The Fourth International Workshop on Image Media Quality and its Applications, IMQA2010
May 13-14, 2010, Tokyo, Japan

VALIDATION OF COMPRESSION DISTORTION METRICS FOR BLOCKINESS, BLUR AND COLOUR BLEEDING

Amal Punchihewa[†], *Donald Bailey*[†], *Alexander Keerl*[††],

[†] Massey University, Palmerston North, New Zealand
[††] University of Applied Sciences and Arts, Göttingen, Germany

g.a.punchihewa@massey.ac.nz

ABSTRACT

There are few objective metrics that distinguish individual compression artefacts which are rapid to evaluate and correlate with human perception. The authors had proposed compression artefact distortion metrics that objectively and separately measure blockiness, blur and colour bleeding artefacts introduced by image compression codecs. The approach used synthetic test patterns specially designed for each artefact. Evaluation of quality in multimedia applications requires that the distortion metrics be correlated to human perception. One limitation of the above three metrics was that no correlation was established to human perception. This paper investigates this correlation, using SSIM as a proxy for human perception, demonstrating a strong correlation. It also highlights the limitation of SSIM in evaluation of quality of compressed colour images.

1. INTRODUCTION

Image and video communication technologies have been advancing at a rapid pace for the last two decades. Mpeg-4 has been deployed in image communication applications yielding higher compression efficiency over mpeg-2 for a given quality [1]. The compression gain is however at the cost of higher encoder and decoder complexities [1]. Ultra high definition television is being researched and developed as a next generation high definition television system. Demonstration systems already operate with image resolution of 4320 x 7680 pixels per frame progressively scanned at 60 frames per second. [2]. 10 bits per pixel would result in a raw data rate of 20gbits/s. The storage and transportation of these signals necessitate a substantial level of compression. There is always need for improved image and video data compression techniques. This need is more pressing with the 3-d television capture and processing now being researched as the next revolution in television broadcasting [3]. Here it is important for the artefacts to be matched between left and right views to minimise stereo fusion instability and viewer fatigue. Scalable and distributed video coding is being researched for future deployments [4], [5]. Thompson corporate research has performed an objective assessment using a software scalable video coder (svc) and observed rate distortion curves [6]. Svc and 3-d television applications demand high performance image and video coders and decoders.
For most digital image and video codecs, increasing bit rate performance has been achieved at the cost of increasing complexity in techniques and implementations. McCann [7], [8] claims that codec development achieves an average reduction in bit rate of 15% per annum. Consumer products that are available today include a broad range of new imaging applications. Jpeg and mpeg codecs are in common use in digital photography, digital tv, photographic quality printers and DVDs. Most of these codecs use block-transform based techniques and consequently produce visually annoying compression artefacts. World wide web (www), television and multimedia networks also deploy numerous codecs. Both service providers and content providers and end users are using codecs to transfer content between production facilities and to deliver, interact and manage content. However there are problems due to errors in colour, coding and optics related to picture quality, especially from the television engineers' perspective, where the picture quality is well below that of conventional broadcast pictures. Although picture quality evaluation is a complex subject from a professional broadcasters' point of view [9], it needs to be performed to ensure an agreed level of service.
Due to limited bandwidth and storage constraints, the quantity of data needs to be reduced for transporting and storage. Source coders are being deployed for data reduction in communication systems. Codecs have been in use for television since its inception in the 1930s. For example, monochrome television broadcasting has an inherent two to one compression due to the use of interlace scanning. In 1960s, colour television broadcasting was developed to be compatible with monochrome television broadcasting. Analogue colour television broadcasting has six to one compression.
In recent times, it is observed that the time period of system improvement from introduction to maturity of an

image codec is about 5~8 years [9]. Furthermore, on average, the long term gain in quality efficiency is about 5~10% per year [9]. Hence, developers need to fine tune and optimise their codecs rapidly to recover the development costs.

2. DESCRIPTION OF THE PROBLEM

In digital television broadcasting, video streaming and other multimedia communications, image and video are the dominant components. With limited communication bandwidth and storage capacity in terminal devices, it is necessary to reduce the quantity of data or the data rate using digital codecs. The sub-sampling techniques and quantisation used in image and video compression codecs introduce distortions known as artefacts. *The digital fact book* defines artefacts as “particular visible effects, which are a direct result of some technical limitation” [10].

