Eurographics Italian Chapter Conference (2006)
G. Gallo and S. Battiato and F. Stanco (Editors)

An Experimental Setup to Evaluate the Performance of Tone Mapping Algorithms

S. Marsi, G. Impoco, A. Ukovich, G. Ramponi

Dipartimento di Elettrotecnica Elettronica e Informatica, University of Trieste, Italy

Abstract

Many algorithms have been proposed in the literature to solve the problem of mapping tones from high dynamic range *sources to low dynamics displays, while preserving the information conveyed by the original scene. Methods and tools to evaluate the performance of tone mapping algorithms are needed. Some experiments have recently been presented involving human observers. In this paper, we propose method not relying on human scoring. Well-illuminated reference pictures of challenging scenes are captured and used as references. Low-quality shots of the same scenes are enhanced using some tone mapping algorithms and the outcome is compared to the reference images using a standard colour distance measure. The results clearly favour the more sophisticated local methods.*

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Viewing Algorithms

1. Introduction

A number of software techniques and hardware devices have recently been devised to acquire pictures whose dynamics can be of several orders of magnitude greater than that of common displays. The dynamics of rendered synthetic scenes can be even higher. Many algorithms have been proposed in the literature to solve the problem of mapping tones from *high dynamic range* (HDR) sources to low dynamics (LDR) displays, while preserving the information conveyed by the original scene i.e., the impression of "being there" [JW97, TT99, FCM00, Ash02, DD02, PY02, FLW02, RSSF02, LSC04, PSUJ04, RGM04, MCR04, IMR05, KMS05]. Although most of the existing techniques rely on common features and produce slightly different results, putting the stress on some specific image features they preserve or enhance certain features more then others. Therefore, there is no all-purpose method to evaluate the outcome of tone mapping algorithms.

Some studies [DMMS02] use preference and similarity between tone-mapped images expressed by observers to rate operators. However, we cannot rely on visual observation to rate the quality of tone-mapped images without any reference scene. Even if HDR displays have been used to evaluate tone mapping techniques [LCTS05], this method can only cope with scenes whose dynamics fits in the displayable range of the HDR device. Moreover, and most important, the display should first be validated against reality. In order to overcome these limitations, tone-mapped images can be evaluated against the corresponding real scene [YBMS05]. However, psychophysical data shows that the immersive visual experience of everyday life is much richer than a displayed scene, however high the dynamics of the display might be. Peripheral effects, of which little is known, can affect the global impression conveyed by a real scene. Acquisition devices do not capture these contributions. Hence, the evaluation rating can be biased by external uncontrollable factors.

Due to the low reproducibility of such experiments and to the difficulty in setting them up, we focused on evaluation methods that do not rely on user observation. General image quality measures compute the similarity between the original and processed images. Although they proved useful for evaluating the loss of information due to lossy compression or noise corruption, they are of little use to compare tone-mapped images. More sophisticated measures try to mimic the behaviour of the human visual system (HVS) [WB02] or evaluate contrast enhancement capabilities of the tone mapping operators [IUR05, UIR05]. In general, there are two main reasons why these image quality measures are not suitable to our purpose. First, they favour algorithms which produce little image variations, since they penalise pixel-wise differences between the original and processed images.

(a) Reference image (I_R).

(b) Badly illuminated image (I_{B_1}).

(c) Worse illuminated image (I_{B_2}).

Figure 1: *Some test pictures used in our experiments.*

However, it is very likely that tone mapping algorithms introduce distortion and vary consistently the overall luminance of the image to preserve image details. Second, the reference image (HDR) and the processed image (LDR) have different dynamics, thus comparing them is not obvious. Moreover, most of these measures are not suited for colour images.

In this paper, we propose an experiment to evaluate tone mapping methods by comparing enhanced test images to a well-illuminated reference image. The CIEDE2000 [LCR01] colour distance measure is used to keep into account human colour perception.

2. Experimental Setup

In our experiment, we consider a number of colour images that contain wide under- and over-exposed areas. For each scene, we acquired a well-illuminated reference image, I_R, and a number badly-illuminated ones, I_{B_i}. All the scenes were acquired by putting the camera in front of a window in daylight, in a room illuminated with a neon diffuse light. A bright light source was also placed on one side of the camera, pointing at the scene. For each scene we took three shots using, respectively, the bright source and the neon light (I_R, Figure 1(a)), the neon light only (I_{B_1}, Figure 1(b)), and no artificial lights at all (I_{B_2}, Figure 1(c)). The acquired pictures were processes using a commercial tool to reduce the effects of the colour temperature of the illuminats. The objective of this experiment is enhancing the images I_{B_i} in order to obtain images I_{P_i} that are as close as possible to I_R. The underlying idea is that an enhancement technique should compensate for the illumination changes all over the image, as if it were acquired under good lighting conditions. The scene in Figure 1 is challenging, since the window has approximately the same illumination in all the images, while the inner illumination is quite different. We expect that locally-adaptive algorithms outperform global methods, since the latter cannot use different illumination corrections for the window and the inner side of the room. The corresponding enhanced images are shown in Figure 2.

