Color-to-grayscale conversion through weighted multiresolution channel fusion

Tirui Wu
Alexander Toet
1 Introduction

Despite the widespread availability of color sensors and color display technologies, color-to-grayscale conversion is still widely used in many applications, including black-and-white printing (e-ink-based book readers), producing reading materials for color blind people, single-channel image processing, and nonphotorealistic rendering with black-and-white media. A good color-to-grayscale conversion should preserve both the overall appearance and the discriminability of the features of the original color image. However, since color-to-grayscale conversion reduces three-dimensional color data to one-dimensional data, loss of information is inevitable. To retain feature discrimination in color-to-gray conversion, the resulting grayscale values should reflect chromatic differences in the input image. Hence, the problem is to find a lower dimension embedding of the original data that optimally preserves the perceptual contrast between data points (pixels) in the original data. This implies that the following constraints should be taken into account:\(^1\)

- Global consistency: pixels with the same color in the input color image should be mapped to the same gray value in the grayscale image.
- Local luminance consistency: luminance gradients should be preserved.
- Preservation of chromatic contrast: different colors should be mapped to different grayscale values.
- Preservation of hue order: the natural ordering of hues should be reflected in their converted grayscale values.

Many different color-to-gray conversion algorithms have been proposed that aim to retain the original discriminability of color images (for some excellent overviews see Refs.\(^2\) and \(^3\)). They are typically divided into two main categories:

- Local and global mappings. In a local mapping method, the color-to-gray mapping of pixel values is spatially varying, depending on the local distributions of colors. Although a local mapping can accurately preserve local features, it may produce inhomogeneous representations of constant color regions when the mapping changes in those regions. In a global mapping method, a spatially uniform color-to-gray mapping is used: the same colors are consistently mapped to the same grayscale values over an image, guaranteeing homogeneous representations of constant color regions. However, it is a challenge to determine a global mapping that preserves local features. The previous studies have shown that the global mappings are typically efficient and fast and ensure that the identical color values are mapped to identical gray values, while local approaches can yield better results but are typically complex and computationally expensive.\(^3,4\)

In this paper, we present a color-to-gray conversion algorithm that retains both the overall appearance and the discriminability of the input image. The algorithm employs a weighted pyramid image fusion scheme to blend the R, G, and B color channels of the input image into a single grayscale image. The use of simple visual quality metrics as weights in the fusion scheme serves to retain visual contrast from each of the input color channels. We demonstrate the effectiveness of the method by qualitative and quantitative comparison with several state-of-the-art methods. © 2014 SPIE and IS&T [DOI: 10.1117/1.JEI.23.4.043004]
pyramid weight maps and summation, the final grayscale converted image is obtained by reconstructing the resulting pyramid. The method is straightforward and automatic (requires no user interaction), has limited complexity, and performs at least as well as the best performing state-of-the-art color-to-grayscale conversion algorithms. The main contribution of this paper is that it formulates the color-to-grayscale conversion problem as a visual saliency-weighted multiscale channel fusion scheme, to achieve a grayscale representation that optimally represents the information contained in each of the channels of an RGB color image.

The rest of this paper is organized as follows. Section 2 briefly reviews related work. Section 3 describes the proposed method. Section 4 demonstrates the effectiveness of the proposed method, and Sec. 5 provides the conclusions.

2 Related Work
A simple and widely used method to convert an input color image into a grayscale image is to first separate its luminance and chrominance channels (e.g., by linearly combining its R, G, and B channels), and then take the luminance channel as its grayscale representation while discarding the chrominance information. A typical approach is, for instance, to take the Y channel in the CIE XYZ color space. In this approach, clearly distinguishable regions with isoluminant colors in the original image will be mapped to the same grayscale and become indistinguishable in the result image.

