An Effective Alternative Structural Similarity Index Algorithm

Khairulnizam Othman and Afandi Ahmad

***Abstract*—Real-time image quality assessment algorithms is an important, outcome research is dedicated to improving this practice. Towards this end, a design of real-time implementable full-reference image or video quality algorithms that are based on the Structural Similarity (SSIM) index and multi-scale SSIM (MS-SSIM) index preferred. The proposed algorithms merged into one single updating process. LIVE image quality database used to evaluate their improvement in form of computational complexity. Experimental results show that the proposed algorithm is an effective alternative for real-time image Structural Similarity with low area cost (time).**

***Index Terms*—Real time, Structural Similarity, effective.**

I. Introduction

Image quality assessment is an emerging field of signal processing. More or less defined as the task of designing an algorithm to automatically judge the perceived “quality” of a photograph, it remains a largely open problem. Latest trends indicate beginning of a new era in digital images and videos, digitized visual information. In addition to the increasing amount of available digital visual data, other factors make the problem of information extraction particularly complicated. First, users ask for more information to be extracted from their datasets, which requires increasingly complicated algorithms. Second, in many cases, the analysis needs to be done in real-time to reap the actual benefits. For instance, a security expert would strive for real-time analysis of the streaming video and audio data in conjunction. Managing and performing run-time analysis on such datasets is appearing to be the next big challenge in computing.

Video quality evaluation is performed to describe the quality of a set of video sequences under study. Video quality can be evaluated objectively by mathematical models or subjectively by asking users for their rating. Also, the quality of a system can be determined offline (i.e., in a laboratory setting for developing new codec’s or services), or in-service to monitor and ensure a certain level of quality. [1], while Full Reference Methods (FR) is FR metrics computes the quality difference by comparing the original video signal against the received video signal. Typically, every pixel from the source is compared against the corresponding pixel at the received video, with no knowledge about the encoding or transmission process in between. More elaborate algorithms may choose to combine the pixel-based estimation with other approaches such as described below. FR metrics are usually the most accurate at the expense of higher computational effort.

The structural similarity (SSIM) index is a method for measuring the similarity between two images. The SSIM index is a full reference metric; in other words, the measuring of image quality based on an initial uncompressed or distortion-free image as reference. SSIM is designed to improve on traditional methods like peak signal-to-noise ratio (PSNR) and mean squared error (MSE), which have proven to be inconsistent with human eye perception.

The difference with respect to other techniques mentioned previously such as MSE or PSNR is that these approaches estimate perceived errors; on the other hand, SSIM considers image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. These dependencies carry important information about the structure of the objects in the visual scene.

A research topic that has attracted a great deal of attention in the past decade is to design novel objective image similarity or dissimilarity measures that correlate well with perceptual image fidelity or distortion [2]. The Structural Similarity (SSIM) index is widely used algorithm in FR image quality assessment applications. A number of algorithms have been derived from SSIM: Multi-scale SSIM (MS-SSIM), Percentile Pooling SSIM (PSSIM) [3], Complex-Wavelet SSIM index (CW-SSIM) [5], Gradient-based Structural Similarity (G-SSIM) [6], and Three-Component Weighted SSIM [7]. All these derivative algorithms aim to improve the accuracy but inevitably increase the computational complexity.

II. Preliminary

A. Single Scale Structural Similarity Index

Based on the trade-offs that the Human Visual system (HVS) is highly adapted for extracting structural information, the SSIM algorithm assesses three terms between two

Manuscript received July 9, 2015.

This work was supported in part by the Malaysian. UTHM under Grant Vot 1301.

Khairulnizam Othman is with the Embedded Computing Research Cluster Microelectronics and Nanotechnology – Shamsuddin Research Centre (MiNT-SRC), University Tun Hussein Onn Malaysia, Johor, Malaysia .

Afandi Ahmad is with the Embedded Computing Research Cluster Microelectronics and Nanotechnology – Shamsuddin Research Centre (MiNT-SRC), University Tun Hussein Onn Malaysia, Johor, Malaysia .

