QUALITY ASSESSMENT OF TONED MAPPED VIDEOS AND IMAGES

A Madhusudhan[1], Rahul.N[2], Shadrach G Samson[3]
[1]Asst Professor, [2,3]B.Tech Student, Department of Electronics and Communication Engineering
CVSR College Of Engineering, Ghatkesar, Medchal (Dist), Hyderabad, Telangana, INDIA

Abstract
With the increasing demand for video-based applications, the reliable prediction of video quality has increased in importance. Numerous video quality assessment methods and metrics have been proposed over the past years with varying computational complexity and accuracy. Methods to assess the visual quality of digital videos as perceived by human observers are becoming increasingly important, due to the large number of applications that target humans as the end users of video. This paper proposes an objective quality assessment of videos by combining structural fidelity and statistical naturalness of the frames. Initially input video is converted into the number of frames and then calculate the mean of the quality of these frames, the frames are converted into gray scale images. Validations using an independent subject rated image, database shows good correlations between subjective ranking scores and proposed tone mapped quality index (TMQI).
Keywords: high dynamic range image, image fusion, video quality assessment, naturalness, statistical naturalness, structural similarity, tone mapping operator

1. INTRODUCTION

Digital images are rapidly finding their way into our daily lives due to the explosion of information in the form of visual signals. These images often pass through several processing stages before they reach to their end-users. In most cases, these end-users are human observers. Through different processing stages, e.g., acquisition, compression, and transmission, images are subjected to different types of distortions which degrade the quality of them. For example, in image compression, lossy compression schemes introduce blurring and ringing effects, which leads to quality degradation. Moreover, in the transmission stage, due to limited bandwidth of the channels, some data might be dropped, which results in quality degradation of the received image.

Our obvious way to implement video quality metric is to apply a still image quality assessment metric on a frame by frame basis. However, a more efficient approach would model the temporal aspects of the human visual system in the design of the metric. Then implemented a video distortion meter by using video quality assessment meter followed by a “cognitive emulator” that temporal effect such as smoothing and temporal masking of the frame quality measure, saturation and symmetric tracking, Van den Braden Lambright eat has extended the HVS modelling into the time dimension by modelling the temporal dimension of the CSF, and generating by two visual streams turned to different temporal aspects of the stimulus from the output of each spatial channel Watson’s digital video quality (DVQ) metric operates in the DCT domain and is therefore more attractive from an implementation point of view since the DCT is efficient to implement and most video coding standards are based on the DCT.

2. RELATED CONTENT

A. Subjective image quality assessment

The most reliable method for assessing the quality of images is through subjective testing, since human observers are the ultimate users in most of the multimedia applications. In subjective testing a group of people are asked to give their opinion about the quality of each image. In order to perform a subjective image quality testing, several international standards are proposed which provide reliable results. Subjective quality assessment methods provide accurate and reliable measurements of the quality of visual signals. However, these methods suffer from different drawbacks that limit their applications:

- They are time consuming and expensive. This is due to the fact that subjective results are obtained through experiments with many observers.
- They cannot be incorporated into real-time applications such as image compression, and transmission systems.
- Their results depend heavily on the subjects' physical conditions and emotional state. Moreover, other factors such as display device and lighting condition affect the results of such experiments.

Therefore, it is necessary to design mathematical models that are able to predict the perceptual quality of visual signals in a consistent manner with subjective evaluations.

B. Objective image quality assessment

The goal of objective IQA is to design mathematical models that are able to predict the quality of an image accurately and also automatically. An ideal objective IQA method should be able to mimic the quality predictions of an average human observer.

Objective IQA methods can also be categorized based on their application scope. General purpose methods are the ones that do not assume a specific distortion type. Therefore, these methods are useful in a wide range of applications. On the other hand, application specific methods are the ones that are designed for specific distortion types. An example of these methods is the algorithms designed for image compression applications. Many quality metrics in image compression are designed for block-DCT or wavelet-based image compression.

