Tone Correction for Pictorial Enhancement

写真画質向上のためのトーン補正

Richard Campbell*

Abstract

The hardcopy industry seeks to reproduce preferred representations of pictorial content. It is often difficult to reproduce the high dynamic range of real world scenes with the relatively low dynamic range of modern display and printing technology. This can lead to poor reproductions due to various limitations in sensing and rendering technology. For uncontrolled environments the automated capture settings (aperture and exposure time) can result in problems such as: under, over, flat, and bi-modal exposures. However, it is possible to design image processing algorithms, which can partially mitigate the effects of these limitations. In this paper, we define the tone correction problem and discuss one solution to produce preferred representations of pictorial content.

ハードコピー産業は，写真画像に対する好ましい表現を再現しようと努めている。現代のディスプレイや印刷技術の比較的低いダイナミック・レンジで，実際のシーンの高いダイナミック・レンジを再現することは多くの場合困難である。これはセンサーやレンダリングの技術における様々な制限のために，不十分な再現を引き起こし得る。制御されていない環境においては，自動キャプチャ設定（レンズの口径および露光時間）が，露光不足，露光過多，低コントラスト，バイ・モーダル，といった問題に帰着し得る。しかしながら，これらの制限の影響を部分的に緩和することができる画像処理アルゴリズムを設計することは可能である。本論文では，トーン補正の問題を定義し，写真画像の好ましい表現を再現するための１つの解決策を述べる。

Introduction

The prevalence of pictorial content in the hardcopy industry has increased with the popularity of consumer digital cameras and the decreased cost of high quality color marking engines. Advances in photographic printers have made high quality photo processing available on the desktop. The effects of this revolution also impact the more traditional hardcopy markets of print and copy.

There exist many levels of realism for any reproduction. For pictorial content, it is often desired to have photo-realism, where the data is processed to produce a reproduction that has the same visual response as the real world scene. This requires knowledge of the human visual system to balance the limitations of the technology used to capture and re-render the scene.

1. Tone Reproduction

The human visual system is able to observe a high dynamic range of luminance information. It is capable of a ten orders of magnitude, candela per meter square, difference between ambient light of star-lit and sun-lit snow scenes, and more than four orders of magnitude difference between shadows and highlights in a typical scene [1] [2]. Capturing and reproducing large dynamic ranges is a difficult problem.

It is often impractical to design technology to capture such a high dynamic range of luminance information. Often it is only reasonable to obtain capture devices whose A/D conversion is 10 to 12 bits, which corresponds to three orders of magnitude. This data is further processed and truncated to 8 bits, or two orders of magnitude, to match the reproducible dynamic range for most displays and

* Sharp Laboratories of America, Inc.

hardcopy.

Such mismatches exemplify the tone reproduction problem, whose aim is to replicate the same visual sensation of a high dynamic range scene with a low dynamic range reproduction.

The process of tone reproduction should be guided by the properties of the human vision system, so that a person viewing the reproduction will feel that it looks natural and conforms to their experience of the real world. However, the human vision system is a complex system that adapts to each scene [1], whose properties are still being discovered.

1•1 Photography

Photographic processing has long studied the process of capturing and reproducing the high dynamic range of real world scenes. This has led to advances in the properties of film and paper, camera design, and film processing. Each step refines the photographic process and the ability to represent a diverse world.

Digital photography contains all the same problems, but utilizes different technology to solve them. Parulski and Spaulding give a good overview of the digital camera color workflow [3] and highlight some of the issues.

One of these issues is subjective preference. Much is known about the human vision system, but what is less well understood is how observers perceive color in complicated scenes and how preference influences the assessment of subjective quality. Often colorimetric accuracy is stressed rather than the preferred reproduction. For pictorial regions, the content of the scene will influence the importance of the tone reproduction for various areas. Given two different tone renderings of the scene, the observer may prefer one over the other based upon the detail and balance of the rendition, rather than accuracy alone.

