Exposure Fusion from a single image

Charles Hessel[1,2] Jean-Michel Morel[1]

[1] CMLA, ENS Cachan, CNRS, Université Paris-Saclay, 94235 Cachan, France
[2] DxO, 3 Rue nationale, 92100 Boulogne-Billancourt

{charles.hessel, jean-michel.morel}@cmla.ens-cachan.fr

Abstract

In this communication we present briefly an adaptation of the Exposure Fusion method to a Retinex usage. Exposure Fusion is a high dynamic range imaging technique to fuse a bracketed exposure sequence into a high quality image. Our proposition takes advantage of this method in the more general context of improving the overall quality of any image, turning Exposure Fusion into a new and simple contrast and color enhancement operator.

Keywords

Exposure Fusion, Tone-Mapping, Retinex

1 Introduction

The dynamic range of real scenes is generally higher than the one of our camera sensors. To capture the entire dynamic range, photographers can take a sequence of image with different exposure times: long times capture information in dark parts of the scene and saturate the brights ones, while a short exposure time captures relevant information in the brights parts. This is called a bracketed exposure sequence. This sequence is then merged into a high dynamic range (HDR) image, which needs to be remapped to the low dynamic range (LDR) of most displays through a tone-mapping operator.

Exposure Fusion [1] was introduced by T. Mertens, J. Kautz and F. Van Reeth in 2009 as an alternative way of constructing a LDR image of a bracketed exposure sequence. This method does not build an intermediate HDR picture. In a nutshell, it directly selects for each pixel the values, among the provided pictures, which should be kept in the final image. As a result, the fused image combines the best areas of the several input images. Although similar techniques already existed [4], this technique brought interesting and successful answers to two crucial questions: how to detect the best pixel from the provided set of images, and how to seamlessly merge those pixels in the final image.

Our proposition simulates the bracketed exposure sequence acquisition from a single LDR image, extending Exposure Fusion to color and contrast enhancement methods. We will first review the original algorithm and then explain our adaptation. The last part will show results and compare it to the state-of-the-art Multiscale Retinex.

2 Exposure Fusion

Exposure fusion first measures the perceptual quality of each pixel in each image of the input sequence. Three pixel-wise metrics are used: the contrast C, saturation S and well-exposedness E. We will denote in the following by ij the position of the pixel in a image, by c the color channel, and by k the position of the image in the input sequence. The *contrast metric* uses the absolute value of a discrete Laplacian filter applied to the grayscale version of the image. Denoting by $K_{\text{Laplacian}}$ a Laplacian kernel, we set

$$C_{ij,k} = \left|\Big(\frac{1}{3}\sum_{c=1}^{3} I_{ij,c,k}\Big) * K_{\text{Laplacian}}\right| . \quad (1)$$

The authors use for $K_{\text{Laplacian}}$ the sum of differences over the four nearest neighbors. The *saturation metric* is the standard-deviation of the pixel's color,

$$S_{ij,k} = \sqrt{\frac{1}{3}\sum_{c'=1}^{3}(I_{ij,c',k} - \frac{1}{3}\sum_{c=1}^{3} I_{ij,c,k})^2} . \quad (2)$$

Finally, the *well-exposedness* metric measure how close the pixel's value is to the median value 0.5 using a Gauss curve:

$$E_{ij,k} = \prod_{c=1}^{3} \exp -\frac{(I_{ij,c,k} - 0.5)^2}{2\sigma^2} , \quad (3)$$

with $\sigma = 0.2$. To account for multiple color channels, this measure is made on each channel separately and the results are multiplied.

The quality measure of each pixel is finally obtained as a product of these three metrics. By using the product, the authors enforce their method to only keep pixels which are acceptable for the three qualities simultaneously. To allow the user to choose the importance given to each quality measure, they added a power function to each one, with parameters ω_c, ω_s and ω_e (by default equal to 1):

$$W_{ij,k} = (C_{ij,k})^{\omega_c}.(S_{ij,k})^{\omega_s}.(E_{ij,k})^{\omega_e} . \quad (4)$$

For the blending process, the resulting weights need to be normalized as

$$\widehat{W}_{ij,k} = \Big(\sum_{k'=1}^{N} W_{ij,k'}\Big)^{-1}.W_{ij,k}\ . \tag{5}$$

At this point, each input image has its normalized weight map. As the authors explain, one could directly use them to fuse the images. But such an operation would lead to strong seams due to the sharp variations in the weights. They instead propose a multiscale fusion, using the method introduced by Ogden *et al.* [3]. This technique builds the Laplacian Pyramid [2] of the output image by blending the Laplacian pyramids of the input images according to the Gaussian pyramid of the weight maps. The fused image is obtained by collapsing the constructed pyramid. We will denote $L\{I\}$ the Laplacian pyramid of the input image I, $G\{W\}$ the Gaussian pyramid of the weights, and l the scale. The blending operation is then:

$$L\{R\}^l_{ij} = \sum_{k=1}^{N} G\{\widehat{W}\}^l_{ij,k}.L\{I\}^l_{ij,k}\ . \tag{6}$$

The algorithm 1 describes the whole process, from the quality measurements to the multiscale fusion.

