Int J Comput Vis
DOI 10.1007/s11263-014-0732-6

Contrast Preserving Decolorization with Perception-Based Quality Metrics

Cewu Lu · Li Xu · Jiaya Jia

Received: 20 February 2013 / Accepted: 10 May 2014

Abstract Converting color images into grayscale ones suffer from information loss. In the meantime, it is one fundamental tool indispensable for single channel image processing, digital printing, and monotone e-ink display. In this paper, we propose an optimization framework aiming at maximally preserving color contrast. Our main contribution is threefold. First, we employ a bimodal objective function to alleviate the restrictive order constraint for color mapping. Second, we develop an efficient solver that allows for automatic selection of suitable grayscales based on global contrast constraints. Third, we advocate a perceptual-based metric to measure contrast loss, as well as content preservation, in the produced grayscale images. It is among the first attempts in this field to quantitatively evaluate decolorization results.

Keywords Decolorization · Color2gray · Conversion · Contrast preservation · Perceptual-based · Quality metrics

1 Introduction

Decolorization, also known as color-to-gray, is a task aiming at converting a color image into a grayscale one, where the latter carries only intensity information. Apart from its fundamental importance in single channel image processing, decolorization is necessary for black-and-white display and printing. For example, many documents in government, companies, and academia are still printed in monotone nowadays to save costs and resource. E-ink monotone display, as a major digital replacement of paper books, can show grayscale figures and characters. These devices, in fact, need effective decolorization algorithms—grayscale versions of a color image can be disappointing when meaningful structures vanish.

Two examples are shown in Fig. 1. The original very conspicuous rectangle in (a) becomes nearly imperceptible after being printed out by a monotone printer, as illustrated in (b). We also tested displaying the chart in (d) on e-ink devices (Amazon Kindle and Barnes and Noble Nook). The effect is also not satisfactory, as shown in (e). There are many similar examples that cannot be well handled even with the most advanced hardware. Thus, research of effective decolorization methods is still in demand.

While decolorization is well-motivated in many practical situations, intuitive methods, such as extracting the lightness channel in the CIELab color space (Sharma and Bala 2002; Fairchild et al. 2005), do not explicitly capture the important appearance features and easily diminish salient structures. To preserve color contrast, recent color-to-gray methods put constraints on spatial intensity variation and require that the grayscale contrast is similar to that of the color input. Signed color difference, as well as the color order for neighboring pixels, is specified (Gooch et al. 2005; Kim et al. 2009) to constrain decolorization based on the Euclidian color distance and on the order of lightness channel. These stringent constraints, in several cases, lose freedom in selecting suitable intensity values and may cause contrast loss. Figure 2a–e contains results of a few color-to-gray methods (Gooch et al.

Communicated by Dr. Srinivas Narasimhan, Dr. Frédo Durand and Dr. Wolfgang Heidrich.

Electronic supplementary material The online version of this article (doi:10.1007/s11263-014-0732-6) contains supplementary material, which is available to authorized users.

C. Lu · J. Jia (✉)
The Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong
e-mail: leojia@cse.cuhk.edu.hk

L. Xu
Lenovo Research and Technology, Hong Kong, China
e-mail: xuli@cse.cuhk.edu.hk

Published online: 01 July 2014

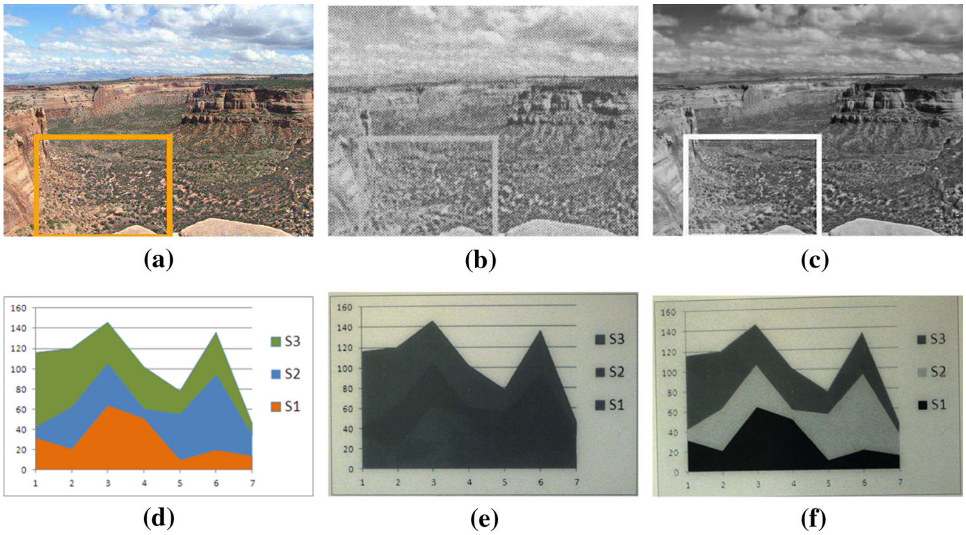


Fig. 1 The problem of color contrast loss. **a** Color input. **b** Printout of (**a**) by a HP Laserjet printer. **c** Our decolorization result. **d** Color input. **e** Image **d** displayed on an Amazon Kindle. **f** Display of our result

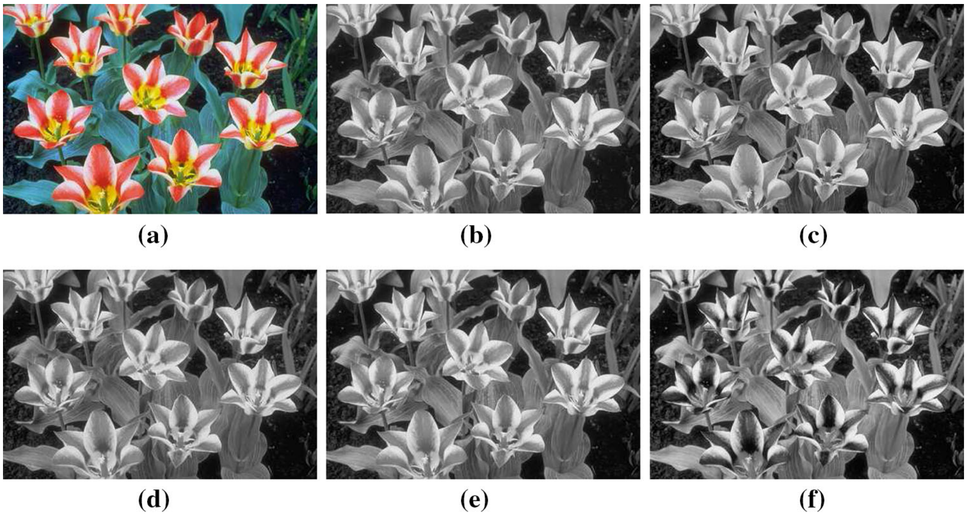


Fig. 2 One example. **a** Color input. **b**–**e** Results of state-of-the-art decolorization methods. **f** Our result with preserved contrast on the petal

2005; Smith et al. 2008; Kim et al. 2009). Although details are mostly preserved, significant color contrast on the petal is not well represented in the grayscale images.

In fact, human visual system does not accurately perceive chrominance and lightness. Instead, their relationship to adjacent context (Lotto and Purves 2002; Corney et al. 2009) plays a vital role. The order of different colors (Wong 2010) also cannot be defined uniquely by people, as evidenced in psychology and medical research. An example is that green becomes brighter than blue in lightness in the CIELab color system. But recent study indicates that people with different culture and language background have different sense of brightness. Someone feels just the opposite (Ozgen 2004; Ke et al. 2010).

In this paper, we relax the rigorous color order constraint and present a new method seeking a better preservation of original color contrast and visual distinctiveness. The advantage of the proposed model lies in the new *weak color order constraint*, which allows for a flexible and practical color-to-gray model. For color pairs without a clear order in brightness, we propose a bimodal distribution, i.e., mixture of two Gaussians, to automatically select suitable orders with respect to the visual context in optimization.

We also address a color-importance problem by introducing a perceptually motivated *Weber–Fechner weight*, which does not excessively depend on the dominant color components. To effectively solve the optimization problem in the new model, we propose a parameterized multivariate polynomial function for color mapping and develop a fixed point iteration solver. Our MATLAB implementation takes less than 1 s to process an one mega-pixel image. Last but not least, we propose a new metric to quantitatively measure the quality of color-to-gray conversion in terms of distinctiveness preservation. Experiments indicate that the new metric is in compliance with human visual perception, providing an automatic way for decolorization assessment.

Our preliminary conference version (Lu et al. 2012) discussed the bimodal objective function. This paper extends it in the following ways. (1) We generalize the mapping model by considering non-local color pairs and combine both the local and global constraints in optimization. (2) We identify the color weighting problem and propose the “Weber–Fechner weight” to ameliorate it. (3) A new solver based on landmark color extraction and fixed point iteration is developed with numerical acceleration. (4) We propose a new quantitative measure CCFR to complement the original CCPR measure and conduct extensive experiments to perceptually verify it.

The rest of the paper is organized as follows. We review related work in Sect. 2. Section 3 presents our parametric decolorization model. The bimodal contrast preserving objective function with the weak order constraint is introduced in Sect. 4, with an approximated energy function presented in Sect. 5. We provide the numerical solver in Sect. 6 and results in Sect. 7. Section 8 concludes this paper.

2 Related Work

General decolorization strategies are to find a fixed and input-independent transform to carry out color-to-gray projection. RGB-to-YUV conversion (Sharma and Bala 2002; Fairchild et al. 2005) can make black-and-white and color TVs compatible based on the same format. The Y channel is a linear combination of RGB channels, used for black-and-white display. Another popular constant mapping to simply extracting lightness in the CIELab (Hunter 1958) color space. Nayatani et al. (1997) proposed a color appearance model, which is still a content-independent mapping. Color contrast loss could happen in these methods.

Other more advanced decolorization methods can be roughly categorized into local and global ones, based on different spatial treatment of pixels. Local methods usually rely on local chrominance edges for result contrast enhancement. Bala and Eschbach (2004) added high-frequency chromaticity components to the lightness channel, in order to enhance edges. Neumann et al. (2007) locally selected consistent color gradients and performed fast 2D integration to get the final grayscale image. Smith et al. (2008) also employed a local sharpening step after obtaining the grayscale image by global mapping. Chrominance edges are enhanced by the adaptively weighted multi-scale unsharp masking. These mechanisms might occasionally alter the appearance of constant-color regions and produce haloing artifacts, as discussed in (Kim et al. 2009).

In global mapping, Bala and Braun (2004) sorted colors in an image according to their lightness. Gooch et al. (2005) enforced color contrast between pixel pairs. For each color pair, color order is assigned based an empirical piecewise function. Rasche et al. (2005) defined constraints directly on different color pairs. A linear color mapping is adopted for acceleration. Ahn et al. (2010) extended the idea of (Gooch et al. 2005) by considering both the global and local contrasts. Grundland and Dodgson (2007) proposed a fast linear mapping algorithm that adds a fixed amount of chrominance to lightness. Parametric piecewise linear mapping is used to convert color to grayscale. Kim et al. (2009) proposed a nonlinear parametric model. The parameters are estimated by minimizing the cost function that aims to preserve the color difference computed in the CIELab color space. Another quadratic objective function computing color difference in modified CIELCH color space was proposed in Song et al. (2010). Ancuti et al. (2011) designed a similar decolorization objective function as (Kim et al. 2009) and took saliency as guidance.

