Saliency-Guided Color-to-Gray Conversion using Region-based Optimization

Hao Du, Shengfeng He, Bin Sheng, Lizhuang Ma, and Rynson W.H. Lau

***Abstract*—Image decolorization is a fundamental problem for many real world applications, including monochrome printing and photograph rendering. In this paper, we propose a new color-to-gray conversion method that is based on a region-based saliency model. First, we construct a parametric color-to-gray mapping function based on global color information as well as local contrast. Second, we propose a region-based saliency model that computes visual contrast among pixel regions. Third, we minimize the salience difference between the original color image and the output grayscale image in order to preserve contrast discrimination. To evaluate the performance of the proposed method in preserving contrast in complex scenarios, we have constructed a new decolorization dataset with 22 images, each of which contains abundant colors and patterns. Extensive experimental evaluations on the existing and the new datasets show that the proposed method outperforms the state-of-the-art methods quantitatively and qualitatively.**

***Index Terms*—Color-to-gray conversion • Saliency-preserving optimization • Region-based contrast enhancement • Dimensionality reduction**

I. Introduction

COLOR-TO-GRAY conversion is widely used in various applications, like monochrome printing, single channel image processing, and stylization. The conventional decolorization method, which directly uses the luminance channel or a weighted sum of the red, green and blue values, fails to preserve the details and features of the original colorful pictures, particularly in isoluminant regions. As a result, advanced algorithms try to capture more details during the conversion.

Decolorization can be regarded as a dimensionality reduction problem that maps a three-dimensional color space to a one-dimensional one. As a result, information loss is inevitable. Thus, the problem of decolorization can be restated as selecting and preserving the discriminative information in the original image. Many decolorization methods have been proposed to address this problem from a human perceptual viewpoint. Most of them aim at constructing linear or non-linear mapping functions to find a criterion for optimizing the decolorization process. For example, some state-of-the-art methods, e.g., [15, 16, 24], propose to preserve contrast in a global manner. However, this strategy may not be able to capture all the details. Some of the locally distinctive colors are often neglected, as they are not globally distinctive. In addition, globally one-to-one color mapping is not sufficient to capture all the details with a single dimension, if the original image contains lots of colors, as shown in the first row of Fig. 1(b).

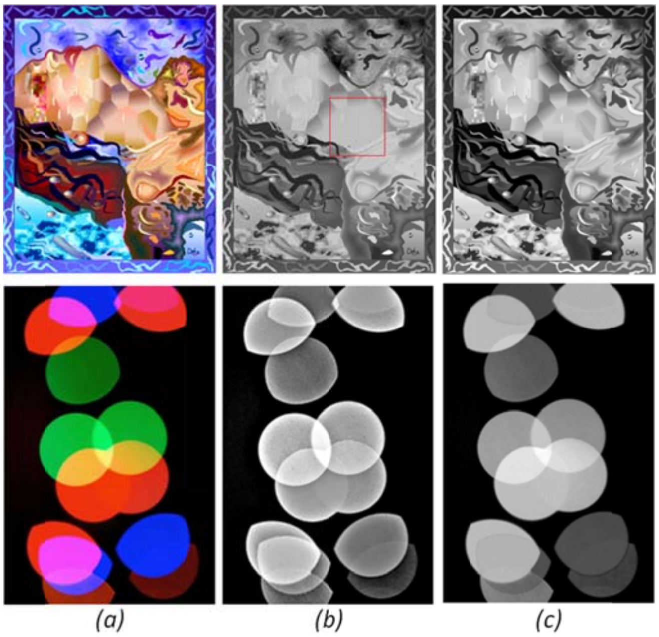


Fig. 1. Comparing to the state-of-the-art global method [16] (first row of (b)) and local method [22] (second row of (b)), our method (c) preserves the original visual perception through the guidance of a region-based saliency model. Hence, it has the advantages of the local method (preserving local discrimination, as shown in the first row) as well as the global method (avoiding inhomogeneous conversion, as shown in the second row).

The key observation behind our approach is that the human visual system tends to see a group of similar pixels as a whole block and is attracted by regions with higher contrast than their surrounding regions. In addition, the perceived color of a region is also affected by its local color contrast [14, 5, 26]. The checker shadow illusion shown in Fig. 3 illustrates this effect.

The authors would like to thank all reviewers for their helpful suggestions and constructive comments. The work is partially supported by the National Natural Science Foundation of China (No.61202154, 61133009), the National Basic Research Project of China (No. 2011CB302203), Shanghai Pujiang Program (No.13PJ1404500), the Science and Technology Commission of Shanghai Municipality Program (No. 13511505000), and the Open Project Program of the State Key Lab of CAD&CG (Grant No. A1401), Zhejiang University. This work was also partially supported by a GRF grant from the RGC of Hong Kong (RGC Ref.: CityU 115112).

D. Hao, B. Sheng (corresponding author) and L. Ma are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China (e-mail: shengbin@sjtu.edu.cn).

S. He and R. Lau are with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mails: shengfeng_he@yahoo.com, rynson.lau@cityu.edu.hk).