(a) I_{B_1}

(b) I_{B_2}

Figure 2: *Images obtained processing the test pictures, using our algorithm.*

3. Tone Mapping Operators

We used the experimental setup described in Section 2 to evaluate the performance of a few tone mapping methods. Here, we present the evaluation results for two simple methods: gamma correction and histogram equalisation. More sophisticated algorithms to reduce the dynamics of pictures also exist [JW97, TT99, FCM00, Ash02, DD02, PY02, FLW02, RSSF02, LSC04, PSUJ04, RGM04, MCR04, IMR05, KMS05]. Most of them are based on the observation that the radiance that hits the lens of an acquisition system is given by the interaction between the illuminant radiation and the reflectance of objects in the scene. Making some simplifications, we can write

$$I(x,y) = L(x,y) \cdot R(x,y) \quad , \tag{1}$$

where $I(x,y)$ is the image acquired by the sensor, $L(x,y)$ is the irradiance of the illuminant, and $R(x,y)$ is the reflectance of the material of objects in the scene. The idea is to reduce the dynamics of $L(x,y)$, while preserving the detail, $R(x,y)$. Since we know only $I(x,y)$, a method to estimate $L(x,y)$ is needed. Observing that the illumination usually varies slowly across the scene, save along strong boundaries, $L(x,y)$ can be estimated from $I(x,y)$ using an edge-preserving smoothing filter. Given $I(x,y)$ and the estimated illumination $L'(x,y)$, the reflectance, $R'(x,y)$, can be computed using Equation 1. The dynamics of the estimated illumination, $L'(x,y)$, is reduced using an attenuation function, such as a gamma correction. The compressed illumination, $\hat{L}(x,y)$, is put together with the reflectance to obtain the low-dynamics image, $\hat{I}(x,y) = \hat{L}(x,y) \cdot R'(x,y)$. The reflectance, $R'(x,y)$, can also be boosted to enhance detail.

Among these algorithms, we test only the one proposed by Marsi et al. (*locally-adaptive-HDR-reduction* algorithm, LAR) [MCR04, IMR05]. We are planning to use our experiment to evaluate the performance of other well known tone mapping operators. In the algorithm by Marsi et al., the illumination is estimated using an IIR edge-preserving smoothing filter. The estimated illumination is gamma corrected.

4. Colour Images

The algorithm described so far is intended for greylevel images, since it processes a single colour channel. In this section, we study some colour mapping techniques and colour spaces in order to process colour images.

Several different colour models have been proposed in the literature [Fai05] but there is still no widely accepted model. Moreover, people working in the field of image enhancing and high dynamic range reduction tend to be vague about how they treat colour in their algorithms. In order to choose a colour mapping technique that fits well our needs, we made some experiments on colour mapping methods. Different colour models were compared using the CIEDE2000 colour distance measure [LCR01]. We favour the colour model showing the minimum distance value. Since the CIEDE2000 measure is point-wise, we compute the mean of the pixel distances in order to have a single value for the whole image.

There are two main methods in the literature to process colour images. Either we process all the colour channels, or a single luminance channel is processed and then the colour channels are added back. One of the most common approaches is to process the three RGB channels separately. Each colour channel is processed using the same algorithm, without using any information from other channels. Then, the three colour channels are simply put together for visualisation. We would expect to get false colours due to colour shifting, since the RGB channels are not statistically uncorrelated. Though, in our experience the colours are represented correctly. Another common approach is to process only the luminance channel, extracted from the RGB triplet. The colour is then added back using one of two methods. Either the chrominance channels are left unchanged (e.g. L*a*b*) or a suitable correction is made, based on the luminance channel. A function to map colours using luminance is:

$$C_i' = \left(\frac{Y'}{Y}\right)^s \cdot C_i \tag{2}$$

where Y and Y' are the input and output luminances, and C_i and C_i' are the input and output tones for the i-th colour channel. Finally, s is a parameter used to tune the gain in the brightness and saturation of colours, and usually ranges in $(0,1]$. This methods is simple and it is faster than the previous ones, since it processes only the luminance rather the three RGB colour channels. Anyway, it is not mathematically sound since, for $s \neq 1$, $Y' = Y$ does not imply $C_i'(x,y) = C_i(x,y)$. This does not happen using luminance/chrominance colour spaces and processing only the luminance, leaving unchanged the chrominance channels. In our experiments, we used the L*a*b* and HSV colour spaces.

method	distance
RGB	91.41
Y	90.14
L*a*b*	86.93
HSV	103.2

Table 1: *Distances measured between the reference image and the output images for different colour mapping methods.*

The mean distance values measured are reported in Table 1. As the table shows, there is no significant difference between the methods tested. Hence, we prefer the L*a*b* space since it is technically sound and does not duplicate burdensome computation as the RGB mapping.