Algorithm 1 Exposure Fusion

Require: input sequence of images I; weights for saturation, contrast and well-exposedness measures $\omega_s, \omega_c, \omega_e$
Ensure: fused image R
 for each image at position $k \in \{1, 2, ..., N\}$ in the input sequence **do**
  Compute contrast metric C using eq. (1)
  Compute saturation metric S using eq. (2)
  Compute well-exposedness metric E using eq. (3)
  Compute weight map W_k of the current image using eq. (4)
 end for
 Normalize weights using eq. (5)
 for each image at position $k \in \{1, 2, ..., N\}$ in the input sequence **do**
  Compute Gaussian pyramid of weights $G\{\widehat{W}\}_k$
  Compute Laplacian pyramid of input images $L\{I\}_k$
  for each coefficient at position ij and scale l **do**
   Update Laplacian pyramid of the output image: $L\{R\}^l_{ij} \leftarrow L\{R\}^l_{ij} + G\{\widehat{W}\}^l_{ij,k}.L\{I\}^l_{ij,k}$
  end for
 end for
 $R \leftarrow$ collapse Laplacian pyramid $L\{R\}$

While the sum of the weights is guaranteed for every pixel to be equal to 1, this does not imply that the reconstructed image belongs to the initial interval. In fact it may well happens that saturations occur in the dark or bright part. Avoiding them is possible by applying an affine rescaling of the image's dynamic to fit it to the standard interval $[0, 255]$. In our experiments, the resulting image generally presented no artifacts. The authors however present a case where the output image suffers from a very low frequency halo, giving an unnatural sensation (see fig. 6 of their paper [1]).

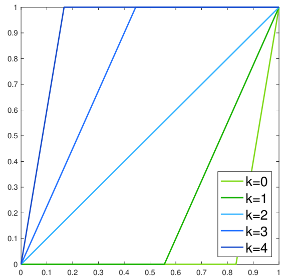

Figure 1: Remapping function used to generate the input sequence, here with parameter $\alpha = 6$ and $N = 4$.

3 Exposure fusion from a single image

The difficulty in local tone-mapping operators is to adapt the contrast modification to different areas and avoid unnatural behaviors at edges such as halo or edge sharpening. Since Exposure Fusion achieves very successfully the similar task of selecting and seamlessly merging areas from images with significant exposure changes, we propose to adapt the algorithm to make it work for a single image. The idea is to generate an input sequence simulating for this sole image its underexposed or overexposed versions. Used in this way, Exposure Fusion becomes a powerful image enhancement operator. The first question we encountered is: how to generate the sequence? We found that the choice of the over-or under-exposure processes is not that critical. Indeed Exposure Fusion metrics are designed to always select the best pixels among the available input images. In other words, Exposure Fusion will measure what correction, among the proposed ones, is the best for each input pixel. It is therefore only necessary to present a sequence which enhances the contrast at all levels of the dynamic.

A captured bracketed image can be written as

$$I_{ij,k} = f(E_{ij}\Delta t_k)$$

where Δt is the exposure time, E the scene irradiance and f is the overall acquisition non-linearity. Here k is the index of the exposure time Δt_k. Writing these exposure times as functions of a reference exposure time Δt_{ref} of I, we therefore have

$$I_{ij,k} = f(E_{ij}\Delta t_{\text{ref}}\alpha_k). \tag{7}$$

Although it is possible to recover f from the sequence of images [8], this is impossible from a single one. The only option is then to make a guess about the form of f and to simulate enough bracketed images compatible with the form of f. Most JPEG images have undergone a multiplicative color balance and a gamma-correction, which is a power function. Thus, approximating f by a power function seems appropriate. Denoting by p the exponent (with range between 1 and around 2.2), we deduce from (7) that the input sequence can be generated by setting

$$I_{ij,k} = I_{ij,\text{ref}}\alpha_k^p\ .$$

To artificially increase the exposure time (there is no reason to decrease it as we can't recover saturated parts) we therefore just have to multiply the image by factors larger than one. We tried other possibilities: adding an offset to the initial image (thus assuming that f was a logarithm) or even powers of the input image (assuming that f was an exponential). We found that using simple multiplications produced good results (see fig. 2).