In short, to preserve color contrast, most previous methods explicitly specify color orders. This strategy may shrink the space to optimally pick grayscale values and lead to less optimal solutions in terms of retaining originally prominent contrast. There is also no good metric that can properly measure the result quality quantitatively.

3 Parametric Decolorization Model

The decolorization function is defined as $g = f(\mathbf{c})$. For each input RGB vector $\mathbf{c} = (c_r, c_g, c_b)$, function f outputs g, the corresponding grayscale value. We adopt a global mapping scheme where pixels with the same color are converted to the same grayscale.

A finite multivariate polynomial function is used. Mathematically, we define the polynomial space of color $\mathbf{c}$ with its degree n as

$$\Pi_n = \mathrm{span}\{c_r^{d_1} c_g^{d_2} c_b^{d_3} : d_i = 0, 1, 2, \ldots, d_1 + d_2 + d_3 \leq n\}, \quad (1)$$

where Π_n is a polynomial space spanned by a family of monomials. The mapping function is thus expressed as

$$f(c_r, c_g, c_b; \omega) = \sum_i \omega_i m_i, \quad (2)$$

where m_i is the ith monomial basis of Π_n. The mapping function is uniquely determined by weights $\{\omega_i\}$. Empirically, n is set to 2, which means the number of $\{\omega\}$ is 9 and the mapping function is a linear combination of elements in $\{c_r, c_g, c_b, c_r c_g, c_r c_b, c_g c_b, c_r^2, c_g^2, c_b^2\}$.

It is noteworthy that the polynomial form, albeit simple, is already sufficiently powerful and flexible. It is actually a *generalization* of common linear and nonlinear color-to-gray mapping functions. We have experiment with the image set (Cadík 2008) by first generating grayscale images using traditional methods, including the lightness channel of CIELab (Hunter 1958), Y channel in the YUV space (Sharma and Bala 2002; Fairchild et al. 2005), and Nayatani model (1997), the last of which is known to be highly nonlinear. Then we

Table 1 Mean and variance of the fitting errors when using our method to approximate other widely adopted color-to-gray models

	Mean	Variance
Nayatani model	1.28×10^{-2}	2.61×10^{-3}
L channel of Lab	3.20×10^{-3}	8.71×10^{-3}
Y channel of YUV	0	0

All grayscale levels are mapped into [0, 1]

fit our model to approximate these grayscale images using quadratic regression, i.e., by minimizing $\|f(\mathbf{c}, \omega) - g'\|^2$, where g' denotes the grayscale results of these widely employed methods. Table 1 lists the fitting errors. They are all very small, indicating that our parametric model is at least capable to work in the same way as previous color-to-gray linear and non-linear mappings.

4 Bimodal Contrast-Preserving Energy Function

We describe in this section our new bimodal color contrast preserving energy function with a weak color order constraint. To begin with, We revisit the energy used in previous approaches for contrast preserving decolorization. The grayscales for pixel pair x and y in a pixel-pair pool $\mathcal{P}$ are denoted as g_x and g_y respectively. They are estimated by solving

$$\min_g \sum_{(x,y)\in\mathcal{P}} (g_x - g_y - \delta_{x,y})^2, \tag{3}$$

where the output image g could be with (as in Kim et al. 2009) or without a parametric form (as in Gooch et al. 2005). $\delta_{x,y}$ is the color contrast, having a signed value indicating the difference of the corresponding color pixels. In the CIELab color space, it can be expressed as

$$|\delta_{x,y}| = \sqrt{(L_x - L_y)^2 + (a_x - a_y)^2 + (b_x - b_y)^2}, \tag{4}$$

which represents the color dissimilarity in the human visual system (Wyszecki and Stiles 2000). The sign of $\delta_{x,y}$ is typically determined by the sign in the L channel, i.e. $\text{sign}(L_x - L_y)$. As discussed in Sect. 1, enforcing specific orders for some color pairs could cause the contrast-loss problem. Also, it may not be in obedience to human visual perception.

4.1 Weak Color Order

Some color pairs obviously can be ordered in terms of brightness. For example, absolute white is always brighter than other colors in common sense. For these color pairs, a single-peak distribution like the one in Eq. (3) is a natural choice, indicating an unambiguous prior. We treat color pairs that satisfy the following constraint as *unwavering*:

$$\mathbf{c}_x \leq \mathbf{c}_y \iff r_x \leq r_y \;\&\; g_x \leq g_y \;\&\; b_x \leq b_y. \tag{5}$$

If Eq. (5) is satisfied, the sign of $\delta_{x,y}$ applies to $g_x - g_y$. Otherwise, we do not specify the sign in prior but instead propose a selection procedure to optimally find the suitable color order. Our unambiguous color pairs keep the order of Helmholtz–Kohlrausch (H–K) effect Nayatani (1997) with a high probability. To verify this, we sample 10K unambiguous color pairs from 250 images in our new COLOR250 dataset. 99.8 % of them are in accordance with the H–K effect.

Combining the above two types of order definition, we can define our novel weak-order constraint. Practically, we construct a map to distinguish between these two cases, which is defined as

$$\alpha_{x,y} = \begin{cases} 1 \text{ if } & r_x \leq r_y, g_x \leq g_y, b_x \leq b_y \\ \quad \text{or} & r_x \geq r_y, g_x \geq g_y, b_x \geq b_y \\ 0.5 \text{ otherwise} & \end{cases} \tag{6}$$

If $\alpha_{x,y} = 1$, we apply prior $G(\delta_{x,y}, \sigma^2)$ for unambiguous color order enforcement. Otherwise, we let the color difference follow a bimodal distribution, which allows for selection of positive and negative signs optimally. There is one special case where $r_x = r_y$, $g_x = g_y$, and $b_x = b_y$. $\alpha_{x,y}$ can be set to any value, as $\delta_{x,y}$ is always zero. The final energy function for pixel pairs in $\mathcal{P}$ is written as

$$\prod_{(x,y)\in\mathcal{P}} \{\alpha_{x,y} G(\delta_{x,y}, \sigma^2) + (1 - \alpha_{x,y}) G(-\delta_{x,y}, \sigma^2)\}. \tag{7}$$

Maximizing Eq. (7) is equivalent to minimizing its negative logarithm, expressed as

$$E_{\mathcal{P}}(g) = -\sum_{(x,y)\ \in\mathcal{P}} \ln\left\{\alpha_{x,y} G(\delta_{x,y}, \sigma^2) + (1 - \alpha_{x,y}) G(-\delta_{x,y}, \sigma^2)\right\}. \tag{8}$$

After taking the parametric polynomial model described in Sect. 3 into Eq. (8), a function is formed consisting of unknown coefficients $\{\omega_i\}$. As the global non-linear mapping is used, only nine parameters need to be estimated.

The difference of two grayscale pixels can then be expressed with respect to unknown parameters $\{\omega_i\}$:

$$\begin{aligned} \Delta g_{x,y} &= g_x - g_y \\ &= f(r_x, g_x, b_x) - f(r_y, g_y, b_y) \\ &= \sum_i \omega_i (m_{ix} - m_{iy}). \end{aligned} \tag{9}$$

We further denote $l_{i(x,y)} := m_{ix} - m_{iy}$, which can be directly computed given the color of pixels x and y. The energy function w.r.t. the parameter set ω is written as

$$E_{\mathcal{P}}(\omega) = -\sum_{(x,y)\in\mathcal{P}} \kappa_{x,y} \cdot \ln\left\{\alpha_{x,y}\exp\left\{-\frac{|\sum_i \omega_i l_{i(x,y)} - \delta_{x,y}|^2}{2\sigma^2}\right\} + (1-\alpha_{x,y})\exp\left\{-\frac{|\sum_i \omega_i l_{i(x,y)} + \delta_{x,y}|^2}{2\sigma^2}\right\}\right\}, \tag{10}$$

where $\kappa_{x,y}$ defines the weight for pixel pairs. It was typically assumed to be constant for all pixel pairs. We will show later that the constant κ setting may not be optimal, especially when dominant color components exist.

The remaining problems in implementing Eq. (10) is to define the set of pairs $\mathcal{P}$ and to set the corresponding weights $\kappa_{x,y}$, which are presented in next section.

5 Local and Non-local Constraints

Ideally, $\mathcal{P}$ should include all Z^2 pixel pairs, where Z is the number of pixels in the image. This typically makes computation costly. In the conference version paper Lu et al. (2012), we set $\mathcal{P} = \mathcal{N}$, where $\mathcal{N}$ is the four-neighbor pair set. It preserves the local color change, but ignores contrast for pixels that are far apart. Here we introduce a new color contrast constraint accounting for both local and non-local color pairs without compromising computation efficiency.

5.1 Approximating Non-local Color Contrast

Evaluating all Z^2 pixel pairs takes much computation. We speed it up given the fact that Eq. (10) depends on the number of *different* color pairs presented in the image. Costs for different pixel pairs with the same colors can be grouped together.

Specifically, we denote by $\delta_{\mathbf{c}_1,\mathbf{c}_2}$ the contrast of a color pair $(\mathbf{c}_1, \mathbf{c}_2)$, $l_{i(\mathbf{c}_1,\mathbf{c}_2)} = m_{i\mathbf{c}_1} - m_{i\mathbf{c}_2}$ the color channel difference, $\alpha_{\mathbf{c}_x,\mathbf{c}_y}$ the per-color-pair coefficient, and by $K_{\mathbf{c}_1,\mathbf{c}_2}$ the overall weight for a color pair. Eq. (10) can be re-written as

$$E_{\mathcal{D}}(\omega) = -\sum_{(\mathbf{c}_1,\mathbf{c}_2)\in\mathcal{D}} K_{\mathbf{c}_1,\mathbf{c}_2} \cdot \ln\left\{\alpha_{\mathbf{c}_1,\mathbf{c}_2}\exp\left\{-\frac{|\sum_i \omega_i l_{i(\mathbf{c}_1,\mathbf{c}_2)} - \delta_{\mathbf{c}_1,\mathbf{c}_2}|^2}{2\sigma^2}\right\} + (1-\alpha_{\mathbf{c}_1,\mathbf{c}_2})\exp\left\{-\frac{|\sum_i \omega_i l_{i(\mathbf{c}_1,\mathbf{c}_2)} + \delta_{\mathbf{c}_1,\mathbf{c}_2}|^2}{2\sigma^2}\right\}\right\}, \tag{11}$$

of which the computation complexity depends only on the cardinality of the the color pair set $\mathcal{D}$ and is regardless of the number of pixel pairs in the image. Fortunately, due to the inherent smoothing property, natural color images are usually with a small number of *main* color components, instrumental in simplifying computation.

5.1.1 Landmark Color Extraction

The goal of landmark color computation is to extract the main color components to form the color pair set $\mathcal{D}$. Similar to (Ahn et al. 2010), for fast estimation, we apply K-means to cluster colors in the image. The main challenge, however, is to determine the number of clusters. We choose to seed landmark colors, and gradually increase them during clustering.

