Image quality evaluation for motion picture compressed by H.264/AVC

Naoya Matsumoto; Graduate School of Advanced Integration Science, Chiba University, Chiba, Japan
Tokiya Abe and Hideaki Haneishi; Research Center for Frontier Medical Engineering, Chiba University, Chiba, Japan

Abstract

When a color video system with highly accurate color reproduction is designed, a good measure for evaluating the image quality including the color reproduction is required. Especially, in codec of motion pictures a good measure is strongly desired to determine a proper compression rate (bit rate). In this paper, we focus on the image quality of motion pictures compressed by H.264/AVC codec which is receiving increased attention. In this study, we investigated the usefulness of S-CIELAB. At first, S-CIELAB color difference calculation was applied to the frames of motion pictures after H.264/AVC codec. The performance was better than CIELAB color difference but not satisfactory. We then limited the region of calculation of CIELAB color difference to the smooth regions where compression error tends to attract attention of observers. Experimental results showing that the modification is promising are obtained.

Introduction

When a color video system with highly accurate color reproduction is designed, a good measure for evaluating the image quality including the color reproduction is required. Especially, in codec of motion pictures a good measure is strongly desired to determine a proper compression rate (bit rate). In this paper, we focus on the image quality of motion pictures compressed by H.264/AVC codec which is receiving increased attention.

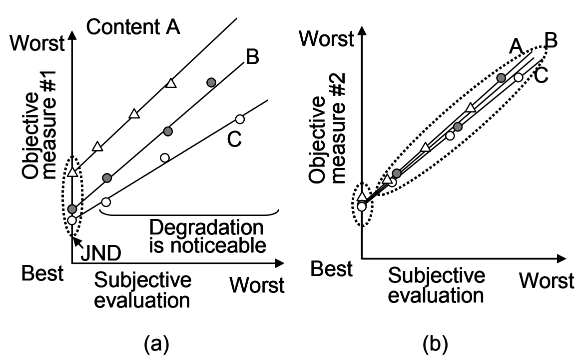


Fig. 1 Schematic illustration of the relationship between subjective and objective evaluation values. (a) Poor objective evaluation meature. (b) Good objective evaluation measure.

The relationship between subjective evaluation and objective evaluation is schematically illustrated in Fig. 1. In this figure, horizontal axis denotes subjective evaluation value (SEV) and the vertical axis denotes objective evaluation value (OEV). The left end of the horizontal axis corresponds to that the observer does not notice the degradation at all, and the righter side the plot is located at, the stronger the observer feels the degradation. The bottom end of the vertical axis corresponds to no degradation, and the upper side the plot is located at, the stronger the degradation is. In this schematic illustration, the measure of OEV is different between (a) and (b). In these figures, three kinds of contents, A, B and C are plotted. Cross section of the approximation line with the vertical axis denotes the OEV for the degradation just noticeable for the observer. We call this value OEVJND (objective evaluation value at just noticeable difference) in this paper. The region righter than the origin denotes that the degradation is noticeable for the observer. We call this region the degradation-noticeable region (DNR) and the OEV in this region OEVDNR in this paper. In Fig. 1(a), both OEVJND and OEVDNR are uneven among the contents. On the other hand, in Fig. 1(b), both values are similar among the contents. So we can say that (b) is a better metric.

OEVJND can be used to determine the compression level at which the degradation is just noticeable and thus is useful in the case that high image quality is required. On the other hand, the OEVDNR is useful to determine the compression level in the case that the required level for the image quality is relatively low. We aim to develop a content-independent image quality metric with respect to both OEVJND and OEVDNR.

There are many previous works in the image quality evaluation [1-7]. In the conventional evaluation methods, MSE (mean square error) or PSNR (peak-signal to noise ratio) has often been used as a measure in the field of video compression. However, it does not necessarily correlate to the subjective image evaluation by human observers. On the other hand, as a conventional measure for color difference, the difference in the CIELAB color space is used. However, in evaluating color difference between images the pixelwise color difference still does not correlate the subjective impression. Fig. 2 shows an evaluation example using PSNR and CIELAB ΔE94. Images (a) and (c) are still images extracted from the original motion pictures which are used in the experiment described later. Images (b) and (d) are H.264/AVC-compressed and decoded images. As visual impression, (d) is heavily degraded while (b) is not degraded so much. However, both the PSNR and ΔE94 show similar values, which means that these measures are not correlated well with human observers.