Local color-to-grayscale conversion methods typically use local chrominance edges for enhancement. Bala and Eschbach add high frequency chromaticity components to the luminance channel to preserve the local distinction between adjacent regions with different colors in the color-to-grayscale conversion. Neumann et al. determine a single-gradient field that is most consistent with both the color and the luminance gradient fields of the input color image and integrate the result to reconstruct a grayscale image. Smith et al. presented a two-step grayscale transformation that combines a global mapping based on perceived lightness with local chromatic contrast enhancement. The method first applies a global mapping based on the Helmholtz–Kohlrausch effect, and then locally enhances chrominance edges using adaptively weighted multiscale unsharp masking. Although the global mapping is image independent, the local enhancement reintroduces lost chromatic discontinuities. However, this method may distort the appearance of constant color regions, and the second step requires user control. Song et al. defined three visual cues by considering the color spatial distribution, the gradient information, and the perceptual priority of the color channels (hue, chroma, and lightness) and formulated color-to-gray conversion as a visual cue preservation procedure.

Global color-to-grayscale conversion methods that aim to minimize the overall perceptual difference between the input color images and the output grayscale images typically involve complex optimization steps. Gooch et al. try to preserve color contrast by minimizing an objective function that is based on the local contrast between all pixel pairs. The method is computationally complex and requires user interaction. Kuk et al. extended the method of Gooch et al. by considering both local and global contrasts. Rasche et al. aim to preserve contrast while maintaining luminance consistency by minimizing an error function based on matching...
perceived grayscale differences to corresponding perceptual color differences. Color quantization is required to reduce the extreme computational costs of the optimization procedure, which results in artifacts in (natural) images with continuous tones. Grundland and Dodgson\textsuperscript{17} presented a global mapping that adds a fixed amount of chrominance to the luminance to enhance grayscale contrast, whereas the original luminance and the color ordering are better preserved by restraining the added amount of chrominance. However, luminance and chrominance contrasts may cancel out on addition and the method does not discriminate between colors perpendicular to the projection axis.\textsuperscript{1} Kim et al.\textsuperscript{1} presented a nonlinear global mapping that preserves feature discriminability, color ordering, and luminance fidelity by minimizing the difference between color and resultant grayscale image gradients. The influence of chromatic contrast on feature discriminability is controlled by the user. Lu et al.\textsuperscript{18} used a bimodal energy function as a more flexible (weaker) color ordering constraint, which enables a fast implementation. Zhu et al.\textsuperscript{12} globally modulate the luminance channel of a color input image with the normalized contrast of the most salient (L, H, or S) channel and adopt the result as the grayscale conversion. Lee et al.\textsuperscript{19} preserve color contrast by adding chromatic contrast to the luminance component. Hsin et al.\textsuperscript{21} combine a global mapping that preserves the natural hue order of the input image with a local mapping that restores local contrast. Lau et al.\textsuperscript{22} define an energy function on a clustered color image. This method is able to perform transformation between different color spaces. Wu et al.\textsuperscript{23} presented an interactive two-scale (global-local) algorithm that first segments the input image, then determines a grayscale for each segment in a global optimization procedure, and finally performs local contrast enhancement to restore local details.

3 Method

Mertens et al.\textsuperscript{24,25} presented a method to fuse a multiple exposure sequence into a single high quality low dynamic range image. To retain as much detail and color as possible, the fusion process selects the visually most salient pixels using simple quality measures (e.g., saturation and contrast) that are computed for each pixel in the multiexposure sequence. The method is related to image fusion techniques used for e.g., depth-of-field enhancement,\textsuperscript{26} multimodal image fusion,\textsuperscript{27-29} and video enhancement.\textsuperscript{30} It is made more flexible than earlier approaches (e.g., Ref. 31) by incorporating adjustable visual saliency metrics. Mertens et al.\textsuperscript{5,24,25} pyramid blending strategy is similar to that of Grundland et al.\textsuperscript{33} but it employs different saliency metrics. In this paper, we present a color-to-grayscale conversion method inspired by Mertens et al.\textsuperscript{24,25} that converts an input color image into a grayscale image by hierarchical saliency weighted fusing of its three color channels.

Following Mertens et al.\textsuperscript{24,25} we first compute a scalar-valued weight map for each of the R, G, and B channels of the input color image. This weight map is composed of three simple visual quality metrics and serves to preserve only the most salient (informative) visual details from each of the input color channels in the final grayscale image.

