

Learning to display high dynamic range images

Guoping Qiu^{a,*}, Jiang Duan^a, Graham D. Finlayson^b

^a*School of Computer Science, The University of Nottingham, Jubilee Campus, Nottingham NG8 1BB, UK*

^b*School of Computing Science, The University of East Anglia, UK*

Received 27 September 2005; received in revised form 10 April 2006; accepted 14 February 2007

Abstract

In this paper, we present a learning-based image processing technique. We have developed a novel method to map high dynamic range scenes to low dynamic range images for display in standard (low dynamic range) reproduction media. We formulate the problem as a quantization process and employ an adaptive conscience learning strategy to ensure that the mapped low dynamic range displays not only faithfully reproduce the visual features of the original scenes, but also make full use of the available display levels. This is achieved by the use of a competitive learning neural network that employs a frequency sensitive competitive learning mechanism to adaptively design the quantizer. By optimizing an L_2 distortion function, we ensure that the mapped low dynamic images preserve the visual characteristics of the original scenes. By incorporating a frequency sensitive competitive mechanism, we facilitate the full utilization of the limited displayable levels. We have developed a deterministic and practicable learning procedure which uses a single variable to control the display result. We give a detailed description of the implementation procedure of the new learning-based high dynamic range compression method and present experimental results to demonstrate the effectiveness of the method in displaying a variety of high dynamic range scenes.

© 2007 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Learning-based image processing; Quantization; High dynamic range imaging; Dynamic range compression; Neural network; Competitive learning

1. Introduction

With the rapid advancement in electronic imaging and computer graphics technologies, there have been increasing interests in high dynamic range (HDR) imaging, see e.g., Ref. [1–17]. Fig. 1 shows a scenario where HDR imaging technology will be useful to photograph the scene. This is an indoor scene of very HDR. In order to make features in the dark areas visible, longer exposure had to be used, but this rendered the bright area saturated. On the other hand, using shorter exposure made features in the bright areas visible, but this obscured features in the dark areas. In order to make all features, both in the dark and bright areas simultaneously visible in a single image, we can create a HDR radiance map [3,4] for the scene. Using the technology of Ref. [3], it is relatively easy to create HDR maps for high dynamic scenes. All one needs is a sequence of low

dynamic range (LDR) photos of the scene taken with different exposure intervals. Fig. 2 shows the LDR display of the scene in Fig. 1 mapped from its HDR radiance map, which has been created using the method of [3] from the photos in Fig. 1. It is seen that all areas in this image are now clearly visible. HDR imaging technology has also been recently extended to video [13,14].

Although we can create HDR numerical radiance maps for high dynamic scenes such as those like Fig. 1, reproduction devices, such as video monitors or printers, normally have much lower dynamic range than the radiance map (or equivalently the real world scenes). One of the key technical issues in HDR imaging is how to map HDR scene data to LDR display values in such a way that the visual impressions and feature details of the original real physical scenes are faithfully reproduced.

In the literature, e.g., Refs. [5–17], there are two broad categories of dynamic range compression techniques for the display of HDR images in LDR devices [12]. The tone reproduction operator (TRO) based methods involve (multi-resolution) spatial processing and mappings not only take into account the

* Corresponding author. Fax: +44 115 951 4254.
E-mail address: qiu@cs.nott.ac.uk (G. Qiu).



Fig. 1. Low dynamic range photos of an indoor scene taken under different exposure intervals.



Fig. 2. Low dynamic display of high dynamic range map created from the photos in Fig. 1. The dynamic range of the radiance map is 488,582:1. HDR radiance map synthesis using Paul Debevec's HDRShop software (<http://gl.ict.usc.edu/HDRShop/>). Note: the visual artifacts appear in those blinds of the glass doors were actually in the original image data and not caused by the algorithm.

values of individual pixel but are also influenced by the pixel spatial contexts. Another type of approaches is tone reproduction curve (TRC) based. These approaches mainly involve the adjustment of the histograms and spatial context of individual pixel is not used in the mapping. The advantages of TRO-based methods are that they generally produce sharper images when the scenes contain many detailed features. The problems with these approaches are that spatial processing can be computationally expensive, and there are in general many parameters controlling the behaviors of the operators. Sometimes these techniques could introduce “halo” artifacts and sometimes they can introduce too much (artificial) detail. TRC-based methods are computationally simple. They preserve the correct brightness order and therefore are free from halo artifacts. These methods generally have fewer parameters and therefore are easier to use. The drawbacks of this type of methods are that spatial sharpness could be lost.

Perhaps one of the best known TRC-based methods is that of Ward and co-workers' [5]. The complete operator of Ref. [5] also included sophisticated models that exploit the limitations of human visual system. According to Ref. [5], if one just wanted to produce a good and natural-looking display for an HDR scene without regard to how well a human observer would be able to see in a real environment, histogram adjustment may provide an “optimal” solution. Although the histogram adjustment technique of Ref. [5] is quite effective, it also has drawbacks. The method only caps the display contrast

(mapped by histogram equalization) when it exceeds that of the original scene. This means that if a scene has too low contrast, the technique will do nothing. Moreover, in sparsely populated intensity intervals, dynamic range compression is achieved by a histogram equalization technique. This means that some sparse intensity intervals that span a wide intensity range will be compressed too aggressively. As a result, features that are visible in the original scenes would be lost in the display. This unsatisfactory aspect of this algorithm is clearly illustrated in Figs. 9–11.

In this paper, we also study TRC-based methods for the display of HDR images. We present a learning-based method to map HDR scenes to low dynamic images to be displayed in LDR devices. We use an adaptive learning algorithm with a “conscience” mechanism to ensure that, the mapping not only takes into account the relative brightness of the HDR pixel values, i.e., to be faithful to the original data, but also favors the full utilization of all available display values, i.e., to ensure the mapped low dynamic images to have good visual contrast. The organization of the paper is as follows. In the next section, we cast the HDR image dynamic range compression problem in an adaptive quantization framework. In Section 3, we present a solution to HDR image dynamic range compression based on adaptive learning. In Section 4, we present detailed implementation procedures of our method. Section 4.1 presents results and Section 4.2 concludes our presentation and briefly discusses future work.

2. Quantization for dynamic range compression

The process of displaying HDR image is in fact a quantization and mapping process as illustrated in Fig. 3. Because there are too many (discrete) values in the high dynamic scene, we have to reduce the number of possible values, this is a quantization process. The difficulty faced in this stage is how to decide which values should be grouped together to take the same value in the low dynamic display. After quantization, all values that will be put into the same group can be annotated by the group's index. Displaying the original scene in a LDR is simply to represent each group's index by an appropriate display intensity level. In this paper, we mainly concerned ourselves with the first stage, i.e., to develop a method to best group HDR values.

Quantization, also known as clustering, is a well-studied subject in signal processing and neural network literature. Well-known techniques such as k -means and various neural network-based methods have been extensively researched [18–21]. Let $x(k)$, $k = 1, 2, \dots$, be the intensities of the luminance component of the HDR image (like many other techniques, we only work on the luminance and in logarithm space, also, we treat

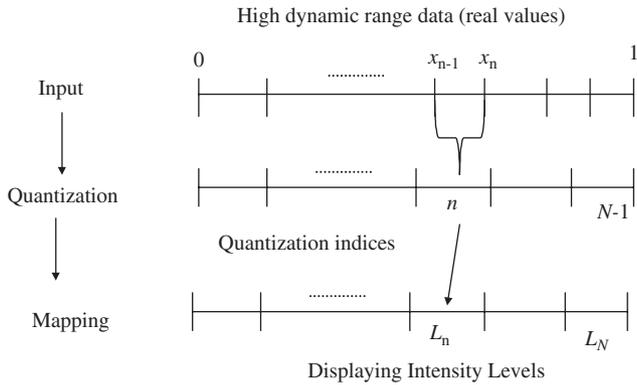


Fig. 3. The process of display of high dynamic range scene (from purely a numerical processing’s point of view).

