Information Fusion 46 (2019) 114–127

Contents lists available at ScienceDirect

Information Fusion

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

Variable augmented neural network for decolorization and multi-exposure fusion

Qiegen Liu[a,b,*], Henry Leung[b]

[a] Department of Electronic Information Engineering, Nanchang University, Nanchang 330031, China
[b] Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada

ARTICLE INFO

Keywords:
Color-to-gray conversion
Multi-exposure fusion
Convolutional neural network
Auxiliary variables
L1-norm
Gradient domain

ABSTRACT

This paper shows how to convert a color image to grayscale using convolutional neural network (CNN), that preserves visual contrast via gradient domain modeling. We propose to explore the auxiliary variable principle to make the input and output variable dimensions to be the same, and use L1-norm error of the image gradients as the loss function criterion. The similarity measure calculates the summation of the gradient correlation between each channel of the color image and the transformed grayscale image. The final gray mapping result is then obtained by reconstruction from a globally initial grayscale image and locally derived gradient images. A weighted objective is proposed to balance the robustness and visual appearance of color images. Furthermore, by revealing the relation between color-to-gray and multi-exposure fusion, the network is applied to multi-exposure fusion. Both quantitative and qualitative evaluations on decolorization and multi-exposure fusion consistently demonstrate the potential of the proposed method against existing state-of-the-art algorithms.

1. Introduction

Color-to-gray conversion techniques are widely used in image processing and computer vision community. The conversion of color image into a grayscale one offers great benefits. The most important one is to assist colorblind persons to better visualize the real world. It also benefits the application of single-channel algorithms on color images in industry field such as black-and-white printing, non-photorealistic rendering with black-and-white media, and so on. Fewer processing operations, faster visualization and costs are the main factors. In addition, decolorization can improve the perceptual quality of an image in advertisement design. These applications prompt the color-to-gray conversion researches in the last decade. However, the color-to-gray conversion is a dimensionality reduction problem that introduces errors in the color-to-grayscale transformation.

The most popular color-to-grayscale transformation is to use the Y channel of CIE XYZ color space (e.g., CIE Y). If the source image is in RGB format, the luminance can be obtained by a linear combination of its R, G and B channels (e.g., the rgb2gray function in Matlab). However, taking the luminance channel only cannot fully represent structures and contrasts in some color images such as those having iso-luminant regions. Many efforts have been made to develop advanced perception-driven decolorization methods [1–17]. They can be broadly categorized as local-based [1–3] or global-based [4–9]. For the local-based approaches, the color-to-gray mapping of pixel values is spatially varying, depending on the local distributions of colors. Although they can preserve local features, constant color regions may be converted inhomogeneously if the mapping changes in the regions. The recently reported saliency-guided region-based optimization method by Du et al. [4] involves a two-stage parametric color-to-gray mapping function, which considers both global and local information to alleviate this problem.

Global-based methods use an objective function that minimizes the differences between mapped gray values and the differences between original color values [5–10]. Gooch et al. proposed color contrast between pixel pairs [5]. Rasche et al. enforced constraints directly on different color pairs and constructed a quadratic function to solve the gray images [6]. Kim et al. [7] proposed a nonlinear parametric model for global mapping. Kuk et al. [8] extended the work of [5] by balancing the gradient among pixels, or between pixels and some predetermined landmarks. Song et al. [9] introduced three visual cues and incorporated them into a global energy function which was optimized by a variational approach. Lu et al. [10] relaxed the strict order constrain and adopted a second order multivariate polynomial parametric model to maximally preserve the original color contrast. Liu et al. [13] presented a gradient correlation similarity (Gcs) measure, and determined the solution with the maximum Gcs value from the linear parametric model inducing candidate images. It is worth mentioning

* Corresponding author:.
E-mail addresses: liuqiegen@ncu.edu.cn, qiegen.liu@ucalgary.ca (Q. Liu), leungh@ucalgary.ca (H. Leung).

https://doi.org/10.1016/j.inffus.2018.05.007
Received 30 August 2017; Received in revised form 9 May 2018; Accepted 26 May 2018
Available online 06 June 2018
1566-2535/ © 2018 Elsevier B.V. All rights reserved.

that many color-to-gray methods adopt the gradient-based techniques to measure the fidelity between the color input and transformed grayscale image [5–14].

Besides accuracy, robustness is another important issue. Calik et al. [18] introduced a psychological evaluation of several color-to-gray algorithms and showed that these conversion methods were founded to perform badly at least one test image, indicating none of them could outperform the others. Song et al. [19] discussed the robustness of existing methods and considered multi-scale contrast preservation by taking advantage of the joint bilateral filter. On the other hand, there exists at least one parameter that has to be determined manually for most of the existing methods. Ma et al. proposed the C2G-SSIM index to predict the perceived quality of converted images and applied it to the parameter tuning of conversion algorithms [20].