The issue of subjective preference is well known in both hardcopy and photographic processing industry. Often color saturation is boosted for business documents and photographs to make the print more vivid.

1•2 Tone Reproduction Defects

Besides the representational problems of mapping high dynamic range to low, the systems used to map the process can themselves fail. Some examples of these failures are; under, over, flat and bi-modal exposures. These capture defects are common to both film and digital photography. Figure 4 shows two digital camera images with poor tone reproduction results.

Under exposure exists when insufficient light is captured by the camera. This failure generally occurs when the exposure time is set too short. This is a result of an error in the exposure calculation or because a longer exposure would produce visible motion blur. The exposure calculation is often used to balance the amount of light with the degree of spatial blurring.

Over exposure occurs when too much light is allowed to enter the camera. This often leads to sensor saturation and clipping of the luminance information. The human eye has the ability to utilize the response compressed data in saturation, by adjusting the gain in the early vision system. This adaptation helps the eye's ability to respond to a high dynamic range of luminance levels [1]. Current CCD and CMOS technology can have problems utilizing the information at or near full well for the sensor due to the non-linearity of the device. The camera's exposure calculation is typically designed so that the device avoids saturation. It has been our experience that these defects are rare for digital photography.

Flat exposure exists when the camera fails to adequately adjust the dynamic range of the capture to the scene. One example occurs when the camera is adjusting to sunny conditions in an outdoor scene. Due to the amount of light available, the camera uses a short exposure time to avoid sensor saturation. The short exposure time prevents the sensor from capturing enough light to resolve detail from the shadows. The result is a tone reproduction of the scene with a loss of contrast.

Bi-modal exposure exists when two or more lighting conditions occur in a scene. A common example of this case is observed when taking an indoor photograph near a window on a sunny day. Depending on the exposure calculation the interior or exterior portions of the image may be correctly reproduced, but often not both. Bi-modal exposures are particularly difficult because the scene often contains a higher dynamic range than single illuminated scenes. Detail consists of luminance changes that carry information like textures on objects or transitions between regions.

1•3 Tone Correction

The failure of a sensor's tone reproduction highlights the difference between tone reproduction and tone correction. Tone reproduction deals with the mapping of the high dynamic range data into the low dynamic range data of

modern displays or prints. This typically involves a calibration process. The complexities of the tone reproduction problem in the photographic process often lead to non-ideal mappings that may benefit from tone correction. Tone correction is applied after data has already been sent through a tone reproduction process. In many cases, the tone correction process starts with low dynamic range data and remaps it to a more accurate or preferred rendering. The unconstrained tone correction problem starts from an unknown source and remaps the result to be a less objectionable and possibly more accurate reproduction of the real world.

1•4　Tone Operators

There exist many tone reproduction operators in the literature. Some are used in closed environments, where light and scene properties are well known, while others are used in less constrained environments. In this paper, we will review only a few of these operators. For a comprehensive overview of tone operators, please refer to Devlin [4].

Curves have been used for tone reproduction in the print and display industries to compensate for nonlinearities. In both cases, curves are used to adjust the input signal to the rendering device, so that the luminance component is correctly reproduced and the output appears as intended to the human observers. Tone reproduction curves are typically formed using a calibration process that measures the output of the device for controlled input sources.

Curves have also been used for correcting tone defects for digital photography [5]. This manual process manipulates the curve shape until a pleasing reproduction of the scene tone is achieved.

An interesting set of experiments [6] run by Braun and Fairchild studied the image dependency of tone reproduction for simulated reduced dynamic range devices. The experiments used trained observers to manipulate the parameterization of the tone reproduction curve in matching experiments between an original tone reproduction and a simulated reduced dynamic range reproduction. The goal was to adjust the tone reproduction curve to best match the reduced dynamic range image to the original. Their results included a simple method for adjusting the tone reproduction based on image and intended display dynamic range.