5. Experimental Evaluation

We tested three enhancing methods: histogram equalisation, gamma correction, and the LAR algorithm. For each method, we compute the (point-wise) CIEDE2000 colour distance with respect to the reference image, I_R. As we do in Section 4, we also compute the mean of the distances computed for single pixels. The results are reported in Table 2, while the corresponding images are shown in Figure 3. The greylevel pixel map encodes the distance between the original and the processed images, using a ramp where dark intensities represent low distances and bright pixels correspond to high distance values. Notice that the range of the distance values is too wide to be visualised as it is. Hence, we clamp low and high values to a suitable range for displaying. In the following, we assume that our images range in $[0,1]$.

The figures show that all the methods perform equally well in the sky area, since it is bright and has low contrast (i.e. no enhancement is needed). At a first glance, the histogram equalisation (Figure 3(a)) seems to produce the best results since the output image is bright and well contrasted. However, false colours are generated, mainly around the window. This colour shift is captured by the distance measure, that correctly scores this method as the worst. Moreover, the outer scene is strongly brighten up, so that many details are lost. The distance images for the other methods are similar. However, the distance image computed for the gamma correction (Figure 3(b)) has a higher contrast than our method. This means that the distance is higher around edges. Hence, either the detail is not well exploited (in fact, it is reduced) or artifact are generated around edges. This effect is less visible in the local method. This suggests that edges and details are better preserved by methods based on intrinsic images. These observations are more evident in the close-ups of Figure 3. Histogram equalisation deviates completely from the reference image, while the other methods are more

(a) Histogram equalisation. The distance is equal to $d = 271.0$.

(b) Gamma correction ($\gamma = 0.55$). Distance: $d = 97.21$

(c) Our algorithm, using our edge-preserving low-pass filter. $d = 91.37$

Figure 3: *Images processed using different methods. The mean distance of the test picture in Figure 1(b) is also shown. Dark intensities indicate a lower distance, thus the darker the better.*

robust. Moreover, the distance image corresponding to our method is more blurred, thus the error around edges is well distributed. This is a nice property because it means that the details are better exploited. The gamma correction does not show this behaviour.

Our experiments show that separating the input image in its illumination and reflectance components worths the computational overhead of estimating the illumination. To give some more evidence of this, we try plot in Figure 4 the distance measure computed using a gamma correction with different values of gamma. Two different plots are shown for low and high frequencies. As the figure suggests, the optimal gamma value used to correct the illumination is significantly lower than the value used to adjust the reflectance component. This is mainly due to the nature of our test images (refer to Figure 1(b)). Basically, the algorithm tries to improve the visibility in dark zones using a strong correction by means of a low gamma (recall that we assume that our images range in $[0,1]$). Conversely, the reflectance should be only slightly enhanced to avoid amplifying the sensor noise. Hence, separating the illumination and reflectance components gives an additional degree of freedom in the choice of parameters, that might yield better results than directly operating on the intensity values of the image.

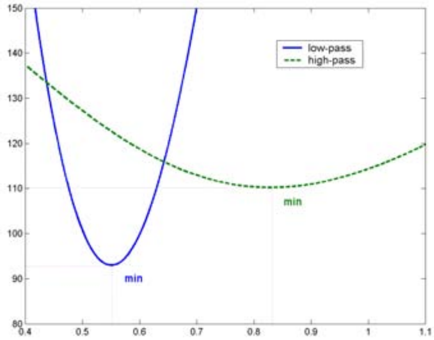

Figure 4: *Plot of the gamma functions that minimise the low-pass and high-pass images. The optimal gamma for low frequencies is significantly lower than the optimal one for high frequencies.*

Finally, Table 2 shows some distance measurements, computed as the mean CIEDE2000 distances between corresponding pixels. We report the mean distances computed for the original image, histogram equalisation, gamma correc-

method	distance (I_{B_1})	distance (I_{B_2})
original	472.0	1112.0
histogram equalisation	271.0	403.0
gamma correction	97.21	233.0
LAR	91.37	197.2

Table 2: *Distances computed between the reference image, I_R, and the output images, I_{B_1} and I_{B_2}, for different methods (see text).*

tion, and the LAR method. Clearly, the local method outperforms global ones. This is even more evident in challenging situations (e.g., the image I_{B_2}).

6. Conclusions

We have presented an experiment to evaluate the performance of tone mapping techniques that does not rely on human observation. Well-illuminated reference pictures of challenging scenes are captured. Then, some low-quality shots are acquired and enhanced using some tone mapping algorithms. The outcome is compared to the reference images using a standard colour distance measure. The results suggest that the chosen local method outperforms simple enhancing methods such as histogram equalisation and gamma correction.