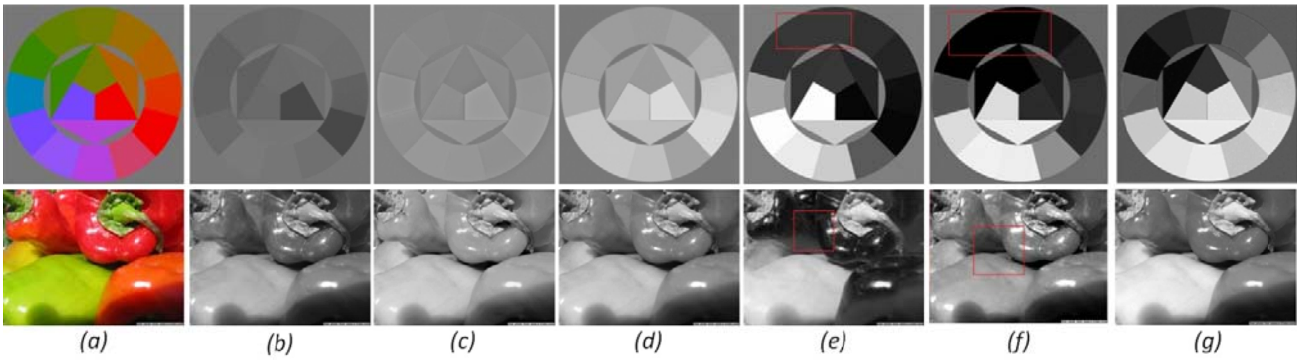


Fig. 2. Comparison with the state-of-the-art methods: (a) original image, (b) rgb2gray(), (c) Smith et al. [22], (d) Kim et al. [11], (e) Lu et al. [16], (f) Song et al. [24]; (g) our method. Our method captures better contrast and color ordering than the other methods.

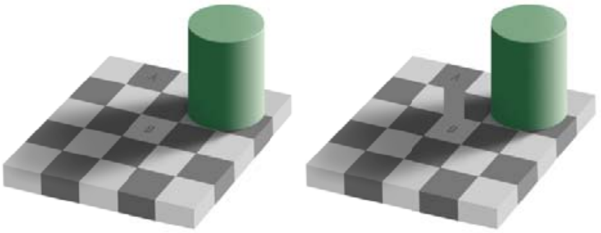


Fig. 3. The checker shadow illusion by Edward Adelson. Blocks A and B appear different in the left image, but are exactly the same after connecting them in the right image.

Two blocks A and B seem different in the left image. Block A looks darker than B, due to the shadow on B. This illusion is dispelled after the two areas are drawn connected. In other words, the visual perception of the target colors can be significantly affected by their neighboring regions. As a result, preserving the local contrast of the input color images is important in order to maintain a similar visual perceptual and level of attention in the output grayscale images.

This observation inspires us to develop a saliency-preserving color-to-gray model. Since pixels of the same color may be perceived differently depending on their surrounding colors, unlike the existing global methods [2, 11, 24], which typically map the same colors to exactly the same grayscale values, our goal is to maintain the same visual perception of the original color image. In addition, as shown in Fig. 1 and 2, the mapping functions of the local approach are more flexible than those of the global approach, as each pixel can have more choices on grayscale values in order to preserve the original details. On the other hand, global information is pivotal to robustness, since pure local methods tend to convert constant color regions in-homogeneously (row 2 of Fig. 1(b)). Although some methods combine global and local conversions [3, 22, 13], they require human interactions, which may be difficult to find optimal parameters to preserve the same perception of the original image. In other words, they lack a visual guidance to automatically generate optimal grayscale images.

To preserve the discriminative details and features in the original image, it is natural to rely on visual saliency detection, which models the human visual system and is applied widely in many applications, such as image segmentation, image retrieval and object recognition. Saliency preservation has been shown effective in decolorization process [2]. However, the saliency model of [2] is not designed for detecting the whole salient object (as shown in Figure. 4(b)). In addition, as the resulted grayscale image preserves saliency according to only a few salient points, it may lead to contrast lost in both salient and non-salient regions.

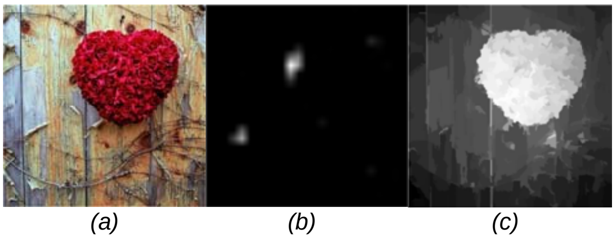


Fig. 4. Given the input image in (a), (b) and (c) show the saliency maps produced by [10] and by our salient object detection method, respectively. Our regional saliency map is more suitable as guidance for decolorization.

In this paper, we propose a robust and perceptually accurate spatial saliency-guided approach for decolorization. First, we propose a two-stage parametric color-to-gray mapping function, which takes into account the global information and the local chromatic contrast. Second, we propose a saliency-preserving optimization approach to preserve the region-based contrast. Spatial information is involved in our saliency model, as to the human visual system, distant pixels should have less effect on each other. Thus, color and spatial information are incorporated both in the mapping function and the saliency model.