http://dx.doi.org/10.15242/IAE.IAE0815017

non-negative signals a and b: luminance *l(a, b)*, contrast *c(a, b)*, and structure *s(a, b)*:

$$l(x,y)=\frac{2\mu_a\mu_b+C_1}{\mu_a^2+\mu_b^2+C_1}.$$

$$c(x,y)=\frac{2\sigma_a\sigma_b+C_2}{\sigma_a^2+\sigma_b^2+C_2}.$$

$$s(x,y)=\frac{\sigma_{ab}+C_3}{\sigma_a\sigma_b+C_3}.$$

The two constant value C_1 and C_2 are defined to avoid the instability when the denominators are very close to zero. These two values are further determined by two subjective selected value K_1, K_2 and the dynamic range of pixel value $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$. Where $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$, and $C_3 = C_2/2$ are small constants; L is the dynamic range of the pixel values, and $K_1 << 1$ and $K_2 << 1$ are scalar constants. The constants C_1, C_2 and C_3 provide spatial masking properties and ensure stability when the denominator approaches zero. Combining the three terms, the general form of SSIM is:

$$SSIM(a,b)=[l(a,b)]^{\alpha}.[c(a,b)]^{\beta}[s(a,b)]^{\gamma}$$

In equation, three component are clearly defined to measure the degree of linear correlation between image a and b. The first one *l(a,b)*, measure how the mean luminance is between the two image while the second *c(a,b)*, estimated the contrast. The third one *s(a,b)*, is the correlation of structure. The parameter α , β and γ can be used to adjust the relative importance of the three component.

By setting $\alpha=\beta=\gamma=1$ and $C_3 = C_2/2$

$$=\frac{(2\mu_a\mu_b+C_1)(2\sigma_{ab}+C_2)}{(\mu_a^2+\mu_b^2+C_1)(\sigma_a^2+\sigma_b^2+C_2)}$$

SSIM index value between a and b generally is defined as the average of all SSIM index values calculated within an 11 x 11 circular symmetric Gaussian weighting function window passed over image, although other "pooling" strategies exist [4]. The Gaussian weighting window prevents artifacts arising from a discontinuous truncation of the image when computing the local values of the SSIM index map.

A. Multi-scale Structural Similarity Index

The distance between the image and the observer affects the observer's perceived image quality. The results of subjective image tests will vary as the viewing distance changes. In addition, images are naturally multi-scale, and both distortions and image features possess multi-scale attributes.

For these reasons, the Multi-scale SSIM (MS-SSIM) index was developed.

In MS-SSIM, quality assessment is performed on multiple scales of the reference and the distorted images. Low-pass filtering and dyadic down-sampling is applied iteratively, and elements of the SSIM index applied at each scale, indexed from 1 (original image) through and the finest scale M obtained after $M-1$ iterations.

At each scale *i*, the contrast and structure terms are calculated: $c_j(a, b)$ and $s_j(a, b)$ respectively. The luminance term is computed only at scale M and represented as $l_M(a, b)$. The overall quality evaluation is obtained by combining the measurement over scales:

$$MS-SSIM(a,b)=[l_M(a,b)]^{\alpha_M}\prod_{j=1}^{M}[c_j(a,b)]^{\beta_j}[s_j(a,b)]^{\gamma_j}$$

where, typically $M=5$, and the exponents $\sigma_M,\beta_j,\gamma_j$ are selected such that $\sigma_M=\beta_j=\gamma_j$ and $\sum_{j=1}^{M}\gamma_j=1$ [3].

III. Issuance Algorithm

The structure term of the SSIM index is independent of the luminance and often plays a less significant perceptual role in predicting visual quality that the other terms. The propose eliminating it to reduce complexity. Another important item is preserve the luminance term since images may suffer from a luminance bias, even if image quality databases do not explicitly include such distortions. Nevertheless, focus on expend as little computation as possible in the luminance term.

The luminance term transform in Fast SSIM with utilizes an *8* x *8* square window, and an integral image technique [9] to compute the luminance similarity between the original and test images.

By utilizing integral image, extracting the mean value of the pixels within a square window can be made quite efficient. As shown in Fig. 1, the value of the integral image at *(a, b)* is the sum of the pixels values above and to the left of *(a, b)*, and including the value at *(a, b)*.

Computing the sum over any rectangular area can be achieved with only two additions and one subtraction. As shown in Fig. 1, the sum of the pixel values within the rectangle *D* can be computed using four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle *A*. The value at location 2 is *L+M*, at location 3 is *L+O*, and at location 4 is *L+M+O+P*. The sum over region *D* can be computed as *'4'+'1'-('2'+'3')* where *'i'* is the value of the integral image at location *i*.