each pixel individually and therefore are working on scalar quantization). A quantizer is described by an encoder Q , which maps $x(k)$ to an index $n \in N$ specifying which one of a small collection of reproduction values (codewords) in a codebook $C = \{c_n; n \in N\}$ is used for reconstruction, where N in our case, is the number of displayable levels in the LDR image. There is also a decoder Q^{-1} , which maps indices into reproduction values. Formally, the encoding is performed as

$$Q(x(k)) = n \quad \text{if } \|x(k) - c_n\| \leq \|x(k) - c_i\| \quad \forall i \quad (1)$$

and decoding is performed as

$$Q^{-1}(n) = c_n. \quad (2)$$

That is (assuming that the codebook has already been designed), an HDR image pixel is assigned to the codeword that is the closest to it. All HDR pixels assigned to the same codewords then form a cluster of pixels. All HDR pixels in the same cluster are displayed at the same LDR level. If we order the codewords that they are in increasing order, that is $c_1 < c_2 < \dots < c_{N-1} < c_N$, and assign display values to pixels belonging to the clusters in the order of their codeword values, then the mapping is order preserving. Pixels belonging to a cluster having a larger codeword value will be displayed brighter than pixels belonging to a cluster with a codeword having a smaller value.

Clearly, the codebook plays a crucial role in this dynamic range compression strategy. It not only determines which pixels should be displayed at what level, it also determines which pixels will be displayed as the same intensity and how many pixels will be displayed with a particular intensity. Before we discuss how to design the codebook, let’s find out what requirements for the codebook are in our specific application.

Recall that one of the objectives of mapping are that we wish to preserve all features (or as much as possible) of the HDR image and make them visible in the LDR reproduction. Because there are fewer levels in the LDR image than in the HDR image, the compression is lossy in the sense that there will be features in the HDR images that will be impossible to reproduce in the LDR images. The question is what should be

preserved (and how) in the mapping (quantization) in order to produce good LDR displays.

For any given display device, the number of displayable levels, i.e., the number of codewords is fixed. From a rate distortion’s point of view, the rate of the quantizer is fixed, and the optimal quantizer is the one that minimizes the distortion. Given the encoding and decoding rules of Eqs. (1) and (2), the distortion of the quantizer is defined as

$$E = \sum_{i,k} \delta_i(k) \|x(k) - c_i\| \quad \text{where} \quad (3)$$

$$\delta_i(k) = \begin{cases} 1 & \text{if } \|x(k) - c_i\| \leq \|x(k) - c_j\| \quad \forall j, \\ 0 & \text{otherwise.} \end{cases}$$

A mapping by a quantizer that minimizes E (maximizes rate distortion ratio) will preserve maximum relevant information of the HDR image.

One of the known problems in rate distortion optimal quantizer design is the uneven utilization of the codewords where, some clusters may have large number of pixels, some may have very few pixels and yet others may even be empty. There may be two reasons for this under utilization of codewords. Firstly, the original HDR pixel distribution may be concentrated in a very narrow range of intensity interval; this may cause large number of pixels in these densely populated intervals to be clustered into a single or very few clusters because they are so close together. In the HDR image, if an intensity interval has a large concentration of pixels, then these pixels could be very important in conveying fine details of the scene. Although such intervals may only occupy a relatively narrow dynamic range span because of the high-resolution representation, the pixels falling onto these intervals could contain important details visible to human observers. Grouping these pixels together will lose too much detail information. In this case, we need to “plug” some codewords in these highly populated intensity intervals such that these pixels will be displayed into different levels to preserve the details. The second reason that can cause the under utilization of codewords could be due to the optimization process being trapped in a local minimum. In order to produce a good LDR display, the pixels have to be reasonably evenly distributed to the codewords. If we assign each codeword the same number of pixels, then the mapping is histogram equalization which has already been shown to be unsuitable for HDR image display [5]. If we distribute the pixel population evenly without regarding to their relative intensity values, e.g., grouping pixels with wide ranging intensity values just to ensure even population distribution into each codeword, then we will introduce too much objectionable artifacts because compression is too aggressive. In the next section, we introduce a learning-based approach suitable for designing dynamic range compression quantizers for the display of HDR images.

3. Conscience learning for HDR compression

Vector quantization (VQ) is a well-developed field [16] (although we will only design a scalar quantizer (SQ), we will borrow techniques mainly designed for VQ. There are

many methods developed for designing vector quantizers. The k -means type algorithms, such as the LBG algorithm [18], and neural network-based algorithms, such as the Kohonen feature map [19] are popular tools. As discussed in Section 2, our quantizer should not only be based on the rate distortion criterion, but also should ensure that there is an even spread of pixel population. In other words, our quantizer should minimize E in Eq. (3) and at the same time the pixels should be evenly distributed to the codewords. Therefore the design algorithm should be able to explicitly achieve these two objectives. In the neural network literature, there is a type of conscience competitive learning algorithm that will suit our application. In this paper, we use the frequency sensitive competitive learning (FSCL) algorithm [20] to design a quantizer for the mapping of HDR scenes to be displayed in low dynamic devices. To understand why FSCL is ideally suited for such an application, we first briefly describe the FSCL algorithm.

The philosophy behind the FSCL algorithm is that, when a codeword wins a competition, the chance of the same codeword winning the next time round is reduced, or equivalently, the chance of other codewords winning is increased. The end effect is that the chances of each codeword winning will be equal. Put in the other way, each of the codewords will be fully utilized. In the context of using the FSCL for mapping HDR data to low dynamic display, the limited number of low dynamic levels (codewords) will be fully utilized. The intuitive result of fully utilized display level is that the displayed image will have good contrast, which is exactly what is desired in the display. However, unlike histogram equalization, the contrast is constrained by the distribution characteristics of the original scene's intensity values. The overall effect of such a mapping is therefore that, the original scenes are faithfully reproduced, while at the same time, the LDR displays will have good contrast.

FSCL algorithm

Step 1: Initialize the codewords, $c_i(0)$, $i=1, 2, \dots, N$, to random numbers and set the counters associated with each codeword to 1, i.e. $f_i(0) = 1$, $i = 1, 2, \dots, N$

Step 2: Present the training sample, $x(k)$, where k is the sequence index, and calculate the distance between $x(k)$ and the codewords, and subsequently modify the distances using the counters values

$$d_i(k) = \|x(k) - c_i(k)\| \quad d_i^*(k) = f_i(k)d_i(k) \quad (4)$$

Step 3: Find the winner codeword $c_j(k)$, such that $d_j^*(k) \leq d_i^*(k) \forall i$

Step 4: Update the codeword and counter of the winner as

$$c_j(k+1) = c_j(k) + \eta(x(k) - c_j(k)) \quad f_j(k+1) = f_j(k) + 1 \quad \text{where } 0 < \eta < 1 \quad (5)$$

Step 5: If converged, then stop, else go to Step 2.

The FSCL process can be viewed as a constrained optimization problem of the following form:

$$J = E + \lambda \sum_i \left(\left(\sum_k \delta_i(k) \right) - \frac{|x(k)|}{N} \right)^2, \quad (6)$$

where λ is the Lagrange multiplier, $|x(k)|$ represents the total number of training samples (pixels), N is the number of codewords (display levels), E and δ are as defined in Eq. (3).

Minimizing the first term ensures that pixel values close to each other will be grouped together and will be displayed as the same single brightness. Minimizing the second term facilitates that each available displayable level in the mapped image will be allocated similar number of pixels thus creating a well contrast display. By sorting the codewords in an increasing order, i.e., $c_1 < c_2 < \dots < c_{N-1} < c_N$, the mapping also ensures that the brighter pixels in the original image are mapped to a brighter display level and darker pixels are mapped to a darker display level thus ensuring the correct brightness order to avoid visual artifacts such as the "halo" phenomenon.