There are many artefact mitigation techniques proposed by researchers which are discussed in reference [12]. However, no rapid objective tools are available to check the effectiveness of those mitigation techniques. There is no world standard yet related to compression artefact distortion measurements despite immense work carried out by the video quality expert group (VQEG). VQEG recognizes the structural similarity index (SSIM) as a perceptual image quality metric. However, the SSIM provides only a single global measure of quality, and is unable to assess the levels of individual artefacts. In summary, there are no objective tools available to carry out rapid testing for individual compression distortions and colorimetry.

The authors proposed, researched and published a framework for the objective evaluation of compression artefacts [13], [14], [15], [16], [17] within this framework, a series of metrics were proposed that were able to assess individual artefacts such as blockiness, ringing, blur, colour bleed. However, one issue with the proposed metrics was the correlation with human perception had not been determined. The purpose of this research was to determine and relate the metric associated with each artefact with the perceived quality of the compressed images.

2.1 Definition of colour components and colour space conversion

In digital image and video, the *R*, *G*, and *B* signals are converted to colour difference signals *Cr* and *Cb*, for use in current video communication interfaces [11]. These are defined as,

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.16875 & -0.33126 & 0.500 \\ 0.500 & -0.41869 & -0.08131 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

The two colour difference signals are used to modulate the colour sub-carrier using quadrature modulation. They can therefore be treated as two components of a vector, where the angle corresponds to the dominant colour, or hue, and the magnitude is the strength of the colour or saturation:

$$hue = h = \tan^{-1}\left(\frac{Cr}{Cb}\right) \quad (2)$$

$$saturatioon = s = \sqrt{Cr^2 + Cb^2} \quad (3)$$

Hue, saturation and luminance defined above in equations (1), (2) and (3) are compatible with current analogue measurement systems. They are based on the human perception system. Hence the red, green and blue components of the test pattern are converted to hue, saturation and luminance prior to the calculation of colour artefact metric. Two colour difference signals defined in (2) and (3) can be modified to suite any other colour space such as NTSC or digital CCIR 601 by changing the two scaling factors in equation (2) and (3).

2.2. Definition of artefact metrics

Colour bleeding is introduced by digital codecs at colour boundaries or edges. Coding colour bleed is identified here as the leakage of colour from one region of colour to another at colour boundaries. Fig. 2 and Fig. 3 show colour bleeding when a digitally coded image having colour is reconstructed.

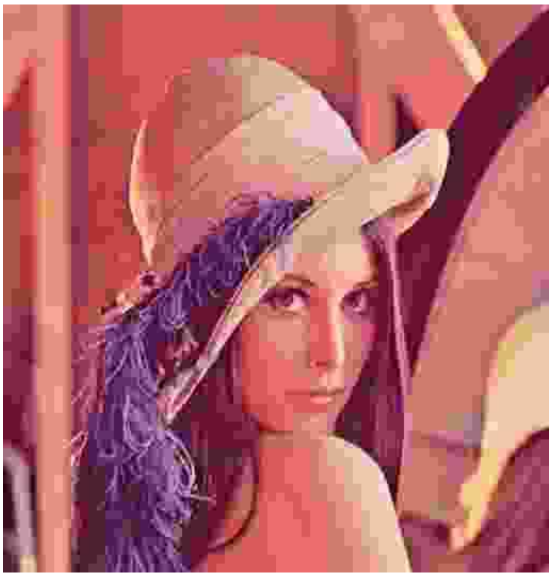

Figure. 2 Coding artefacts resulting from a JPEG codec at a compression ratio of 113 on Lena test image