Our parametric color-to-gray mapping function determines the grayscale value of each pixel by two factors: the luminance value of the pixel and the sum of the distance weighted chromatic differences between the pixel and its surrounding pixels. As human tends to consider a group of similar pixels as a whole, our saliency model is region-based, instead of pixel-based. This leads to a more efficient process and helps avoid disturbance from outliers. Comparing to the saliency map predicted from traditional saliency detection as shown in Fig. 4(b), which is not object-aware and only highlights some salient pixels in the image, our regional saliency as shown in Fig. 4(c) is obviously more suitable for estimating perceptual differences.

The main contributions of this work are as follows:

1. We propose a two-stage color-to-gray mapping function. Instead of ensuring global consistence, it considers global information and distance weighted color differences between surrounding pixels to preserve local visual perception. Thus, more details can be captured.

2. We use a saliency model to guide the decolorization process, by computing regional saliency rather than predicting sparse saliency points to better describe the perceptual differences between the original color image and the output grayscale image, and avoid negative influence from outliers.
3. We have derived a saliency-guided optimization algorithm to obtain the best decolorization result.

Existing decolorization methods are usually evaluated using Cadik's dataset [4]. Although this dataset contains various types of images, some of them are very simple and contain only a few colors. Real world scenes, especially for outdoor environments, usually involve abundance of colors and patterns. As a complement, we have constructed a new dataset, called Complex Scene Decolorization Dataset (CSDD), to study the performance of the proposed method. Extensive qualitative and quantitative evaluations on the two datasets show that the proposed method outperforms the state-of-the-art methods.

II. Related Work

Recently, many decolorization methods have been proposed to address the problem of loss of details and features during the decolorization process. These techniques can be roughly categorized into global and local methods.

Local decolorization methods treat pixels of the same colors differently to enhance local contrast, according to different attributes like local chrominance edges. Bala and Eschbach [3] enhance contrast by adding high frequency components of chromaticity to the lightness channel. Neumann et al. [20] propose a local color gradient-based method to obtain the best perceptual grayscale image measured in their Coloroid color space. Smith et al. [22] use Helmholtz-Kohlrausch (H-K) predictors to obtain an interim grayscale image and adjust the contrast with the Laplacian pyramid. They then use adaptively weighted multi-scale unsharp masks to enhance the chrominance edges. Gooch et al. [8] find the best matched gray values by minimizing the differences between the chrominance and luminance values of nearby pixels. Although these methods can preserve local features, they may occasionally distort the appearance of constant color regions, as discussed in [11]. In contrast, the proposed method uses a two-stage parametric mapping function to take into account both global and local information.

Global decolorization methods consider the entire image as a whole and apply some consistent mapping functions to all the pixels in the image. As a result, pixels of the same colors will be converted into the same grayscale values, which is different from the local methods. Kuk et al. [13] extend the idea in [8] by considering both local and global contrasts in the image and encoding them using an energy function. Rasche et al. [21] use a linear color mapping function to obtain the optimal conversion, which is determined directly by imposing constraints on different color pairs to maintain a constant luminance. Grundland and Dodgson [9] propose a linear mapping algorithm that preserves the original lightness and color order by adjusting the grayscale value with the chrominance. Kim et al. [11] propose a non-linear global mapping function to feature discriminability and reasonable color ordering. Queiroz et al. [6] propose a method that applies high frequency details to the grayscale image based on a wavelet transform.

Our work has a similar spirit to the work by Ancuti et al. [2]. They obtain the color difference by merging luminance and chrominance to enhance chromatic contrast with the guidance of saliency. Three offset angles are selected for good decolorization. The main difference is that their decolorization process is guided by only a few salient points, which is not enough to preserve the contrast of the entire image. In addition, their saliency model cannot produce a continuous saliency map for objects in the image and the saliency maps may be sparse, which may not represent the visual perception of the entire image, as only a few pixels are marked as salient. In contrast, our saliency model aims at detecting the entire salient object, which provides better guidance for optimization.

All these global methods, however, is not able to handle images with a lot of colors, as they cannot fully capture the details in a globally unified manner. On the other hand, to the human visual system, using a visual cue as guidance is more effective than minimizing the color difference in a particular color space. These limitations motivate us to use local information to enhance contrast and global information to obtain a global consistent mapping under the guidance of saliency.

III. Our Approach

Our color-to-gray algorithm is designed based on the following principles to preserve the visual perception of a color image:

- *Chromaticity Contrast*: Besides the luminance contrast, the chromaticity distinctiveness of pixels in the color image should also be considered.
- *Region-based Contrast*: The image is perceived based on the visual contrast of regions, instead of individual pixels.
- *Distance-dependent Contrast*: The visual contrast between two objects/regions should be inversely proportion to their distance.
- *Saliency Preservation*: Saliency in the input color image should be preserved in the output grayscale image.