In the first place, we form the initial color set $\mathcal{C}$ as containing main colors in the image by removing those occupying less than 1 % of the total pixels. We perform K-means clustering with s centers on $\mathcal{C}$ and obtain a cluster center set $\mathcal{C}_s$. We then generate an approximation of $\mathcal{C}$ from centers $\mathcal{C}_n$, denoted as $\bar{\mathcal{C}}$, by substituting each color in $\mathcal{C}$ with its clustering center in $\mathcal{C}_s$, and test the difference between $\mathcal{C}$ and $\bar{\mathcal{C}}$. If it is large, we increase s and repeat this process (see Algorithm 1). The increasing factor of s is set to $\alpha = 2 - \exp\{-\|\mathcal{C} - \bar{\mathcal{C}}\|^2/\sigma_c\}$. It makes s increase fast in the beginning and slow down towards the end. For natural images, s is typically smaller than 100; for charts and forms, the number is much smaller. Our following method uses the estimated color in $\mathcal{C}_s$.

5.1.2 Weber–Fechner Weight

We discuss the role of the weights $K_{\mathbf{c}_1,\mathbf{c}_2}$ in Eq. (11). By definition, it should be the summation of weights $\kappa_{x,y}$ over all pixel pairs (x, y) that with color $(\mathbf{c}_1, \mathbf{c}_2)$. If all pixel pairs in the images are treated equally, $K_{\mathbf{c}_1,\mathbf{c}_2}$ should be $N_{\mathbf{c}_1} N_{\mathbf{c}_2}$, where $N_{\mathbf{c}_1}$ and $N_{\mathbf{c}_2}$ are the numbers of pixels that are with color $\mathbf{c}_1$ and $\mathbf{c}_2$ respectively. They make color contrast between large regions be best preserved to reduce energy. It however downplays the importance of color contrasts between small-size regions, as illustrated in Fig. 3, which may not appropriate. According to the Weber–Fechner law (Reber 1985), which explains that the relationship between real-world stimuli, such as weight, length, and size, and human perception follows a logarithmic correspondence. Intuitively, in our case, the importance of regions should not grow linearly with size. Figure 4 shows the total number of instances for the three color pairs in Fig. 3. Directly minimizing Eq. (10) with equal weights generates the result shown in Fig. 3b.

Algorithm 1 Color Set Approximation

```
input: initial color set $\mathcal{C}$
initialization: $s = 2$
repeat
    partition $\mathcal{C}$ into $s$ clusters using K-means, with $s$ centers stored in $\mathcal{C}_s$.
    reconstruct $\mathcal{C}$ using $\mathcal{C}_s$, denoted by $\bar{\mathcal{C}}$.
    $s \leftarrow \alpha \cdot s, \alpha > 1$
until $\|\mathcal{C} - \bar{\mathcal{C}}\|^2 \leq \varepsilon$
output: color set $\mathcal{C}_s$
```

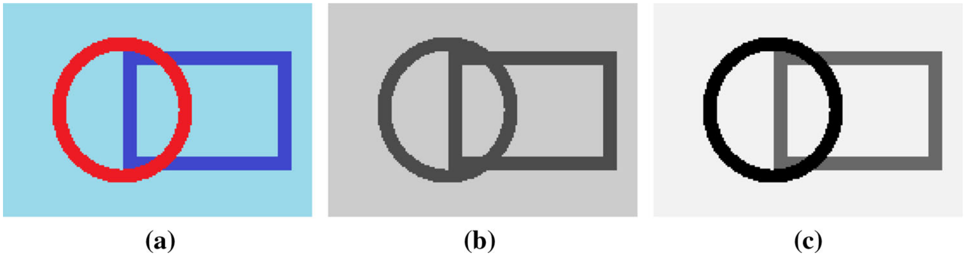


Fig. 3 Region size v.s. color contrast. **a** Input. **b** Result with equal weights for each pixel pair. **c** Result after applying the Weber–Fechner weights

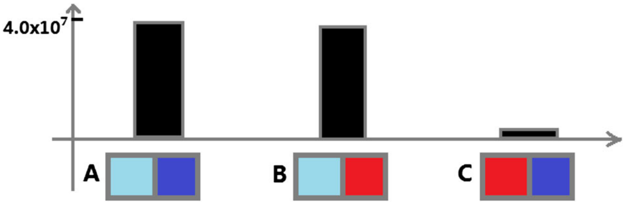


Fig. 4 Numbers of instances for the three color pairs in Fig. 3a

We update the weights of color pairs as

$$K_{\mathbf{c}_1,\mathbf{c}_2} = \log\left(\frac{N_{\mathbf{c}_1}N_{\mathbf{c}_2}}{N_0}\right), \tag{12}$$

to follow the Weber–Fechner law. N_0 is an empirical value defining the threshold. Color pairs with the instance number below it can be neglected.

If the instances of one color pair are less than $N_0 = 0.001^2(Z)^2$, we treat the pair as not important and set its weight $K_{\mathbf{c}_1,\mathbf{c}_1}$ to zero. The Weber–Fechner weight has an effect to attenuate excessively large N. This is equivalent to setting $\kappa_{x,y} = \frac{K_{\mathbf{c}_x,\mathbf{c}_y}}{N_{\mathbf{c}_x}N_{\mathbf{c}_y}}$ in Eq. (10), which reduces weights of pixel pairs with dominant colors. Better results can generally be accomplished, as exemplified in Fig. 3c.

By substituting the Weber–Fechner weight into Eq. (10) and employing the landmark color set $\mathcal{C}_s$, we obtain a new energy form w.r.t. parameter ω

$$\begin{aligned} E_{\mathcal{D}_s}(\omega) = -\sum_{(\mathbf{c}_1,\mathbf{c}_2)\in\mathcal{D}_s} K_{\mathbf{c}_1,\mathbf{c}_2} \cdot \\ \ln\left\{\alpha_{\mathbf{c}_1,\mathbf{c}_2}\exp\left\{-\frac{|\sum_i \omega_i l_{i(\mathbf{c}_1,\mathbf{c}_2)} - \delta_{\mathbf{c}_1,\mathbf{c}_2}|^2}{2\sigma^2}\right\} + \right. \\ \left. (1-\alpha_{\mathbf{c}_1,\mathbf{c}_2})\exp\left\{-\frac{|\sum_i \omega_i l_{i(\mathbf{c}_1,\mathbf{c}_2)} + \delta_{\mathbf{c}_1,\mathbf{c}_2}|^2}{2\sigma^2}\right\}\right\}, \end{aligned} \tag{13}$$

with $K_{\mathbf{c}_1,\mathbf{c}_2}$ defined in Eq. (12) and $\mathcal{D}_s = \{(\mathbf{c}_1,\mathbf{c}_2)|\mathbf{c}_1 \in \mathcal{C}_s \;\&\; \mathbf{c}_2 \in \mathcal{C}_s\}$, the set of color pairs drawn from $\mathcal{C}_s$.

5.2 Combining Local and Non-local Constraints

We call $E_{\mathcal{D}_s}(\omega)$ in Eq. (13) the non-local energy as pixels that are not neighboring are considered. However, for the sake of efficient computation, some color components have been dropped during clustering, which possibly lead to detail loss in practice, as shown in Fig. 6. To reduce this problem, we combine the non-local constraint $E_{\mathcal{D}_s}(\omega)$ with its local counterpart $E_{\mathcal{N}}(\omega)$ that involves only neighboring pixels in the set $\mathcal{N}$ to form the final objective

$$E(\omega) = E_{\mathcal{N}}(\omega) + \lambda E_{\mathcal{D}_s}(\omega), \tag{14}$$

where λ is a scalar balancing two terms. In defining $E_{\mathcal{N}}(\omega)$, we simply set $\kappa_{x,y}$ to 1 in Eq. (10). The efficacy of both the non-local and local constraints will be validated in Sect. 7.

6 Numerical Solution

Energy function (14) involves two parts representing respectively the local and non-local contrast constraints. Both constraints can be uniformly expressed in the form of Eq. (10), which is still non-convex in general. We in this section devise an efficient iterative numerical solver.

In Eq. (14), taking partial derivatives with respect to $\{\omega_j\}$ and setting them to zeros yield an equation system. To facilitate presentation, we define

$$\begin{aligned} \beta_{x,y} &:= \frac{\alpha_{x,y}G(\delta_{x,y},\sigma^2)}{\alpha_{x,y}G(\delta_{x,y},\sigma^2) + (1-\alpha_{x,y})G(-\delta_{x,y},\sigma^2)}, \\ \beta_{\mathbf{c}_1,\mathbf{c}_2} &:= \frac{\alpha_{\mathbf{c}_1,\mathbf{c}_2}G(\delta_{\mathbf{c}_1,\mathbf{c}_2},\sigma^2)}{\alpha_{\mathbf{c}_1,\mathbf{c}_2}G(\delta_{\mathbf{c}_1,\mathbf{c}_2},\sigma^2) + (1-\alpha_{\mathbf{c}_1,\mathbf{c}_2})G(-\delta_{\mathbf{c}_1,\mathbf{c}_2},\sigma^2)}. \end{aligned} \tag{15}$$

With a few algebraic operations, the partial derivative on ω_j, i.e. $\frac{\partial E(\omega)}{\partial \omega_j}$, can be expressed as

$$\begin{aligned} &\lambda \sum_{(\mathbf{c}_1,\mathbf{c}_2)\in\mathcal{D}_s} K_{\mathbf{c}_1,\mathbf{c}_2} \cdot \\ &\left(\sum_i \omega_i l_{i(\mathbf{c}_1,\mathbf{c}_2)} l_{j(\mathbf{c}_1,\mathbf{c}_2)} + (1-2\beta_{\mathbf{c}_1,\mathbf{c}_2}) l_{j(\mathbf{c}_1,\mathbf{c}_2)} \delta_{\mathbf{c}_1,\mathbf{c}_2}\right) \\ &+ \sum_{(x,y)\in\mathcal{N}} \left(\sum_i \omega_i l_{i(x,y)} l_{j(x,y)} + (1-2\beta_{x,y}) l_{j(x,y)} \delta_{x,y}\right) = 0. \end{aligned} \tag{16}$$

By setting all $\frac{\partial E(\omega)}{\partial \omega_j}$ to zeros, we obtain a total of 9 equations. The difficulty in solving it stems from the terms β,

Algorithm 2 Contrast-Preserving Decolorization

```
input: color image $\mathbf{c} = (r, g, b)$
initialize $\omega_i^0$, $t \leftarrow 0$
compute $\delta_{x,y}$ and $l_{i(x,y)}$ for each neighboring pixel pair in $\mathcal{N}$
compute $\delta_{\mathbf{c}_1,\mathbf{c}_2}$ and $l_{i(\mathbf{c}_1,\mathbf{c}_2)}$ for each color pair in $\mathcal{D}_s$
repeat
    compute $\beta^t_{x,y}$, $\beta^t_{\mathbf{c}_1,\mathbf{c}_2}$ given $\omega^t$
    solve for $\omega^{t+1}$
    $t \leftarrow t + 1$
until $t > t_{max}$
$g = f(\mathbf{c}; \omega^t)$
map $g$ back to the range $[\min(\mathbf{c}), \max(\mathbf{c})]$
output: grayscale image $g$.
```

which contain nonlinear functions about ω. We apply the fix-point iteration strategy on ω to linearize the corresponding equations. Specifically, to solve for ω_i^{t+1} in the $(t+1)$th iteration, we use the previously estimated ω_i^t to generate the nonlinear term β^t, which yields equations

$$\sum_i \omega_i^{t+1}\left(\sum_{(x,y)\in\mathcal{N}} l_{i(x,y)} l_{j(x,y)} + \lambda \sum_{(\mathbf{c}_1,\mathbf{c}_2)\in\mathcal{D}_s} K_{\mathbf{c}_1,\mathbf{c}_2} l_{i(\mathbf{c}_1,\mathbf{c}_2)} l_{j(\mathbf{c}_1,\mathbf{c}_2)}\right)$$
$$= \sum_{(x,y)\in\mathcal{N}} (2\beta^t_{x,y} - 1) l_{j(x,y)} \delta_{x,y}$$
$$+\lambda \sum_{(\mathbf{c}_1,\mathbf{c}_2)\in\mathcal{D}_s} K_{\mathbf{c}_1,\mathbf{c}_2} (2\beta^t_{\mathbf{c}_1,\mathbf{c}_2} - 1) l_{j(\mathbf{c}_1,\mathbf{c}_2)} \delta_{\mathbf{c}_1,\mathbf{c}_2}. \tag{17}$$

Now the unknowns only exist in the left hand side of Eq. (17) and we have 9 equations in the form of Eq. (17) by varying j. In each iteration, the system is linear w.r.t. $\{\omega_i^{t+1}\}$ and can be solved easily. In implementation, we use the backslash operator "\" in MATLAB.