C. High Dynamic Range Image (HDR)

High dynamic range image (HDRI or HDR) is a set of techniques used in imaging and photography to reproduce a greater dynamic range of luminosity than possible using standard digital imaging or photographic techniques. HDR images can represent more accurately the range of intensity levels found in real scenes, from direct sunlight to faint starlight, and is often captured by way of a plurality of differently exposed pictures of the same subject matter. Because in the display devices are not capable to display the high dynamic range images.

D. Tone Mapping

Tone mapping is a technique used in image processing and computer graphics to map one set of colours to another in order to approximate the appearance of high dynamic range images in a medium that has a more limited dynamic range. More dynamic range means high contrast and brightness of the images. Some of the computer print outs and CRT, LCD monitors are the limited dynamic range values only doesn't exceeds the limited range of the dynamic range. Different methods of tone mapping operators have been developed in the previous years. They all can be divided in two main types Global (or spatiality uniform) operators and Local (or spatial varying) operators. The effect of the algorithm changes in each pixel according to the local features of the image. These algorithms are high difficulty than the global operators. They can show halo effect and ringing and the output can look unrealistic, but they can provide the best performance, since human vision is mainly sensitive to local contrast.

3. PROPOSED METHOD

Assessment of quality

During the process of conversion of HDR videos to LDR videos, Tonned mapped operators cannot preserve all the information in HDR videos because of sudden reduction in dynamic range. While viewing the particular LDR videos by human visual system this loss of information can't be observed. For making the LDR videos to human visual system in a pleasant way Structural similarity index matrix (SSIM) plays a critical role. On the later end, Structural fidelity does not give whole quality measurement; therefore it cannot alone sufficient to provide overall quality measurement. In order to provide a visually good quality LDR video a unique combination of preservation of structural similarity statistics and naturalness' statistics must be needed, but combing this two critical factors (structural similarity statistics and naturalness statistics) is still a complex factor(in some cases).

Structural similarity index matrix (Structural fidelity)

Measurement of structural similarities between two different videos for accurate quality measurement needs novel techniques to in order to perform this innovative task, one such a technique is proposed in literature is Structural similarity index matrix (SSIM). Structural fidelity mainly applied on local features and the comparison between three main components mainly luminance, contrast and brightness. Since Toned mapped operators are intended to change the intensity of local features and contrast. So it is unacceptable to compare the intensity of local features and contrast. So let us consider the typical example of structural fidelity, let take two patches 'x' and 'y' from HDR video and as well as LDR video respectively. Then the structural fidelity measuring between these two patches is given as

$$S_{local}(x,y) = \frac{2\sigma'_x\sigma'_y + C1}{\sigma'^2_x + \sigma'^2_y + C1} \cdot \frac{\sigma_{xy} + C2}{\sigma_x\sigma_y + C2}$$

Where local standard deviations and the HDR video patches cross correlation with LDR video patches is given as $\sigma'_x\sigma'_y$ and $\sigma_{xy}$and C1 ,C2 are positive stabilizing constants. Here in comparison process luminance part is missing while comparison in SSIM, the signal strength based on SSIM is further modified based on two innovative considerations, one is difference of HDR video patches to LDR video patches signal strength is correct by calculations getting on two results, one is above threshold and another is below threshold (both must be significant) and the second case is where one video patch is significant and the second one is in significant. Original SSIM is quite different from the modified one as shown in above discussion.

In order to know the difference between the significant video patches from the insignificant video patches, as shown in equation Sigma parameter and sigma transpose parameter we derive the significant values to 1 and insignificant values to 0.This particular process is called non linear mapping and the non linear mapping tends to visualize the fixed threshold in contrast feature, but Human visual system does not have any fixed threshold length for contrast feature in practical. Now in order to know the significant and insignificant video patches a detection probability is needed and that detection probability is Psychometric function. This psychometric function evaluates the 50% of detection probability; most commonly used detection probability function is normal cumulative distribution function. The normal cumulative distribution function is given by

$$p(s) = \frac{1}{\sqrt{2\pi\theta_s}} \int_{-\infty}^{s} \exp\left[-\frac{(x-\tau_s)^2}{2\theta_s^2}\right] dx$$

