A New Quality Metric based on FFT Transform

Mohamed Ben Amor
Laboratory of Electronics and Information
Technologies, Sfax
National school of Engineering
University of Sfax, Tunisia

Fahmi Kammoun
Laboratory of Electronics and Information Technologies, Sfax National school of Engineering University of Sfax, Tunisia

Nouri Masmoudi
Laboratory of Electronics and Information Technologies, Sfax National school of Engineering University of Sfax, Tunisia

ABSTRACT
The study of the human visual system (HVS) is very interesting to quantify the quality of an image or to predict perceived information. The contrast sensitivity function (CSF) is one of the main ways to incorporate the HVS properties in an imaging system. It characterizes its sensitivity to spatial and temporal frequencies. In this paper we are interested in establishing a metric with full reference to the image and video. We realize in our algorithm, the FFT transformation to apply the CSF function. Our method is applicable to any size of image and video sequence by increasing its size at powers of two. This increase is achieved by adding "mirror image". The experimental results show that our method keeps better the different frequency components. She is more efficient than the method of "zero padding" and returns results very close to those of the DFT transformation.

General Terms
Signal processing, image quality metric.

Keywords
PSNR "Peak Signal-to-Noise Ratio";CSF "contrast sensitivity function"; HVS " human visual system "; FFT " Fast Fourier Transform "; zero padding

1. INTRODUCTION
Recently, human visual system's characteristics widely used measure the visibility of errors between a reference image and another deformed. An objective quality measures should provide strong correlation with those provided by observers. There are many methods to evaluate the quality of a degraded image. They can be classified into three categories according to the data they need to produce their measurement:

- Quality criteria with full reference (FR) require the provision of the original image [1, 3, 6].

- Quality criteria with reduced reference (RR) need only a description of the original image. The quality score is determined from the degraded image and the reduced reference. The choice of descriptors is a key element in the development of such a criterion. They are used when it is impossible to have two images together (transmission constraint). Thus, their performance is lower than the criteria with full reference [17, 18].

- Quality criteria without reference (NR) need to know anything about the original image. They also apply when we have transmission constraints, and used with systems where the damage cause is known in advance [19, 20].

The image quality metrics can be also classified according to their complexity into four categories:

- The mathematical metrics: These are the most common criteria for quality assessment. They are based on simple mathematical measures such as distances or measures derived from signal processing such as mean squared error (MSE) or signal to noise ratio (SNR).

- Approaches based on structural fidelity: The criteria for adopting this approach are more recent. They are not based on properties of low vision, but on high properties on the human visual system response to a degraded image. The main hypothesis is that our perception is particularly suitable for the extraction of structural information of an image. The idea is to measure the degradation of this structural information [1, 2].

- Metrics based on modeling mono-channel: The first models of the human visual system based on an approach with a single channel. The visual system is then seen as a spatial filter whose characteristics are defined by a function of contrast sensitivity [3, 4, 5].

- Metrics based on modeling multi-channel: These metrics can characterize the effects of masking better than those with single-channel because they can partially integrate some specificity of receptive fields in the human visual system. In recent years, these criteria become even more popular [6, 7, 8].

In fact, this paper present a novel quality metric based on CSF, with a new scheme to achieve the FFT transform on image and video.

This paper is organized as follows; in section 2 contrast sensitivity function is briefly introduced. The generation of sequences degraded is introduced in section 3. Section 4 shows the proposed models and the new scheme to achieve the FFT transform of any size. The results are discussed in section 5. Finally; section 6 is devoted to conclusion.

2. CONTRAST SENSITIVITY FUNCTION
The contrast sensitivity function known (CSF), is based on the ability of visual system to detect differences in luminance, thus determine the existence of edges between homogeneous surfaces. It expresses the sensitivity variation of the human visual system to the contrast versus different spatial frequencies.

Several studies have led to define the sensitivity as a function of spatial frequencies. The resulting curve is the CSF function of the examination subject taking into account the experiment conditions (the nature of the stimulus used, viewing distance, viewing angle, monocular or binocular vision, etc...).