Spatial CIELAB (S-CIELAB) [4] is a measure for calculating the color difference between images which takes spatial sensitivity characteristics of the human visual system (HVS) into account and therefore has better correlation with subjective evaluation. Wang et al are proposing sophisticated methods taking advantages of known characteristics of the HVS [3]. However, those methods seem to be weighted on the luminance.

In this study, we investigated the usefulness of S-CIELAB. At first, S-CIELAB color difference calculation was applied to the frames of motion pictures after H.264/AVC codec. The performance was better than CIELAB color difference but not satisfactory. We then limited the region of calculation of S-CIELAB color difference to smooth regions where observers tend to notice the compression error. The result was better than the simple S-CIELAB color difference. Details of this study are described in this paper.

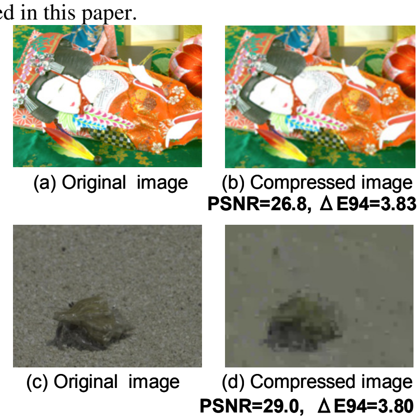


Fig. 2 Examples of objective evaluation using PSNR and DE94.

Color difference in S-CIELAB

S-CIELAB is now well-known in the field of color science and technology. The processing flow of calculation the S-CIELAB color difference is briefly shown in Fig. 3. Original R, G, B images and degraded R, G, B images are respectively transformed to the opponent color components, A, T, D [8]. Then, the spatial frequency filtering modeling the contrast sensitivity function (CSF) of HVS as shown in Fig. 4 is performed in A, T, and D component, respectively. Those filtered images are transformed to X, Y, Z colorimetric values and then L*, a*, b* values. Finally the color difference is calculated pixel-by-pixel and then the mean value or the median is calculated.

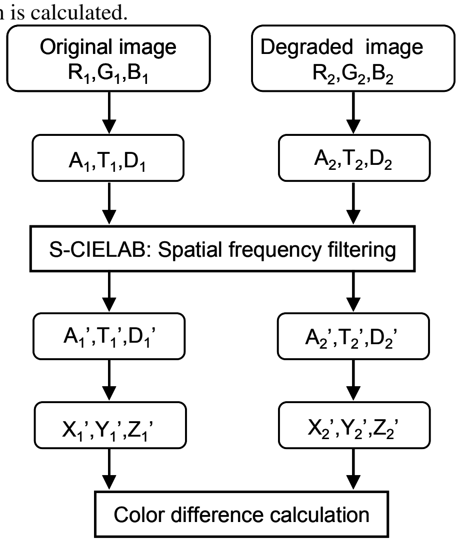


Fig. 3 Flow of calculation of S-CIELAB color difference.

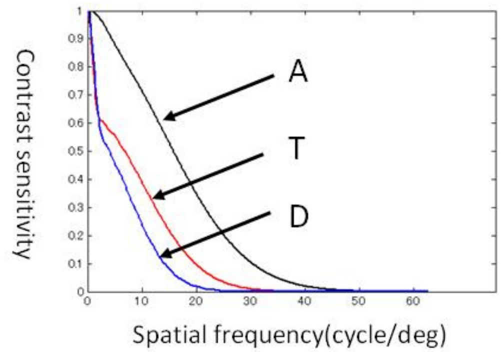


Fig. 4 Contrast sensitivity function model for the opponent color component of human observer. These shapes are used as spatial frequency filter in calculation of S-CIELAB color difference.