Our computation framework is sketched in Algorithm 2. Figure 5 shows an example where the grayscale image is updated in iterations. We begin with a simple initialization $\{\omega_i^0\}$ as $\{0.33, 0.33, 0.33, 0, 0, 0, 0, 0, 0\}$. The corresponding 9 coefficients $\{\omega_i\}$ in iterations are listed in Table 2.

To make the resulting grayscale image g viewable, we linearly scale values with respect to the largest and smallest values $\max(\mathbf{c})$ and $\min(\mathbf{c})$ in the original color image. To process a one-megapixel color image, our MATLAB implementation spends 0.9s on a desktop PC equipped with an Intel i3 CPU and 4GB Memory.

7 Evaluation and Comparisons

We conduct extensive experiments to evaluate our method and compare it with other prior work using the benchmarking dataset (Cadík 2008) and an extra COLOR250 dataset containing much more examples. The parameters of the methods to compare are the same as those used in Cadík (2008) and are fixed for all images. For our method, σ is set to 0.2 and the weight $\lambda = 0.05\frac{|\mathcal{N}|}{\sum_{(\mathbf{c}_1,\mathbf{c}_2)\in\mathcal{D}_s} K_{\mathbf{c}_1,\mathbf{c}_2}}$, where $|\mathcal{N}|$ is the cardinality of the neighboring pixel pair sets used for normalization.

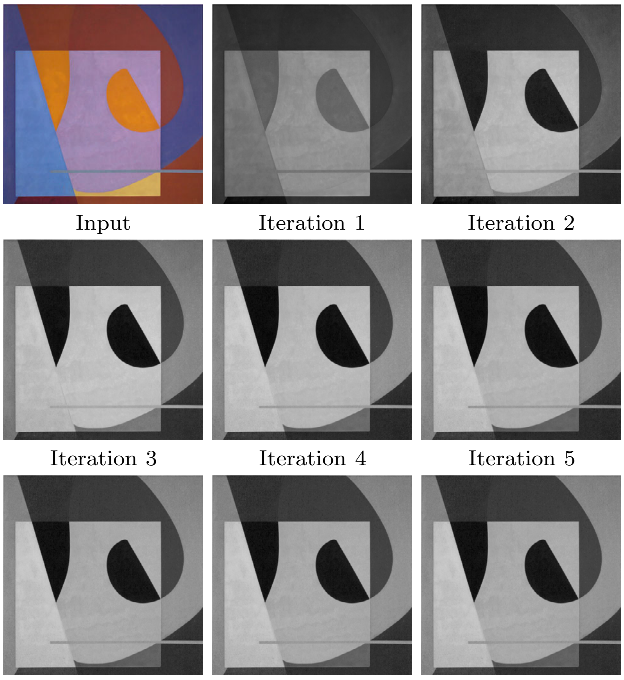

Input
Iteration 1
Iteration 2
Iteration 3
Iteration 4
Iteration 5
Iteration 13
Iteration 14
Iteration 15

Fig. 5 Results updated in iterations

7.1 Model Validation

As given in Eq. (14), our model consists of local and non-local terms for contrast preservation. Both of them are important and effective in handling different types of structures. We respectively evaluate them.

First, we experiment with the model including and without including the local contrast term $E_{\mathcal{N}}(\omega)$ in Eq. (14). A few results are shown in Fig. 6. The results shown in (b) are produced by optimizing $E_{\mathcal{D}_K}$ only. They are already visually compelling. But compared to (c), a bit less details are available, indicating the importance of also enforcing neighboring contrast. We note that our method is a *global mapping* one that projects one color to a grayscale, regardless the position in the image. The "local" and "non-local" constraints here are only to represent neighboring pixel colors and those set apart, in computing the global mapping.

Second, we compute results using only the local color constraint $E_{\mathcal{N}}(\omega)$ in Eq. (14), which corresponds to the method presented in our conference paper. Figure 7 shows the results. Compared to our prior model, the new one can capture the color difference for regions that are not adjacent to each other, as illustrated in (b) and (c). This property is particularly important for region distinctiveness representation and is vastly beneficial for charts, forms, clip-arts, and tables. The

Table 2 Computed coefficients ω in different iterations

Iter.	r	g	b	rg	rb	gb	r^2	g^2	b^2
1	0.33	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00
2	0.97	0.91	0.38	−3.71	2.46	−4.01	−0.42	4.00	0.79
3	1.14	−0.25	1.22	−1.55	1.53	−3.51	−1.18	3.32	0.69
4	1.33	−1.61	2.10	1.35	−0.36	−1.61	−1.69	1.70	0.29
5	1.52	−2.25	2.46	2.69	−1.38	−0.30	−1.95	0.79	−0.02
6	1.64	−2.59	2.65	3.50	−1.99	0.59	−2.13	0.18	−0.27
11	1.94	−3.21	2.98	5.61	−3.22	2.55	−2.80	−1.33	−0.87
12	1.96	−3.26	3.00	5.80	−3.31	2.70	−2.87	−1.46	−0.92
13	1.98	−3.29	3.02	5.94	−3.38	2.81	−2.91	−1.56	−0.96
14	1.99	−3.31	3.03	6.03	−3.42	2.89	−2.95	−1.62	−0.98
15	2.00	−3.32	3.04	6.10	−3.45	2.94	−2.98	−1.67	−1.00

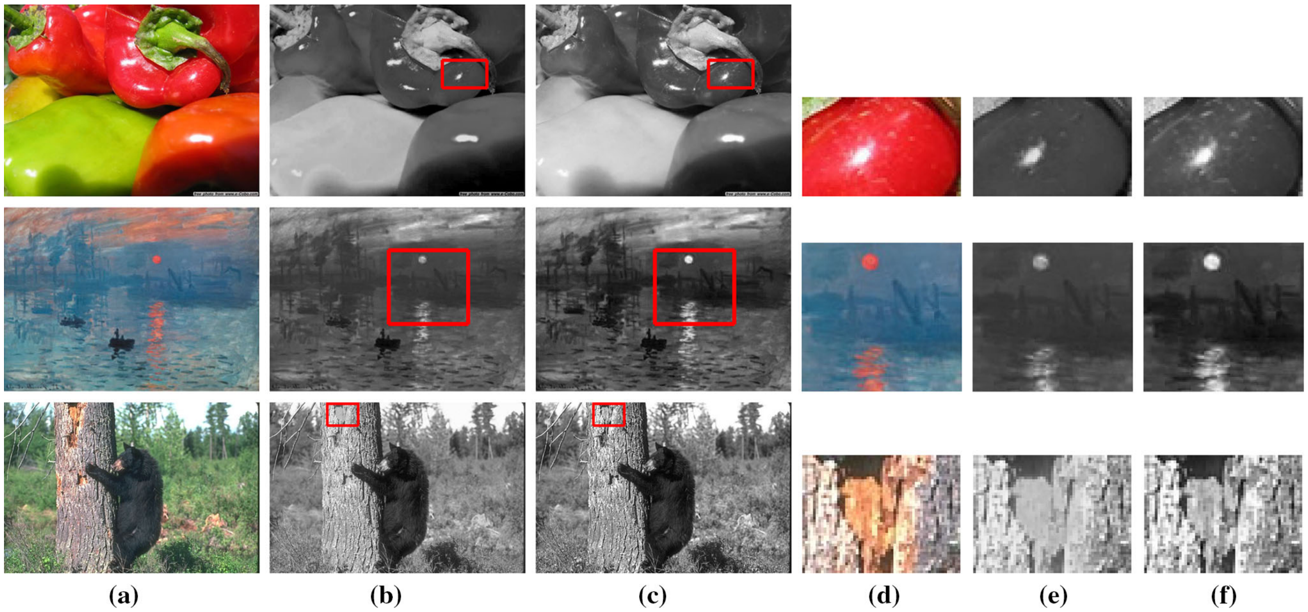


Fig. 6 Results produced with and without local contrast constraint. **a** Color inputs. **b** Results using the non-local contrast term only. **c** Results combining both local and non-local color contrast preserving terms. **d**–**f** Close-ups

first example in Fig. 7 also illustrates the importance of the *Weber–Fechner weight* in retaining contrast between small color regions. The three color curves contain only a small portion of pixels in the chart. Their contrast in our result, however, is not diminished even with the small region sizes.

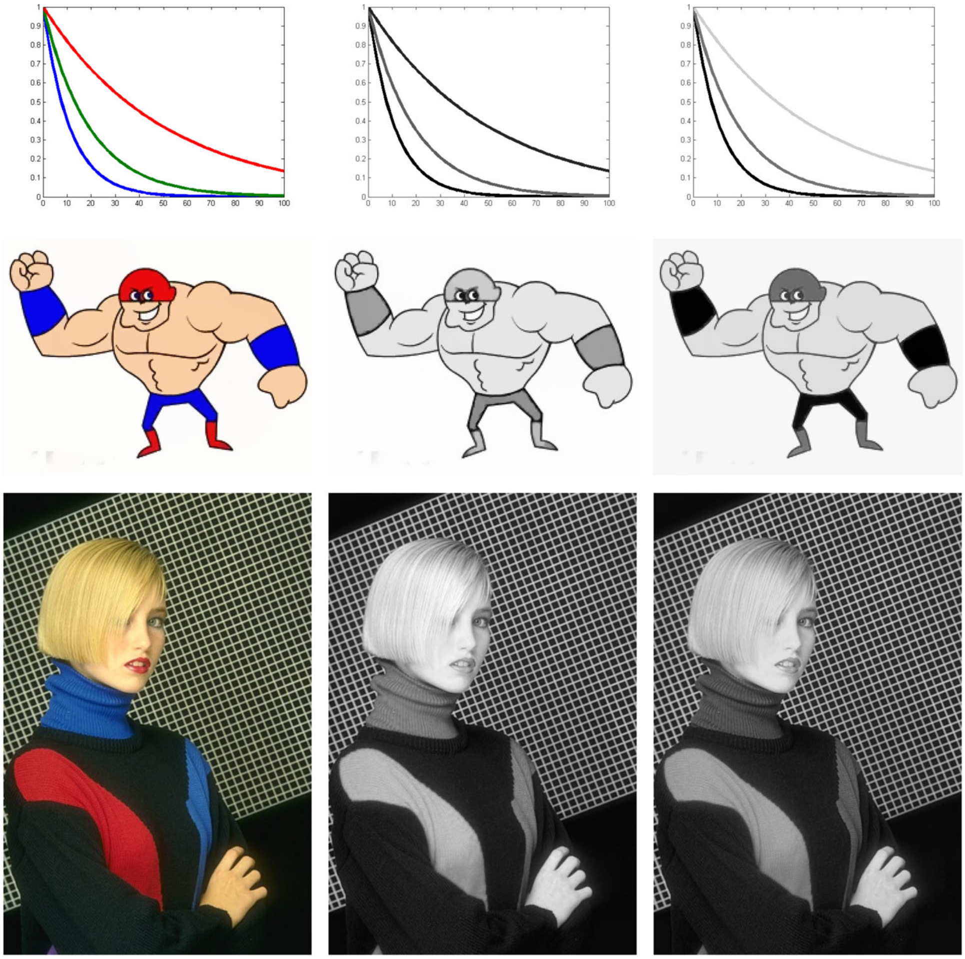
7.2 Qualitative Evaluation

We compare our method with those of (Gooch et al. 2005; Smith et al. 2008; Kim et al. 2009), which are representative local and global mapping color-to-gray approaches.