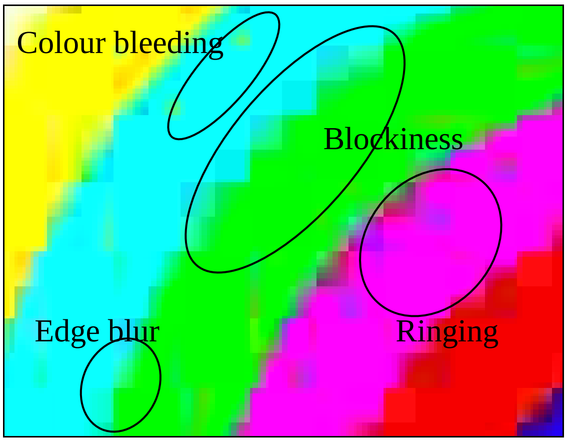

Figure. 3 Coding artefacts around the colour edges resulting from a JPEG codec on 1:120 compression

The colour bleeding therefore appears as a consequence of spreading of hue angle. The higher the leakage of colour, the higher the visibility of colour error and value of the coding colour bleed. The coding colour bleed is defined as the error in hue as follows.

Consider a test pattern containing N distinct colours. Let the mean hue value of colour region r in the original pattern be $\overline{H_r}$ and the mean hue value of the corresponding colour in the reconstructed pattern be $\hat{H_r}$, then the coding hue bleed can be defined as,

$$CHB = \frac{\sum_{r=1}^{N} \left| \overline{H_r} - \hat{H_r} \right|}{N} \qquad (4)$$

Ringing always occurs at edges and blur generally occurs at edges. Since the concern is on the blur occurring at edge, the rest of the paper concentrates on edge blur. Ringing is an undesirable visible effect around edges. Many codecs transform the pixel values into the frequency domain where the transformed coefficients are then quantised. Quantisation errors resulting from this approach give rise to ringing around sharp discontinuities in the test pattern.

An ideal sharp edge contains components at all frequencies. Any change in the amplitude of any of these components will result in ripples in the pattern with amplitude corresponding to the error. As a result of energy compaction in a codec, many of the high frequency components are very small, and get quantised to zero. This loss of high frequency components leads to blur in reconstructed image.

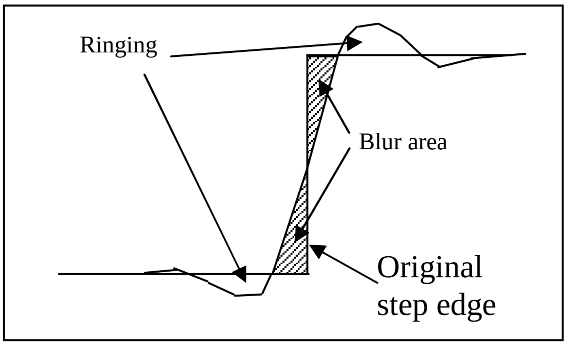

Figure. 4 Ringing and edge blur at an edge of a one-dimensional signal.

Ringing and edge blur are defined in Figure 4. The region between the first crossings on each side of the edge transition is defined as the edge blur region. Outside of this, from the start of the first overshoot on each side, the errors are classified as ringing. To obtain a measure of blur, consider the shaded area in Figure 4. The greater the blur, the larger will be the shaded area. By dividing the area by the step height, a measure of average blur width can be obtained. In a similar manner, the area between the ringing signal and ideal signal provides a measure of the severity of ringing. With sampled data, an ideal step edge would involve a transition between two pixels.

The edge blur and ringing are therefore quantified as

$$\text{Edge blur} = \frac{\sum_{\text{blur\_region}} |\text{Error}|}{\text{Step Size}} \qquad (5)$$

$$\text{Ringing} = \frac{\sum_{\text{ringing\_region}} |\text{Error}|}{\text{Step Size}} \qquad (6)$$

Blockiness can be expressed as the discontinuity in amplitude per block boundary pixel in the test pattern. The higher the value of the blockiness, the higher the visibility of block structure.

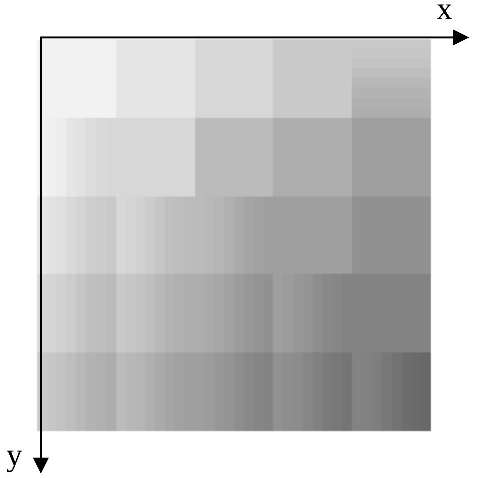

Figure. 5 Blockiness resulting from JPEG codec at high compression ratio.