First evaluation experiment and result

Method

We conducted two kinds of subjective evaluation experiments. One is to find the OEVJND and the other is to find the OEVDNR. In the OEVJND finding experiment, we first prepare several levels of degradation by compression so that a just noticeable compression level (JNCL) exists between those levels. After conducting the observer evaluation experiment and estimating the JNCL, the OEV is calculated in each content. If the calculated OEVJND would be similar each other between contents, it could be used as a measure giving the JNCL independent from video content. Fig. 5 shows the flow to investigate the performance of measure.

In the OEVDNR finding experiment, many degraded images with different bit rates in which observers notice the image degradation are prepared. In the rating experiment with those images, the double stimulus continuous quality scale (DSCQS) method was used. SEV is calculated from the experimental data. On the other hand, OEV is calculated in each content. If the OEVDNR vs SEV characteristics would be similar each other between contents, correlation coefficient between two values becomes high.

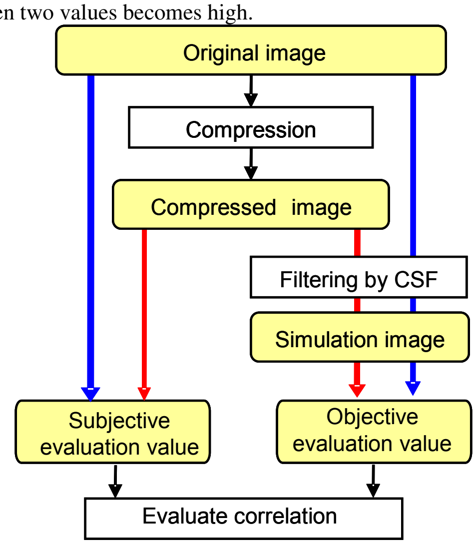


Fig. 5 Flow to investigate the performance of objective evaluation measure.

Evaluation experiment

In the OEVJND finding experiment, four kinds of movies were used. Some frames of the contents are shown in Fig. 6.Each movie has 720x540 pixel image size. The length of the original movie is 180 or 120 frames. Features of each content are as follows.

"glass" (180 frames) A table on which four glasses are put is slowly rotating. The composition is relatively simple. There are many smooth regions in a sense that there is no spatially abrupt and large change of pixel values.

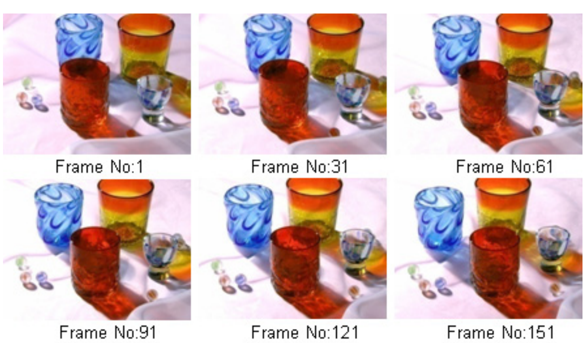
"japan" (180 frames) A battledore in which a colorful picture is painted is slowly panned from left to right. The composition itself is complex but the target keeps still. So the motion itself is simple.

"crab" (120 frames) A helmet crab is slowly walking on the beach. A wave comes to the crab halfway. The color is poor through the content. The camera is fixed.

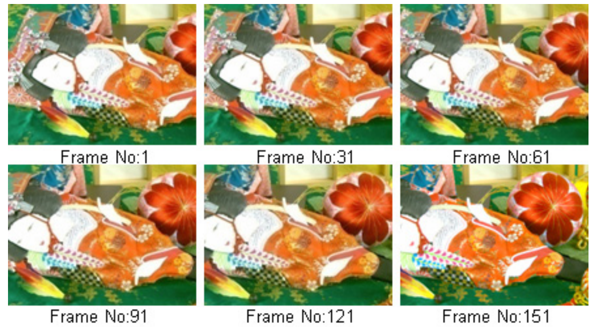
"okinawa" (120 frames) Women wearing traditional costume are dancing outside in relatively fast motion.