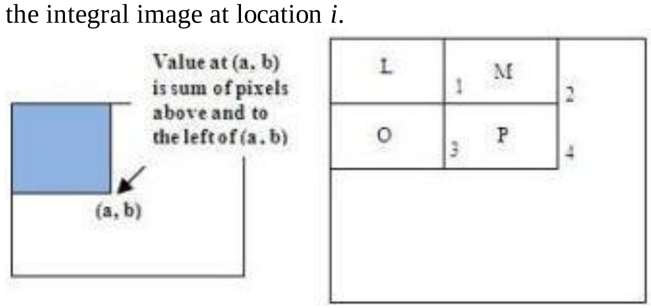

Fig. 1. Left: Integral Image. Right: How To Compute Sum Value Over Region P In Integral Image Domain.

Using the integral image and a square window, the complexity of computing the luminance term is reduced considerably [8]. Assuming the window size is *n* x *n*, the standard SSIM index algorithm (using a Gaussian weighted window) requires n^2 multiplies and (n^2-1) additions to calculate the mean value, while the proposed Fast SSIM algorithm only requires 3 additions and 1 subtraction.

IV. Decryption Algorithm

The computation of the variance term is the most time consuming part of the SSIM algorithm. In order to lower the complexity, we substitute a gradient value in Fast SSIM. Following Field, while images of real-world scenes vary greatly in their absolute luma and chroma distributions, the gradient magnitudes of natural images generally obey heavy tailed distribution laws [9]. Indeed, some no-reference image quality assessment algorithms [10-12] use the gradient image to assess blur severity. Similarly, the performance of the Gradient-based SSIM index [6] suggests that applying SSIM on the gradient magnitude may yield higher performance. The gradient is certainly responsive to image variation. Moreover, the gradient magnitude has low complexity and is amenable to integer-only implementation.

We generate the gradient image using the Roberts gradient templates. The Roberts Cross operator performs a simple, quick to compute, 2-D spatial gradient measurement on an image.

The gradient magnitude is approximated by

$$|\nabla I| = \max\{|\nabla i|, |\nabla j|\} + (\frac{1}{4})\min\{|\nabla i|, |\nabla j|\}$$

where ∇i and ∇j are the Roberts template responses in the two orthogonal directions. This approximation is based upon a simple expansion of the gradient. The contrast *c(x, y)* and structure *s(x, y)* terms of the Fast SSIM index algorithm are then defined:

$$c(x,y) = \frac{\left(2\mu_{Gx}\mu_{Gy} + C_2\right)}{\left(\mu^2{}_{Gx} + \mu^2{}_{Gy} + C_2\right)}.$$

$$s(x,y) = \frac{\left(\mu_{GxGy} + C_3\right)}{\left(\mu_{Gx}\mu_{Gy} + C_3\right)}.$$

where $C_3 = C_2/2$, and

$$\mu_{Gx} \cdot = \frac{1}{N}\sum_{i=1}^{N}|\nabla x_i|$$

$$\mu_{GxGy} \cdot = \frac{1}{N}\sum_{i=1}^{N}|\nabla x_i||\nabla y_i|\ |\nabla y_i|$$

where $|\nabla x_i|$ and $|\nabla y_i|$ are the gradient magnitude values of the images x and y at spatial coordinate *i*, estimated using the approximation (1).

The Fast SSIM index between signals x and y is then:

$$Fast-SSIM(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\mu_{GxGy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\mu_{Gx}^2 + \mu_{Gy}^2 + C_2\right)}$$

In MS-SSIM, the contrast and structural terms are calculated over multiple scales. Therefore, the Fast MS-SSIM index between signal x and y is defined as:

$$Fast-MS-SSIM(x,y) = [l_M(x,y)]^{\alpha_M}\prod_{j=1}^{M}[c_j(x,y)]^{\beta_j}[s_j(x,y)]^{\gamma_j}$$

where *M = 5*, the exponents $\sigma_M, \beta_j, \gamma_j$ are selected such that $\sigma_M = \beta_j = \gamma_j$ and $\sum_{j=1}^{M}\gamma_j = 1$.

Another modification that we make is to use an integer approximation to the Gaussian weighting window. In this way, the computation of the contrast and structural terms uses only integer operations. *8* x *8* windows. The simplifications include: all computations reduced to integer operations, with square roots eliminated.

V. Optimization Algorithm

Optimization is an essential process when implementing an algorithm for real time applications. We propose to apply parallel computing and sub-sampling on the Fast SSIM index algorithm in order to achieve the best performance.

Data-level parallelization and frame–level parallelization are adopted in this study to optimize the parallel computing. After optimizing computation of the luminance part, about 86% of the computation is consumed on the contrast and structure terms. Since most operations in Fast SSIM are integer-only, Fast SSIM is amenable to Single Instruction Multiple Data (SIMD) optimization. Also, since Fast SSIM does not currently use any dependency between frames, it is natural to conduct frame level parallelization.