Consider an *m*x*n* luminance information *Y*, reconstructed from a 8x8 block coded test pattern having *M* rows and *N* columns. As shown on Fig. 5, both vertical and horizontal edges can be observed at regular pixel interval of 8 because of the 8x8 block processing. Consider row y, along line y, the horizontal blockiness can be calculated as,

$\sum_{x} \left| Y[x,y] - Y[x+1,y] \right|$ where x= 8, 16, 24,…,(N-8). This computation is repeated for all rows from y=1 to M. The total of the vertical blockiness *VB* can be written as,

$$VB = \sum_{y=1}^{M} \sum_{x} \left| Y[x,y] - Y[x+1,y] \right| \tag{7}$$

This results from $\frac{(N-8)}{8}.M$ block boundary pixels. Similarly, the horizontal blockiness *HB*,

$$HB = \sum_{x=1}^{n} \sum_{y} \left| Y[x,y] - Y(x,y+1] \right| \tag{8}$$

Results from $\frac{(M-8)}{8}.N$ block boundary pixels.

Both the *HB* and the *VB* can be combined and normalised by dividing the number of boundary pixels. Hence the blockiness per boundary pixel *B* can be expressed as,

$$\begin{aligned} B &= \frac{HB+VB}{\frac{N-8}{8}.M + \frac{M-8}{8}.N} \\ &= \frac{4.(HB+VB)}{NM - 4(M+N)} \end{aligned} \tag{9}$$

For blockiness calculations, the luminance component derived from *R*, *G* and *B* primary colour components is used from both original and reconstructed test patterns.

Formal subjective tests are very expensive due to the resources required for the experiments (for example using the ITU standard test procedures), time, and human involvement. However, the SSIM provides a HVS based image quality metric that has been shown to correlate well with subjective assessment in the form of a mean opinion score (MOS). This enables the SSIM to be used as a proxy for the MOS, rather than using subjective testing. Like the metrics proposed in publications [13], [14], [15], the SSIM is also based on a fully known reference image quality measurement method. However, the SSIM is unable to distinguish individual artefacts.

This paper presents the validation of three artefact distortion metrics proposed by the authors. This was done by using the accompanying synthetic test patterns that were designed previously in [13], [15], [16] and shown in figure 9 of the appendix by carrying out series of experiments using the Matlab™ JPEG, Irfanview™ JPEG and Irfanview™ JPEG2000 codecs. A data set consisting of 128 synthetic test images were deployed with above three image codecs in the experiments.

2.3 Structural Similarity Based Image Quality Assessment

The SSIM for pair of images is based on three components [18], [19] a luminance comparison, a contrast comparison, and a structure comparison. The luminance comparison function between two images i_1 and i_2,

$$l(I_1, I_2) = \frac{\left(2\mu_{I_1}\mu_{I_2} + C_1\right)}{\left(\mu_{I_1}^2 + \mu_{I_2}^2 + C_1\right)} \tag{10}$$

Where μ_i is the mean of image *i*, and c_1 is a constant to avoid instability with division by zero. The contrast comparison function between the two images is

$$c(I_1, I_2) = \frac{\left(2\sigma_{I_1}\sigma_{I_2} + C_2\right)}{\left(\sigma_{I_1}^2 + \sigma_{I_2}^2 + C_2\right)} \tag{11}$$

Where σ_i is the standard deviation of image *i*, and c_2 is again a constant to avoid instability with low contrast images. The structure comparison function is

$$s(I_1, I_2) = \frac{\left(\sigma_{I_1 I_2} + C_2/2\right)}{\left(\sigma_{I_1}\sigma_{I_2} + C_2/2\right)} \tag{12}$$

The SSIM is then given as the product of these three terms [18], [19]

$$SSIM(I_1, I_2) = \frac{\left(2\mu_{I_1}\mu_{I_2} + C_1\right)\left(2\sigma_{I_1 I_2} + C_2\right)}{\left(\mu_{I_1}^2 + \mu_{I_2}^2 + C_1\right)\left(\sigma_{I_1}^2 + \sigma_{I_2}^2 + C_2\right)} \tag{13}$$

For colour images, colour components are used in place of luminance component.