For above four contents, four movies with difference compression levels were generated by H.264/AVC SDK (MainConcept). Thirty observers with normal vision participated in the experiment. A high-vision master monitor, HTM-1980 (Ikegami) with the resolution of 1920x1080 pixels, and non-compression video recorder, UDR-2E (Keisoku Giken) were used. In the evaluation experiment, a H.264/AVC codec image and an original image were sequentially presented to observers. The order of those images was at random and the observer was asked to select the image which he or she felt as degraded image. Short movies of four or six seconds were displayed repeatedly. Observation time was not limited. For each compression level, the ratio that observers selected a correct image was calculated. For each degraded image, mean color differences measured by the CIELAB and S-CIELAB are calculated. We defined the JNCL by 75% correctly answered ratio and calculated the corresponding color difference by the probit analysis.

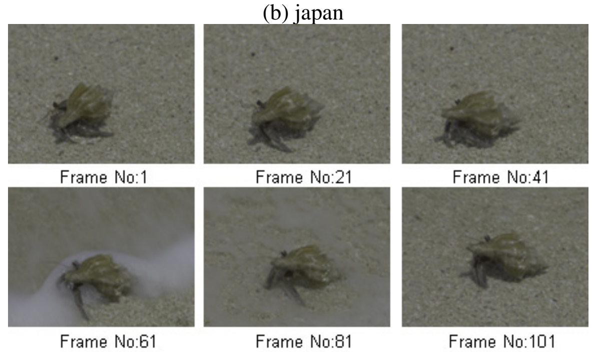
In the OEVNDR finding experiment, the same apparatus was used as the OEVJND finding experiment. However, the following things are different. (1) Two more contents were added, i.e. six contents were used. (2) 12 observers participated in this experiment. (3) The DSCQS method was used.



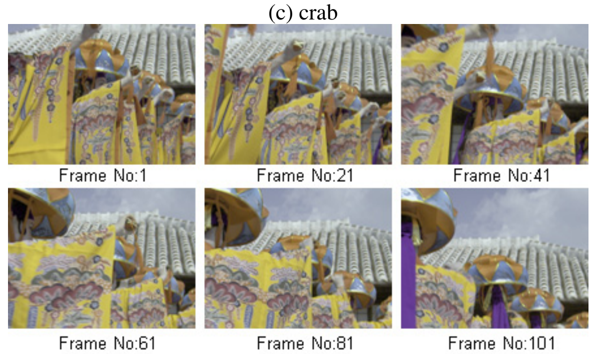
(a) glass



(b) japan



(c) crab



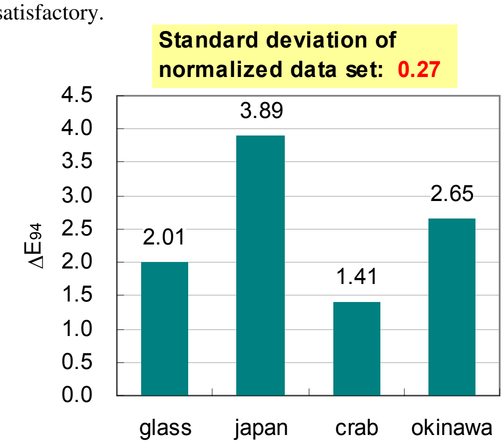
(d) okinawa

Fig. 6 Motion pictures used in the OEVJND finding experiment. Six frames are extracted from the original movie and presented.