In order to achieve these goals, we first formulate a parametric color-to-gray mapping function, which is presented in details in Section III.A. In Section III.B, we propose a saliency model for decolorization by considering local color and grayscale contrasts. Finally, we present our optimization process to select the optimal grayscale image for output in Section III.C.

There are two parameters, k and l, in the parametric color-to-gray mapping function, and different grayscale images can be obtained by applying the mapping function with different k and l values on the color images. Here, we denote the grayscale image converted by parameters k and l as G_{kl}. A saliency model is used as a guidance to obtain the best G_{kl}. The overview of our algorithm is shown in Fig. 5.

A. Parametric Color-to-gray Mapping Function

Our parametric color-to-gray mapping function maps the color of a pixel in the input color image I to a grayscale value in the grayscale image G. It is formulated with two terms:

$$gray_p = L_p + C_p, \tag{1}$$

where $gray_p$ is the output gray value of a pixel p in I. L_p is the

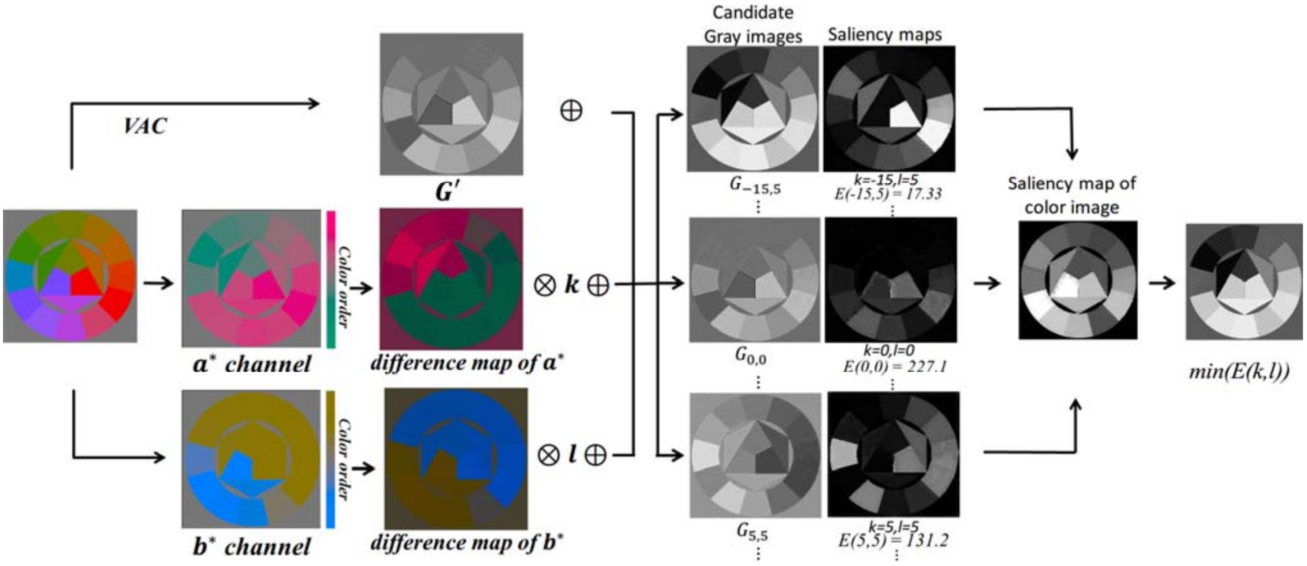


Fig. 5. Overview of the proposed method. The two-stage mapping function first computes an initial gray image by the VAC approach, and then enhances the local contrast by extracting the difference maps from the A and B channels of the CIE LAB color space with different parameters *k* and *l*. The final gray image is obtained by minimizing the saliency difference energy among the candidate gray images.

location-independent luminant value of p. C_p is the location-dependent luminant value, which is used to adjust L_p according to the local color differences of p. In order to achieve the lightness fidelity [11], L_p is the dominant contributor in this mapping function.

To compute L_p of a color pixel p, a straightforward solution is to extract the luminance Y channel of the CIE XYZ color space. However, the CIE Y channel does not consider the Helmholtz–Kohlrausch (H-K) effect [18], which states that a more saturated object/region looks brighter than a less saturated one even if they have the same luminance, as shown in Fig. 6(c).

The H-K effect can be measured by two approaches, Variable-Achromatic-Color (VAC) and Variable-Chromatic-Color (VCC) [19]. In this paper, we use the VAC method to model L_p, as VAC is more accurate than VCC in distinguishing very bright isoluminant colors and thus more suitable for color-to-gray conversion [22]. Hence, L_p is calculated as follows:

$$L_p = L_p^* + (-0.1340\, q(\theta) + 0.0872 K_{Br}) s_{uv} L_p^* \,, \qquad (2)$$

where L_p^* is the lightness component p in the CIE LAB color space, s_{uv} is the chromatic saturation, $q(\theta)$ is the quadrant metric, and K_{Br} is a constant (see [22] for details). We can see from Fig. 6 that the H-K predictor preserves the visual details and contrast well, while the CIE Y channel cannot capture the distinctiveness among the pixels with similar luminance.