Where p is the function of detection probability, s is sinusoidal stimulus amplitude and τ_s is modulation threshold and θ_s is the standard deviation of normal distribution that controls the continuous variation is detection probability function, and the ratio is given as

$$k = \frac{\tau_s}{\theta_s}$$

According to the Croziers law, the K term in equation represents a constant, and its range is in between 2.3 to 4, and if the value is 3 then the probability of occurrence of false detection is low and it is considerably low. Contrast sensitivity function is one function that collects all data which is psychological functions is used to know the visual contrast sensitivity.

$$A(f) \approx 2.6[0.0192 + 0.114f]\exp[-(0.114f)^{1.1}]$$

Where f denotes the spatial frequency, this function is normalized have peak value 1.thus provides the relative sensitivity function of frequency function. In our proposed framework usage of Kelly CSF measurement function combining the Kelly CSF function with above equation we got as follows

$$\tau_s(f) = \frac{1}{\lambda A(f)}$$

By using the above equation, calculation of contrast threshold function is done assuming the pure sinusoidal stimulus. In order to convert it to the signal strength, two factors taken into account. one contrast sensitivity and later is mean signal intensity, for this signal strength threshold is measured using the standard deviation of the signal. And the threshold value measured on standard deviation of signal is calculated as follows

$$\tau_\sigma(f) = \frac{\bar{\mu}}{\sqrt{2}\lambda A(f)}$$

Where $\bar{\mu}$mean intensity value is obtained by combing the mean signal intensity and standard deviation, then based croziers law equation is as follows

$$\theta_\sigma(f) = \frac{\tau_\sigma(f)}{k}$$

Then, we can mapping can be between sigma and sigma transpose is done as follows

$$\sigma' = \frac{1}{\sqrt{2\pi}\theta_\sigma}\int_{-\infty}^{\sigma} exp\left[-\frac{(x-\tau_\sigma)^2}{2\theta_\sigma^2}\right]dx$$