3 VALIDATION OF DISTORTION METRICS

Three series of experiments were carried out to evaluate SSIM and blockiness, blur, and colour bleeding distortion metrics over the full range of compression available with three image codecs. Each of the experiments is described below.

3.1 Validation of blockiness artefact metric

The sine-squared grey scale-radial test pattern was designed to highlight blockiness artefacts due to compression [13]. The SSIM was computed on the same set of images used to calculate the blockiness by exercising the three image codecs over the full possible range of quality factors from 1 to 100. A scatter plot was made between SSIM and blockiness metric [13] as shown in figure 6.

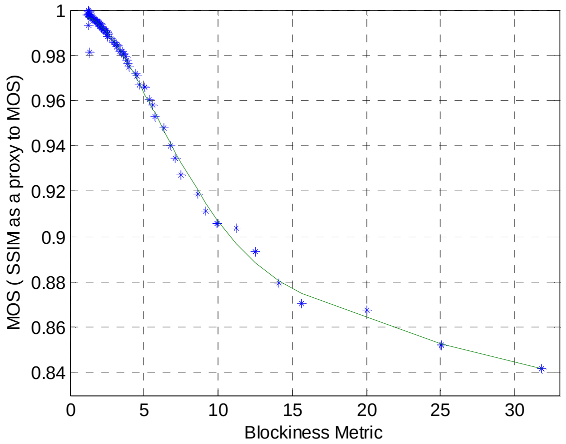

Figure 6 Scatter plot of ssim against the blockiness metric b1 for matlabtm jpeg codec on sine-squared grey scale-radial test pattern

Figure 6 shows that blockiness metric strong correlation to SSIM except for extremely high compression. However, this level of compression is not practical for images as they introduce distortions beyond acceptable levels for image communication. This means that, blockiness metric computed on the test pattern correlates well with human perception (as indicated by the structural similarity index).

3.2 Validation of blur artefact metric

The monochrome-rings test pattern was designed to highlight blur artefact due to compression [13]. The SSIM was computed on the same set of images used to calculate the blur by exercising the three image codecs over the full possible range of quality factors. A scatter plot was made between SSIM and blur metric as shown in figure 7.

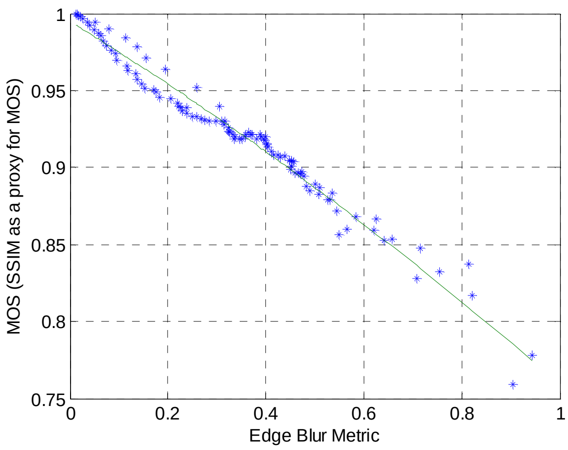

Figure 7. scatter plot of ssim against the blur metric for Matlab™ JPEG codec on monochrome-rings test pattern

Figure 7 shows strong correlation between the blur metric and SSIM except for extremely high compression. As blur metric proposed shows correlation for most of possible compressions, blur metric provides a good measure of blurring artefact visible to the human visual system.

3.3 Validation of colour bleeding artefact metric

The honeycomb test pattern was designed to highlight colour bleeding artefacts due to compression [15], [16]. The SSIM was computed on the same set of images used to calculate the hue spread by exercising the three image codecs over the full possible range of quality factors. A scatter plot was made between SSIM and hue spread metric as shown in figure 3.