Although VAC is able to distinguish the brightness among various chromatic colors with a constant luminance, the location-dependent contrast between p and its neighboring pixels is not taken into account. Hence, we introduce a location-dependent luminance factor C_p to adjust the final grayscale value to reflect this location-dependent contrast. Motivated by the chromaticity-based contrast, we measure chrominance differences by combining two distance-weighted color differences in the A and B channels of the CIE LAB color space. Given a pixel p in image I, we define C_p as follows:

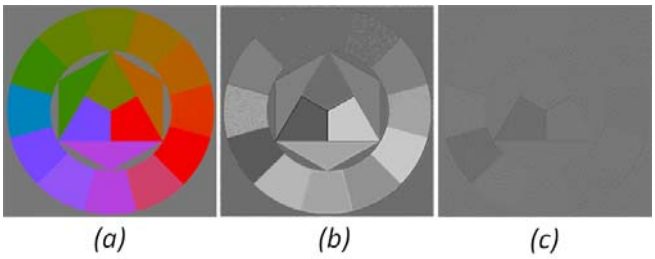


Fig. 6. Comparison of the VAC conversion (b) and the CIE Y channel (c) of the input image (a). The proposed method utilizes VAC conversion, which considers the H-K effect and preserves chromatic contrast better than the CIE Y channel.

$$C_p = k \sum_{q \in \Omega} (a_p - a_q) \mathrm{e}^{-\frac{\Delta D_{pq}}{\sigma^2}} + l \sum_{q \in \Omega} (b_p - b_q) \mathrm{e}^{-\frac{\Delta D_{pq}}{\sigma^2}} \,, \qquad (3)$$

where a_p and a_q are the A channel values of pixels p and q, respectively. b_p and b_q are the B channel values of p and q. ΔD_{pq} denotes the Euclidean distance between p and q. We use the exponential form $e^{-\frac{\Delta D pq}{\sigma^2}}$ to approximate the location-dependent contrast. In our experiments, σ^2 is set to $0.4 \times (width \times height)$, where $width \times height$ is the image size. k and l are unknown variables to be optimized based on the saliency preservation criteria.

Parameters k and l are related to the A and B channels of the CIE LAB colorspace, respectively. The CIE LAB colorspace has two color axes: the A and B channels are defined based on the fact that a color cannot be both red and green, or both blue and yellow, because these colors oppose to each other. The L channel closely matches human perception of lightness. The L, A, B channels are three linearly independent vectors in the 3D colorspace. According to this color opponency theory, we believe that the color differences can be precisely described by the combination of the A and B channels. Thus, we intuitively separate these two channels to calculate the color differences between a pixel and its neighbors, and then add these color