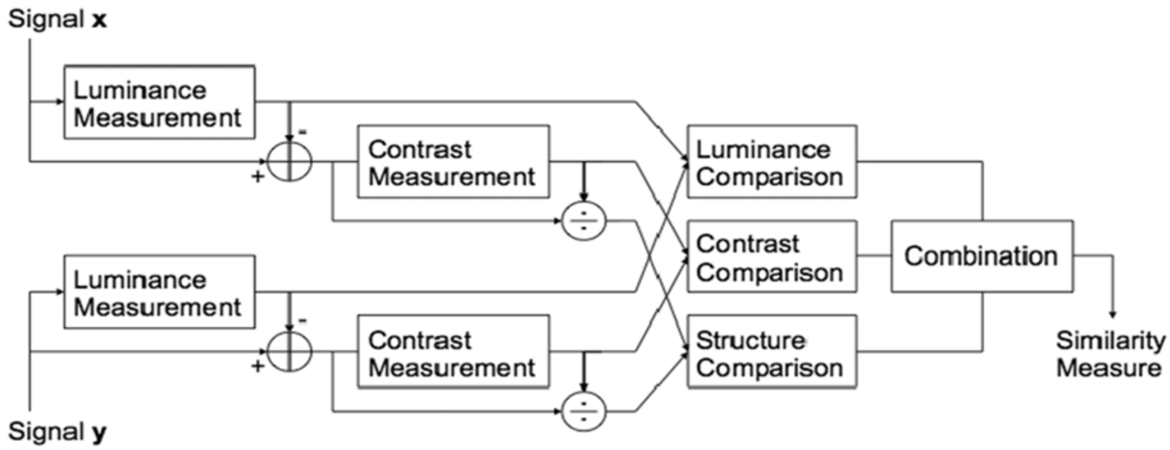


Figure 1: Proposed method

Then the mapped versions of sigma and sigma transpose in x and y patches are evaluated by 0 and 1. 0 represents the insignificant value and 1 represents the significant value. Visibility of video is depends on the distance between the video the resultant observer and this distance is calculated based on the sampling density of the video. Based on the capability of human visual system multi scale approach is introduced in order to evaluate the weighted SSIM values and the to down sampled the video by low pass filtering method and then create the pyramid structure .By using the two different TMO's on HDR video and on LDR video local structural fidelity map is generated. In this process loss of information may attain in LDR video compared to the HDR video ,in some scenarios bright regions may miss, in some regions dark regions may miss, but it cannot be observed in LDR video. Then single score attain by pooling algorithm is as follows

$$S_l = \frac{1}{N_l}\sum_{i=1}^{N_l} S_{local}(x_i, y_i)$$

Where xi and yi are HDR and LDR video patches. Where Nl is the number of patches. Then overall structural fidelity is attained by combining the scale fidelity score is as follows

$$S = \prod_{l=1}^{L} S_l^{\beta_l}$$

First the structural fidelity performance is measured. Second the fidelity is checked at each stage and then by using the window overall standard deviation is obtained. Third distance of video from resultant user is measured by applying the CSF. Fourth main intensity values are attained by setting the mean of dynamic range of LDR video values. Then by combing all this measures we get overall psychological experiments. In order to perform this on RGB video it has to convert to Yxy space using the Y component.

Statistical quality assessment (natural scene)

The high quality of LDR videos does not mean that it has all information same as in HDR video. Subjective approach has more drawbacks regarding cost of process and time consuming while processing on contrast ,brightness etc. By using Objective approach we can get better correlation than subjective it can be compared with subjective approach ,in naturalness of scene in HDR to LDR is evaluated as follows

$$P_m(m) = \frac{1}{\sqrt{2\pi}\sigma_m}\exp\left[-\frac{m-\mu_m}{2\sigma_m^2}\right]$$

Above equation the intensity and contrast values which is used for quality measurement of global contrast and intensity values of tone mapping videos, Then HDR to LDR in natural scenes is evaluated as follows as

$$P_d(d) = \frac{(1-d)^{\beta_d - 1} d^{a_d - 1}}{B(a_d, \beta_d)}$$

And the statistical measurement is as follows

$$N = \frac{1}{K} P_m P_d$$

Quality assessment model

In much wide variety of applications, users preferred the single score by combining the overall score, IQA (Video quality assessment) is measured as follows

$$Q = aS^{\propto} + (1 - a)N^{\beta}$$

Alpha and beta parameters play a crucial role in quality assessment .The objective assessment is innovative method which overcomes the all drawbacks we faced in literature, and by using the innovative methods quality assessment of videos is done in ease way by IQA.

4. RESULTS

The Structural fidelity is done in multi scale approach as shown in following figures, here ‘S’ is total multi scale score and ‘s1’’s2’’s3’’s4’’s5’ are number of multi scales used in Structural fidelity. ‘Q’ is total quality score and where as ‘N’ is scene naturalness. The objective approach mainly relies on structural fidelity “S” and Scene naturalness measurement “N”.