Figure 8 shows that colour bleeding measured as *hue spread* has a strong correlation to SSIM for all possible compressions. This tends to suggest that all 128 compressed images are good and near-equal quality. However, informal subjective tests revealed that more than 20 of images are poor quality and about 9 of them are very poor quality. The main problem with the SSIM for this application is that it is based on luminance. The honeycomb test image has constant luminance, so the SSIM is unable to detect the distortions. This shows that SSIM metric requires more modifications before it applies to the colour image quality assessment.

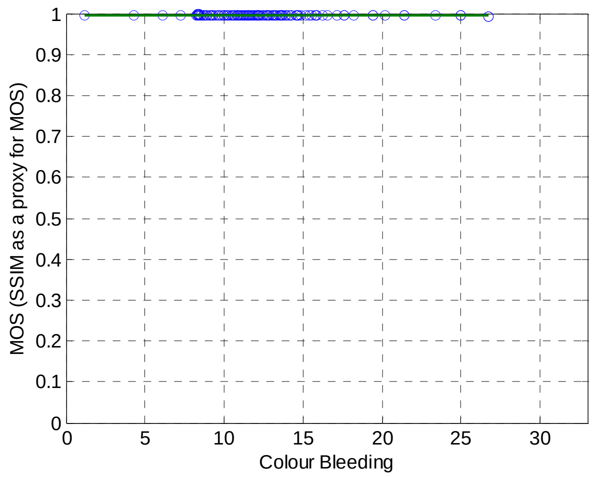

Figure 8. scatter plot of mos against the colour bleeding metric (hue spread) with the data set based on honeycomb test pattern

4 SUMMARY AND CONCLUSIONS

Compression artefacts metrics were evaluated using Matlab[tm] JPEG, Irfanview[tm] JPEG and Irfanview[tm] JPEG2000 codecs. The test patterns *sinesquared grey scale-radial monochrome*, *monochrome rings and honeycomb* (these are like traditional colour bar test pattern) were deployed to stress the codecs. In the evaluation of artefact distortion metrics, it is observed that distortion metrics increase with an increase in compression ratio. Artefact detection and measurement algorithms compute only the artefact under study. This contrasts with the SSIM which provides a single, global, distortion measure.

Blockiness and blur metrics were validated using the human perception based SSIM as proxy for mos. Both blockiness and blur metrics showed a very strong level of correlation. This enables the proposed two test patterns and distortion metrics to be used for codec performance evaluation in conjunction with synthetic test patterns.

The colour bleeding metric gave a SSIM value of almost unity, regardless of the level of distortion. The limitation here is in the SSIM, rather than the colour bleeding metric. Future research will focus on how to improve the SSIM metric for colour image evaluation, enabling it to be used to validate the colour bleed metric.

5. Appendix

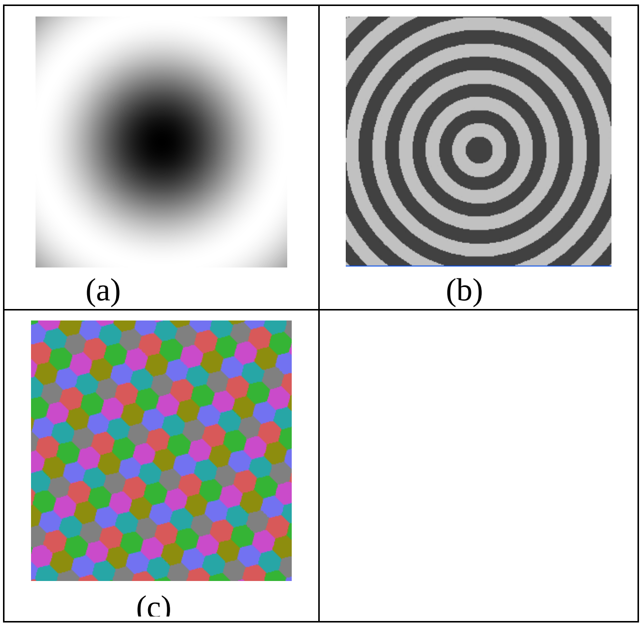

Figure 9. Test patterns deployed during the experiments (a) sinesquared grey scale-radial monochrome [13] (b) monochrome rings [13] and (c) honeycomb [16]

6. REFERENCES

[1] P. Sunna. “AVC/H.264 – an advanced video coding system for SD and HD broadcasting”, *The EBU Technical Review Journal*, Volume 302, pp. 1-6, 2005.