Figures 8 and 9 contain a few images selected from the benchmarking dataset (Cadík 2008). Our results, shown in the second column, preserve the most contrast presented in the input color images. For the images shown in the first, third and the fifth rows in Fig. 8 and second row in Fig. 9, our method produces results with very different color orders compared with others. It bears out the fact that in decolorization, color difference visualization for neighboring pixels is much more important in visual perception than strictly keeping the intensity order, which contrarily could be ambiguous by nature. To validate the bimodal function, we also show results produced by replacing the bimodal function by a single-model one in Fig. 10.

7.2.1 COLOR250 Dataset

The dataset produced in Cadík (2008) contains 24 images, most of which are synthesized. To validate our method on



(a) **(b)** **(c)**

Fig. 7 Effect of non-local contrast preserving. **a** Color inputs. **b** Results only using the local contrast constraint. **c** Our results produced from the new model

more natural images usable for different applications, we introduce a much larger dataset to test decolorization performance. It contains natural images, as well as digital charts, logos, and illustrations, which are ubiquitous in document printing.

There are two packages in our dataset. One is 200 nature images selected from the Berkeley Segmentation Dataset (BSD) (Martin et al. 2001) and saliency detection dataset (SD) (Achanta et al. 2009). Although these datasets are designed for other purposes, the contained images are natural with foreground and background. We manually select 100 most colorful images from BSD and SD separately and get a total of 200 natural images based on the fact that decolorization methods perform similarly on colorless or grayish images in general.

We also add five main types of chart images in our dataset, which include pie, curve, and bar charts, logos and maps. They are used most often in document printing. We show in Fig. 11 a few examples and compare our results with others. The difference is clear.

7.3 Quantitative Evaluation

Decolorization lacks quantitative evaluation for understanding the performance of different methods. We thus contribute perceptually motivated quantitative evaluation. Two basic metrics are proposed, measuring respectively the color contrast preserving and content similarity to the color input. They are motivated by the fundamental requirement of color-to-gray conversion to maintain as much as possible color change, meanwhile not generating new edges that do not find any correspondences in the original color images.

Our final measure combines the two metrics to form an E-score, similar to the widely used F-measure for “precision-recall” in the area of statistics. We perform quantitative evaluation using the user study data (Cadík 2008), which contain human labeled information from various subjects.

7.3.1 Color Contrast Preserving Ratio (CCPR)

One metric is based on the phenomenon of color perception—that is, when the Euclidean distance of two colors in the

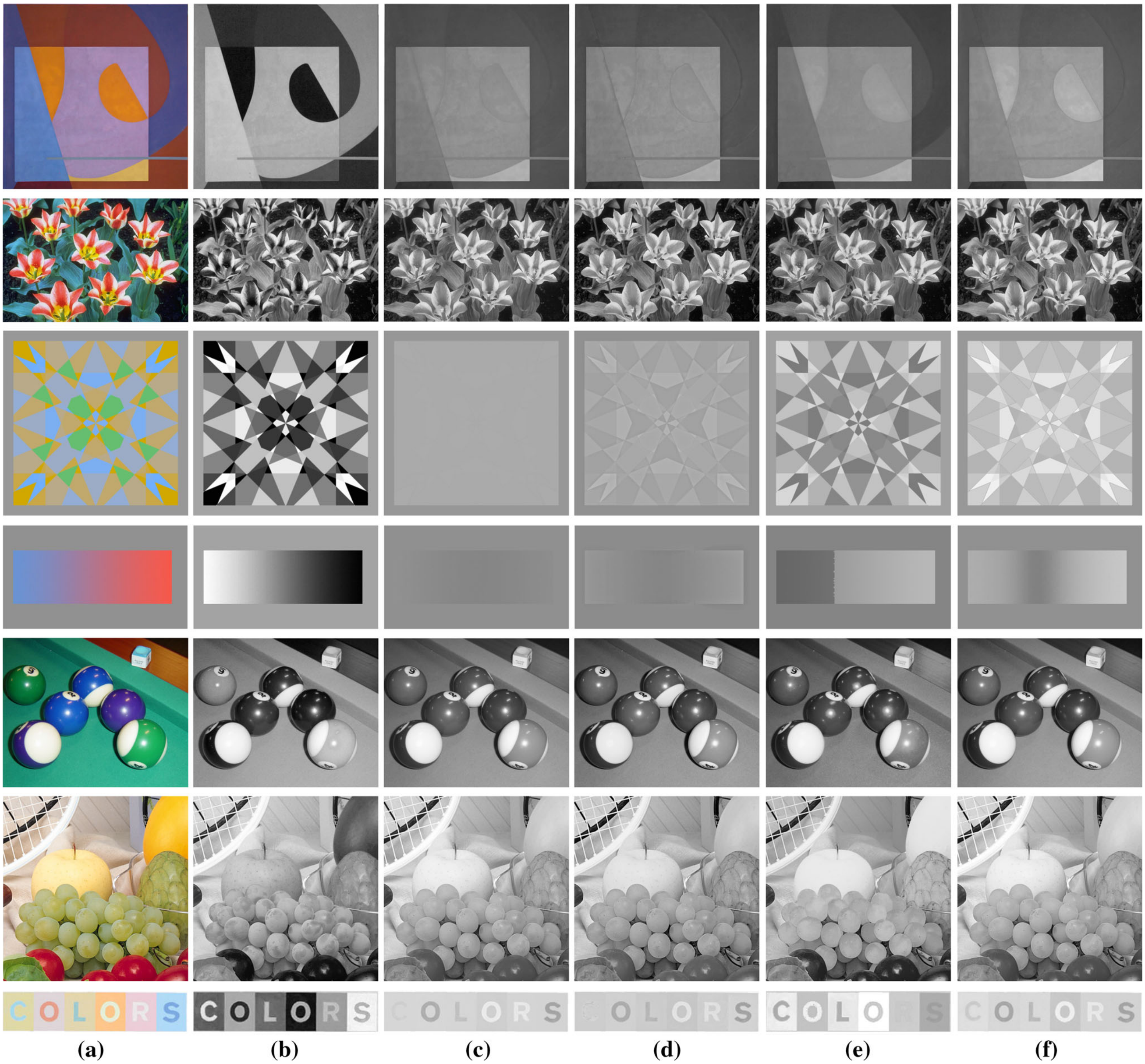


Fig. 8 Comparison with other decolorization methods

Lab space δ is smaller than a certain value τ, the difference is generally invisible in human visual system.[1] The task of contrast-preserving decolorization is therefore to only maintain color contrast that is perceivable by human. Considering the set containing all pixel pairs in the image $\mathcal{P}$, we define the color contrast preserving ratio (CCPR) as

$$\mathrm{CCPR} = \frac{\#\{(x, y)|(x, y) \in \Omega, |g_x - g_y| \geq \tau\}}{\|\Omega\|}, \tag{18}$$

where $\Omega \subseteq \mathcal{P}$ is the sub-pixel pair set with $\delta_{x,y} \geq \tau$. $\|\Omega\|$ denotes the number of pixel pairs in Ω. Here, $\#\{(x, y)|(x, y) \in \Omega, |g_x - g_y| \geq \tau\}$ is the number of pixel pairs in Ω that are still distinct after decolorization.

As $\mathcal{P}$ contains Z^2 pixel pairs. Exhaustive search for CCPR is time consuming. We resort to random sampling (following a uniform distribution) from set $\mathcal{P}$ to calculate CCPR. $10Z$ pairs are sampled in practice for the sake of efficiency. To increase robustness, we repeat evaluation for 15 times and

[1] It is suggested in Chen and Wang (2004) that $\tau < 6$ makes structures imperceivable to human visual system.

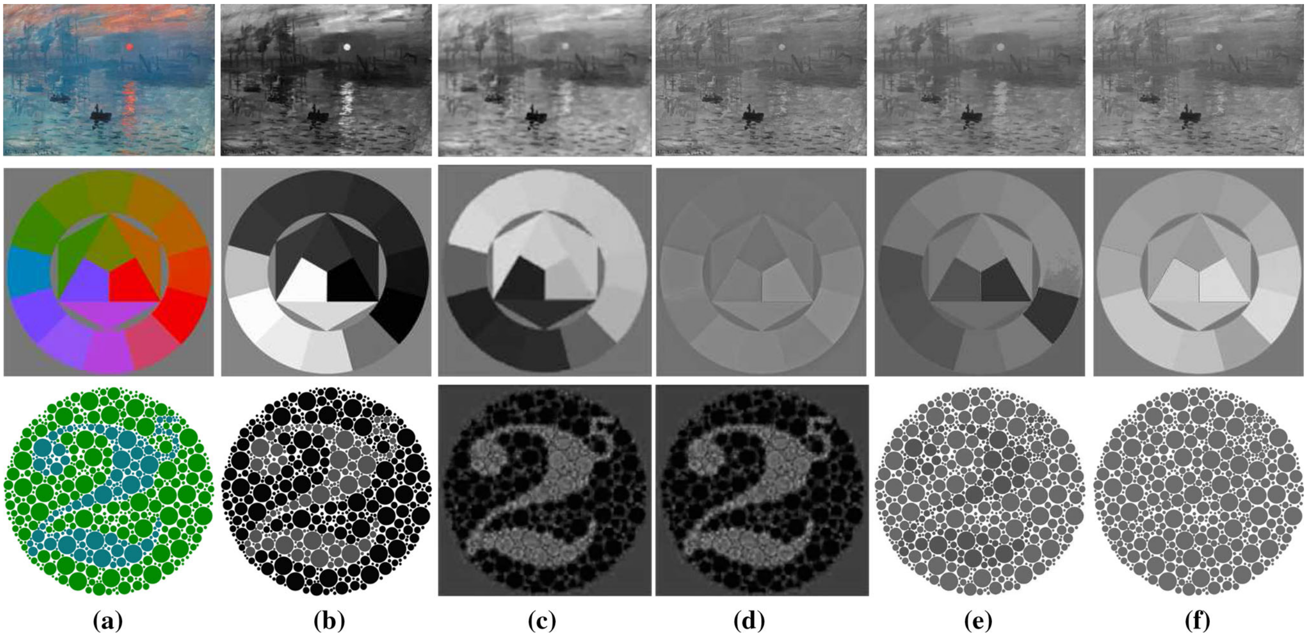


Fig. 9 Comparison with other decolorization methods

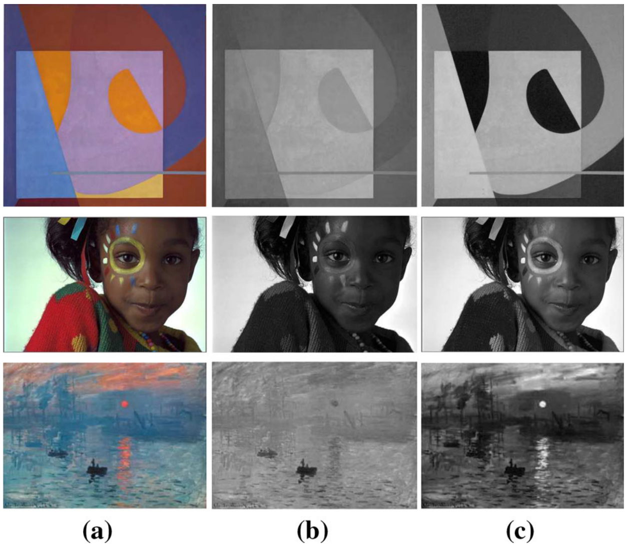


Fig. 10 Using bimodal and non-bimodal functions. **a** Inputs. **b** Results from the one-model function. **c** Our results

compute the mean CCPR, which should be close to the true value as illustrated in Table 3, where the variances of CCPR are small enough.