VIDEO 1

Frame 1

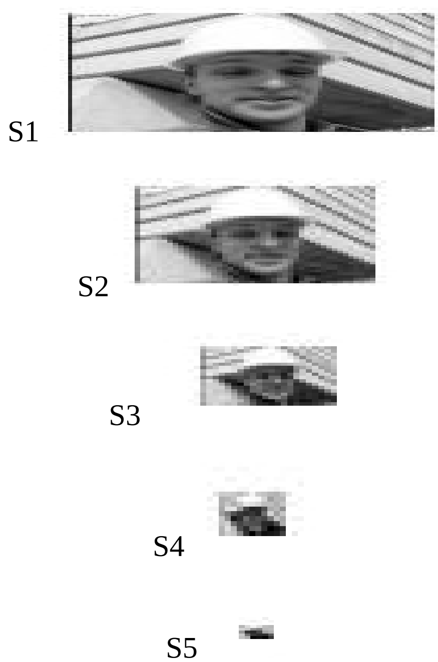


Figure 2: Frame 1 multi scales used in Structural fidelity

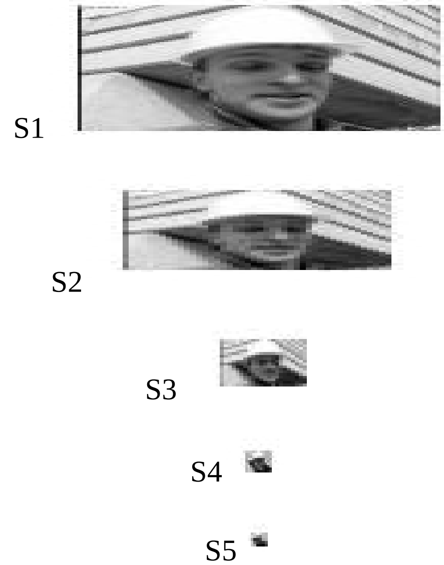


Figure 3: Frame 2 multi scales used in Structural fidelity

Frame Number	**Overall Quality(SSIM & NSS)**
1	0.83532
2	0.83508
3	0.83442
4	0.83503
5	0.83631
6	0.83723
7	0.83823
8	0.83899
9	0.84051
10	0.84107

Table 1: Overall Quality Score

VIDEO 2

Frame 2

S1

S2

S3

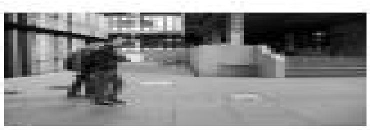
S4

S5

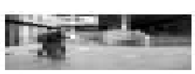
Figure 4: Frame 1 multi scales used in Structural fidelity

Frame 2

S1



S2



S3

S4

S5

Figure 5: Frame 2 multi scales used in Structural fidelity

Frame Number	Overall Quality(SSIM & NSS)
1	0.75976
2	0.75962
3	0.76031
4	0.76117
5	0.76195
6	0.76188
7	0.76204
8	0.76238
9	0.76265
10	0.76247

Table 2: Overall Quality Score

video1

Frame 1

Figure 6: LDR images generated with different parameter S= 0.89905(S1=0.77153 ,S2=0.85502 , S3=0.89341, S4=0.93916 , S5=0.98948) , N= 0.18311, Q= 0.83532

Frame 2

Figure 7: LDR images generated with different parameter S= 0.90043 (S1=0.76923 , S2=0.8567 , S3=0.89546 , S4=0.94105 , S5=0.98908) , N= 0.1805, Q= 0.83508

Video 1:- average of video quality score 0.837243

NO. OF	CONSTANT VALUES			SSIM	NSS	Q=(a*(S^α) + (1-a)*(N^β))
Frames	A	A	B	S	N	Q
1	0.8012	0.3046	0.7088	0.89905	0.18311	0.83532
2	0.8012	0.3046	0.7088	0.90043	0.1805	0.83508
3	0.8012	0.3046	0.7088	0.9001	0.17801	0.83442
4	0.8012	0.3046	0.7088	0.90275	0.17767	0.83503
5	0.8012	0.3046	0.7088	0.90483	0.18083	0.83631
6	0.8012	0.3046	0.7088	0.9062	0.18324	0.83723
7	0.8012	0.3046	0.7088	0.90748	0.18615	0.83823
8	0.8012	0.3046	0.7088	0.90787	0.18901	0.83899
9	0.8012	0.3046	0.7088	0.91173	0.19124	0.84051
10	0.8012	0.3046	0.7088	0.91251	0.19281	0.84107