The first quality measure $C$ represents contrast and is computed by taking the square value of a Laplacian-filtered version of the input channel. It tends to assign a large weight to visually important elements such as edges and texture. The second measure $S$ is defined for a given channel as one minus the absolute difference between that channel and the mean of all three channels (e.g., for the R channel, the $S$ measure is $1.0 - |R - [(R + G + B)/3]|$; assuming pixel values ranging from 0 to 1). This measure assigns larger weights to pixels with a smaller deviation from the mean of the three channels. The rationale for this assumption is the fact that the color channels for naturalistic images are highly correlated in RGB space. Therefore, pixels near the mean of the three

![Fig. 2 Comparison of the results of our method with those of (CIEY), (Smith08), (Neumann07), (Grundland07), (Rasche05), (Gooch05), (Bala04) for each of the 24 test images. Input and result images of other methods are courtesy of Cadik.](image-url)
channels are assumed to be well defined, whereas pixels far from this mean are assumed to represent outliers. The last quality measure $E$ (corresponding to Mertens et al.'s well exposedness) is computed by weighting each pixel intensity $i$ with its distance from 0.5 (the mean pixel value) using a Gaussian weighting function: $\exp[-(i - 0.5)^2 / 2\sigma^2]$. Following Mertens et al., we adopt $\sigma = 0.2$. This weighting ensures that pixels that are either under- (near 0) or over- (near 1) exposed are weighted less than pixels that are well exposed (near 0.5). This metric prioritizes information from intermediate valued image regions that typically provide a more articulated representation of image details.

The three quality measures are then combined into a single-weight map through multiplication:

$$W_{ij;k} = C_{ij;k} \times S_{ij;k} \times E_{ij;k},$$  \hspace{1cm} (1)

where $C$, $S$, and $E$ represent, respectively, contrast, saturation and well-exposedness, with corresponding “weighting” exponents $\omega_C$, $\omega_S$, and $\omega_E$. The subscripts $ij, k$ refer respectively to a pixel with coordinates $(i, j)$ in channel $k$, $k \in \{R, G, B\}$.

After computing a weight map for each of the (R, G, and B) input channels, the three weight maps are normalized such that they sum to one for each pixel:

$$\bar{W}_{ij;k} = \left[ \sum_{k'}^N W_{ij;k'} \right]^{-1} W_{ij;k}. \hspace{1cm} (2)$$

Straightforward application of the normalized weight maps to fuse the three color channels of the input color image may result in undesirable halos around edges and may spill information across object boundaries. This problem can be solved by using a pyramidal image fusion scheme as suggested by Mertens et al. The hierarchical weighted fusion scheme we propose for color-to-grayscale conversion is as follows (see Fig. 1). First, a multiresolution weight map is constructed for each input channel by applying a Gaussian pyramid transform to its corresponding weight map. Next, the three input channels are decomposed into Laplacian pyramids, which basically contain band-pass filtered versions at different spatial scales. Then, these Laplacian pyramids are multiplied by their corresponding Gaussian pyramid weight maps and the results are summed. The final grayscale converted image is then obtained by reconstructing the resulting pyramid. This color-to-grayscale conversion method is straightforward and automatic (it requires no user interaction) and has limited complexity. In Sec. 4, we will show that it performs at least as well as the best performing state-of-the-art color-to-grayscale conversion algorithms.

4 Experimental Results

To evaluate the performance of our algorithm, we applied it to a set of 24 color images (first column in Fig. 2) that are part of a publicly available color-to-gray benchmark dataset, which also includes the results of several state-of-the-art color-to-grayscale conversion algorithms. In all cases, we used the parameter setting $\omega_C = \omega_S = \omega_E = 1$. Figure 2 shows the results of our color-to-grayscale conversion algorithm (second column) and those of the CIE-Y conversion, Smith et al., Neumann et al., Grundland and Dodgson, Rasche et al., Gooch et al., and Bala and Eschbach, for each of the 24 input color images. These results show that our method preserves the overall visual appearance and color contrast of all the input images just as well or even better than the best performing state-of-the-art methods. Figure 3 shows the results for the fifth input image from Fig. 2 together with an enlarged part of each of the results. This figure clearly shows the ability of our algorithm to preserve both local and more global consistencies.