Therefore, there are as many ways as CSF examination. Usually, only two results obtained by the same procedure will be comparable. Studies have led find analytical formula close to the experimental results. Among the best known, we can cite the model of Mannos and Sakrison [3] the model of Nill [9] and the model of Ngan [10]:

$$\text{CSF}_{M \, et \, S}(f) = 2.6(0.0192 + 0.114f) e^{-(0.114f)^{1.1}} \quad (1)$$

$$\text{CSF}_{\text{Nill}}(f) = (0.2 + 0.45f) e^{-0.18f} \quad (2)$$

$$\text{CSF}_{\text{Ngan}}(f) = (0.31 + 0.69f) e^{-0.29f} \quad (3)$$

3. GENERATION OF DEGRADED SEQUENCES

Fig 1. (a) Original sequence. (b) Sequence degraded with QP = 26 (c) Sequence degraded with QP = 32. (d) Sequence degraded with QP = 44

To evaluate the performance of the proposed method, we used a set of sequences CIF (352,288). The generation of degraded sequences was performed with JM 10.1, free code encoder H264 [11].

The first image is encoded in intra mode to build the frame I. As a result, all other images are coded in inter mode with motion estimation for macro blocks of size 16×16 (P frame). This step is performed for different quantization factors QP. Thus, for each original sequence, we obtain set degraded sequences.

Figure 1-a represents an original sequence, Figure 1-b is a sequence degraded with QP = 26, Figure 1-c is a sequence degraded with QP = 32, Figure 1-d is a sequence degraded with QP = 32.

4. THE PROPOSED MODEL

The proposed model is represented in Figure 2.
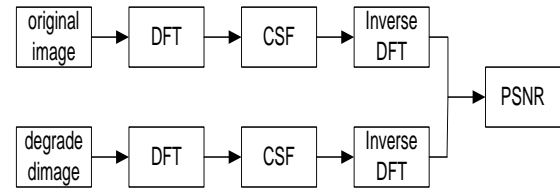
Fig 2. The proposed model

Firstly, a two dimension Discrete Fourier transform (DFT 2d) is applied to both original and degraded image. Then each spatial frequencies horizontal and vertical (f (u),f (v)) is converted to (cycle /degree). This operation is performed according to the expressions (4 and 5) [12]:

$$f(u) = (u - 1)/(\Delta N) \quad (4)$$

$$f(v) = (v - 1)/(\Delta N) \quad (5)$$

Where N is the number of frequencies and Δ = 0.25 mm (the dot pitch) and u, v =1, 2, 3…N.

$$f_s \, (cycle/degee) = f_i \, (cycle/pixel) \times f_n \, (pixel/degree)$$

$$= \frac{\pi}{180 \times \arcsin\left(\frac{1}{\sqrt{1+dis^2}}\right)} \times \sqrt{f(u)^2 + f(v)^2} \quad (6)$$

Where, "dis" is the viewing distance in millimetres.

The filtering operation is performed by multiplying each resulting value of the DFT (real part and imaginary part) by the coefficient of contrast sensitivity functions corresponding:

$$f_s(\text{filtred}) = f_s \times \text{CSF}(f_s) \quad (7)$$

Based on our previous works [13], we found that Nill filter [9] is best suited to our application and it gives better results in terms of correlation with the human visual system (see fig. 3).
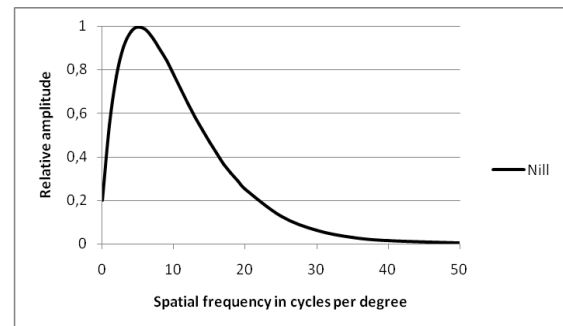
**Fig 3. Contrast sensitivity function (CSF)**

There are two ways to filter an image from a CSF space (see fig. 4) [14]. The first is based on the CSF normalization to the frequency peak [15]. For frequencies below the peak value, the coefficient 1 is applied which preserve the signal.

