ARTICLE IN PRESS

Computer Vision and Image Understanding xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Clustering based content and color adaptive tone mapping

Hui Li[*,a], Xixi Jia[b,a], Lei Zhang[*,a]

[a] Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[b] School of Mathematics and Statistics, Xidian University, Xi'an, China

ARTICLE INFO

Keywords:
High dynamic range
Tone mapping
Clustering

ABSTRACT

By extracting image luminance channel and separating it into a base layer and a detail layer, the Retinex theory has been widely adopted for tone mapping to visualize high dynamic range (HDR) images on low dynamic range display devices. Many edge-preservation filtering techniques have been proposed to approximate the base layer for Retinex image decomposition; however, the associated tone mapping methods are prone to halo artifacts and false colors because filtering methods are limited in adapting the complex image local structures. We present a statistical clustering based tone mapping method which can more faithfully adapt image local content and colors. We decompose each color patch of the HDR image into three components, *patch mean, color variation* and *color structure*, and cluster the patches into a number of clusters. For each cluster, an adaptive subspace can be easily learned by principal component analysis, via which the patches are transformed into a more compact domain for effective tone mapping. Comparing with the popular edge-preservation filtering methods, the proposed clustering based method can better adapt to image local structures and colors by exploiting the image global redundancy. Our experimental results demonstrate that it can produce high-quality image with well-preserved local contrast and vivid color appearance. Furthermore, the proposed method can be extended to multi-scale for more faithful texture preservation, and off-line subspace learning for efficient implementation.

1. Introduction

The dynamic range (i.e., the ratio of maximum to minimum irradiance) of a natural scene is usually very high, approximately 14 orders of magnitude (Duan et al., 2010; Reinhard et al., 2010). However, the generic camera sensors have limited dynamic range, often resulting in under-exposure or over-exposure regions in a captured picture. High dynamic range imaging (HDR) has thus been an important topic in the field of computer vision and computational photography. One widely used strategy to extend the camera dynamic range is to take a sequence of images under different exposures (Ma et al., 2017; 2015b; Wu et al., 2016). With this strategy, there are two categories of approaches to obtain the HDR-like images: multi-exposure image fusion (MEF) (Wu et al., 2016) in image domain, and HDR content reconstruction in radiance domain (Badki et al., 2015; Debevec and Malik, 1997; Mitsunaga and Nayar, 1999).

MEF directly fuses the sequence of images into one image, which is easy to operate but suffers from the problem of ghosting artifacts and the severe dependency on the selection of exposure sequences (Mahmoudabadi et al., 2017; Wu et al., 2016). HDR content reconstruction methods first establish the radiance map by recovering the camera response function (CRF) and fuse the pixel values in radiance domain (Debevec and Malik, 1997). However, the calculation of CRF is complex and prone to reconstruction errors (Chakrabarti et al., 2014). With the improvement of sensor response sensitivity, high-end cameras can directly generate high-bit raw data without the recovery of CRF.

With the high-bit HDR image available, one important issue is how to display the HDR data. The standard display devices such as LCD, CRT, projectors and printers mostly have a low dynamic range (LDR) and cannot display HDR images directly. To fill in the gap between HDR data and LDR display, techniques have been developed to compress the dynamic range of HDR data for effective display, which are called tone mapping or tone reproduction (Drago et al., 2003; Fattal et al., 2002). A good tone mapping algorithm should faithfully preserve the image detailed features and colors while reducing the irradiance level. In the past two decades, a number of studies have been conducted to develop effective tone mapping algorithms. Generally speaking, the tone mapping methods fall into two primary categories: global tone mapping methods (Drago et al., 2003; Tumblin and Rushmeier, 1993) and local tone mapping methods (Fattal et al., 2002; Reinhard et al., 2002).

Due to the limited computational resources, early studies (Drago et al., 2003; Larson et al., 1997; Tumblin and Rushmeier, 1993; Ward,

* Corresponding author.
E-mail addresses: xiaohui102788@126.com (H. Li), xixijia@stu.xidian.edu.cn (X. Jia), cslzhang@comp.polyu.edu.hk (L. Zhang).
URL: http://www4.comp.polyu.edu.hk/~cslzhang/ (L. Zhang).

http://dx.doi.org/10.1016/j.cviu.2017.11.001
Received 14 February 2017; Received in revised form 28 June 2017; Accepted 1 November 2017
1077-3142/

1994) focus on designing simple global tone mapping operators. Tumblin and Rushmeier (1993) proposed a non-linear tone mapping algorithm according to the brightness perception of human visual system. Ward (1994) compressed image contrast instead of absolute luminance using a simple linear compression function. Larson et al. (1997) applied histogram adjustment to tone mapping by preserving the histogram distribution of the original HDR data. The adaptive logarithmic mapping in Drago et al. (2003) compresses the dynamic range with different logarithmic bases. The higher irradiance is compressed via log_2, whereas the lower irradiance via log_{10}, to achieve desirable contrast and detail preserving. Reinhard and Devlin (2005) proposed a simple and practical *s* curve for global tone mapping in independent channels. The global operators are computationally efficient without halo artifacts. However, the local contrast and visibility of details in the produced LDR images are not satisfactory.