7.3.2 Color Content Fidelity Ratio (CCFR)

CCPR is effective to measure the contrast loss after color-to-gray conversion. It, however, does not tell how the grayscale image is faithful to the color input in terms of structures. A special case is that a noise map shown in Fig. 12b could have a large CCPR. To complement it, similar to the recall measure in statistics that measures whether the relevant instances are retrieved or not, we propose a color content fidelity ratio (CCFR) to incorporate this type of information. It is similarly important to quantitatively measure decolorization. The structure inconsistency presented in Fig. 13 can be detected and reflected in CCFR.

We define the CCFR metric complementary to CCPR as

$$\mathrm{CCFR} = 1 - \frac{\#\{(x, y)|(x, y) \in \Theta, \delta_{x,y} \le \tau\}}{\|\Theta\|}, \tag{19}$$

where Θ is the set containing pixel pairs with $|g_x - g_y| > \tau$, corresponding to structures with the least contrast. If the original pixel difference is small, i.e., $\delta_{x,y} \le \tau$, the ratio $\#\{(x, y)|(x, y) \in \Theta, \delta_{x,y} \le \tau\}/\|\Theta\|$ measures the occurrence of unwanted "artifacts" in the result.

7.3.3 E-score

We introduce the E-score to jointly consider CCPR and CCFR. It is the harmonic mean of the two measures, similar to the F-measure in statistics. It is written as

$$\mathrm{E{-}score} = \frac{2 \cdot \mathrm{CCPR} \cdot \mathrm{CCFR}}{\mathrm{CCPR} + \mathrm{CCFR}}. \tag{20}$$

Figure 12 lists the CCPR, CCFR and E-score beneath each result generated by state-of-the-arts. It is noticeable that the E-score basically complies with human perception. Ideally, the highest E-score is achieved when all the color contrast

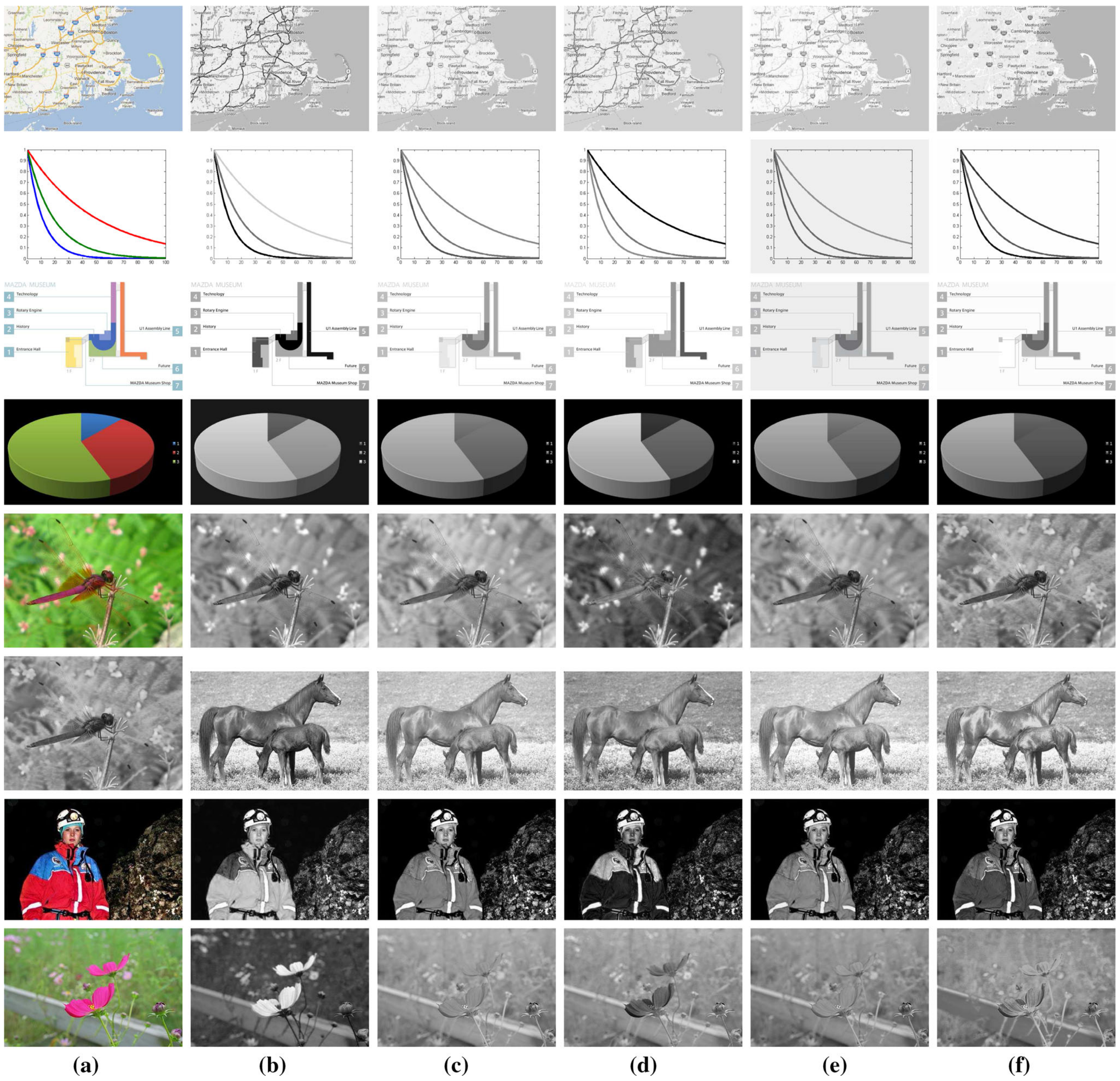


Fig. 11 Comparison with other decolorization methods on COLOR250

Table 3 Verification of the sampling strategy to compute CCPR

Image no.	1	2	3	4	5	6	7	8	9	10
CCPR (all pairs)	0.8900	0.9868	0.8618	0.9174	0.9373	0.9298	0.9156	0.9387	0.8322	0.8664
CCPR (sampling pairs)	0.8902	0.9868	0.8618	0.9178	0.9375	0.9297	0.9153	0.9388	0.8321	0.8665
Variance (sampling pairs)$\times 10^{-4}$	6.0938	1.5914	5.6992	4.8793	5.7371	2.1276	12.6871	9.0362	6.0970	4.8001

The first ten images in dataset Cadík (2008) are used. “CCPR (all pairs)” refers to the ground truth CCPR computed using all pixel pairs. “CCPR (sampling pairs)” is the mean CCPR of 15 rounds of random sampling. The last row shows the variance of the 15 runs

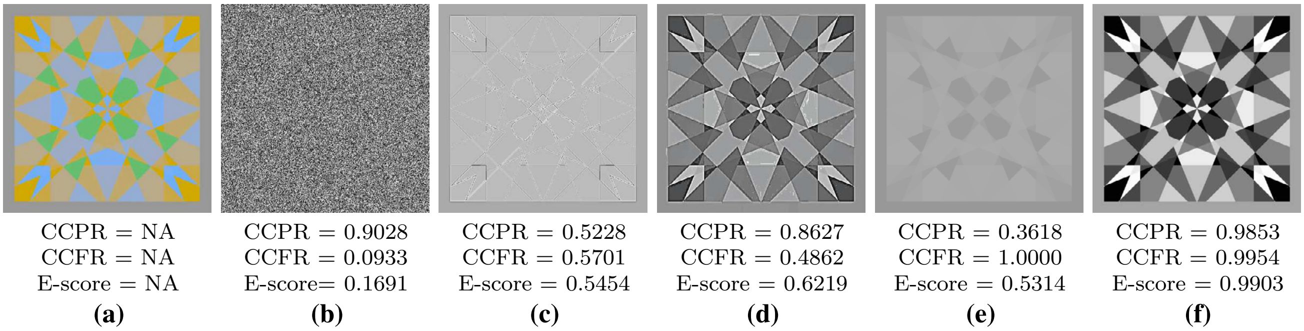


Fig. 12 CCPR, CCFR, and E-score with $\tau = 5$ for different grayscale results. **a** Input image. **b** A noise map. **c**–**d** Results of Bala and Eschbach (2004) and Smith et al. (2008) respectively. **e** Output from the MATLAB *rgb2gray* function. **f** Our result

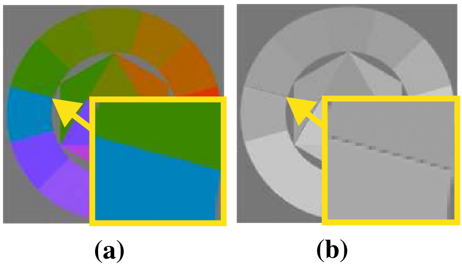


Fig. 13 Inconsistent edges may be generated in the grayscale image (**b**) given the input (**a**)

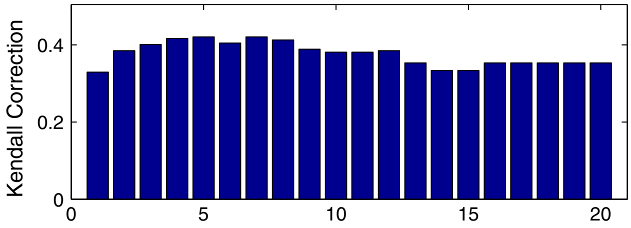


Fig. 14 Kendall correction w.r.t. τ

is preserved while no new edge is created in the grayscale result, which make both the CCPR and CCFR equal to one.

7.3.4 Perceptual Validation

Cadík (2008) evaluated seven main color-to-gray methods on 168 images. The preference scores are obtained from 119 subjects with 20,328 responses, which can be used also to validate the usefulness and correctness of the E-score. Ideally, if a preference score is high, the corresponding E-score should be large. Given several results that are produced by different methods, we rank them respectively according to the E-score and preference scores. The two sets of ranks give a correlation to measure if similar ranks are obtained. To statistically estimate the similarly, we adopt the Kendall rank correlation coefficient (Nelsen 2001).