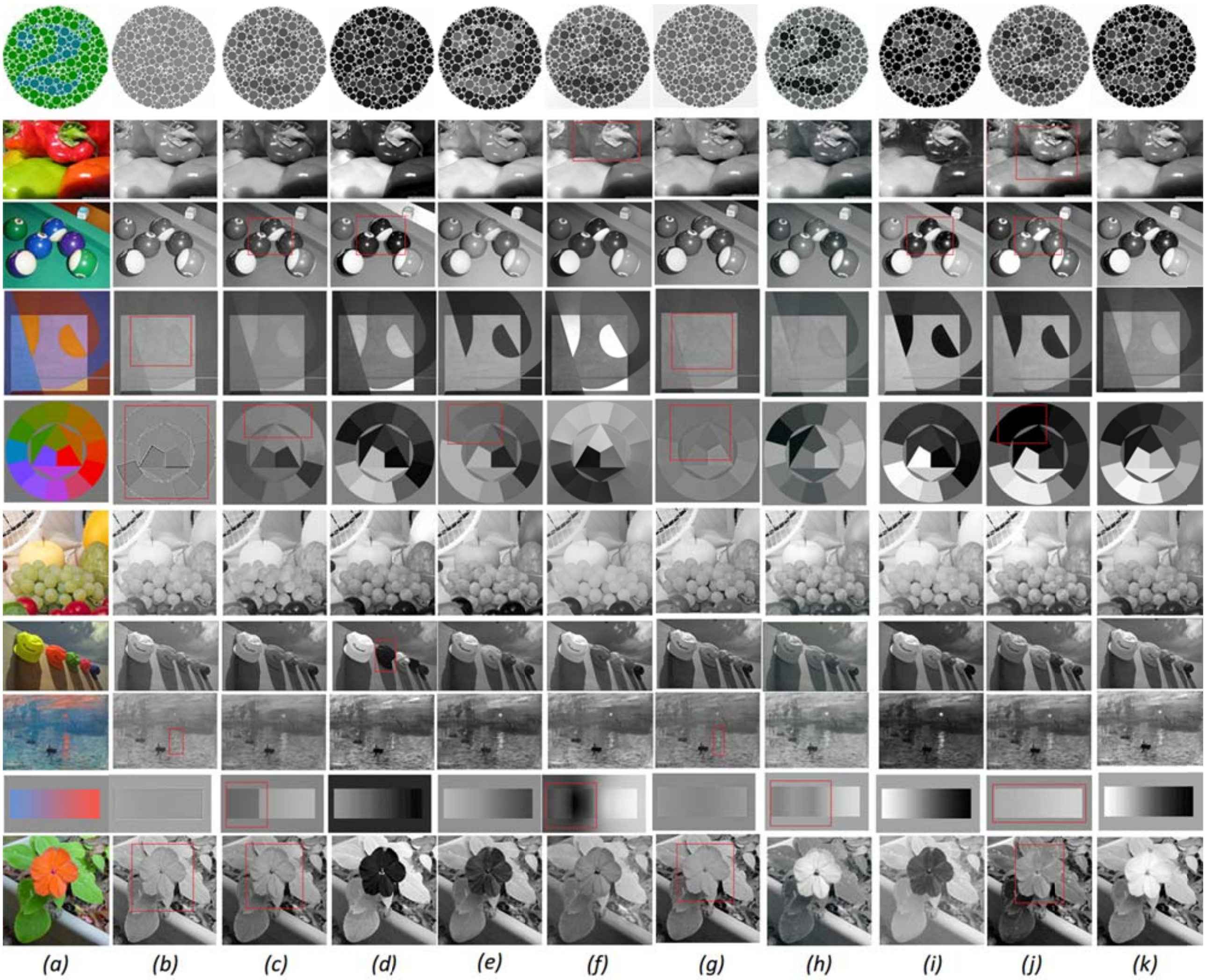


Fig. 7. Qualitative Evaluation on Cadik's dataset [4]: (a) original image, (b) Bala et al. [3], (c) Gooch et al. [8], (d) Rasche et al. [21], (e) Grundland et al. [9], (f) Neumann et al. [20], (g) Smith et al. [22], (h) Ancuti et al.[2], (i) Lu et al. [16], (j) Song et al. [24], and (k) our method. Red rectangles indicate regions of inaccurate representation of contrast. Results show that our method preserves contrast, color order, and perceptually accuracy better than the state-of-the-art methods. The quantitative evaluation is shown in Fig. 9. The red rectangles indicate regions of loss of contrast, compared with the original images.

differences to the initial grayscale value to recover the lost contrast of the chroma. In addition, k refers to the maximum degree of impact of the differences between a pixel and its neighbor pixels in channel A. Parameter l has a similar meaning.

In our implementation, we normalize the A and B channels while computing the color differences in Eq. (3), and the sum of the color differences to keep $\sum_{q\in\Omega}(a_p - a_q)e^{-\frac{\Delta D_{ij}}{\sigma^2}}$ and $\sum_{q\in\Omega}(b_p - b_q)e^{-\frac{\Delta D_{ij}}{\sigma^2}}$ within ±1.

Fig. 5 shows how our mapping function converts a color image into candidate grayscale images. For some of the candidate images (such as k=-15, l=5), the location-dependent luminance factor C_p is able to recover the details and contrast back to the VAC converted grayscale image L_p. It is worth noting that some of the candidates in Fig. 5 preserve color order better than L_p. This is the advantage of using distance- weighted color differences in A and B channels in Eq. (3).

B. *Regional Contrast Saliency Model*

For the human visual system, regions with higher contrast than neighboring regions tend to attract higher attention. We model image saliency according to the relative importance of different image regions, and use the saliency model as an evaluation criterion to determine contrast magnitude, which is then optimized in the next stage.

Cognitive psychology and neurobiology studies [25, 12, 17] divide visual attention into two different types, bottom-up and top-down fashions, based on how visual information is processed. Bottom-up visual attention is subconscious and stimulus-driven. Thus, it is a subjective feeling triggered by objective stimuli. Top-down visual attention is task-driven and based on knowledge. For example, finding a particular person in a group photo is a top down visual attention, whereas a red flower among yellow flowers will attract attention in a bottom up manner. Here, we predict bottom-up saliency by computing local contrast of objects/regions over their surroundings.