Because the mapping not only favors an equally distributed pixel distribution but also is constrained by the distances between pixel values and codeword values, the current method is fundamentally different from both traditional histogram equalization and simple quantization-based methods. Histogram equalization only concerns that the mapped pixels have to be evenly distributed regardless of their relative brightness, while simple quantization-based mapping (including many linear and nonlinear data independent mapping techniques, e.g., Ref. [2]) only takes into account the pixel brightness value while ignoring the pixel populations in each mapped level. As a result, histogram equalization mapping will create visually annoying artifacts and simple quantization-based method will create mappings with under utilized display levels which often resulting in many features being squashed and become invisible. With FSCL learning, we could achieve a balanced mapping.

4. Implementation details

Because the objective of our use of the FSCL algorithm in this work is different from that of its ordinary applications, e.g., vector quantization, there are some special requirements for its implementation. Our objective is not purely to achieve rate distortion optimality, but rather should optimize an objective function of the form in Eq. (6) and the ultimate goal is, of course, to produce good LDR displays of the HDR images. In this section, we will give a detailed account of its implementation.

4.1. Codebook initialization

We work on the luminance component of the image only and in logarithm space, from the floating point RGB pixel values of the HDR radiance map, we calculate

$$L = \log(0.299 * R + 0.587 * G + 0.114 * B), \\ L_{\max} = \text{MAX}(L), \quad L_{\min} = \text{MIN}(L). \quad (7)$$

In using the FSCL algorithm, we found that it is very important to initialize the codewords to linearly scaled values. Let the codebook be $C = \{c_i; i = 1, 2, \dots, N\}$, N is the number of codewords. In our current work, N is set to 256. The initial

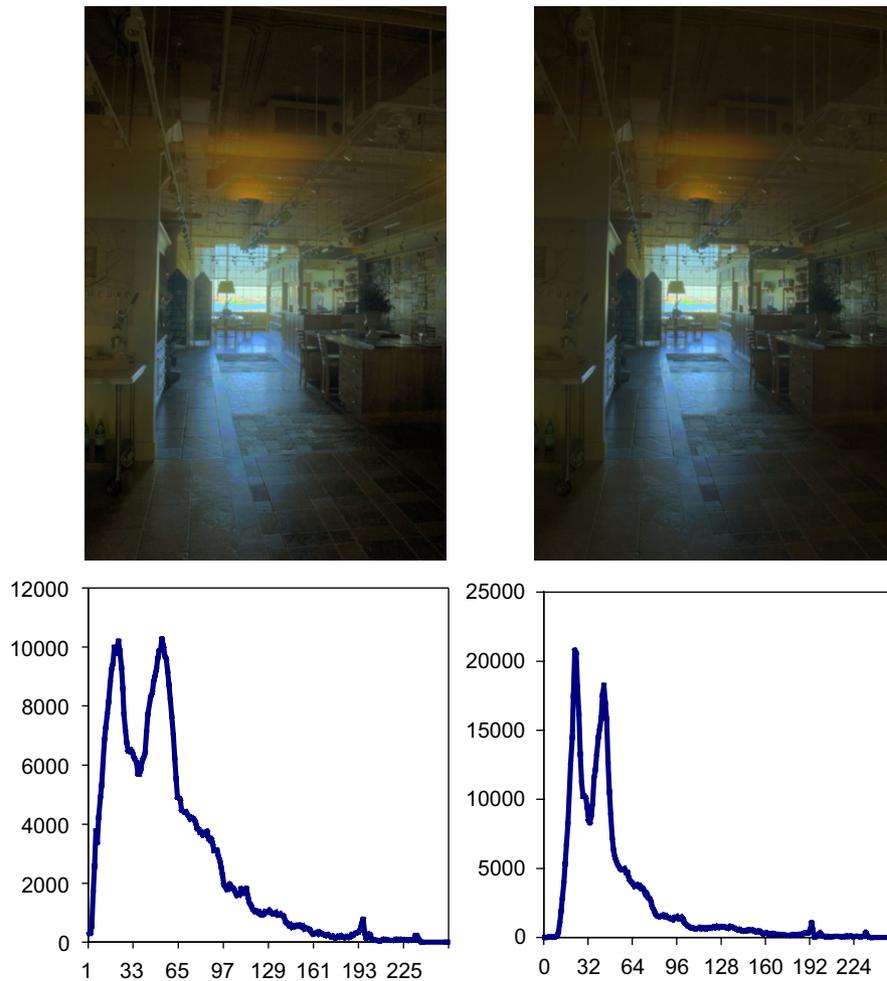


Fig. 4. Competitive learning without conscience mechanism will produce a mapping close to linear scaling. Left column: competitive learning without conscience mechanism. Right column: linear scaling. Radiance map data courtesy of Fredo Durand.

values of the codewords are chosen such that they are equally distributed in the full dynamic range of the luminance:

$$c_i(0) = L_{\min} + \frac{i}{256}(L_{\max} - L_{\min}). \quad (8)$$

Recall that our first requirement in the mapping is that the display should reflect the pixel intensity distribution of the original image. This is achieved by the optimization of E in Eq. (3) with the initial codeword values chosen according to Eq. (8). Let us first consider competitive learning without the conscience mechanism. At the start of the training process, for pixels falling into the interval $c_i(0) \pm (L_{\max} - L_{\min})/512$, the same codeword c_i will win the competitions regardless of the size of pixel population falling into the interval. The pixels in the intensity intervals that are far away from c_i will never have the chance to access it. In this case, distance is the only criterion that determines how the pixels will be clustered and pixel populations distributed into the codewords are not taken into account. After training, the codewords will not change much from their initial values because each codeword will only be updated by pixels falling close to it within a small range. In the encoding

or mapping stage, since the codewords are quite close to their corresponding original positions which are linearly scattered in the FDR, the minimum distortion measure will make the mapping approximately linear and extensive simulations demonstrate that this is indeed the case. As a result, the final mapped image will exhibit the pixel intensity distribution characteristics of the original image. Fig. 4(a) shows a result mapped by competitive learning without the conscience mechanism, compared with Fig. 4(b), the linearly scaled mapping result, it is seen that both the images and their histograms are very similar which demonstrate that by initializing the codewords according Eq. (8), optimizing the distortion function E in Eq. (3) has the strong tendency to preserve the original data of the original image. However, like linear compression, although the resulting image reflects the data distribution of the scene, densely populated intensity intervals are always over-quantized (too much compression) which causes the mapped image lacking contrast.

With the introduction of the conscience mechanism, in the training stage, competition is based on a modified distance measure, d^* in Eq. (4). Note that if a codeword wins the competitions frequently, its count and consequently its modi-