Luminance histograms have also been used to study and correct tone reproduction problems [7]. Histogram equalization and histogram shaping in general have been used as reproduction operators. Both operators make assumptions on the desired tone reproduction. Histogram equalization, for example, assumes that all tone values are equally likely, a poor hypothesis for most scenes.

Homomorphic filtering [7] takes advantage of a general property of most scenes. The recorded luminance values can be separated into the product of two components: an illumination component and a reflection component. The illumination component contains the large variations of scene light sources, and is assumed to change less frequently. The reflection component contains most of the detail. To preserve or enhance detail in a scene, the high frequency components associated with the reflection values are maintained or boosted, while the lower frequency components are suppressed. This process tends to enhance the perceived detail in shadows and highlights. Setting the degrees of attenuation or enhancement for the various frequency components is an area of active study.

A more recent method [2] is based upon the conceptual framework of Ansel Adam's photographic Zone System. The Zone System measures the mid-gray, highlight and shadow luminance in the scene and predicts how each of these will map to print. These measurements, along with experience, guide the reproduction. The new method takes a user-specified mid-gray value for each image and produces a new tone reproduction for the image. The operator has both a global and spatially varying luminance mapping. The spatially varying component operates similarly to photographic dodging and burning to bring out local detail.

2.　Neural Network Tone Correction

It is common for pictorial data to have already passed through a tone reproduction process before hard copy. Sensor, time and computational limitations, along with other restrictions, all contribute to the tone defects that can occur in digitally captured images. Correcting these defects may result in noticeable image quality improvements when the images are displayed. Some of the techniques used for tone reproduction may not be suitable for implementation in standard hardcopy workflows. Time, memory, and computational power constraints all affect the cost and market of hardcopy devices. Our proposed solution provides dramatic benefit, while remaining practical for embedded implementation in capture or rendering devices.
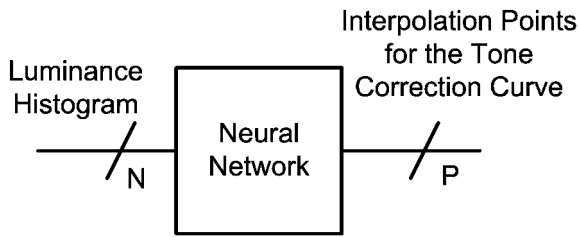
Fig. 1   Neural network tone correction.

Our tone correction technology accepts a low dynamic range reproduction from the capture device and remaps it to a more preferable low dynamic range reproduction. The input reproduction has already been mapped from the high dynamic range of the real world to the low dynamic range of the reproduction. This loss of precision implies that it is not always possible or practical to obtain the optimal tone reproduction by correcting the capture device reproduction. Instead, our technology seeks to reproduce the preferred reproduction.

Our algorithm utilizes training data that includes subjective reasoning that balances the benefits of tone correction with its undesired side effects, such as increased noise and/or contouring. The subjective reasoning also balances the benefit based upon scene content. The strength of the correction is dependent on the scene content. One example of this is the main subject. When the main subject is under exposed, the preferred strength of the correction is often stronger than other similarly under exposed areas, even at the expense of increasing noise. The preferred correction places more emphasis on recovering the main subject's tone relative to the scene and improving contrast within the subject.

Our technology utilizes a training set of preferred corrections to learn the mapping between an image statistic and the correction. Most of the computationally intensive processing is done off-line during training. The resulting tone operator requires a simple image statistic and a minimal amount of calculation to produce a correction lookup table. This allows the system to perform well in embedded environments where computational power and time constraints are critical.

2•1   Neural Network Tone Operator

Our proposed solution uses a neural network to estimate the tone correction, by controlling the shape of a tone correction curve (**Fig. 1**). The curve defines a luminance transformation that can be applied using a lookup table.

During training, the preferred correction is limited to a tone correction curve. This approach constrains the problem and minimizes the computational impact. The tone correction curve maps the input luminance values to a new output luminance values. To ensure a smooth natural-looking transformation, the curve will form a monotonically increasing function in luminance. The monotonicity of the curve preserves the order of the luminance relationships between input and output values.