Let e_i be the E-score for the result produced by the ith method and p_i be the preference score for the same result. If two pairs (e_i, e_j) and (p_i, p_j) are with the same order (i.e., $(e_i - e_j)(p_i - p_j) > 0$), the pair (i, j) is concordant. Otherwise, it is discordant. Kendall rank correction coefficient is accordingly defined as

$$R = \frac{\#\{\text{concordant pair}\} - \#\{\text{discordant pair}\}}{\frac{1}{2}n(n-1)}, \tag{21}$$

where n denotes the number of results to rank for each example. Because we evaluate seven methods, including CIELab and the methods of (Smith et al. 2008; Gooch et al. 2005; Bala and Eschbach 2004; Neumann et al. 2007; Rasche et al. 2005; Grundland and Dodgson 2007), n is set to 7 in our case. $\#\{\cdot\}$ is the operator to return the set size. R ranges in $[-1, 1]$. If two rankings primarily agree with each other, we get $R > 0$ ($R = 1$ refers to the situation that the orders are exactly the same). Otherwise, $R < 0$. When $R = 0$, the ranks can be treated as independent from the statistics perspective.

We plot the Kendall correction coefficients for the E-score and preference score ranks under different τ in Fig. 14. The coefficient R is positive for all τ we tested, indicating that the ranks are mostly similar. Empirically, $R > 0.4$ means that the two sets of ranks are positively related with sufficient significance.

In the experiments described in Cadík (2008), there is a u score, which follows the Thurstone law to describe the subjects' agreement degree for each case. When the decolorization results have an obvious quality order, u is high. We show two examples in Figs. 15 and 16. They are with large and small u scores respectively. They indicate when human subjects cannot agree on the result orders, the preference scores are not quite meaningful.

Thus for the example in Fig. 15, human can easily know which result looks good. Our E-scores are close to the preference scores. For the one in Fig. 16, subjects tend to chose their preference in a nearly random order. It explains why the Kendall correction is low for this example. There is basically no correct order suggested by human subjects.

By selecting images from the 24 image dataset with the highest u scores, we compute Kendall correction. Scores with

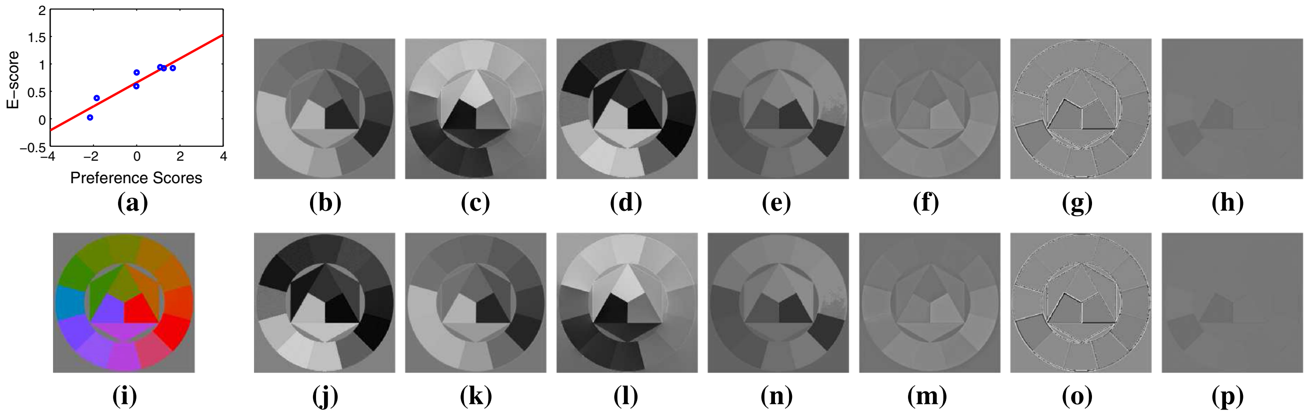


Fig. 15 Example with a large subject agreement score ($u = 0.5670$). **a** Our E-scores versus preference scores. They are primarily corresponding. **i** Input image. **b–h** Results of 7 methods shown according to the preference scores in an ascending order. **j–p** Results of 7 methods shown according to E-scores in an ascending order

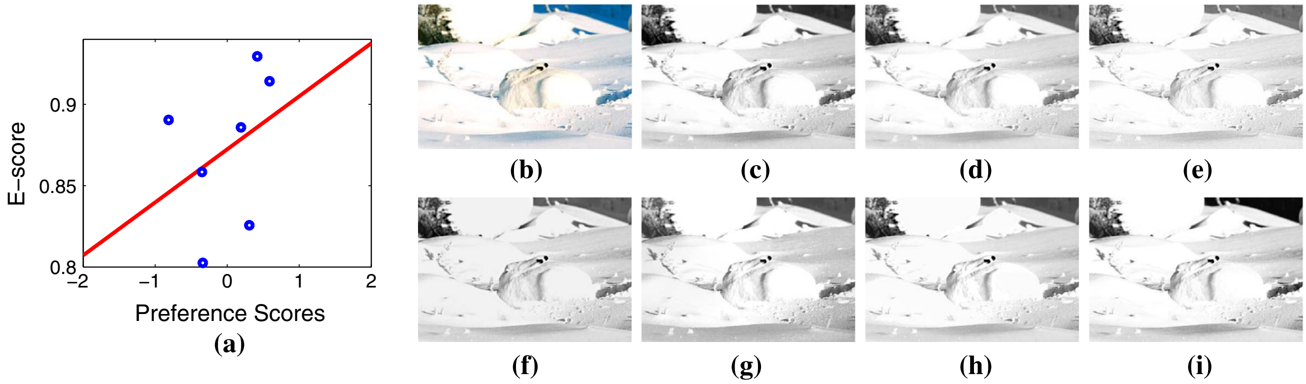


Fig. 16 Example with a small subject agreement score ($u = 0.103$). **a** E-score versus preference scores. They are not consistent for this example. **b** Input image. **c–i** Results of seven methods shown according to the preference scores in an ascending order

different image subsets that contain 4–24 images are listed in Table 4. We also evaluate consistency of E-scores and accuracy scores provided by Cadík (2008) with the same setting. The results are listed in Table 5. The quantities are large for subsets containing a few images with the highest u, manifesting similar human preference for these examples. E-scores are consistent with both the accuracy and preference scores. This validates the usefulness of the E-score measure in decolorization.

7.3.5 Evaluation based on CCPR, CCFR, and E-score

We quantitatively evaluate our results based on CCPR, CCFR and the E-score on the 24-image dataset (Cadík 2008) and COLOR250 dataset.

We plot the scores on the 24 images in Fig. 17. It implies that content-independent decolorization, such CIE Y channel, in general yields low CCPR. A few local approaches, on the other hand, could generate unwanted artifacts or overly enhance color edges, which lead to low CCFR. It is noteworthy that a higher CCPR or CCFR does not necessarily correspond to a better result. Only the harmonic mean of them, i.e. the E-score, determines the final quality. CCPR, CCFR, and E-score with $\tau = 5$ for different grayscale results of Fig. 8 are listed in Table 6.

We show in Fig. 18 the score plots on the COLOR250 set. The statistics also indicate that our method can achieve satisfactory decolorization performance in terms of the E-score measure.

7.3.6 User Study

A user study is carried out using the metrics of Cadík (2008). We invited 57 observers (30 males and 27 females) to join our user study. Following the setting of (Cadík 2008), we compute the accuracy and preference scores. Twenty-nine participants performed the accuracy experiment and 28 subjects took part in the preference experiment. We randomly select 30 images from our COLOR250 dataset as testing data. Six methods with codes or executable available online are evaluated. They are CIE Y, Grundland and Dodgson (2007), Smith et al. (2008), Gooch et al. (2005), Lu et al. (2012), and

Table 4 Average Kendall correction (with preference scores) for examples with the highest u scores

#	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4
$\tau = 4$	0.42	0.42	0.42	0.43	0.43	0.44	0.44	0.46	0.46	0.49	0.48	0.50	0.50	0.52	0.54	0.59	0.63	0.66	0.63	0.66	0.69
$\tau = 5$	0.42	0.43	0.43	0.44	0.43	0.44	0.44	0.46	0.46	0.48	0.48	0.49	0.50	0.52	0.55	0.61	0.65	0.67	0.65	0.68	0.71
$\tau = 6$	0.40	0.41	0.41	0.42	0.41	0.42	0.42	0.43	0.43	0.44	0.44	0.46	0.45	0.47	0.50	0.57	0.62	0.63	0.60	0.60	0.64
$\tau = 7$	0.42	0.43	0.43	0.44	0.44	0.44	0.44	0.46	0.46	0.47	0.47	0.49	0.48	0.51	0.51	0.59	0.64	0.65	0.66	0.67	0.67
$\tau = 8$	0.41	0.43	0.43	0.43	0.43	0.43	0.43	0.45	0.45	0.47	0.48	0.49	0.50	0.53	0.54	0.61	0.67	0.69	0.67	0.68	0.69
$\tau = 9$	0.40	0.40	0.40	0.41	0.40	0.40	0.41	0.43	0.43	0.46	0.47	0.49	0.49	0.54	0.55	0.62	0.68	0.70	0.68	0.70	0.70

The number of images # varies from 4 to 24. We set $\tau \in \{4, 5, 6, 7, 8, 9\}$ for robustness

Table 5 Average Kendall correction (with accuracy scores) for examples with the highest u scores

#	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4
$\tau = 4$	0.40	0.40	0.41	0.42	0.42	0.42	0.43	0.44	0.45	0.46	0.47	0.49	0.49	0.51	0.52	0.58	0.61	0.63	0.64	0.65	0.69
$\tau = 5$	0.41	0.42	0.42	0.42	0.42	0.42	0.42	0.45	0.45	0.45	0.46	0.48	0.48	0.50	0.54	0.58	0.62	0.65	0.65	0.65	0.69
$\tau = 6$	0.38	0.40	0.40	0.40	0.40	0.41	0.41	0.42	0.42	0.42	0.42	0.45	0.45	0.46	0.48	0.56	0.61	0.62	0.62	0.62	0.62
$\tau = 7$	0.40	0.40	0.40	0.41	0.43	0.43	0.43	0.43	0.45	0.46	0.47	0.47	0.48	0.48	0.49	0.57	0.61	0.64	0.64	0.64	0.66
$\tau = 8$	0.38	0.40	0.42	0.42	0.42	0.42	0.43	0.43	0.43	0.45	0.46	0.46	0.48	0.52	0.53	0.59	0.65	0.66	0.66	0.66	0.67
$\tau = 9$	0.38	0.39	0.39	0.39	0.39	0.39	0.40	0.41	0.42	0.44	0.44	0.46	0.47	0.53	0.53	0.59	0.66	0.69	0.69	0.69	0.70

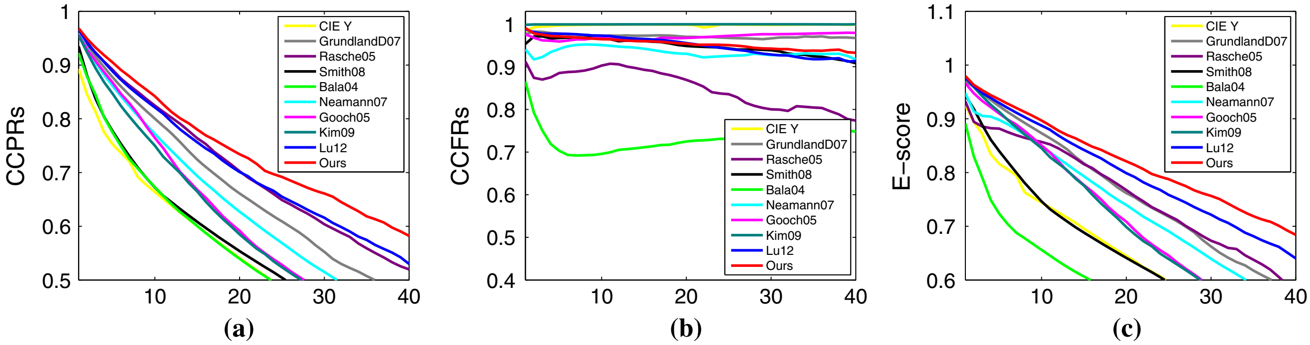


Fig. 17 CCPR, CCFR, and E-score plots on the 24-image dataset (x-axis: τ value; y-axis: CCPR, CCFR, and E-score)

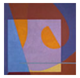
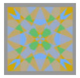
the proposed method. We required each observer to perform 180 trials (pair-wise comparison), which is an acceptable quantity for one observer without experiencing exhaustion and loss of concentration. A total of 57 observers completed 10260 human responses. The overall accuracy and preference scores are listed in Table 7. Our accuracy and preference scores are large, proving that our grayscale result can represent color inputs nicely. For our new E-score metric, we show its kendall correction with accuracy and preference scores in Table 8. They indicate that ranks of E-score and user labels are positively related with sufficient significance.