Regarding sub-sampling, we suggest that the contrast and structure terms need not be computed at the original scale in Fast MS-SSIM [15]. Since humans are less sensitive to higher spatial frequencies, skipping computation of the contrast and structure terms at the first scale appears to not lower performance, but it does increase the computation speed dramatically. The experiment results shown in the next section support this assumption.

VI. Experimental And Result

We test a number of image quality assessment algorithms using the LIVE database [13]. The database includes DMOS subjective scores for each image and 6 types of distortions. The distortions in the first stage, the database contained 460 images, where 116 were original images and the rest 344 are JPEG and JPEG2000 compressed. Two sample images (cropped from *768* x *512* to *256* x *192* for visibility) are shown below. Note that quantization in JPEG and JPEG2000 algorithms often results in smooth representations of fine detail regions (e.g., the tiles in the upper image and the trees in the lower image). Compared with other types of regions, these regions may not be worse in terms of point wise difference measures (as shown in the absolute error map). However, since the structural information of the image details are nearly completely lost, they exhibit poorer visual quality. Close piece-by-piece comparison of the SSIM index and the absolute error maps, we observe that the SSIM index is more consistent with perceived quality measurement. Note: in either distortion or quality maps, brighter means better quality.

Fast-SSIM was evaluated against the LIVE DMOS scores using the Spearman Rank Order Correlation Coefficient (SROCC). The performance numbers on speed were tested on a *768* x *512* video with 250 frames. All experiments were conducted on a Intel® Core™ i7-5500U 2.4GHz platform,

http://dx.doi.org/10.15242/IAE.IAE0815017

except the experiment result on multi-threading optimization, which was run on an Intel® Core™ i7-4790 Processor (3.6 GHz, 8 MB cache, 4 cores) platform [14].

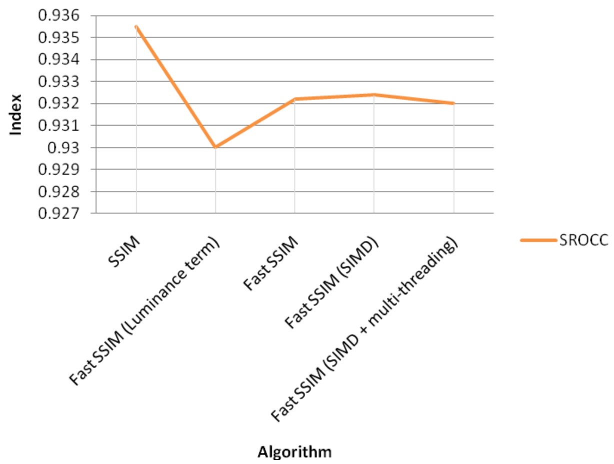

Fig. 2. SROCC SSIM Algorithms Evaluation.

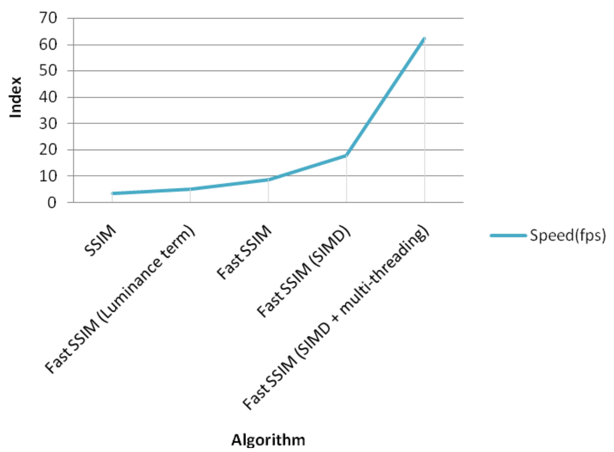

Fig. 3. Speed(Fps) SSIM Algorithms Evaluation.

