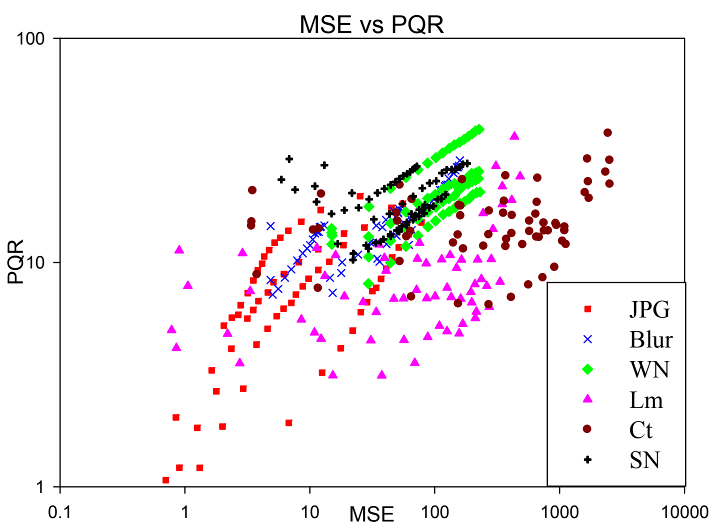

Figure 6. Relationship between PQR and MSE in logarithmic scale.

CWSSIM metric has clear distribution in the logarithmic scale in the Fig. 7, which depended on noise types. If we separate the noise into three groups, structure (JPG and Blur), luminance (Lm and Ct) and gaussian noise (WN and SN), the order of structure group distribution from high to low in the Fig. 7 was V5, V11, V13 and V3 respectively. We could find the coefficient c that was the same sequence at structure group and the gaussian noise group was similar, too. The trend of gaussian noise group was approach linearly in the logarithmic scale, so we could compare the coefficient b between WN and SN. The coefficient b really was similar to each other between WN and SN in the table 4. CWSSIM has a little disadvantage that the worse cases of Ct and Lm have the close estimation value, but they were notably different at human vision.

Table 3. Parameters of fitting curve after PQR estimation.

	JPG			Blur			WN		
	a	b	c	a	b	c	a	b	c
V3	-4.47E-01	1.37E+00	2.94E-01	2.23E+00	-3.45E+00	2.28E+00	-5.38E-02	5.92E-01	4.98E-01
V5	-3.90E-01	1.21E+00	2.58E-01	9.71E-01	-2.34E+00	2.40E+00	4.42E-02	1.39E-01	8.40E-01
V11	-2.12E-01	1.06E+00	6.79E-02	2.40E-01	-3.66E-01	1.06E+00	2.85E-01	-7.45E-01	1.57E+00
V13	1.71E-01	3.01E-01	6.56E-03	-3.42E-03	7.64E-01	-2.23E-01	4.18E-01	-1.24E+00	1.96E+00
	Lm			Ct			SN		
	a	b	c	a	b	c	a	b	c
V3	1.18E-01	-2.59E-01	1.11E+00	X	X	X	6.66E-01	-1.84E+00	2.56E+00
V5	1.15E-01	-2.46E-01	9.58E-01	X	X	X	3.95E-01	-1.19E+00	2.15E+00
V11	1.62E-01	-2.84E-01	7.43E-01	X	X	X	7.34E-01	-2.09E+00	2.57E+00
V13	1.68E-01	-2.77E-01	6.41E-01	X	X	X	6.53E-01	-1.94E+00	2.53E+00

Table 4. Parameters of fitting curve after CWSSIM estimation.