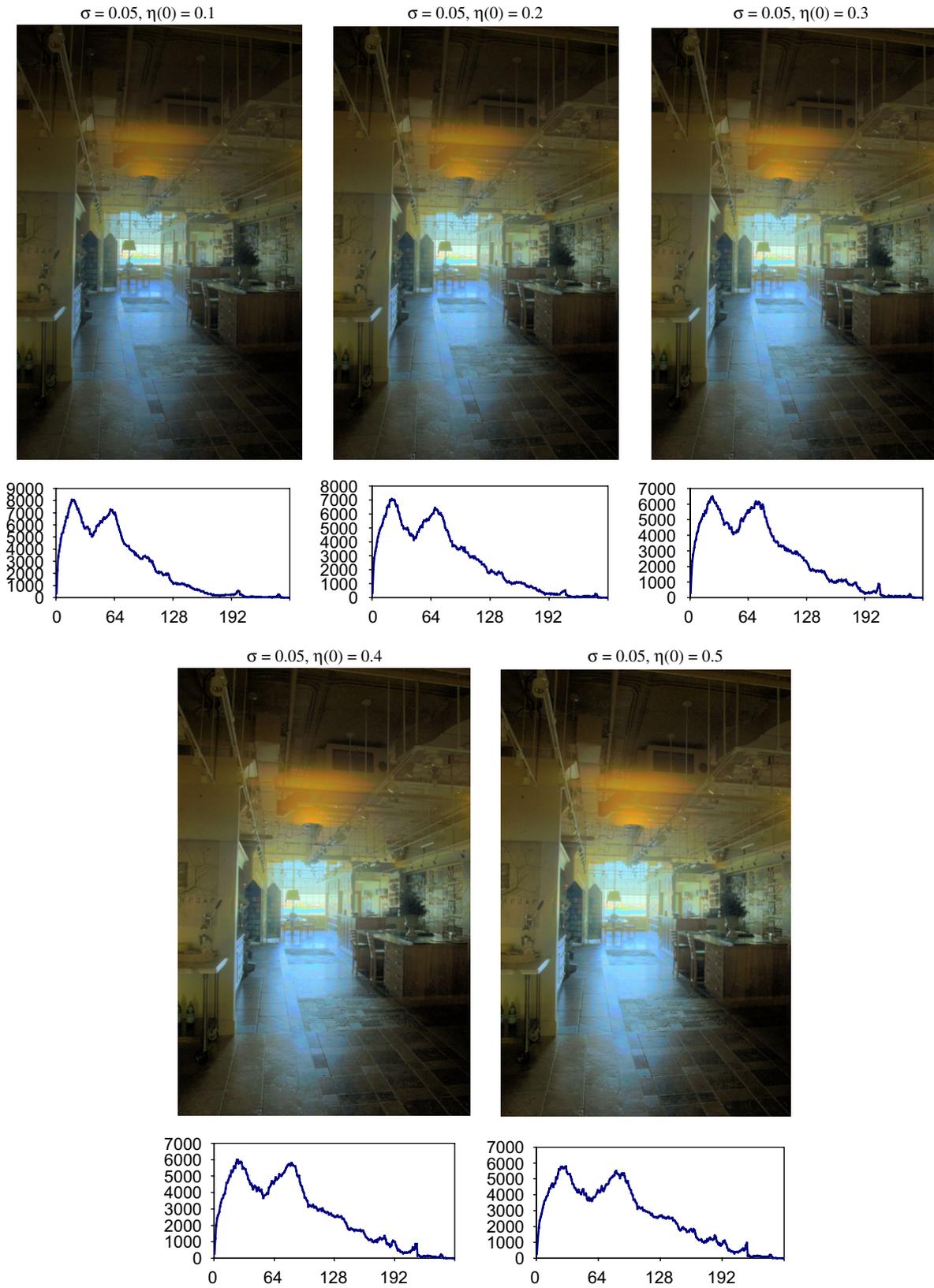


Fig. 5. Final mapped images and their histograms for fixed σ and various $\eta(0)$. Radiance map data courtesy of Fredo Durand.

fied distance increases, thus reducing its chance of being the winner and increasing the likelihood of other codewords with relatively lower count values to win. This property is exactly

what is desired. If a codeword wins frequently, this means that the intensity interval it lies in gathers large population of pixels. In order to ensure the mapped image having good

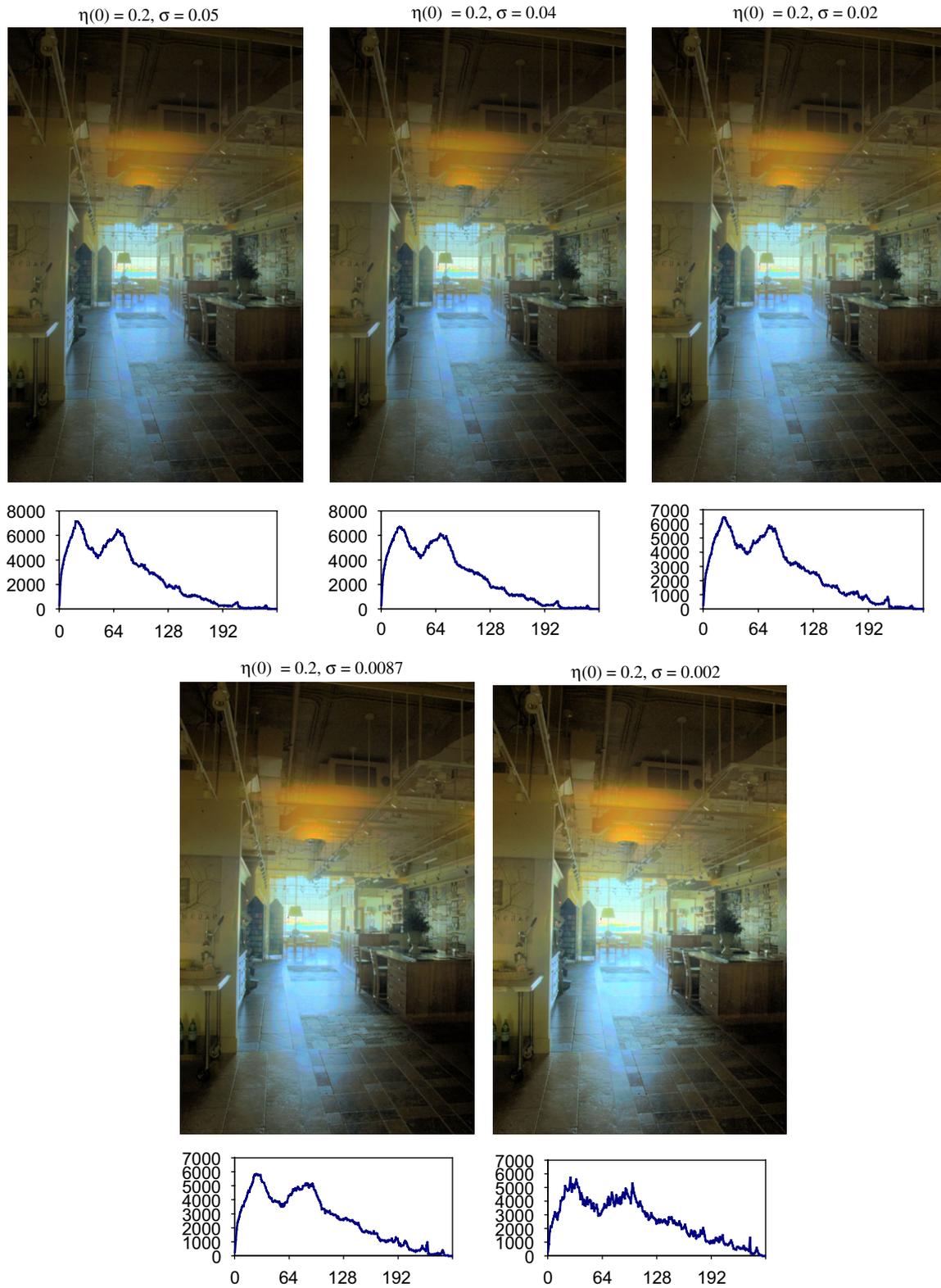


Fig. 6. Final mapped images and their histograms for fixed $\eta(0)$ and various σ . Radiance map data courtesy of Fredo Durand.

contrast, these densely populated intensity intervals should be represented by more codewords, or equivalently, more display levels. The incorporation of the conscience mechanism makes the algorithm conscience of this situation and passes the

chances to codewords with lower count values to win the competitions so that they can be updated towards the population dense intensity intervals and finally join in the quantization of these intervals.

The training starts from linearly scattered codewords, with the introduction of the conscience mechanism, the codewords are moved according to both their values and the pixel population distribution characteristics. At the beginning of the training phase, the mapping is linear, and the mapped image's histogram is narrow, as training progresses, the mapped image's histogram is gradually wider, and when it reaches an equilibrium, it achieves an optimal mapping that preserves the appearance of the original scene and also has good visibility and contrast.

Although initializing the codewords according to Eq. (8) may lead to larger final distortion E than other (random) methods for codeword initialization, this does not matter. Unlike conventional use of FSCL, our objective is not to achieve rate distortion optimality, but rather, the aim is to cluster the HDR image pixels in such a way that the groupings not only reflect the pixel distribution of the original scene, the pixel populations are also evenly distributed to the codewords such that the image can be displayed in LDR devices with good visibility and contrast.