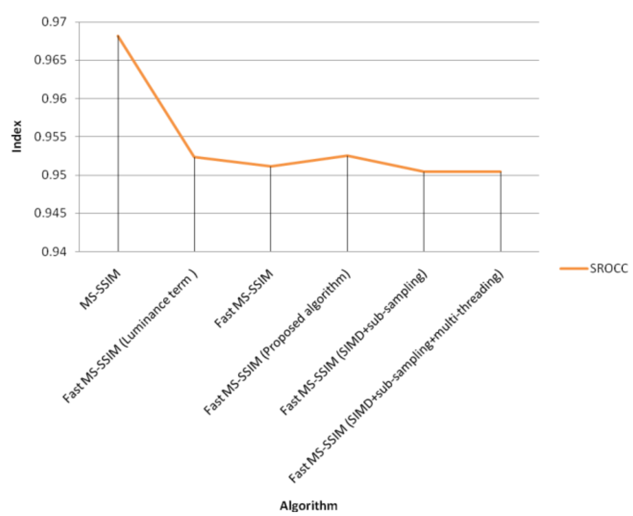

Fig. 4. SROCC MS-SSIM Algorithms Evaluation.

Fig. 2 to Fig. 5 makes it clear that the Fast SSIM and Fast MS-SSIM algorithms suffer no performance loss in terms of their subjective scores to predict the perceptual quality of test images capability; the SROCC scores are very close. Nevertheless, looking at the performance improvement on speed, the improvement from SSIM to Fast SSIM is 155% (from 3.44 fps to 8.76 fps). Thus Fast SSIM is 2.68 times faster than SSIM. For optimization, Intel SSE2 instructions were implemented to calculate the mean and correlation of the radiant images, to demonstrate the improvement on applying

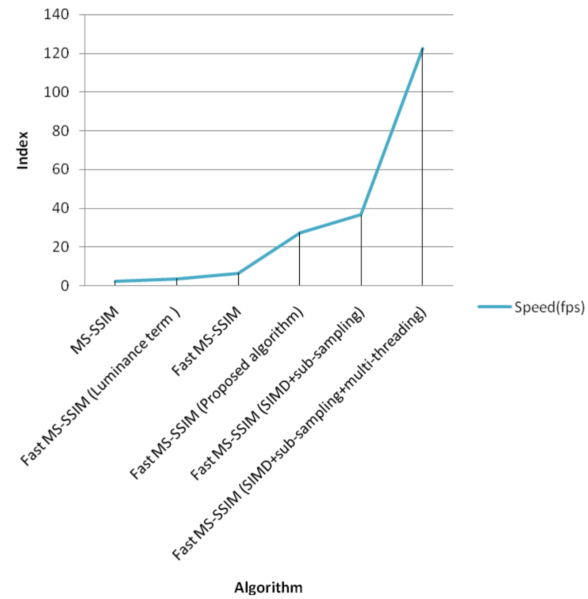

Fig. 5. Speed(fps) MS-SSIM algorithms evaluation.

data-level parallelization. As shown in Table 1, Fast SSIM with SIMD enhances the performance from 8.76 fps to 17.83 fps. Finally, with multithreading optimization, Fast SSIM reaches 62.23 fps on an Intel® Core™ i7-4790 Processor (3.6 GHz, 8 MB cache, 4 cores) platform, which qualifies the algorithm for real-time application.

The performance numbers for Fast MS-SSIM are shown in Table 2. The modifications in Fast MS-SSIM are the same as the modification in Fast SSIM, except that we propose to skip the analysis on contrast and structural terms on the original scales. Table 2 shows that the SROCC scores of Fast MS-SSIM and Fast MS-SSIM with sub-sampling are very close, but both are a little lower than MS-SSIM. However, if we compare the performance of Fast MS-SSIM with Fast MS-SSIM with sub-sampling, Fast MS-SSIM with sub-sampling yields better performance for assessing image quality, at speeds adequate for real-time application.

VII. Conclusion Remarks

Proposed Fast SSIM and Fast MS-SSIM index algorithms are verified their performance. The experimental results show that the proposed algorithms not only have competitive performance with SSIM and MS-SSIM for assessing image quality, but have much lower computational complexity. Indeed, the proposed algorithms achieve real-time performance with simple optimization.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments. To my parents, wife and kids for giving me love supports and encouragement. Nothing would have been possible without Allah, the Creator and Lord of the universe.

REFERENCES

(Periodical style)

[1] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison. Broadcasting," *IEEE Transactions on,* vol. 57(2), pp. 165-182, 2011.
http://dx.doi.org/10.1109/tbc.2011.2104671

[2] E. Y. Chan, W. K. Ching, M. K. Ng, and J. Z Huang, "An optimization algorithm for clustering using weighted dissimilarity measures." *Pattern recognition,* vol. 37(5), pp. 943-952, 2004.
http://dx.doi.org/10.1016/j.patcog.2003.11.003

[3] M. Seufert, M. Slanina, S. Egger, and M. Kottkamp, "To pool or not to pool: A comparison of temporal pooling methods for HTTP adaptive video streaming," *In Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on IEEE,* pp.. 52-57, July 2013.
http://dx.doi.org/10.1109/qomex.2013.6603210

[4] C. Li, and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Processing: Image Communication,* vol. 25(7), pp. 517-526, 2010.
http://dx.doi.org/10.1016/j.image.2010.03.004

[5] K. Okarma, D. Frejlichowski, P. Czapiewski, P. Forczmański, and R. Hofman, "Similarity estimation of textile materials based on image quality assessment methods. *In Computer Vision and Graphics," Springer International Publishing,* pp. 478-485, 2014.