Our choice to avoid human scoring is justified by the difficulties of setting up experiments involving human observers and by other technical considerations. We plan to use our experimental setup to evaluate the most common state-of-the-art tone mapping operators.

Acknowledgements

This work was supported by the PRIN project No.2003098078 and has been conducted in the framework of the SCHEMA NoE (IST-2001-32795).
We thank one of the anonymous reviewers who detected a mistake in our first draft.

References

[Ash02] ASHIKHMIN M.: A tone mapping algorithm for high contrast images. In *13th Eurographics Workshop on Rendering* (2002), Eurographics Association.

[DD02] DURAND F., DORSEY J.: Fast bilateral filtering for the display of high-dynamic-range images. In *SIGGRAPH '02* (2002), ACM Press, pp. 257–266.

[DMMS02] DRAGO F., MARTENS W., MYSZKOWSKI K., SEIDEL H.-P.: *Perceptual Evaluation of Tone Mapping Operators with Regard to Similarity and Preference*. Research Report MPI-I-2002-4-002, MPI, Saarbrücken, Germany, August 2002.

[Fai05] FAIRCHILD M. D.: *Color Appearance Models, 2nd Edition*. Imaging Science and Technology. Wiley-IS&T, 2005.

[FCM00] FUNT B. V., CIUREA F., MCCANN J. J.: Retinex in MatLab. In *Color Imaging Conference* (2000), pp. 112–121.

[FLW02] FATTAL R., LISCHINSKI D., WERMAN M.: Gradient domain high dynamic range compression. *ACM Transanctions on Graphics 21*, 3 (2002), 249–256.

[IMR05] IMPOCO G., MARSI S., RAMPONI G.: Adaptive reduction of the dynamics of hdr video sequences. In *ICIP 05* (11-14 September 2005).

[IUR05] IMPOCO G., UKOVICH A., RAMPONI G.: Exploiting the local contrast to evaluate image dynamics reduction algorithms. In *COST-05* (May 2005).

[JW97] JOBSON D., WOODELL G.: Properties of a center/surround Retinex. *IEEE Transactions on Image Processing 6*, 3 (March 1997), 451–462.

[KMS05] KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Lightness perception in tone reproduction for high dynamic range images. In *EUROGRAPHICS 2005* (Dublin, Ireland).

[LCR01] LUO M., CUI G., RIGG B.: The development of the CIE 2000 colour-difference formula : CIEDE2000. *Color Research and Application 26*, 5 (Oct. 2001).

[LCTS05] LEDDA P., CHALMERS A., TROSCIANKO T., SEETZEN H.: Evaluation of tone mapping operators using a high dynamic range display. *ACM Trans. Graph. 24*, 3 (2005), 640–648.

[LSC04] LEDDA P., SANTOS L. P., CHALMERS A.: A local model of eye adaptation for high dynamic range images. In *AFRIGRAPH2004* (Nov. 2004), ACM Press.

[MCR04] MARSI S., CARRATO S., RAMPONI G.: Image contrast enhancement using recursive rational filters. In *IST-2004* (May 2004).

[PSUJ04] PAL C., SZELISKI R., UYTTENDAELE M., JOJIC N.: Probability models for high dynamic range imaging. In *CVPR04* (2004), pp. II: 173–180.

[PY02] PATTANAIK S., YEE H.: Adaptive gain control for high dynamic range image display. In *SCCG '02: 18th Spring Conf. on Computer Graphics* (2002), ACM Press, pp. 83–87.

[RGM04] RIZZI A., GATTA C., MARINI D.: From Retinex to automatic color equalization: Issues in developing a new algorithm for unsupervised color equalization. *J. of Electronic Imaging 13* (Jan 2004), 75–84.

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. In *SIGGRAPH '02* (2002), ACM Press, pp. 267–276.

[TT99] TUMBLIN J., TURK G.: LCIS: a boundary hierarchy for detail-preserving contrast reduction. In *SIGGRAPH '99* (1999), pp. 83–90.

[UIR05] UKOVICH A., IMPOCO G., RAMPONI G.: A tool based on the co-occurrence matrix to measure the performance of dynamic range reduction algorithms. In *IST-05* (May 2005).

[WB02] WANG Z., BOVIK A.: A universal image quality index. *IEEE Signal Processing Letters 9* (March 2002), 81–84.

[YBMS05] YOSHIDA A., BLANZ V., MYSZKOWSKI K., SEIDEL H.-P.: Perceptual evaluation of tone mapping operators with real-world sceness. In *IS&T/SPIE's 17th Symposium on Electronic Imaging* (San Jose, USA, January 2005), pp. 192–203.