The other way is based on the normalization of the CSF to the reference frequency (spatial frequency of the stimulus reference). In this case, the coefficient of 1 is applied for frequencies below the reference value.
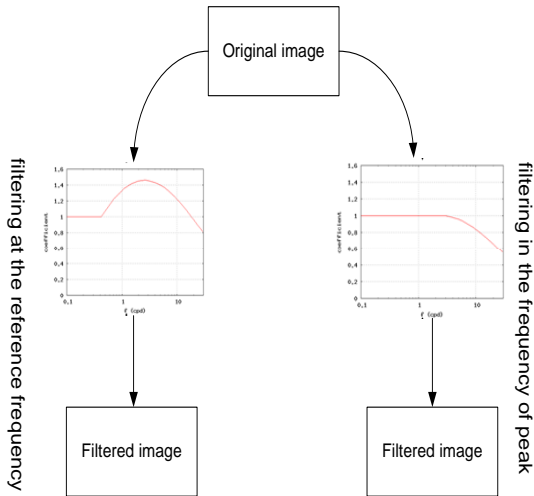
Fig 4. CSF conventional filtering approaches

In our study, we use the first method with a peak frequency equal to 5 cycles / degree (see fig. 5).
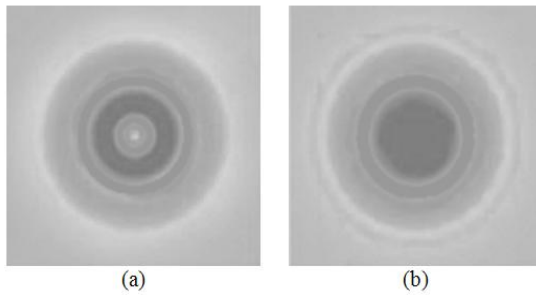
Fig 5. (a) 2D CSF. (b) 2D CSF for luminance

After the filtering operation, we realized the inverse DFT in order to reconstruct the original image and degraded. Finally, we calculate the PSNR between two images using the following expressions:

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(I_{ori}(i,j) - I_{deg}(i,j) \right)^2 \qquad (8)$$

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} (dB) \qquad (9)$$

So we have for each viewing distance a different PSNR for the same image. In fact, a distortion visible for a distance of 500 mm can be invisible for a distance of 3000 mm (see fig.6).
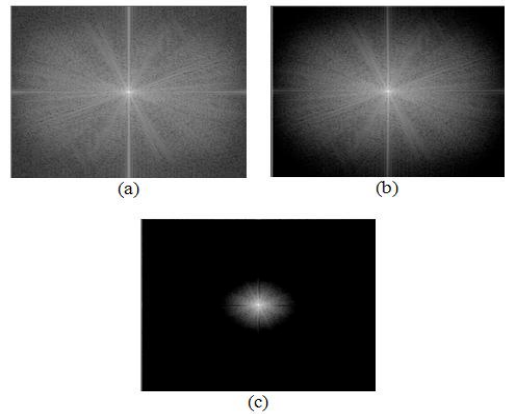
Fig 6. Fourier spectrum of the luminance component (a) unfiltered. (b) filtered to 500 mm. (c) filtered to3000 mm

But the problem is that the DFT is very heavy in terms of computation time especially for video sequences. Therefore, we must find a way to increase speed. As a solution, we use the FFT (Fast Fourier Transform) instead of DFT. There are many FFT algorithms that require having an image size of power of two. So, we have to increase the image size (512,512). The traditional way is the "zero padding" [16]: the neighborhood outside of the image is regarded as a set of pixels of null values (see fig.7).
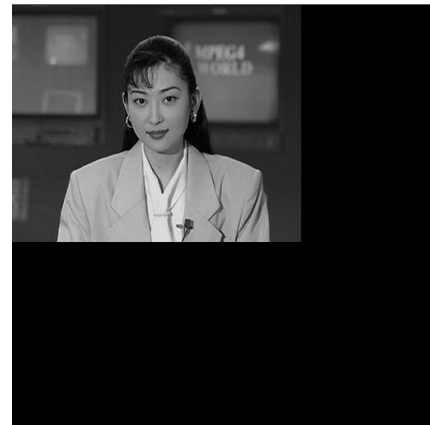
Fig 7. Zero padding

Our idea is to use a sort of "image Mirrors" which consist to copy the pixels from the image itself rather than complete nulls with pixels (see fig.8).

Fig 8. Mirrors of image

Our method keeps the DC component because the filling with zeros "zero padding" tends to decrease this component. On the other hand, our method better preserves the nonzero components. In fact, the method of zero padding introduces zones that do not include these non-zero frequency components and reduces their importance. Our method tends to keep the influence of these frequency components by copying different parts of the original image.