The luminance transformation is controlled by a function approximation neural network. This family of neural networks can approximate complex functional relationships. For this application, the neural network influences the shape of the tone correction curve by controlling the location of a set of interpolation points as shown in **Fig. 2**. A spline parameterization defines the tone correction curve and resulting correction lookup table. The parameterization can be controlled to enforce the monotonicity of the curve, resulting in an increasing function throughout the input luminance range.

The network is driven by the information contained in the luminance histogram of the pictorial region. The histogram is a measure of the number of pixels in a pictorial region whose luminance values lie within a range of values. The number of pixels within each tone range is a strong indicator of the type of exposure defect. Figure 3 is the histogram of an image with a bi-modal exposure defect. The histogram shows that the image contains a high distribution of dark tones and light tones with less in the middle.

The luminance histogram contains no information about spatial context or scene content. This missing information
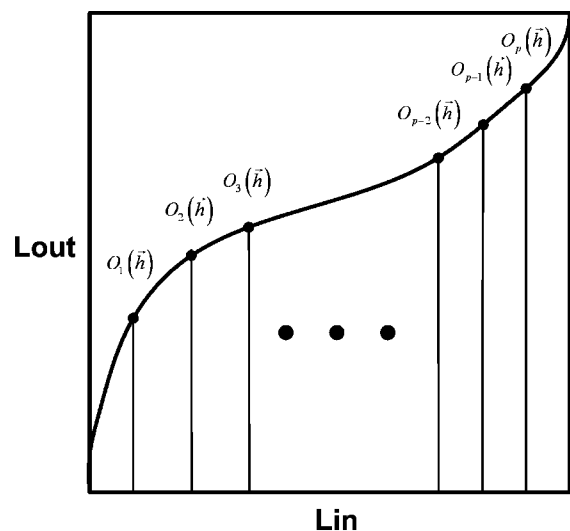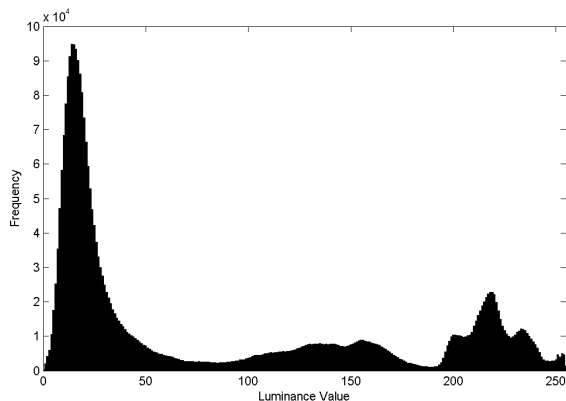

Fig. 2   Tone correction curve.

Fig. 3　Luminance histogram for image in Fig. 4 c).

can be important in the perception of the scene and thus should influence the correction. This implies that the luminance histogram is not a sufficient statistic to measure all the complexities necessary to reproduce the ideal preferred correction. Even with its limitations, the luminance histogram is a strong indicator of the type and magnitude of a correction that improves most pictorial content.

2•2　Learning

During training, the goal is to have the network generalize the relationship between luminance histogram distribution and correction curve shape. The mapping should improve the tone distribution of pictorial regions, while minimizing the introduction of objectionable artifacts. This implies that we desire a network with conservative corrections, unless the gathered information suggests that a strong correction is appropriate. This can be accomplished by training a network to have smooth performance. The performance of these networks does not vary greatly with only slight deviations in the input values. Large deviations can occur when a network has been over trained.