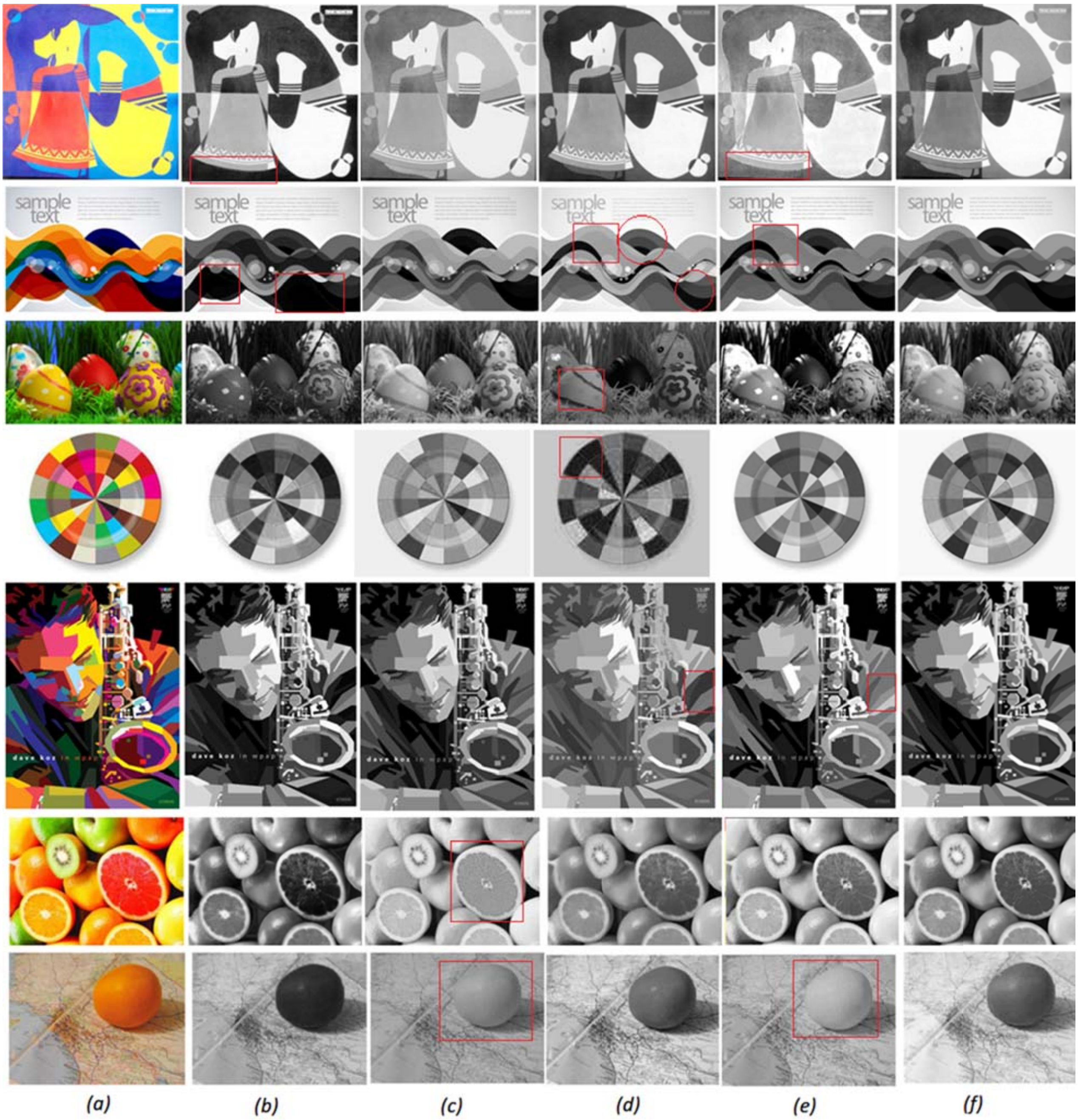


Fig. 8. Qualitative evaluation on the CSDD dataset: (a) original image, (b) Grundland et al. [9], (c) Smith et al. [22], (d) Lu et al. [16], (e) Song et al. [24], and (f) our method. Even in the more complex scenarios, our method is able to preserves contrast according to visual perception of the original image and produces both perceptually accurate and detail-retained results. (Quantitative evaluation is shown in Fig. 10.) The red rectangles indicate the regions of loss of contrast, compared with the original images. The red circles indicate regions that the algorithms wrongly enhance the contrast.

Pixelwise regional contrast is computational expensive, and usually noisy. To be able to extract contrast efficiently while avoiding outliers, we compute the saliency in segment level, instead of pixel level. We use the simple linear iterative clustering (SLIC) method [1] to obtain image segments. The SLIC method is simple, parameter-free, and can produce a desired number of regular and compact superpixels at low computational cost. Although the SLIC method does not guarantee a perfect segmentation result, the segmentation errors appeared in regions with similar colors (e.g., two similar regions belonging to different objects misclassified as the same region) would not destroy the original perception after decolorization.

After we have segmented the image, we compute the saliency of each segment s_i in the original color and the output grayscale images as follows:

$$Sc(s_i) = \sum_{s_i \in S} N(s_i) \cdot e^{-\frac{\Delta D_{ij}}{\sigma^2}} \cdot \Delta C_{ij}/3, \qquad (4)$$

$$Sg(s_i) = \sum_{s_i \in S} N(s_i) \cdot e^{-\frac{\Delta D_{ij}}{\sigma^2}} \cdot \Delta G_{ij}, \qquad (5)$$

where S is the set of segments in the image. $Sc(s_i)$ and $Sg(s_i)$ are the saliency values of segment s_i in the color and grayscale images, respectively. $N(s_i)$ is the number of pixels in s_i, so that a superpixel with a large number of pixels contributes more than one with a small number of pixels. ΔD_{ij} indicates the Euclidean distance between the centers of segments s_i and s_j. ΔC_{ij} is the color contrast computed as the L2 difference between s_i and s_j in the CIE LAB color space. Similarly, ΔG_{ij} is the grayscale contrast between the two segments. We calculate the spatial distance as an exponential weight so that the contributions of different segments depend on their distances. σ^2 is related to the size of the image. It is defined as $\sigma^2 = 0.4 \times (width \times height)$. The computed saliency will be used as a guidance in the optimization step. All the saliency values are normalized to [0, 1].