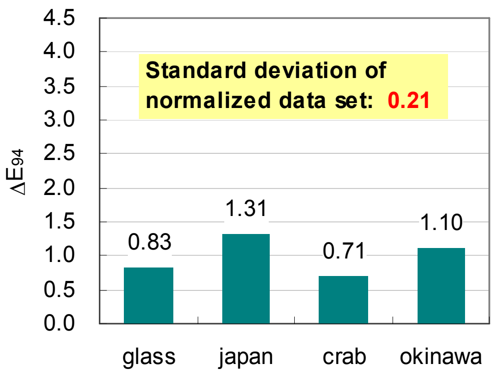
Result

(1) OEVJND finding experiment

For four video contents, JNCLs were determined and the color differences of those compressed images and original images were calculated in CIELAB and S-CIELAB color space. The results are shown in Fig. 7. In Fig. 7(a), the CIELAB color difference ranges from 1.41 to 3.89. All these values were divided by the maximum value 3.89 and then the standard deviation was calculated. The result was 0.27. This value represents the relative variation of the color differences between contents. Fig. 7(b) shows the result in the case of S-CIELAB. All values are reduced from the CIELAB color difference. This is natural because the low pass filtering by CSF of HVS produces smoothed images for both original and codec images then the degree of similarity increases. The standard deviation of the normalized data set was 0.21. This is smaller than the CIELAB case. Especially in "japan" case, the color difference changes from 3.89 to 1.31. "japan" has many complex elements in the frames. Large color differences caused in the CIELAB calculation would have decreased markedly in turn in the S-CIELAB case. This result shows the effectiveness of the use of S-CIELAB in some degree. However, it still has variation and not satisfactory.



(a) CIELAB



(b) S-CIELAB

Fig. 7 Color differences corresponding to JNCL. Standard deviation of normalized data set is also shown.

(2) OEVDNR finding experiment

The correlation coefficient between OEVs and SEVs was calculated. All plots for six video contents were used in the calculation. In this experiment, not only CIELAB but also MSE and PSNR were calculated and compared with S-CIELAB. The result is shown in Table 1. As shown, S-CIELAB shows much better performance than the other three measure.

Table 1 Correlation between OEVDNR and subjective evaluation value.

Measure	Correlation coefficient
MSE	0.54
PSNR	0.55
CIELAB	0.58
S-CIELAB	0.83

Modification of S-CIELAB calculation

Method

Although S-CIELAB performs better than CIELAB, MSE, PSNR, it is still not satisfactory. Thus we tried the modification of the S-CIELAB. According to the comments of observers, they notice the degradation by H.264/AVC image compression in the region where pixel values are spatially smooth, i.e. there is no strong edge. It means that the difference in such region rather than the whole image directly correlates to the observer's notice. So, we limited the region of calculation of color difference to the smooth region.

The smooth region was extracted by the following processing. The achromatic component of the original color image is first extracted. Then, LOG (Laplacian of Gaussian) filtering is carried out. In this filtering, a two-dimensional Gaussian kernel given by the following form is fist convolved.

$$G(x,y) = \frac{1}{\sqrt{2\pi\sigma}} \exp[-(x^2+y^2)/\, 2\sigma^2]$$

Here σ was set to 5 [pixel]. The kernel size was 31 x 31 pixels. After the above filtering, Prewitt edge enhancement was performed both horizontally and vertically. After taking an average of two directional edge enhanced images, pixel values are normalized linearly so that the maximum and minimum of the filtered images become 255 and 0, respectively. Finally, the image was binarized by a proper threshold. In this study, the threshold was set to 15 empirically. The results are shown in Fig. 8. Here black parts represent edge regions, and white parts represent smooth regions.

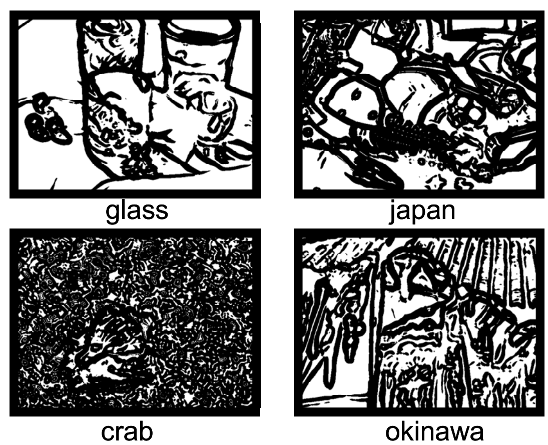


Fig. 8 Extracted smooth regions in frame in each content (white part).