5. RESULTS AND DISCUSSION

5.1 Results with DFT

We apply our model for Akiyo, foreman and mobile sequence with a quantization factor of (44, 32, and 26); the results of PSNR are shown in Table 1.

Table1 PSNR at different viewing distances

	Dist (mm)	500	1000	1500	2000	2500	3000
Akiyo	Qp=26	46.63	49.81	51.53	52.61	53.42	54.06
	Qp=32	42.10	45.27	47.28	48.75	49.84	50.68
	Qp=44	33.17	35.87	37.80	39.34	40.61	41.69
Foreman	Qp=26	45.43	49.36	51.37	52.61	53.51	54.11
	Qp=32	40.72	44.38	46.65	48.28	49.53	50.51
	Qp=44	31.23	33.74	35.64	37.26	38.62	39.75
mobile	Qp=26	45.70	49.95	51.82	52.86	53.58	54.09
	Qp=32	40.17	44.99	47.28	48.77	49.77	50.51
	Qp=44	29.24	33.77	36.26	38.04	39.45	40.26

These results are expected: when gets enlarged viewing distance we have an amelioration of PSNR. That means that when the observation distance increases, there are forms of distortion that HVS cannot detect.

5.2 Comparison of FFT padding and FFT mirror with DFT in the original sequences

We use firstly a comparison based on the original sequence comparing our method and the zero padding method to the DFT scheme (see fig.9).
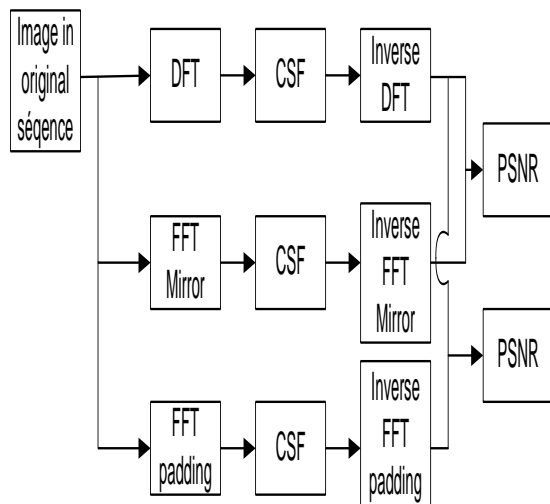
Fig 9. Comparison of FFT padding and FFT mirror with DFT in the same sequence

After the filtering operation, we return to the initial size of the image (352, 288). The results are shown in the table 2.

**Table 2 comparison PSNR DFT / FFT padding and DFT / FFT mirror**

	Distance (mm)	500	1000	1500	2000	2500	3000
Akiyo	DFT/FFT padding	32.33	32.26	31.42	30.60	29.89	29.27
	DFT/FFT mirror	38.62	35.75	34.07	32.81	31.82	31.02
Foreman	DFT/FFT padding	30.11	29.01	27.56	26.35	25.35	24.51
	DFT/FFT mirror	36.37	34.47	33.90	33.45	32.67	32.43
Mobile	DFT/FFT padding	31.27	30.45	29.46	28.59	27.85	27.21
	DFT/FFT mirror	39.03	37.69	36.42	35.43	34.68	34.07

These results show that the PSNR is higher for the DFT/FFT mirror comparison. The FFT mirror is closer to DFT.

The superiority of our method manifests primarily on the short distances. Indeed, the distortions are better perceived by the human visual system and therefore the contrast sensitivity function is less selective at high frequencies. These results show that our method better keeps the high frequencies that the method of zero padding.