8 Concluding Remarks

We have presented a new decolorization method that can well maintain the original color contrast. We leverage a weak color constraint to allow for very flexible and optimal grayscale representation, based on the fact that human perception has limited ability to determine ordering of color with respect to brightness. So rather than intuitively defining the sign of grayscale difference, we propose a bimodal objective function to increase the search space in optimization. This strategy enables automatically finding suitable grayscales to best preserve significant color change. Our contribution also includes incorporating local and non-color color pairs for contrast maintenance and a new COLOR250 image dataset for decolorization evaluation.

It is also notable that we have proposed a well-motivated E-score metric to quantitatively evaluate the decolorization results, in compliance with human perception. Both the quantitative and qualitative experiments on the two datasets validate the effectiveness of the proposed method.

Table 6 CCPR, CCFR, and E-score with $\tau = 5$ for different grayscale results of Fig. 8

Images		CIE Y	Grundland and Dodgson (2007)	Rasche et al. (2005)	Smith et al. (2008)	Bala and Braun (2004)	Neumann et al. (2007)	Gooch et al. (2005)	Kim et al. (2009)	Ours
	CCPR	0.743	0.853	0.916	0.741	0.769	0.835	0.822	0.783	**0.922**
	CCFR	0.998	0.990	0.965	0.977	0.452	0.957	0.994	**0.999**	0.97
	E-score	0.852	0.916	0.94	0.843	0.570	0.892	0.900	0.878	**0.945**
	CCPR	0.892	0.875	0.912	0.899	0.894	0.872	0.891	0.897	**0.935**
	CCFR	0.997	0.986	0.970	0.960	0.763	0.978	0.982	**0.999**	0.961
	E-score	0.941	0.927	0.940	0.928	0.824	0.922	0.934	0.945	**0.948**
	CCPR	0.364	0.692	0.681	0.439	0.522	0.850	0.875	0.767	**0.985**
	CCFR	**1.000**	0.999	0.838	0.993	0.570	0.997	1.000	1.000	0.995
	E-score	0.533	0.818	0.752	0.609	0.545	0.918	0.933	0.868	**0.99**
	CCPR	0.000	0.813	0.879	0.110	0.083	0.879	0.908	0.744	**0.933**
	CCFR	1.000	1.000	0.373	1.000	0.433	0.753	0.908	1.000	**1.000**
	E-score	0.000	0.897	0.524	0.198	0.139	0.811	0.908	0.853	**0.965**
	CCPR	0.890	**0.910**	0.904	0.891	0.883	0.860	0.884	0.901	0.908
	CCFR	0.994	0.978	0.921	0.913	0.474	0.973	0.759	**0.998**	0.997
	E-score	0.939	0.943	0.912	0.902	0.617	0.913	0.817	0.947	**0.951**
	CCPR	0.842	0.852	**0.886**	0.834	0.813	0.793	0.815	0.841	0.866
	CCFR	0.995	0.954	0.894	0.959	0.657	0.989	0.861	**0.998**	0.925
	E-score	0.912	0.900	0.890	0.893	0.727	0.880	0.838	**0.913**	0.894
	CCPR	0.331	0.752	0.790	0.333	0.467	0.814	0.806	0.452	**0.926**
	CCFR	**1.000**	0.997	0.722	0.984	0.638	0.772	0.982	1.000	0.869
	E-score	0.497	0.857	0.754	0.498	0.539	0.792	0.885	0.622	**0.897**

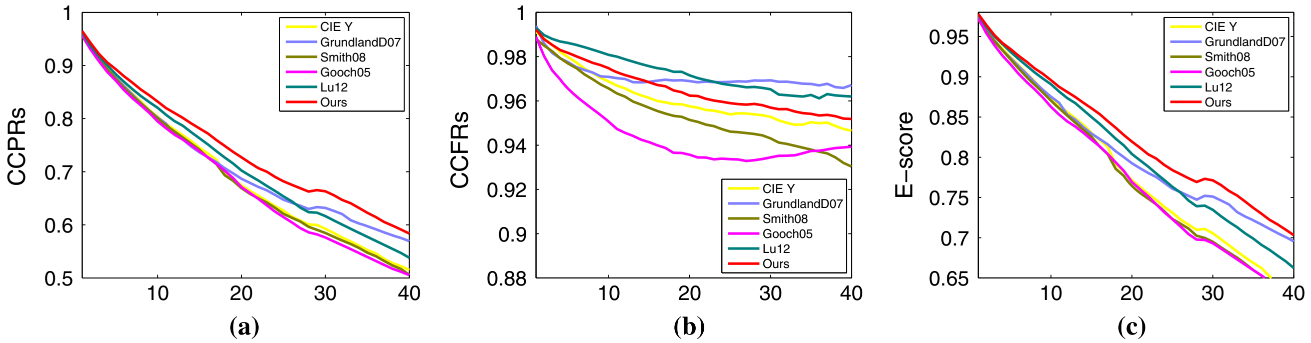


Fig. 18 CCPR, CCFR, and E-score plots on the COLOR250 dataset (x-axis: τ value; y-axis: CCPR, CCFR, and E-score)

Table 7 Overall accuracy and preference scores.

	CIE Y	Gooch et al. (2005)	Smith et al. (2008)	Grundland and Dodgson (2007)	Lu et al. (2012)	ours
Accuracy	−0.87	−0.27	−0.19	0.26	0.49	0.58
Preference	−0.79	−0.35	−0.17	0.40	0.44	0.46

Table 8 Average Kendall correction

τ	4	5	6	7	8	9
Accuracy	0.69	0.69	0.70	0.71	0.71	0.71
Preference	0.67	0.67	0.67	0.68	0.69	0.69

First row: average Kendall correction between accuracy score and E-score. Second row: average Kendall correction between preference score and E-score. We set $\tau \in \{4, 5, 6, 7, 8, 9\}$ to evaluate the robustness

Acknowledgments The authors would like to thank the editor and all the anonymous reviewers for their time and effort. This work is supported by a grant from the Research Grants Council of the Hong Kong SAR (Project No. 413110) and by NSF of China (key Project No. 61133009).

References

Achanta, R., Hemami, S. S., Estrada, F. J., & Susstrunk, S. (2009). Frequency-tuned salient region detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*

Ahn, J. H., Kuk, J. G., & Cho, N. I. (2010). A color to grayscale conversion considering local and global contrast. In *Asian Conference on Computer Vision (ACCV)*.

Ancuti, C. O., Ancuti, C., & Bekaert, P. (2011). Enhancing by saliency-guided decolorization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*

Bala, R., & Braun, K. (2004). Color-to-grayscale conversion to maintain discriminability. In *Proceedings of SPIE*, pp. 196–202.

Bala, R., & Eschbach, R. (2004). Spatial color-to-grayscale transform preserving chrominance edge information. In *Color Imaging Conference*.

Cadík, M. (2008). Perceptual evaluation of color-to-grayscale image conversions. *Computer Graphics Forum*, pp. 1745–1754.

Chen, H. C. & Wang, S. J. (2004). The use of visible color difference in the quantitative evaluation of color image segmentation. In *International conference on acoustics, speech, and signal processing (ICASSP)*, vol. 3, pp. 593–596.

Corney, D., Haynes, J. D., Rees, G., & Lotto, R. B. (2009). The brightness of colour. *PLoS ONE*, *4*(3), e5091.

Fairchild, M. D. (2005). *Color appearance models*. Chicheste: Wiley.

Gooch, Amy Ashurst, Olsen, Sven C., Tumblin, Jack, & Gooch, Bruce. (2005). Color2gray: Salience-preserving color removal. *ACM Transactions on Graphics (TOG)*, *24*(3), 634–639.

Grundland, Mark, & Dodgson, Neil A. (2007). Decolorize: Fast, contrast enhancing, color to grayscale conversion. *Pattern Recognition*, *40*(11), 2891–2896.

Hunter, R. S. (1958). Photoelectric color difference meter. *Journal of the Optical Society of America*, *48*(12), 985–993.

Kim, Y., Jang, C., Demouth, J., & Lee, S. (2009). Robust color-to-gray via nonlinear global mapping. *ACM Transactions on Graphics (TOG)*, *28* (5)

Lotto, R. B., & Purves, D. (2002). A rationale for the structure of color space. *Trends in Neurosciences*, *25*(2), 84–89.

Lu, C., Xu, L., & Jia, J. (2012). Contrast preserving decolorization. In *International Conference on Computational Photography (ICCP)*.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *International Conference on Computer Vision (ICCV)*.

Nayatani, Y. (1997). Simple estimation methods for the helmholtz kohlrausch effect. *Color Research and Application*, *24*, 385–401.

Nelsen, R. B. (2001). *Kendall tau metric*. Berlin: Springer.

Neumann, L., Cadík, M.,& Nemcsics, A. (2007). An efficient perception-based adaptive color to gray transformation. *Computational Aesthetics*

Ozgen, E. (2004). Language, learning, and color perception. *Current Directions in Psychological Science*, *13*(3), 95–98.

Rasche, K., Geist, R., & Westall, J. (2005). Detail preserving reproduction of color images for monochromats and dichromats. *IEEE Computer Graphics and Applications*, pp. 22–30.

Reber, A. S. (1985). *The Penguin dictionary of psychology*. London: Penguin Books.

Sharma, G., & Bala, R. (2002). *Digital Color Imaging Handbook*. Boca Raton: CRC Press.

Smith, K., Landes, P. E., Thollot, J., & Myszkowski, K. (2008). Apparent greyscale: A simple and fast conversion to perceptually accurate images and video. *Computer Graphics Forum*, *27*(2), 193–200.

Song, M., Tao, D., Chen, C., Li, X., & Chen, C. W. (2010). Color to gray: Visual cue preservation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, *32*(9), 1537–1552.

Wong, B. (2010). Points of view: Color coding. *Nature Methods*, *7*(8), 573.

Wyszecki, G., & Stiles, W. S. (2000). *Color science: Concepts and methods. Quantitative data and formulas*. New York: Wiley-Interscience.

Zhou, K., Mo, L., Kay, P., Kwok, V., Ip, T. N., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. In *Proceedings of the National Academy of Sciences*.