Multi-exposure image fusion (MEF) is a problem that is close to color-to-gray conversion. Given a source image sequence with different exposure levels as input, MEF is devoted to synthesizing them as a single output such that more informative and perceptual details can be reproduced. Similar to the decolorization problem, the methods in the literatures try to combine the sequence images with weighting map under different criteria [21–31]. These approaches can be roughly classified into two classes: the signal decomposition strategy and the edge-preserving strategy. In the first group, Burt and Kolczynski applied Laplacian pyramid decomposition to MEF [21]. They selected the local energy of pyramid coefficients and the correlation between pyramids within the neighborhood to calculate the weights. Mertens et al. [22] adopted proper contrast, high saturation and well exposure as quality measures to guide the fusion process in a multiresolution fashion. Li et al. [23] enhanced the details of a given fused image by solving a quadratic optimization problem. Recently, Ma et al. [24] proposed a patch decomposition approach to MEF. The adoption of signal direction enables the method to handle dynamic scenes. The class of edge-preserving strategy consists of gradient optimization techniques and edge-preserving filters. For instance, bilateral filter was used in [25] to calculate edge information, which was subsequently employed to compute the weights. Song et al. [26] estimated the initial image by maximizing the visual gradient, and synthesized the fused image by suppressing reversals in image gradients. Zhang and Cham [27] constructed visibility and consistency measures from gradient information and used them as the weighting factors. A similar gradient-based MEF method was proposed in [28]. Li et al. [29] employed the guided filter to control the roles of pixel saliency and spatial consistency in constructing the weighting map.

In this paper, we propose a global and local fusion method that incorporates global algorithms with a new locally gradient-based model using a deep learning architecture [32–34]. We investigate the feasibility of utilizing supervised learning to produce robust and visually perceptual color-to-gray mapping. Motivated by the Gcs method [13], we use L1-norm of image gradient errors as a criterion fast and robust color-to-gray conversion. The similarity measure is conducted between each channel of the input color image and resulting grayscale images to reflect the degree of preserving feature discriminability and color ordering in color-to-gray conversion. The proposed mapping result is then obtained by intermediate reconstruction from gradient domain. A weighted objective is proposed to balance the global and local visual appearance of color images. We also investigate the relationship between decolorization and MEF, and apply the network to fuse a source sequence at three exposure levels, via viewing each exposure level as a signal channel.

The reminder of this paper is organized as follows. Section 2 summarizes some conventional gradient-based measures for decolorization. Then L1-norm guided neural network is developed to deal with the color-to-gray conversion. The detailed implementation of the DecolorNet network and its application to multi-exposure image fusion dubbed FusionNet are given in Section 3. In Section 4, the proposed method is compared with conventional approaches. It is shown that feature discriminability is better preserved on a variety of color images and multi-exposure video sequences. Discussions and conclusions are given in Sections 5 and 6, respectively.

2. Proposed DecolorNet model

2.1. Review of gradient-based metrics

In order to retain feature discriminability in color-to-gray conversion, a common strategy is to minimize the distance of pixel differences between the input color and the resulting grayscale images. Assume that the input color image is in the RGB format, where indices *R, G, B*stand for the RGB channels. Let $\delta_{x,\,y}$ ($|\delta_{x,y}|=\sqrt{\sum_{c=\{R,G,B\}}(I_{c,x}-I_{c,y})^2}$) be the color contrast having a signed value indicating the difference of a color pair and $g_x - g_y$ denote the gray difference value between pixels g_x and $g_y$respectively, then the classical L2-norm based energy function [5, 7] is defined by

$$\min_g \sum_{(x,y)\in P} (g_x - g_y - \delta_{x,y})^2 \tag{1}$$

The estimated image *g* could be with [7] or without [5] a parametric form. *P* stands for a pixel pair pool that contains the local and nonlocal candidates. Incorporating the differences among distant pixels into the energy function makes the model able to deal with both pixels at nearest neighbors and long-scale contrast regions [8,10,12]. To alleviate the difficulty caused by the sign of $\delta_{x,\,y}$, Lu et al. [10] proposed a bimodal contrast-preserving objective function and employed a finite multivariate polynomial function for mapping.

Different from the conventional gradient error (GE) norm-based measures used in refs. [5,7,9,10,12], Liu et al. [13] proposed the Gcs measure that calculates the gradient correlation for each channel rather than all channels at one time. It computes the overall pixel-wise similarity between the gradient magnitudes in each channel of the original color image and the resulting grayscale image, i.e.

$$\min_g -\sum_{(x,y)\in P} \frac{2|I_{c,x}-I_{c,y}||\nabla g_{x,y}|}{|I_{c,x}-I_{c,y}|^2+|\nabla g_{x,y}|^2} \tag{2}$$

The mathematical notation ∇ represents the nonlocal gradient operator at pixel pair (*x, y*) as $\nabla g_{x,y} = g_x - g_y$. Note that the structure preservation is described by the Gcs, partly alleviating the shortcoming of the GE measure that excessively focuses on penalizing the pixels with large gradient magnitude value.

2.2. DecolorNet

We consider here mapping color images to gray images via convolutional neural network (CNN). Intuitively and obviously, using the loss function in Eq. (1) directly with regard to the input $\delta_{x,\,y}$ and output$g_x - g_y$ is not appropriate as it comes across the difficulty of determining the sign of$\delta_{x,\,y}$. Even if the network is perfectly trained, the L2-norm derived GE measure excessively focuses on penalizing the pixels with large magnitude values and may ignore the pixels with small magnitude values.

Another consideration is to follow the dimension reduction idea by means of deep learning. i.e., input $\nabla I_{c,\,xy}$ and output$g_x - g_y$. However, designing a neural network for dimension reduction is complicated. As stated in many references [34–36], the number of multi-layer should be large such that the reduction process can be exhibited robustly. For instance, the encoder-decoder system in [23] needs layer-by-layer pretraining and fine-tuning. By the way, until now there is no state-of-the-art dimension reduction method proposed in the field of decolorization. For instance, the principal component analysis (PCA) has been applied to decolorization in [5,6,12], but the results are not satisfactory. This indicates that directly employing dimension reduction techniques to color-to-gray is not a promising direction.

Download English Version:

https://daneshyari.com/en/article/6937841

Download Persian Version:

https://daneshyari.com/article/6937841

Daneshyari.com