5.3 Comparison of FFT padding and FFT mirror with DFT by using the original and the degraded sequences

To compare the two methods we use the following chart (see fig 10).
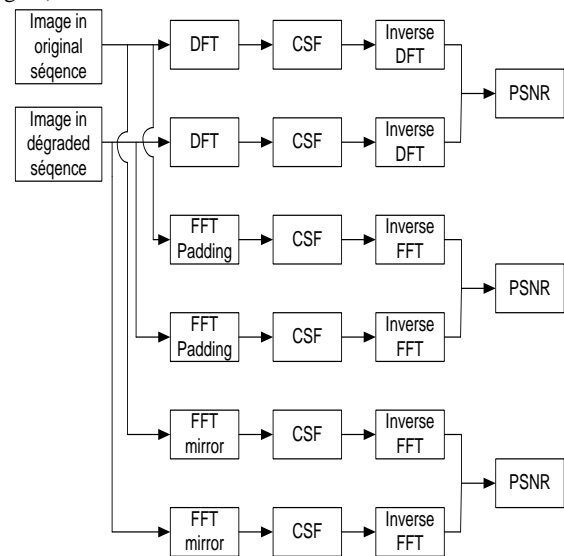
Fig 10. Comparison of FFT padding and FFT mirror with DFT by using the original and the degraded sequences

The results are shown in the table 3. These results show that generally, the PSNR resulting is closer to the PSNR of the DFT scheme. For example, we notice that the PSNR is the same for both methods for the sequence mobile with a distance of 500 mm, Qp=32 and 26. We test our method on several types of sequences (Foreman, Akiyo and Mobile) and several values of Qp (26, 32 and 44) and we found similar results.

5.4 Speed in execution time

Our method was developed with Microsoft visual C++ 6.0 using a Pentium 4 machine (Horologe frequency: 2GHz, 1 GB of RAM).

For a single image, the time needed to establish the DFT operation is 52921 ms. For cons, the FFT operation requires only 1750 ms. So we have an algorithm 30 times faster. As a result, this model is applicable for images as well as for video sequences.

Table 3 results of both methods

		Distance(mm)	500	1000	1500	2000	2500	3000
Akyio	**PSNR (Qp=44)**	DFT	33.17	35.87	37.80	39.34	40.61	41.69
		FFT zero padding	33.28	35.93	37.87	39.39	40.64	41.71
		FFT Mirror	33.19	35.86	37.78	39.30	40.53	41.57
	PSNR (Qp=32)	DFT	42.10	45.27	47.28	48.75	49.84	50.68
		FFT zero padding	42.13	45.28	47.33	48.80	49.93	50.78
		FFT Mirror	42.11	45.22	47.25	48.70	49.80	50.65
	PSNR (Qp=26)	DFT	46.63	49.81	51.53	52.61	53.42	54.06
		FFT zero padding	46.63	49.84	51.59	52.68	53.45	54.08
		FFT Mirror	46.64	49.81	51.51	52.63	53.37	54.00
Foreman	**PSNR (Qp=44)**	DFT	31.23	33.74	35.64	37.26	38.62	39.75
		FFT zero padding	31.29	33.76	35.63	37.22	38.56	39.67
		FFT Mirror	31.24	33.73	35.56	37.13	38.45	39.55
	PSNR (Qp=32)	DFT	40.72	44.38	46.65	48.28	49.53	50.51
		FFT zero padding	40.76	44.44	46.73	48.36	49.67	50.62
		FFT Mirror	40.71	44.37	46.61	48.26	49.47	50.46
	PSNR (Qp=26)	DFT	45.43	49.36	51.37	52.61	53.51	54.11
		FFT zero padding	45.47	49.40	51.47	52.75	53.59	54.24
		FFT Mirror	45.41	49.35	51.33	52.62	53.46	54.11
Mobile	**PSNR (Qp=44)**	DFT	29.24	33.77	36.26	38.04	39.45	40.26
		FFT zero padding	29.28	33.84	36.34	38.15	39.57	40.47
		FFT Mirror	29.23	33.75	36.22	37.98	39.37	40.50
	PSNR (Qp=32)	DFT	40.17	44.99	47.28	48.77	49.77	50.51
		FFT zero padding	40.23	45.02	47.50	49.06	50.13	50.92
		FFT Mirror	40.17	44.90	47.32	48.80	49.86	50.61
	PSNR (Qp=26)	DFT	45.70	49.95	51.82	52.86	53.58	54.09
		FFT zero padding	45.75	50.09	52.04	53.11	53.90	54.51
		FFT Mirror	45.70	49.96	51.89	52.96	53.68	54.24

6. CONCLUSION

In this paper we present a quality metric based on the contrast sensitivity function CSF. Our idea is that we cannot give a quality factor at all distances of observation, but what is visible from a distance may be invisible to another longer. So for each viewing distance we have a factor that gives the image quality.