Because of the special way in which the initial values of the codewords are set, some codewords may scatter far away from densely populated intensity intervals where more codewords are needed to represent the pixels. In order to achieve this, we let the fairness function, f_i 's in Eq. (5) of the FSCL algorithm, accumulate the counts throughout the entire training process, thus increasing the chances to obtain more contrast.

4.2. Setting the training rate η

The training rate η in Eq. (5) is an important parameter in the training of the quantizer. It not only determines the convergence speed but also affects the quality of the final quantizer. In the neural network literature, there are extensive studies on this subject. In our work, we use the following rule to set the training rate as a function training iterations

$$\eta(k) = \eta(0) \exp(-\sigma k), \quad (9)$$

where $\eta(0)$ is the initial value set at the beginning of the training, σ is a parameter controlling how fast the training rate decays as training progresses.

In the experiments, we found that larger initial training rate values and slower decreasing speeds (smaller σ in Eq. (9)) led to higher contrast images when the algorithm achieved an equilibrium state. The result is expected. Larger initial values and slower decreasing speed of the training rate make η remain relatively large throughout. Once a codeword, which may be far away from densely populated intensity intervals, wins a competition, the larger training rate would drag it closer to the densely populated intensity intervals and thus increasing its chance to win again the next time round. The final result is that densely populated intensity intervals are represented by more codewords or more display levels and the displayed images will have more details. The effects for smaller training rates are almost the opposite, small training rate will result in

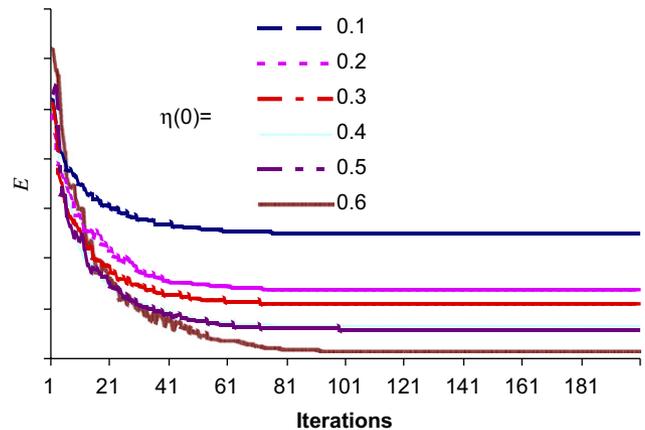


Fig. 7. Training curves for fixed $\sigma = 0.05$ and various $\eta(0)$ corresponding to those in Fig. 5.

the image having a narrower histogram (lower contrast) and more similar to the effect of linear compression. This is again expected. Because we initialize the codewords by linear scaling and small η will make the final codewords closer to their initial values when the training achieves an equilibrium. Fig. 5 shows examples of mapped images and their histograms with different $\eta(0)$ and fixed σ . Fig. 6 shows examples of mapped images and their histograms with different σ and fixed $\eta(0)$. By changing the training rate settings, the mapped image's histogram can be controlled to have narrower shapes (lower contrast images) or broader shapes (higher contrast images). This property of the training rate provides the users with a good guidance to control the final appearance of the mapped HDR images.

According to the above analysis, in order to control the final appearance of the mapped image, we can fix one parameter and change the other. To achieve the same contrast for the final mapping, we can either use a smaller $\eta(0)$ and a smaller σ , or, a larger $\eta(0)$ and a larger σ . However in the experiments we found that, with a smaller $\eta(0)$ and a smaller σ , the algorithm took longer to converge to an equilibrium state. Fig. 7 shows the training curves for the training rate settings corresponding to those in Figs. 5 and 8 shows the training curves for the training rate settings corresponding to those in Fig. 6. The image in Fig. 5 mapped with $\eta(0) = 0.5$ and $\sigma = 0.05$ and the image in Fig. 6 mapped with $\eta(0) = 0.2$ and $\sigma = 0.002$ have similar contrast. However, as can be seen from the training curves in Figs. 7 and 8, for $\eta(0) = 0.5$ and $\sigma = 0.05$, the training converged within 100 iterations and for $\eta(0) = 0.2$ and $\sigma = 0.002$ the training took more than 400 iterations to converge. Thus as a guide for the use of the training algorithm, we recommend that the control of the final image appearance be achieved by fixing a relatively aggressive σ and adjusting $\eta(0)$ because of the fast convergence speed under such conditions. Setting $\eta(0) = 0 \sim 1$ and $\sigma = 0.05$, and train the quantizer for 100 iterations worked very well for all images we have tried. After 50 iterations, E changed very little by further iterations, therefore for most images, less than 50 iterations suffice. Through experiments, we also found that it was only necessary to use

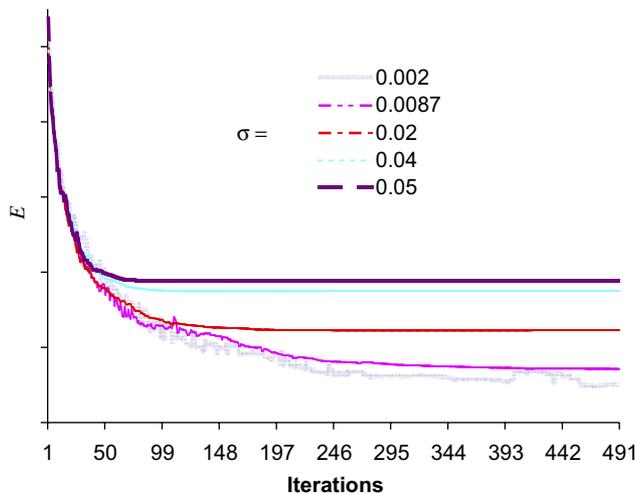


Fig. 8. Training curve for fixed $\eta(0) = 0.2$ and various σ corresponding to those in Fig. 6.

about a quarter of the pixels in an image to train the mapping quantizer.

4.3. Pixel clustering and mapping for display

Once the codebook has been designed and the codewords ordered in an increasing order, a pixel x is assigned a display value L_d , according to

$$L_d = \arg(\min_i \|x - c_i\|). \quad (10)$$

Although not a focus of this paper, it is important to mention how we render our quantized images for display. So far, we have taken an HDR image and quantized it into 256 levels. How should we use these quantized values to drive a CRT? This is a complicated issue, and in this paper, we take a simple (not necessarily the best) approach: All the pixels in a group indicated by index n is mapped to display intensity level n .

The rationale for this comes from what is known about lightness perception in our own vision. Using the CIE standards, it is known that if we are to display a uniform gray scale on a monitor then we choose uniform steps in ‘lightness’ and then display these ‘lightness’ values (by carrying out the correct calibration). Let us consider our quantized values to be quantized ‘lightness’ values (since we wish each level to be visually equal importance). In the CIE standard, lightness is defined to be the cube root of a linear luminance signal. It follows then, that to linearize our processed data we would have to raise it to the power of 3. However, we note that the transfer function of a typical monitor is often modeled as a power around 2.2–2.4. That is, the monitor transforms our perceptual measures into almost the correct linearized signal for display. Of course, if the target for images is a medium other than a CRT monitor then we may have to carry out a calibration and convert our processed images accordingly. Finally, the LDR image is

rendered as

$$R_d = \left(\frac{R_{in}}{L_{in}}\right)^\gamma L_d, \quad G_d = \left(\frac{G_{in}}{L_{in}}\right)^\gamma L_d, \quad B_d = \left(\frac{B_{in}}{L_{in}}\right)^\gamma L_d, \quad (11)$$

where L_{in} and L_d are luminance values before and after compression, γ controls display color (setting it between 0.4 and 0.6 worked well).

4.4. Implementation summary

One of the problems of neural network training is that the final results are greatly influenced by the (random) initial settings. In our application, the initial codewords are set in a deterministic way, i.e., linearly sampled from the full dynamic range of the images’ intensity. We therefore remove one of the stochastic elements of the training algorithm. We further show that setting the initial codeword values according to Eq. (8) not only removes one of the nondeterministic elements of the algorithm but also fits perfectly into the tone mapping requirements of our application. Through empirical investigations, we further developed principles for selecting the training rate parameters for faster convergence and for the use of the training rate to control the final appearances of the mapped images.