[2] M. Sugawara. “Super Hi-Vision –research on a future ultra-HDTV system”, *The EBU Technical Review Journal*, Volume 2008-Q2, pp. 1-7, 2008.

[3] K. Tanioka, “Integral 3D Television”, Broadcast Technology, *Journal of Science and Technical Research Laboratories of NHK*, ISSN 1345-4095 No. 33, p. 19, 2007.

[4] M. Mark, M. Grgic, S. Grgic, "Scalable Video Coding for Network Applications", *4th EURASIP-IEEE Region 8 International Symposium on Video/Image Processing and Multimedia Communications*, Croatia, pp. 205-211, 2002.

[5] O. Cohen, Unified Engine for Scalable Video Coding (SVC), *The Proceedings of the Broadcast-Asia 2008 Conference*, Singapore, pp 1-6, 2008.

[6] A. Kouadio, M. Clare, L. Noblet, V. Botteau, “ SVC – a highly scalable version of H.264/AVC”, *The EBU Technical Review Journal*, Volume 2008-Q2, pp. 1-7, 2008.

[7] MPEG Industry forum, "Home Page", http://www.m4if.org/mpeg4/, Visited August 2004 and October 2008.

[8] Centre National Research Technologies in France, "Multimedia page", http://tim.irisa.fr/veille/TVadsl/Broadband%20TV/Broadband%20TVSEM/Presentations/Day%202%20morning/MPEG-4%20Video/H264_broadbandTV.pdf, Visited August 2004 and October 2008.

[9] D. Wood, Everything you want to know about Video Codecs, *EBU Technical review*, European Broadcasting Union, Geneva, Volume 295, pp 1-8, 2003.

[10] B. Pank, *The Digital Fact Book*. Berkshire: Quantel Limited, p. 25, 2008.

[11] M Ghanbari, Standard Codecs: Image Compression to Advanced Video Coding, Institution of Electrical Engineers, 2003, p.119.

[12] A. Punchihewa and D.G. Bailey, “Artefacts in Image and Video Systems: Classification and Mitigation”, *Proceedings Image and Vision Computing New Zealand (IVCNZ 2002)*, Auckland, New Zealand, pp. 197-202, 2002

[13] G.A.D. Punchihewa, D.G. Bailey and R.M. Hodgson, "Benchmarking Image Codecs by Assessment of Coded Test Images: The development of Test Images and New Objective Quality Metrics", *Journal of Telecommunications and Information Technology (JTIT2006)*, 1/2006, pp 11-16, 2006.

[14] A. Punchihewa, J Armstrong, S. Hangai, T Hamamoto, "Objective Evaluation of Components of Colour Distortions due to Image Compression”, *The IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Japan, EA92(A) No. 12, pp 3307-3312, 2009.

[15] G.A.D. Punchihewa, D.G. Bailey and R.M. Hodgson, "The Development of a Synthetic Colour Test Image for Subjective and Objective Quality Assessment of Digital Codecs", *11th Asia Pacific Conference on Communications (APCC2005)*, Perth, Australia, pp. 881-885, 2005.

[16] G.A.D. Punchihewa, D.G. Bailey and R.M. Hodgson, "The Development of a Novel Image Quality Metric and a Synthetic Colour Test Image for Objective Quality Assessment of Digital Codecs", *The IEEE Telecommunication Conference Region 10, (IEEE Tencon 2005)*, Melbourne, Australia, pp. 1-6, 2005.

[17] A. Punchihewa, D. Salvador, “An Integrated Environment for Objective Evaluation of Digital Codecs using a Small set of Novel Image Quality Metrics and Synthetic Test Images”, *The IEEE International Conference on Information and Automation (ICIA 2005)*, Sri Lanka, pp.1-6, 2005.

[18] Z. Wang, A. Bovik, *Modern Image Quality Assessment*, Morgan Publishers, 2003.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity" IEEE Transactions on Image Processing, vol. 13, no. 4, Apr. 2004.