We also present the method of mirror image, which allowed us to increase the size of the image to the power of two to use the fast Fourier transform FFT instead of the discrete Fourier transform DFT. The goal is to increase the speed of the metric that allows us to use it for movies.

7. REFERENCES

[1] Z.Wang , A.C.Bovik, H.R. Seikh, and E.P.Simoncelli, « Image quality assessment : from visibility to structural similarity», In IEEE Transactions on Image Processing, volume 13, 2004.

[2] D.M. Chandler and S.S. Hemami, « VSNR: A wavelet-based visual signal-to-noise ratio for natural images», in IEEE Transactions on Image Processing, 16(9): pages 2284–2298, 2007, doi :10.1109/TIP.2007.901820. 126, 145.

[3] J. Mannos and D. Sakrison, « The effects of a visual fidelity criterion on the encoding of images », In IEEE Transactions on Information Theory, volume IT-20, p.525-536, 1974.

[4] C. F. Hall and F. Hall, « A Nonlinear Model for the Spatial Characteristics of the Human Visual System », IEEE Transaction Systems, Man and Cybernetics, vol. SMC-7, No. 3, p. 161-170, 1977.

[5] J. O. Limb, « Distortion Criteria of the Human Viewer » , IEEE Transaction on Systems, Man and Cybernetics, vol. SMC-9, No. 12, p. 778-793, 1979.

[6] S. Daly, « the Visible Differences Predictor: an Algorithm for the Assessment of Image Fidelity », in Digital Images and Human Vision, edited by A. B. Watson, MIT press; p. 197-206. 1993.

[7] S. A. Karunasekera and N. G. Kingsbury, «A Distortion Measure for Coding Artifacts in Images: Implementation Aspects, Internal Report», Signal Processing Group, Department of Engineering, University of Cambridge, 1995.

[8] P. Le callet, «Critères objectifs avec référence de qualité visuelle des images couleurs», PhD thesis, Polytechnic University of Nantes, 2001. 117, 140.

[9] N.B. Nill, «A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment», IEEE Transactions on communications, Vol. COM-33, No. 6, pp. 551-556, 1985.

[10] K. Ngan, K. Rao, and H. Singh; Cosine transform coding incorporating human visual system model; Presented at SPIE fiber' 86, 1986.

[11] ITU-T Rec. H.264 / ISO/IEC 11496-10, «Advanced Video Coding», Final Committee Draft, Document JVTF100, December 2002.

[12] K. Veeraswamy, S. Srinivaskumar, B.N. Chatterji, «Designing Quantization Table for Hadamard Transform based on Human Visual System for Image Compression», ICGST-GVIP Journal, Volume 7, Issue 3, November 2007.

[13] M. Ben Amor, A. Samet, F. Kammoun, N. Masmoudi, «exploitation des caractéristiques du système visuel humain dans les métriques de qualité», Cinquième workshop AMINA 2010, pp 123-130.2010

[14] V. Rosselli, M.C. Larabi, C. Fernandez-Maloigne, «Métrique de différence couleur basée sur le seuil de perception» on COROSA COmpression et REprésentation des Signaux Audiovisuels, Montpellier, 8-9 novembre 2007.

[15] A. Stoica, M.C. Larabi, C. Fernandez Maloigne, «Amélioration de la qualité visuelle d'images couleur dans le cadre du standard de compression JPEG2000», traitement du signal 2004_volume 21_numéro spécial L'image numérique couleur (2004)

[16] A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[17] M. Chen and G. Bailey « Image Quality Assessment Using Data Hiding for Performance Evaluation of Visual Communication Networks», fourth international working conference performance modelling and evaluation of heterogeneous networks HET-NETs, p. 1-6 September 2006.

[18] Z. Wang and E.P. Simoncelli, «Reduced-reference image quality assessment using a wavelet-domain natural image statistic model», Human Vision and Electronic Imaging X, 5666, Jan. 2005.

[19] Y. Horita, SH. Arata and T. Murai « No-reference image quality assessment for jpeg/jpeg2000 coding », XII European Signal processing conference EUSIPCO Vienne Austria, p. 1301-1304 September 2004.

[20] M.G. Choi, J.H. Jung and J.W. Jean « No-Reference Image Quality Assessment using Blur and Noise », proceedings of world academy of science, engineering and technology volume 38 february 2009 issn:2070-3740.