Table 3: Calculate SSIM & NSS Score

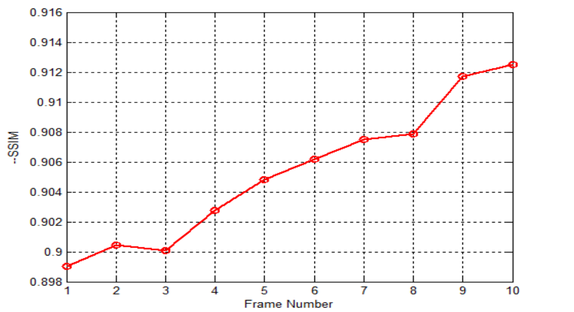


Figure 8: Structural Similarity Index Matrix Values Graph

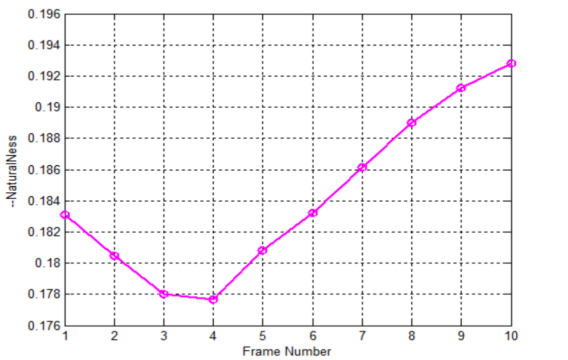


Figure 9: Naturalness scene statistics values Graph

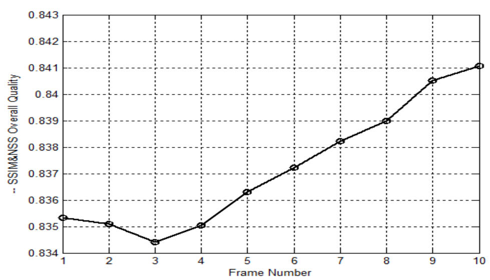


Figure 10: Overall Quality score Graph

video2

frame1

Figure 10: LDR images generated with different parameter S= 0.78967 (S1=0.61727 ,S2= 0.72682,S3= 0.7938, S4=0.84413 , S5=0.8967), N= 0.024061, Q= 0.75976

frame2

Figure 11: LDR images generated with different parameter S= 0.78968 (S1=0.61301 ,S2=0.7255 , S3=0.79521 , S4=0.84593, S5=0.89857) , N= 0.023722, Q= 0.75962

Video 2:- average of video quality score 0.761423

NO. OF	**CONSTANT VALUES**			**SSIM**	**NSS**	**Q=(a*(S^α) + (1-a)*(N^β))**
Frames	**a**	**A**	**B**	**S**	**N**	**Q**
1	0.8012	0.3046	0.7088	0.78967	0.024061	0.75976
2	0.8012	0.3046	0.7088	0.78968	0.023722	0.75962
3	0.8012	0.3046	0.7088	0.79192	0.02428	0.76031
4	0.8012	0.3046	0.7088	0.7943	0.02428	0.76117
5	0.8012	0.3046	0.7088	0.79663	0.024546	0.76195
6	0.8012	0.3046	0.7088	0.79656	0.024414	0.76188
7	0.8012	0.3046	0.7088	0.79726	0.027314	0.76204
8	0.8012	0.3046	0.7088	0.79843	0.024338	0.76238
9	0.8012	0.3046	0.7088	0.79946	0.024277	0.76265
10	0.8012	0.3046	0.7088	0.79897	0.024191	0.76247