To investigate the relative contribution of the different factors in the fusion weight map we also tested seven different combinations of the parameters $\{\omega_C, \omega_S, \omega_E\}$: [1 0 0], [0 1 0], [0 0 1], [1 1 0], [1 0 1], [0 1 1], and [1 1 1]. Figure 4 shows an example of the results of our algorithm with each of these parameter settings on image 5 from Fig. 2. This figure clearly shows that the elimination of the contrast measure from the weight map ($\omega_C = 0$; images 2, 3, and 6 in Fig. 4) results in grayscale images with significantly reduced luminance contrast. Figure 5 shows a comparison of the relative effect of our three quality measures. In this example, color-to-grayscale conversion is performed with either contrast, saturation, or well-exposedness as a weighting factor. This example shows that the contrast factor C preserves local luminance gradients, whereas the relative saliency factor S and the well-exposedness factor E both serve to preserve, respectively, chrominance and global luminance contrast to some extent.

To enable an objective quantification of the performance of a color-to-grayscale conversion method, we adopt the normalized cross-correlation NCC between the resulting grayscale image and the R, G, and B color channels of the original input image as a conversion quality metric.
where $I_i$ represents one of the three (R, G, or B) channels of the color input image, $I_g$ represents the grayscale output image, and $x$, $y$ represent the image coordinates. The rationale for using this metric is the requirement that the color variations (features) in each of the individual color channels should be optimally preserved in the resulting grayscale image. We computed the NCC metric for each of the conversions shown in Fig. 2. The results shown in Fig. 6 demonstrate that our new method performs better than or equal to the other methods for all images tested here.

5 Conclusions

We presented a color-to-gray conversion algorithm that employs a weighted pyramid image fusion scheme to blend the R, G, and B color channels of the input image into a single-grayscale image. The use of simple visual quality metrics as weights in the fusion scheme serves to retain contrast details from each of the input color channels. Visual (subjective) and quantitative (objective) comparisons of the results of our method with those produced by several state-
of-the-art methods demonstrate that our new method retains as many or even more features from the original color image in the final grayscale fused image than the best performing state-of-the-art methods.

5.1 Limitations of the Method

For most of the examples tested and for a single-fixed parameter setting ($\omega_c = \omega_a = \omega_g = 1$; i.e., equal weights for each of the quality measures), our algorithm already retains more information in the final grayscale image representation than several of the state-of-the-art methods. As for any color-to-gray-scale conversion algorithm there may, of course, be exceptions to this finding. For instance, the relatively strong contribution of the contrast factor to the composite weight map may result in an over enhancement of edges (e.g., image id3 in Fig. 2). This can easily be remedied by adjusting the relative weights of the individual quality measures.

The relative luminance contribution of the individual channels to the final grayscale image is weighted with their respective saliency maps (composed of the three feature quality factors C, S, and E). In this way, the method converts the input color image into a grayscale representation that retains visually salient information at each level of detail from the individual color channels. A limitation of this approach is the fact that the perceptual hue order may be lost in this process.

Since there are currently no validated objective metrics available to quantify the performance of color-to-grayscale conversion methods, we adopted the normalized cross-correlation NCC between the resulting grayscale image and the R, G, and B color channels of the original input image as the computational conversion quality metric. This metric measures to what extent information from the individual input color channels is retained in the final result. Since this metric has no direct relation to human visual perception, it may, in some cases, assign a quality ranking that is not strictly related to human judgment (e.g., the visual quality of the CIE-Y results for images 3 and 5 in Fig. 2 is actually quite low, whereas their NCC values are relatively high).

Acknowledgments

Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA8655-11-1-3015. The U.S. Government is authorized to reproduce and distribute reprints for governmental purpose notwithstanding any copyright notation thereon.

References