[6] G. H. Chen, C. L. Yang, and S. L. Xie, "Gradient-based structural similarity for image quality assessment," *In Image Processing, 2006 IEEE International Conference on IEEE,* pp. 2929-2932, October 2006.
http://dx.doi.org/10.1016/j.image.2010.03.004

[7] C. Li, and A. C. Bovik, "Three-component weighted structural similarity index," *In IS&T/SPIE Electronic Imaging -International Society for Optics and Photonics*, pp. 72420Q-72420Q, January 2009.

[8] D. M. Rouse, and S. S. Hemami, "Understanding and simplifying the structural similarity metric," *In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on IEEE,* pp. 1188-1191, October 2008.
http://dx.doi.org/10.1109/icip.2008.4711973

[9] X. Huang, and S. Forchhammer, "Improved side information generation for distributed video coding," *In Multimedia Signal Processing, 2008 IEEE 10th Workshop on IEEE*, pp. 223-228, October 2008.

[10] E. Cohen, and Y. Yitzhaky "No-reference assessment of blur and noise impacts on image quality," *Signal, image and video processing,* vol. 4(3), pp. 289-302, 2010.

[11] F. Meyer, "Color image segmentation. In Image Processing and its Applications," *International Conference on IET,* pp. 303-306, 1992.

[12] A. P Carleer, O. Debeir, and E. Wolff, "Assessment of very high spatial resolution satellite image segmentations," *Photogrammetric Engineering & Remote Sensing,* vol. 71(11), pp. 1285-1294, 2005.
http://dx.doi.org/10.14358/PERS.71.11.1285

[13] H.R. Sheikh, Z. Wang, L.K. Cormack, A.C. Bovik, "LIVE image quality assessment database," Release 2, [online]: Available at: http://live.ece.utexas.edu/research/quality/subjective.htm

[14] A. K. Moorthy, and A. C. Bovik, "Visual importance pooling for image quality assessment," *Selected Topics in Signal Processing, IEEE Journal,* vol. 3(2), pp. 193-201, 2009.
http://dx.doi.org/10.1109/JSTSP.2009.2015374

[15] D. M., Rouse, and S. S. Hemami, S. S. (2008, February). Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM. In Electronic Imaging 2008 (pp. 680615-680615). International Society for Optics and Photonics.

Khairulnizam Othman received the M.S. degree in Mechatronic engineering from the University Malaysia Perlis, Perlis, Malaysia, in 2010. He is currently pursuing the PhD in Electrical and Electronic Engineering at The University Tun Hussein Onn Malaysia at Johore, malaysia. He is currently Lectrure the Embedded Computing Research Cluster Microelectronics and Nanotechnology – Shamsuddin Research Centre (MiNT-SRC), The University Tun Hussein Onn Malaysia. His research interests include FPGA, image and video quality assessment, image and embedded medical image processing.

Afandi Ahmad, PhD is a lecturer at Universiti Tun Hussein Onn Malaysia (UTHM) within the Department of Computer Engineering in the Faculty of Electrical and Electronic Engineering. He joined UTHM as a tutor in May 2002 and appointed as a lecturer in December 2003. He obtained his BSc in Electrical Engineering (Computer Technology) from Universiti Teknologi Malaysia (UTM-ITTHO) and MSc in Microelectronics from Universiti Kebangsaan Malaysia (UKM) in 2002 and 2003, respectively. He received his PhD in Electronic and Computer Engineering from Brunel University, West London in 2010. He has been awarded a number of prizes and travel grants and has published in impact factor's journals as well as five star international conferences and actively contributes as a reviewer in IEEE TCVT, Springer and Hindawi. He is a graduate member of the IEM, professional member of ACM, member of the IEEE, IET, ASEE and IAENG. His research interests include: Embedded computing systems, 3-D/4-D medical image analysis and diagnosis, partial and dynamic reconfigurable architectures, 3-D/4-D compression and also engineering education.