	JPG			Blur			WN		
	a	b	c	a	b	c	a	b	c
V3	3.45E-01	6.13E-01	-3.32E+00	-4.44E+00	2.14E+01	-2.61E+01	-9.30E-02	9.37E-01	-2.52E+00
V5	-9.12E-01	2.37E+00	-1.91E+00	-3.99E+00	9.59E+00	-6.26E+00	-7.96E-02	5.79E-01	-1.24E+00
V11	-6.89E-01	1.96E+00	-1.92E+00	-1.45E+00	6.02E+00	-6.66E+00	-7.94E-02	5.41E-01	-1.37E+00
V13	-5.05E-01	1.95E+00	-2.54E+00	-2.89E+00	1.30E+01	-1.50E+01	-8.83E-02	7.48E-01	-1.92E+00
	Lm			Ct			SN		
	a	b	c	a	b	c	a	b	c
V3	1.90E-01	1.83E-01	-3.28E+00	1.10E-01	1.15E-01	-3.43E+00	-4.15E-02	8.16E-01	-2.54E+00
V5	4.04E-02	1.39E-01	-2.05E+00	3.67E-02	1.17E-01	-2.18E+00	-3.28E-02	4.97E-01	-1.20E+00
V11	7.68E-02	4.75E-02	-2.18E+00	4.35E-02	1.14E-01	-2.38E+00	-1.19E-01	7.21E-01	-1.57E+00
V13	2.06E-01	-6.00E-02	-2.66E+00	2.89E-02	2.94E-01	-2.94E+00	-7.44E-02	8.13E-01	-2.18E+00

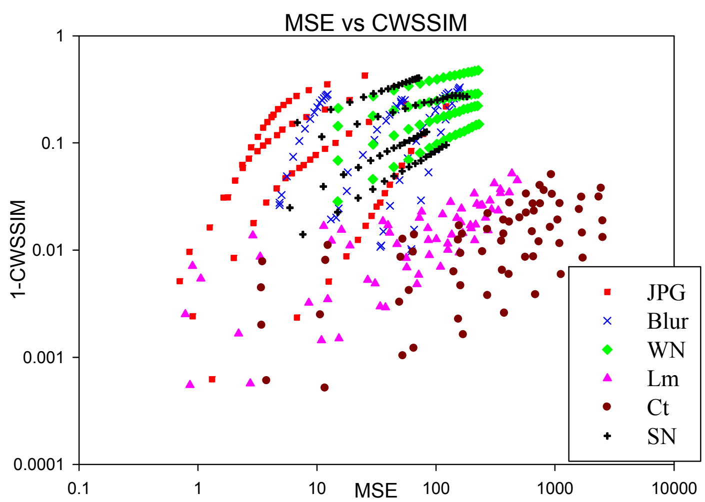

Figure 7. Relationship between 1-CWSSIM and MSE in logarithmic scale.

The goal of objective metrics is exactly to estimate human perception. Although the objective metrics are not perfect, they still have advantages at the structure or luminance estimation.

5. CONCLUSION

In this paper, we proofed the objective metrics on the several images with several noise types. Some objective metrics depend on noise, some depend on content. All metrics have advantage by themselves. We fitted the curves of evaluation value in the results and CWSSIM metric had clear distribution. We try to fit the curves of CWSSIM, but the coefficients of fitting results were not perfection. If we could separate the curves into content and noise factors, we could forecast the estimation value of human vision by MSE. In the future, we will find the fitting method to perfectly forecast the estimation value of human vision.

REFERENCE

[1] M. H. Pinson, and S. Wolf, “New standardized method for objectively measuring video quality,” Ieee Transactions on Broadcasting, 50(3), 312-322 (2004).

[2] Greg Cermak, Philip Corriveau, Mylène C.Q. Farias et al., “FINAL REPORT FROM THE VIDEO QUALITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF VIDEO QUALITY ASSESSMENT, ” (2003).

[3] M. P. Sampat, Z. Wang, S. Gupta et al., “Complex Wavelet Structural Similarity: A New Image Similarity Index,” Ieee Transactions on Image Processing, 18(11), 2385-2401 (2009).

[4] 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, 26(1), 98-117 (2009).

[5] “Methodology for the subjective assessment of the quality of television pictures,” ITU-R BT.500-11 (2002).

[6] B. W. Andrew, “Toward a perceptual video-quality metric.” 3299, 139-147 (1998).

[7] A. B. Watson, J. Hu, and J. F. McGowan, “Digital video quality metric based on human vision,” Journal of Electronic Imaging, 10(1), 20-29 (2001).

[8] A. C. Brooks, X. N. Zhao, and T. N. Pappas, “Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions,” Ieee Transactions on Image Processing, 17(8), 1261-1273 (2008).

[9] W. Zhou, and E. P. Simoncelli, “Translation Insensitive Image Similarity in Complex Wavelet Domain.” Proc. IEEE, 2, 573-576 (2005).

[10] L. Mei Hwan, O. Ee Ping, L. Weisi et al., “Comparison of Video Quality Metrics on Multimedia Videos.” Proc. IEEE, 457-460 (2006).