Therefore, our implementation procedure is deterministic. The algorithm starts from a linear scaling mapping and gradually adds contrast to the mapped image as training progresses. By setting the training rate parameters appropriately, we can control the final appearances of the mapped image. In fact, in all our experiments, we fixed $\sigma = 0.05$ and only varied $\eta(0)$. Therefore our algorithm only has one adjustable variables $\eta(0)$, which is used to control the final contrast of the mapped images.

5. Experimental results

We have applied our method to many HDR graphics and photographs. Similar to work of other authors, we only map the dynamic range of the luminance, and color correction on these images was not performed. Our method belongs to the global (spatially invariant) TRC category of technique [12]. We first compare our technique with one of the well-known TRC-based method of Ward and co-workers [5]. We thank Greg Ward for making his software available for creating the results of his technique.

Fig. 9 shows the mapping curves and LDR images of our new method and those of histogram equalization and the method of Ref. [5] for the bathroom image. It is seen that the mapping curve of our method and that of Ref. [5] have a similar shape, but it is also seen that at the very bright part of the image, our mapping has less aggressive compression. This is reflected in the mapped images. It is seen that in areas surrounding the lamp, our mapping has retained more details than the method of Ref. [5] and histogram equalization. This can be explained by inspecting the mapping curves, where it is seen that our mapping curve has reserved more display levels for the bright part of the intensity intervals. The reason that the method of Ref. [5] compresses the dynamic range in this interval so aggressively

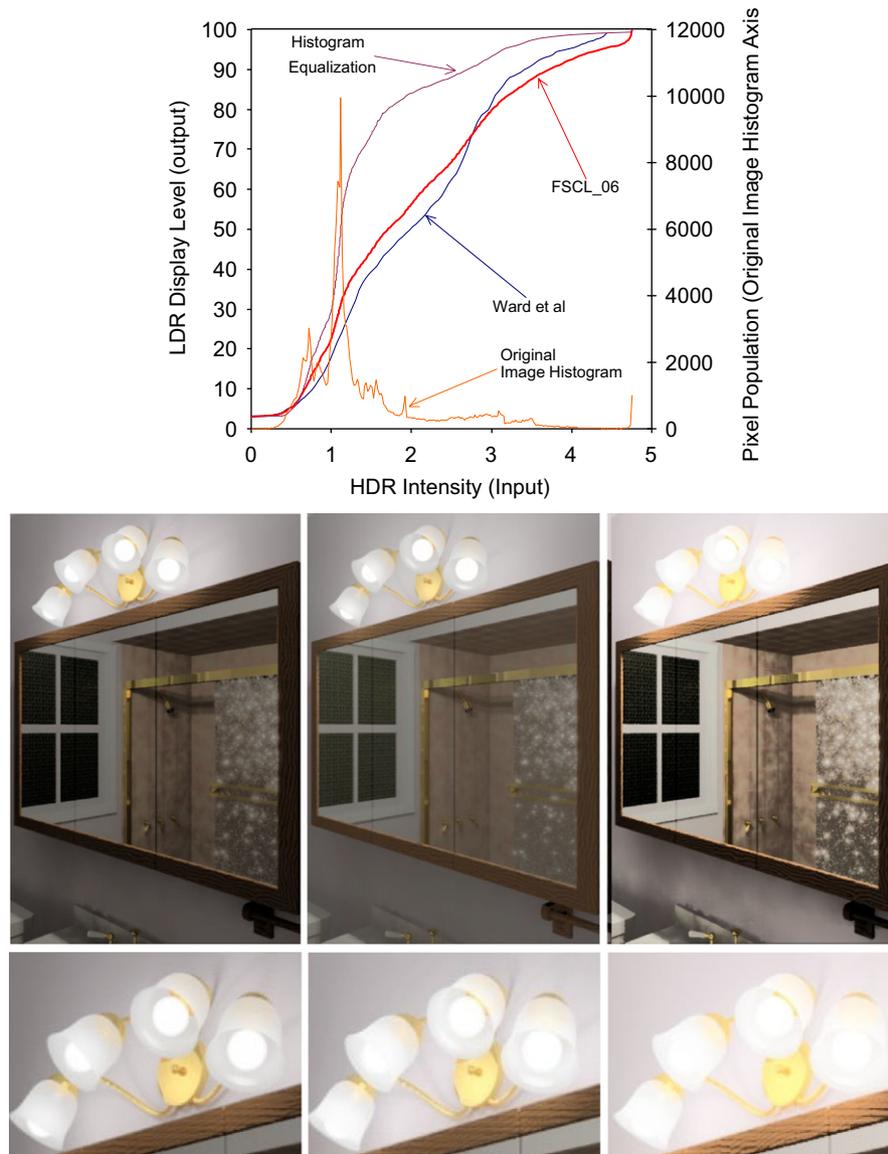


Fig. 9. Top row: original HDR image's histogram and display mapping curves for the bathroom image, the HDR input axis is in logarithmic scale and the LDR display level output axis is normalized to 100. Compared with Ward's method, our compression was less aggressive in the bright intensity intervals, this is reflected in the mapped LDR image where our display has retained more details than both Wards method and histogram equalization. Middle row: left to right, images mapped by the new method for $\eta(0) = 0.6$, $\sigma = 0.05$, iterations = 80, Ward's method and histogram equalization. Bottom row: amplified regions of the images above them. Radiance map courtesy of Greg Ward. The dynamic range of the image is 56,731:1.

is because, the method is based on histogram equalization (by accumulating pixel population) to produce its mapping curve, however, pixel population in this interval is sparse, the display produced in this region by the mapping curve will have low contrast and in such case the algorithm of Ref. [5] does nothing but approximately follows histogram equalization mapping. What a good mapping should do is to be able to adjust the number of display levels assigned to such sparsely populated intervals and it is seen that our method can do just that.

Fig. 10 shows the mapping curves and LDR images of our new method and those of histogram equalization and the method of Ref. [5] for the Memorial Church image. It is again observed that our mapping has been able to assign display

levels more sensibly. For example, in the areas around the windows, both displays of the method of Ref. [5] and histogram equalization have become very saturated, while our display has retained more details. This is because our mapping has assigned more display levels to these bright but sparsely populated intensity intervals whilst histogram equalization and the method of Ref. [5] have compressed this interval very aggressively because the pixel populations in these intensity intervals are very sparse.

Fig. 11 shows yet another example. Here we show that our method can achieve satisfactory results for different $\eta(0)$'s. It is again seen very clearly that our method was able to assign display levels more sensibly than the method of Ref. [5]. It

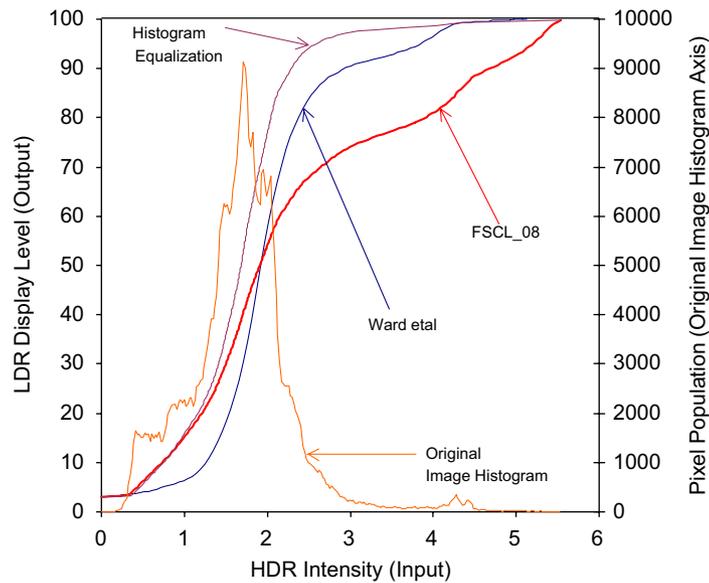


Fig. 10. Top row: original HDR image's histogram and display mapping curves for the Memorial Church image, the HDR input axis is in logarithmic scale and the LDR display level output axis is normalized to 100. Compared with Ward's method, our compression was less aggressive in the brighter intensity intervals where pixel population is sparse, this is reflected in the mapped LDR image where our display has retained more details than both Ward's method and histogram equalization. Middle row: left to right, images mapped by the new method for $\eta(0) = 0.8$, $\sigma = 0.05$, iterations = 80, Ward's method and histogram equalization. Bottom row: amplified regions of the images above them. Radiance map courtesy of Paul Debevec. The dynamic range of the image is 340,016:1.

is seen that because the method of Ref. [5] compresses the sparsely populated bright intensity intervals so aggressively, the areas surrounding the lamps have become completely saturated. In contrast, our method has been able to distribute display levels more appropriately and in our displays the lamps are clearly visible.