Result

(1) OEVJND finding experiment

The result of calculation of CIELAB color difference in the limited regions is shown in Fig. 9. The standard deviation of the normalized data set was reduced to 0.17. Especially, in "japan" color difference reduces from 1.31 to 1.07 and in "Okinawa", it reduces from 1.10 to 0.96. These changes contribute to the uniformity of the evaluation values. In these movies, large color differences exist in the complex pattern. By limiting the calculation region, these values are eliminated and average color differences would have decreased.

For comparison, the same limitation was applied to CIELAB case and the result was 0.25 although the graph is not shown here. As a result the region-limited S-CIELAB color difference was an effective measure among the tried ones in this study.

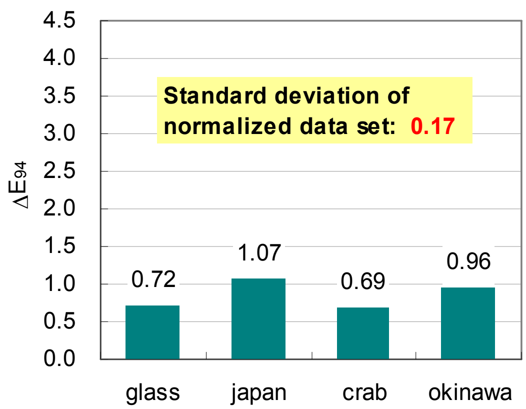


Fig. 9 Color differences corresponding to JNCL. Standard deviation of normalized data set is also shown.

(2) OEVDNR finding experiment

The effect of region limitation of calculation was also studied in the OEVDNR. Both CIELAB and S-CIELAB, the region limitation shows the improvement of the performance. However, the degree of improvement remains small and further modification of the measure is needed.

Table 2 Comparison between whole region calculation and limited region calculation.

Measure		Correlation coefficient
CIELAB	whole region	0.58
	limited region	0.63
S-CIELAB	whole region	0.83
	limited region	0.86

Conclusion

We have studied on measures for evaluating the image quality of motion picture compressed by H. 264/AVC. Mainly S-CIELAB and its modification were investigated. Criterion of goodness of measure was the correlation with the subjective evaluation value by human observers. We first applied the S-CIELAB color difference calculation to the frames of motion pictures after H.264/AVC codec. The performance was better than CIELAB color difference but not satisfactory. We then limited the region of calculation of the CIELAB color difference to the smooth regions where compression error tends to attract attention of observers. We obtained experimental results showing that the modification is promising.

References

[1] Elaine W. Jin, Xiao-Fan Feng and John Newell, IS&T's 1998 PICS Conference , pp.154-158(1998)

[2] Garrett M. Johnson, Mark D. Fairchild, Color Research and Application ,vol. 28 ,No. 6, pp. 425-435 (2003)

[3] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, IEEE Trans. on Image Processing, vol.13, No.4, pp.600-612 (2004)

[4] X.Zhang, B.A.Wandell, SID Journal, vol.5, pp.61-63 (1997)

[5] X.Tong, D.Heeger and C.van den B.Lambrecht , Part of the IS&T/SPIE Conference on Human Vision and Electronic Imaging Ⅳ, vol.3644 , pp.185-196(1999)

[6] S. Winkler, "A perceptual distortion metric for digital color video," Proc. SPIE, vol. 3644, pp. 175–184, 1999.

[7] S. Daly, "The visible differences predictor: an algorithm for the assessment of image quality," SPIE Vol. 1666, pp.2-15 (1992)

[8] S. L. Guth, R. W. Massof and T.Benzschawel, J. Opt. Soc. Am., Vol.70, No. 2(1980)

Author Biography

Naoya Matsumoto received his BS degree in 2008 from Chiba University, Japan. He is currently a master course student in Graduate School of Advanced Integration Science, Chiba University, Japan.

Hideaki Haneishi received his M.S. degree in 1987 and his Ph.D. degree in 1990 from the Tokyo Institute of Technology, Japan. Since 1990, he has been working with the Department of Information and Computer Sciences, Chiba University, Chiba, Japan. He was a visiting research scientist at the Department of Radiology, University of Arizona, from 1995-1996. He is currently a professor at the Research Center for Frontier Medical Engineering, Chiba University. His research interests include color image processing, image reconstruction and medical image processing.