Recent studies focus more on local tone mapping techniques. Fattal et al. (2002) designed a novel local tone mapping operator based on gradient attenuation. They compressed the drastic irradiance changes by reducing the large gradients under a multi-scale framework. Reinhard et al. (2002) classified the dynamic range of display devices into 11 zones according to the different irradiance in HDR data. Li et al. (2005) put forward a multi-resolution image decomposition method using symmetrical analysis-synthesis filter banks for local tone mapping. The gain map of each subband is calculated to alleviate the halo artifacts. Shan et al. (2010) developed a globally local optimization method with a locally linear model, where the guidance map is constructed via local statistical information. Gu et al. (2012) replaced the linear assumption (Shan et al., 2010) with the local non-linear gamma correction. Ma et al. (2015a) designed a tone mapped image quality index (TMQI) and performed dynamic range compression by optimizing this index. Chen et al. (2005) segmented the HDR image into different regions via the earth movers distance (EMD), and applied local tone mapping operation on each component. Ferradans et al. (2011) proposed a two-stage tone mapping method: human visual system based global tone mapping, followed by optimization based local contrast enhancement. Duan et al. (2010) improved the tone mapping performance of Larson et al. (1997) by applying adaptive local histogram adjustment on non-overlapped blocks. In general, local tone mapping methods are spatially adaptive, and can reproduce the local details and contrast well. However, these local operators have higher computational cost and are prone to producing halo artifacts (Li et al., 2005) and ringing effect (Shibata et al., 2016).

In recent years, researchers have been focusing on the design of various edge-preserving filters for tone mapping. The main principle is to decompose an HDR image into a detail layer and a base layer, and impose different operations on the two layers. In particular, the base layer image can be obtained by filtering the HDR data. Tumblin and Turk (1999) made the first attempt to design edge-preserving filters by using anisotropic diffusion to replace Gaussian filtering based on the Retinex theory (Jobson et al., 1997). Durand and Dorsey (2002) developed a fast implementation of bilateral filtering for tone mapping, which can efficiently generate smoothed images while preserving the edges. Based on this framework, many subsequent works (Farbman et al., 2008; Guarnieri et al., 2011; He et al., 2013; Kou et al., 2015; Li and Zheng, 2014; Li et al., 2015; Xu et al., 2011) have been proposed to better remap the HDR data. In Farbman et al. (2008), a weighted least squares based global optimization method was proposed to smooth the HDR data, where a larger weight is given to local details and contours, while a smaller weight is distributed to strong edges. An iterative method was proposed in Guarnieri et al. (2011) to improve the solving of weighted least squares. By minimizing the global gradient of an HDR image, Xu et al. (2011) used the l_0 norm as the regularizer to smooth the HDR image. He et al. (2013) proposed a guided filtering based method for edge preservation. A linear relationship is assumed between the guided image and the image to be filtered to avoid large edge loss. Some works (Kou et al., 2015; Li and Zheng, 2014; Li et al., 2015) introduce the gradient information as the weight to balance the data term and regularizer term in a local window, which share the similar idea to global weighted least squares.

The luminance edge-preservation filtering based tone mapping algorithms mentioned above can improve the visual quality of tone mapped image; however, the nonlinear filters used by them are not flexible and adaptive enough to fit the various edges and structures in natural images, resulting in halo artifact and false colors. Different from those luminance filtering based methods, in this paper we develop a statistical clustering based tone mapping method to more effectively exploit the image local and global redundancy. We do not separate an image into luminance and chrominance channels to process; instead, we work on image patches, and decompose a color patch into three components: patch mean, color variation and color structure. It is well-known that there exist repetitive patterns/structures in natural images (Dong et al., 2011; Zhang et al., 2010). Based on the color structure component, we group similar patches into clusters, and use statistical signal processing tools such as principal component analysis (PCA) to define a subspace of the patches in a cluster. Consequently, we can project each patch into a more compact domain, where the tone mapping operation can be more effectively performed. Compared with the edge-preservation filtering based methods, our proposed statistical clustering based method is more local content and color adaptive and robust since it exploits the image global redundancy to decompose local structures.

The main contributions of our paper lie in the following aspects. (1) Instead of using the deterministic edge-preserving filters, we leverage statistical clustering methods to better represent the local color structures of HDR images. Each patch will be adaptively processed based on its cluster. (2) We perform tone mapping in the PCA transformed domain other than the intensity domain, where the coefficients have explicit physical meanings and can be more effectively compressed. (3) Different from previous methods which extract luminance channel and perform layer separation on it, we do not extract luminance channel but process image luminance and chrominance information simultaneously.

The rest of our paper is organized as follows. Section 2 presents the proposed method in detail. Section 3 presents extensive experimental results and discussions. Section 4 concludes the paper.

2. Content and color adaptive tone mapping

2.1. The proposed tone mapping framework

Most previous tone mapping methods process luminance and chrominance separately. A typical framework of conventional tone mapping methods is shown in Fig. 1(a). Given an HDR image in RGB format, the luminance channel is first extracted as $L = 0.2126 \cdot R + 0.7152 \cdot G + 0.0722 \cdot B$ for the XYZ color space (Fattal et al., 2002), or $L = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$ for the YUV color space (Li and Zheng, 2014). In some literature (Gu et al., 2013), the average of R, G, B channels $L = 1/3(R + G + B)$ is employed as the luminance. After dynamic range compression on luminance, the chrominance is processed based on the compressed luminance to reproduce the tone mapped image. The widely used color processing operation is $C_{out} = \left(\frac{C_{in}}{L_{in}}\right)^s \cdot L_{out}$, where C represents the chrominance channel, L_{in} and L_{out} denote the luminance before and after HDR processing, and s adjusts the color saturation of the tone mapped image. The empirical value of s is between 0.5 and 0.9 (Gu et al., 2013).

In our proposed method, we do not separate image into luminance and chrominance channels to process. Instead, we propose a very different approach, whose framework is shown in Fig. 1(b). We partition the input RGB image into overlapped color patches, and decompose each patch into three nearly uncorrelated components. The color patches are clustered into a number of clusters, and statistical analysis is used to compress each HDR patch to an LDR one. The flowchart of the