Figs. 12 and 13 show comparisons of mapped images of our method with other two methods in the literature. It is seen that our method is clearly better than the simple quantization method and comparable to photograph tone reproduction approach.

It turns out that the models can be pre-trained, that is, the mapping models trained using one image (or a group of images) can be used to mapped those images that are not part of the training data. Figs. 14 and 15 show results of mapped displays using pre-trained (using other images) models. In the neural network literature, this is known as the generalization capability [21]. With such a capability, it is possible to collect training images that are similar to the images we want to display to pre-train the mapping models (quantizers). Therefore, an image can be mapped directly using models pre-trained using similar images without necessarily having to train the mapping model

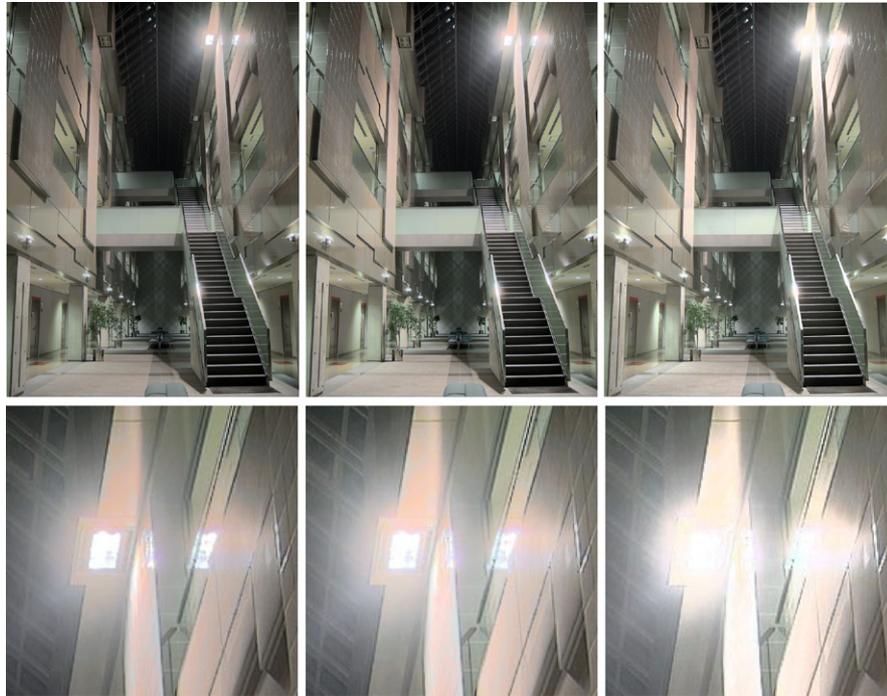
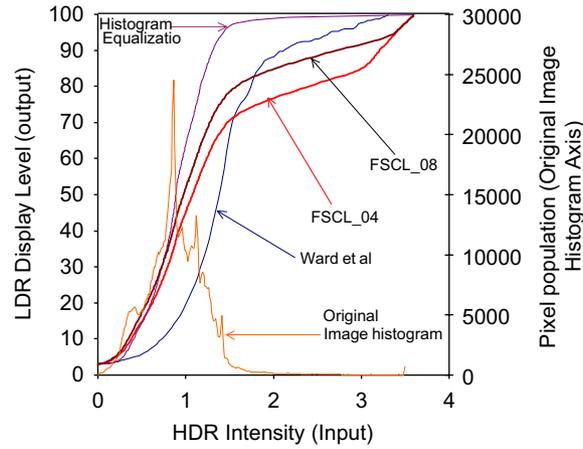


Fig. 11. Top row: original HDR image's histogram and display mapping curves for the Atrium image, the HDR input axis is in logarithmic scale and the LDR display level output axis is normalized to 100. Compared with Ward's method, our compression is less aggressive in the bright intensity intervals where pixel population is sparse, this is reflected in the mapped LDR images where our displays have retained more details than Ward's method. Middle row: left to right, images mapped by the new method for $\eta(0) = 0.4$, $\sigma = 0.05$, iterations = 80, the new method for $\eta(0) = 0.8$, $\sigma = 0.05$, iterations = 80 and Ward's method. Bottom row: amplified regions of the images above them. It is seen clearly that Ward's method has made the lamps and their surrounding areas completely saturated. Radiance map courtesy of Greg Ward. The dynamic range of the image is 3,900:1.



Fig. 12. Left: result of the new method ($\eta(0) = 0.3$, $\sigma = 0.05$, iterations = 80). Right: results of Schlick's rational sigmoid quantization technique [2]. Radiance map courtesy of Program of Computer Graphics, Cornell University.



Fig. 13. Left: result of the new method ($\eta(0) = 0.4$, $\sigma = 0.05$, iterations = 80). Right: result of [9], image courtesy of Erik Reinhard. Radiance map courtesy of Paul Debevec.

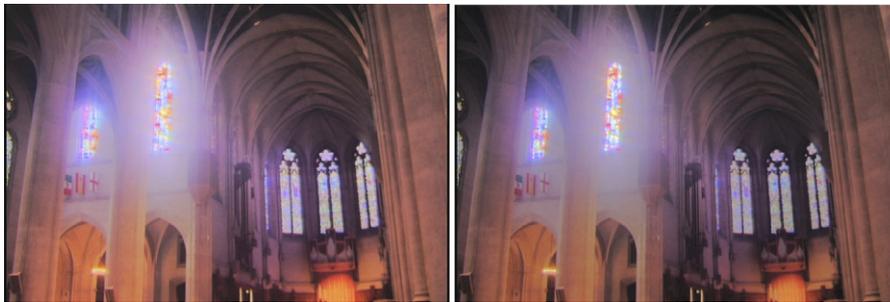


Fig. 14. Left: result of the new method, mapping model trained on its own data ($\eta(0) = 0.5$, $\sigma = 0.05$, iterations = 80). Right: result of the new method, mapping model trained on data from the image in Fig. 13 ($\eta(0) = 0.5$, $\sigma = 0.05$, iterations = 80). Radiance map courtesy of Paul Debevec.



Fig. 15. Result of the new method. The left image's mapping model was trained on the HDR data from the right image ($\eta(0) = 0.8$, $\sigma = 0.05$, iterations = 100). The right image's mapping model was trained on the HDR data from the left image ($\eta(0) = 0.5$, $\sigma = 0.05$, iterations = 100). Data courtesy of Greg Ward.

online, thus improving the mapping speed. Fig. 16 shows more examples of our results.

If an image is to be mapped using its own model and the quantizer is to be trained online, the training of the quantizer will take up most of the computing time. Once the model is trained, mapping is very fast. Our programs were written in C, which had not been optimized for speed. For a 512×768 image, using a quarter of the pixels and training for 80 iterations, it took a Pentium 3 GHz processor PC 18.6 s to train the

mapping quantizer. For most images, fewer than 50 iterations suffice. Once the quantizer has been designed, it took 0.67 s to map a 512×768 image. This training speed can be improved by introducing fast training algorithms and optimized training parameters.