Table 4: Calculate SSIM & NSS Scores

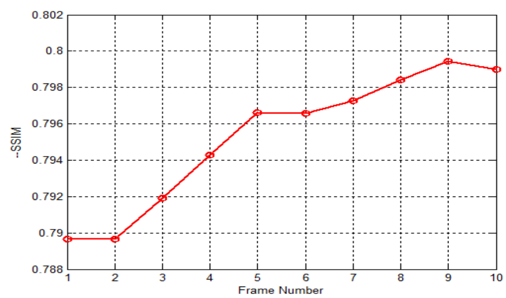


Figure 12: Structural Similarity Index Matrix Values Graph

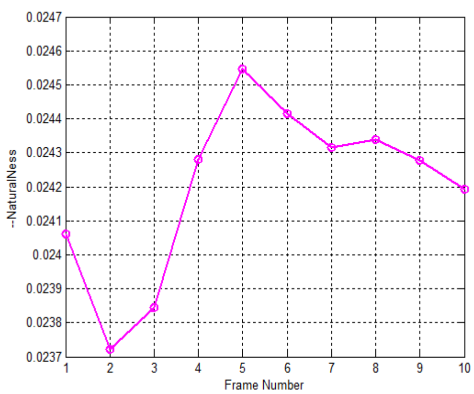


Figure 13: Naturalness scene statistics values Graph

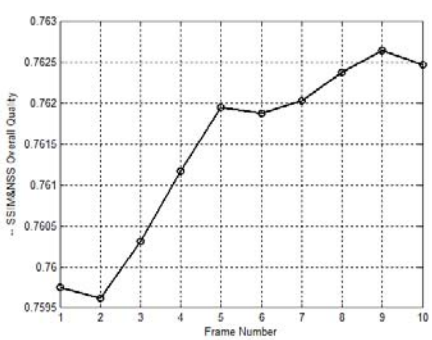


Figure 14: Overall Quality score Graph

CONLUSION

The Objective video quality assessment by using mean of the objective quality assessment of images (or frames) of input video, here quality score of the frames are nothing but the quality of the video. Our experiment show that reasonably correlated by TMQI with evaluation subject values of frames, the current procedure is applicable and tested using natural videos only. Assessment of videos is a time consuming process, as video size increases simultaneously time to read all frames in a video also increases. Till now assessment of videos is done on subjective approach, which is quite expensive and time consuming, so in near future extension of proposed method done on videos in objective approach which is more reliable.

REFERENCES

[1] F. Drago, W. L. Martens, K. Myszkowski, and Seidel H. P. Perceptual evaluation of tone mapping operators.In Proc.Of the SIGGRAPH Conf. Sketches and Applications, 2003.

[2] Martin _Cad__k and Pavel Slav__k. The naturalness of reproduced high dynamic range images. In IV '05: Proceedings of the Ninth International Conference on Informa-tionVisualisation, pages 920{925, Washington, DC, USA, 2005. IEEE Computer Society.

[3] T. O. Aydm, R. Mantiuk, K. Myszkowski, and H. Seidel.Dynamic range inde- pendent image quality assessment. In SIGGRAPH'08: International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH, 2008.

[4] F. Drago, K. Myszkowski, T. Annen, and N. Chiba.Adaptive logarithmic mapping for

displaying high contrast scenes. Computer Graphics Forum, 22(3):419{426,2003.

[5] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, “Evaluation oftone mapping operators using a high dynamic range display,” ACMTrans. Graph., vol. 24, no. 3, pp. 640–648, 2005.

[6] P. J. Burt and E. H. Adelson.The Laplacian pyramid as a compact image code. IEEE Trans. Communications, 31:532{540, April 1983.

[7] A. Yoshida, V. Blanz, K. Myszkowski, and H. Seidel, “Perceptualevaluation of tone mapping operators with real-world scenes,” Proc.SPIE, Human Vis. Electron.Imag., vol. 5666, pp. 192–203, Jan. 2005.

[8] M. ˇ Cadík, M. Wimmer, L. Neumann, and A. Artusi, “Image attributesand quality for evaluation of tone mapping operators,” in Proc. 14thPacific Conf. Comput. Graph. Appl., 2006, pp. 35–44.

[9] Martin _Cad__k, Michael Wimmer, Laszlo Neumann, and Alessandro Artusi. Image attributes and quality for evaluation of tone mapping operators. In Proceedings of the 14th Paci_c Conference on Computer Graphics and Applications, pages 35{ 44, Taipei, Taiwan, 2006. National Taiwan University Press.

[10] A. J. Kuang, H. Yamaguchi, G. M. Johnson, and M. D. Fairchild,“Testing HDR image rendering algorithms,” in Proc. IS T/SID ColorImag. Conf., 2004, pp. 315–320.