Each used multiplier generates a pair of images. Indeed, applying a multiplier $\alpha > 1$ creates saturation. In order to prevent this loss of information, we propose two functions: f_{dark} multiplies by α and saturates the image in the dark parts, while f_{bright} saturates it in the bright parts. The important parameters thus left to the user are the maximal multiplicative factor α applied to the input image, and the total number N of images to generate. Denoting t an intensity, the remapping function are:

$$f_{\text{dark}}(t,k) = \max\{0, \Big((\frac{2k}{N}-1)^2(\alpha-1)+1\Big)(t-1)+1\}$$

$$f_{\text{bright}}(t,k) = \min\{1, \Big((\frac{2k}{N}-1)^2(\alpha-1)+1\Big)t\}$$

The factor $((\frac{2k}{N}-1)^2(\alpha-1)+1)$ applied to the intensity vary between 1 and α, with shorter increments around 1 – which resembles the way cameras do, since exposure time is generally a power of two. We drew these functions for the various values of k (denoting the position in the generated input sequence) in figure 1. All factors are equal or superior to 1 to guarantee that the fused image does not loose contrast.

The pseudo-code 2 describes the very simple steps of our algorithm: first, the generation of the input sequence, and then the application of Exposure Fusion.

Algorithm 2 Exposure fusion from a single image

Require: Method parameters: input image I, number of images to generate N (even), mapping functions f_{dark} and f_{bright}; Exposure fusion parameters: $\omega_s, \omega_c, \omega_e$
Ensure: output fused image R
  for $k \in \{0, 1, 2, ..., N\}$ **do**
    if $k < N/2$ **then**
      $\widehat{I}_{ij,k} \leftarrow f_{\text{dark}}(I_{ij}, k)$
    else
      $\widehat{I}_{ij,k} \leftarrow f_{\text{bright}}(I_{ij}, k)$
    end if
  end for
  $R \leftarrow$ Apply exposure fusion to sequence $\widehat{I}$ with parameters $\omega_s, \omega_c, \omega_e$

4 Results

Our experiments indicate that this method challenges the well known and very effective Multiscale Retinex [5, 6, 7]. It seems indeed able to increase both the lighting and contrast in dark areas, thus revealing information in the shadows. Furthermore, even the bright parts of the input image are improved. This is particularly relevant as that Multiscale Retinex tends to compress details in the bright areas, and generally gives grayish skies. These observations are confirmed by figure 2. Concerning the colors, exposure fusion from a single image shows more saturation.

An IPOL article is under preparation. A workshop is already available at `http://ipolcore.ipol.im/demo/clientApp/demo.html?id=77777444001`, letting the user try the two presented methods on his own images and explore the effect of each parameter.

Annex

Image credit: NASA `http://dragon.larc.nasa.gov/retinex/`

References

[1] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum*, volume 28, pages 161–171. Wiley Online Library, 2009.

[2] Peter Burt and Edward Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983.

[3] J.M. Ogden, E.H. Adelson, J R. Bergen, and P.J. Burt. Pyramid-based computer graphics. *RCA Engineer*, 30(5), 1985.

[4] P. J. Burt, K. Hanna, and R. J. Kolczynski. Enhanced image capture through fusion. *Proceedings of the Workshop on Augmented VisualDisplay Research*, pages 207–224.NASA – Ames Research Center., Dec. 1993.

[5] E. Land and J. McCann, Lightness and retinex theory, *Journal of the Optical Society of America*, 61 (1971), pp. 1–11.

[6] D.J. Jobson, Z. Rahman, and G.A. Woodell, A multiscale retinex for bridging the gap between color images and the human observation of scenes, *IEEE Transactions on Image Processing*, 6 (1997), pp. 965–976.

[7] Ana Belén Petro, Catalina Sbert, and Jean-Michel Morel, Multiscale Retinex, *Image Processing On Line*, (2014), pp. 71–88.

[8] Paul E Debevec and Jitendra Malik, Recovering high dynamic range radiance maps from photographs, *ACM SIGGRAPH 2008 classes*

Figure 2: Tone-Mapping with the "Single Image Exposure Fusion": original (left) and tone-mapped (center). Comparison with Multiscale Retinex on the intensity channel [7] (right)

Figure 3: Tone-Mapping with the "Single Image Exposure Fusion": generated input sequence (top row) and the corresponding weights (bottom row).