As in many other tone mapping works in the literature, our evaluations of results are mainly based on visual assessment of displays on LDR devices. Due to the subjective nature of the problem, a quantitative measure of tone mapping quality is



Fig. 16. More examples mapped by the new method (using own models). Radiance maps courtesy of Paul Debevec. The parameters used for the above images are: top: $\eta(0) = 0.4$, $\sigma = 0.05$, iterations = 80; Bottom: $\eta(0) = 0.8$, $\sigma = 0.05$, iterations = 80.

still very difficult to obtain. With the advent of HDR display systems [22], tone mapping operators for HDR images can be evaluated against HDR displays in the future.

6. Conclusions and future work

In this paper, we have presented a new method, which learns to map high dynamic range image data to be displayed in low dynamic range devices. The frequency sensitive competitive learning (FSCL) algorithm, which ensures that all the available levels in the low dynamic range display are fully utilized, has been shown to give very good performances. We have presented a very detailed procedure that removes the stochastic elements of neural network training and turns the implementation into a deterministic procedure where there is only one variable which has also been used to control the appearances of the final display. Our contributions in this paper are the introduction of a novel learning-based method and its detailed, deterministic and practicable implementation procedure for the display of high dynamic range image.

Our current method was strongly motivated from a computational perspective. Indeed, the contribution of this work should be viewed as the introduction of a novel learning-based computational method for the display of high dynamic range image.

Our future work would investigate issues such as (i) ways to incorporate perceptual aspects into the algorithm, for example, taking inspirations from the works of Refs. [5,16,17], (ii) better methods to render the output of the quantizer to produce better display and (iii) fast training and rendering algorithms.

Acknowledgments

We would like to sincerely thank following researchers for their help. Greg Ward has kindly provided us his radiance software package. Fredo Durand, Raanan Fattal, Sumanta Pattanaik, Erik Reinhard, Jack Tumblin, and Greg Ward have patiently explained many technical questions. We also thank various authors for making their data available on the Internet for experiment.

References

- [1] K. Chiu, et al., Spatially nonuniform scaling functions for high contrast images, in: Proceedings of Graphics Interface'93, pp. 245–53.
- [2] C. Schlick, Quantization techniques for the visualization of high dynamic range pictures, in: Proceedings of Fifth Eurographics Workshop on Rendering, June 1994.
- [3] P.E. Debevec, J. Malik, Recovering high dynamic range radiance maps from photographs, in: Proceedings of SIGGRAPH'97, pp. 369–378.
- [4] T. Mitsunaga, S.K. Nayar, High dynamic range imaging: spatially varying pixel exposures, Proceedings of CVPR'2000, vol. 1, 2000, pp. 472–479.
- [5] G. Ward, H. Rushmeier, C. Piatko, A visibility matching tone reproduction operator for high dynamic range scenes, IEEE Trans. Visualization Comput. Graphics 3 (1997) 291–306.
- [6] J. Tumblin, H.E. Rushmeier, Tone reproduction for realistic images, IEEE Comput. Graphics Appl. 13 (1993) 42–48.
- [7] J. Tumblin, J.K. Hodgins, B.K. Guenter, Two methods for display of high contrast images, ACM Trans. Graphics 18 (1999) 56–94.
- [8] R. Fattal, D. Lischinski, M. Werman, Gradient domain high dynamic range compression, in: Proceedings of SIGGRAPH 2002, 2002, pp. 249–256.
- [9] E. Reinhard, M. Stark, P. Shirley, J. Ferwerda, Photographic tone reproduction for digital images, in: Proceedings of SIGGRAPH 2002, 2002, pp. 267–276.
- [10] F. Durand, J. Dorsey, Fast bilateral filtering for the display of high dynamic range image, in: Proceedings of SIGGRAPH 2002, 2002, pp. 257–265.
- [11] M. Ashikhmin, A tone mapping algorithm for high contrast images, in: Rendering Techniques '02 (Proceedings of the 13th Eurographics Workshop on Rendering), June 26–28, 2002, pp. 145–156.
- [12] J. DiCarlo, B. Wandell, Rendering high dynamic range images, Proc. SPIE 3965 (2001) 392–401.
- [13] S.B. Kang, M. Uyttendale, S. Winder, R. Szeliski, High dynamic range video, ACM Trans. Graphics 22 (3) (2003) 319–325.
- [14] R. Mantiuk, G. Krawczyk, K. Myszkowski, H-P Seidel, Perception-motivated high dynamic range video encoding, Proceedings of SIGGRAPH 2004, 2004, pp. 733–741.
- [15] G. Ward, A contrast-based scalefactor for luminance display, in: Graphics Gems IV, Academic Press, New York, 1994, pp. 415–421.
- [16] S.N. Pattanaik, J.A. Ferwerda, M.D. Fairchild, D.P. Greenberg, A multiscale model of adaptation and spatial vision for realistic image display, Proceedings of SIGGRAPH'98, Orlando, July 1998, pp. 287–298.
- [17] J.A. Ferwerda, S.N. Pattanaik, P. Shirley, D.P. Greenberg, A model of visual adaptation for realistic image synthesis, Proceedings of SIGGRAPH'96, pp. 249–258.

- [18] A. Gersho, R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.
- [19] T. Kohonen, *Self-organization and Associative Memory*, Springer, Berlin, 1989.
- [20] S.C. Ahalt, et al., Competitive learning algorithms for vector quantization, *Neural Networks* 3 (1990) 277–290.
- [21] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall International, Inc., Englewood Cliffs, 1999.
- [22] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, A. Vorozcovs, High dynamic range display systems, *ACM Trans. Graphics (SIGGRAPH 2004)* 23 (3) (2004) 760–768.

About the Author—GUOPING QIU is an Associate Professor and Reader in Visual Information Processing at the University of Nottingham, UK. He received his B.Sc. in Electronic Measurement and Instrumentation from the University of Electronic Science and Technology of China, in 1984 and his Ph.D. in Electronic Engineering from the University of Central Lancashire, UK, in 1993. He has been performing research in the fields related to image processing for more than 15 years and has authored more than 90 publications. His research interests include high dynamic range imaging, learning-based image processing, content-based image indexing and retrieval, and interactive computational visual information processing using optimization. More of his research can be found at <http://www.viplab.cs.nott.ac.uk/>.

About the Author—JIANG DUAN received the B.E. degree in Mechanical Engineering from Southwest Jiaotong University, China, in 2001, the M.Sc. degree in Colour Imaging Science from the Colour & Imaging Institute, the University of Derby, UK, in 2002. From 2002 to present, he is a Ph.D. research student in the School of Computer Science and Information Technology at the University of Nottingham, UK. Between May 2005 and September 2005, he worked at Xerox Research Centre Europe (XRCE), Grenoble, France. At XRCE he received the First Prize for his Internship Project. His research interests include high dynamic range image tone mapping and display.

About the Author—GRAHAM D. FINLAYSON received the B.Sc. degree in Computer Science (with honors) from the University of Strathclyde, Glasgow, Scotland, in 1989. He received the M.Sc. and Ph.D. degrees from Simon Fraser University, Vancouver, Canada, respectively, in 1992 and 1995. His Ph.D. dissertation was awarded a Dean's medal for academic excellence. From 1995 to 1997, he was a lecturer in the Department of Computer Science at the University of York in the United Kingdom. In 1997, he was appointed to a readership at the University of Derby, where he was a founding member of the Colour & Imaging Institute. In 1999, he left Derby to become a full professor in the School of Computing Sciences at the University of East Anglia. Professor Finlayson is interested in how color can be used to solve problems in computer vision and allied disciplines such as image processing and digital photography. He has made many contributions toward solving the color constancy problem (removing color bias due to illumination from images) and his algorithms have been implemented in commercial cameras. His current research interests include dynamic range compression, the automated removal of shadows in images, photometric invariance, and the application of computational techniques to the understanding of human vision. He is an Associate Editor of *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, an Associate Editor of the *Journal of Electronic Imaging* and was a Guest Editor of the *Journal of Computer Vision and Image Understanding (CVIU)*.