Several learning techniques have been proposed to improve the ability to generalize solutions and prevent over fitting. One such learning technique is Bayesian Regularization developed by David MacKay [8]. Introducing regularization into the training process penalizes the parameters for being too complex. The difficulty with using regularization is choosing the performance ratio that weights the desire to minimize the error versus penalizing the solution for being complex. MacKay introduced a framework that optimized the selection of the performance ratio using a Bayesian

inference technique. Another difficulty with any training procedure is determining the stopping criteria. Bayesian regularization also predicts the number of network parameters that are being effectively used in the generation of the tone correction. It is often better to use convergence of the number of effective parameters as a stopping condition, rather than sum of squared error (a measure of performance) or sum of squared weights (a measure of complexity). For further details see [8].

2•3　Method

The testing and training procedures have been designed to measure both performance and generalization. The procedures use a set of pictorial images and their corresponding tone corrections. This complete set was divided into a test set and a training set. The training set contained 3 times as many examples as the test set. The sets are chosen randomly, and this process can be repeated, to generate multiple train and test sets. Each train and test set pair was used to train and test the network and the results were compared. Correspondence between the network performances for the random sets serves as an indicator of good network architecture, a sufficient set of tone correction examples, and a training procedure that is able to generalize the solution.

The pictorial samples were derived from multiple sources, so that the results were not biased towards a specific exposure calculation. The samples include a distribution of tone defects with various degrees of correction required. These samples were hand-corrected by trained operators to construct a tone correction curve for each sample. A tool was provided to aid in the visualization, correction, and collection of the data.

2•4　Results

Several different network structures were tested. These networks varied in the number of units, the types of transfer functions, and network architecture. The final configuration was chosen based upon the network testing error and its complexity. For most problems, if a network is overly complex, its training performance improves but its testing error increases, indicating poor generalization and over-training. Our experiments showed that the networks with more units performed better in some cases but not significantly. To minimize calculation and to aid in identifying a network with smooth performance, the simplest network was chosen whose testing error was close

to the best performer. In our case, we found that a network with a small number of units was sufficient to obtain good tone correction results. Since computational complexity was one of our design criteria we chose the simpler solution.

Figure 4 shows two examples of neural network tone correction on digital camera images. **Fig. 4** a) is a bi-modal exposure taken late in the day with strong shadows, while **Fig. 4** c) shows another bi-modal exposure taken under a tree canopy overlooking the ocean. Both benefit from redistributing the tone from the dark and light-tones into the mid-tones. Most dramatic improvement is visible in the shadow textures (**Fig. 4** b, d).

Conclusion

Tone reproduction is a difficult problem and an area of future study. Pictorial content is particularly difficult and often contains defects in the original tone reproduction attributable to the capture device. We proposed a tone correction technique to improve overall tone reproduction for display or print. The tone correction outlined in this paper produces the preferred tone reproduction of pictorial content. The process is suitable for embedded implementation and provides significant benefits, while

minimizing the introduction of objectionable artifacts.

References

[1]   J. A. Ferwerda, "Elements of Early Vision for Computer Graphics," IEEE Computer Graphics and Applications, vol. 21, pp. 22-33, 2001.

[2]   E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic Tone Reproduction for Digital Images," presented at SIGGRAPH, 2002.

[3]   K. Parulski and K. Spaulding, "Color image processing for digital cameras," in Digital Color Imaging Handbook, G. Sharma, Ed. New York: CRC, 2003.

[4]   K. Devlin, "A review of tone reproduction techniques," Department of Computer Science, University of Bristol, Technical Report November 2002.

[5]   K. Eismann, Photoshop Restoration & Retouching, Second ed. Boston: New Riders, 2004.

[6]   G. J. Braun and M. D. Fairchild, "Image lightness rescaling using sigmoidal contrast enhancement functions," Journal of Electronic Imaging, vol. 8, pp. 380-393, 1999.

[7]   R. C. Gonzalez and R. E. Woods, Digital Image Processing: Addison-Wesley Publishing Company, Inc., 1992.

[8]   D. MacKay, Information Theory, Inference and Learning Algorithms: Cambridge University Press, 2003.

a) Image with poor tone reproduction

b) Neural network tone correction result

c) Image with poor tone reproduction

d) Neural network tone correction result

Fig. 4   Tone correction examples.