C. Decolorization Using Optimization

With the parametric color mapping function and the region-based saliency model, we may combine them to find the best parameter values for *k* and *l* in Eq. (3), in order to preserve the visual perception to the original image.

First, we precompute the location-independent luminance factor L_p in Eq. (2) and the difference maps of the *A* channel $\sum_{q\in\Omega}(a_p - a_q)e^{-\frac{\Delta D_{ij}}{\sigma^2}}$ and the *B* channel $\sum_{q\in\Omega}(b_p - b_q)e^{-\frac{\Delta D_{ij}}{\sigma^2}}$ in Eq. (3). We then vary *k* and *l* to generate different candidate grayscale images G_{kl}. The saliency maps of these candidate images can be computed by Eq. (5) and the saliency map of the original color image by Eq. (4).

Our goal is to maintain the same degree of visual perception after decolorization. In other words, the saliency map of the output image should be as close to that of the original image as possible. To preserve the visual saliency in the decolorization process, we minimize the difference of saliency values between the input color image and the output grayscale image. The energy function is defined by:

$$E(k,l) = \sum_{s_i \in S} |Sc(s_i) - Sg(s_i)|^2. \tag{6}$$

To solve this minimization problem, we rewrite Eq. (6) as:

$$E(k,l) = \|M - Ndiag(A_{kl}B)\|^2, \tag{7}$$

where:

$$M = \begin{bmatrix} Sc(s_1) \\ Sc(s_2) \\ \vdots \\ Sc(s_n) \end{bmatrix}$$

$$A_{kl} = \begin{bmatrix} 0 & |L_1^{kl} - L_2^{kl}| & \dots & |L_1^{kl} - L_n^{kl}| \\ |L_2^{kl} - L_1^{kl}| & 0 & \dots & |L_2^{kl} - L_n^{kl}| \\ \vdots & \vdots & \ddots & \vdots \\ |L_n^{kl} - L_1^{kl}| & |L_n^{kl} - L_2^{kl}| & \dots & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & e^{-\frac{\Delta D_{12}}{\sigma^2}} & \cdots & e^{-\frac{\Delta D_{14}}{\sigma^2}} \\ e^{-\frac{\Delta D_{21}}{\sigma^2}} & 0 & \cdots & e^{-\frac{\Delta D_{24}}{\sigma^2}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\frac{\Delta D_{n1}}{\sigma^2}} & e^{-\frac{\Delta D_{n2}}{\sigma^2}} & \cdots & 0 \end{bmatrix}$$

$$N = \begin{bmatrix} N(s_1) & 0 & 0 & 0 \\ 0 & N(s_2) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & N(s_n) \end{bmatrix}$$

Matrices *M* and *B* are fixed for an input color image. Hence, they can be precomputed. Matrix N_{kl} is determined by *k* and *l*. L_i^{kl} is the grayscale value of a segment in grayscale image G_{kl}. Parameters *k* and *l* are in fact used to control the amount of gray value added to L_p, so that unexpected artifacts (such as too bright or too dark) will not appear in the output image. In our implementation, we set $k \in \{-15, \dots, 0, \dots, 15\}$ and $l \in \{-15, \dots, 0, \dots, 15\}$ for finding the best combination.

IV. Experiments And Discussion

We have implemented our method using C++, and conducted all experiments on a PC with an i7 2.8GHz CPU and 4GB RAM. For an input image of resolution 800×600, our program takes 2.48 seconds to get the result. We have evaluated the proposed method using Cadik's decolorization dataset [4] as well as our Complex Scene Decolorization Dataset (CSDD). (We will release this dataset to the public.) For Cadik's dataset, we are able to compare with a lot of existing methods which results are available publicly. However, for the CSDD dataset, we can only compare with existing methods with source codes available. All the corresponding results are produced using the default parameter settings for fair comparison. In addition, we have also conducted a user experiment to evaluate the proposed method subjectively.

A. Qualitative Evaluation

1) Evaluation on Cadik's Dataset

We have compared our method with 9 state-of-art methods [3, 8, 21, 9, 20, 22, 2, 16, 24] on Cadik's dataset. For [24], we compare with their latest works, instead of their preliminary works in [23]. The qualitative comparison on 10 representative images is shown in Fig. 7. Results of the proposed method (shown in the last column) perform favorably or similarly against the state-of-the-art methods in most of the images. Specifically, the proposed method is able to capture better color ordering and visual distinction in hue, as demonstrated in the fifth row. This merit is due to the fact that our color-to-gray mapping function adjusts the luminance by the *A* and *B* channels in the CIE LAB color space, which can discriminate different hues more accurately. The proposed method also preserves edges well. As shown in the eighth row, results of (b), (d) and (g) fail to capture the red sun and its reflection. In the ninth row, the proposed method successfully captures the gradually changing color, due to handling the H-K effect, while most of the other methods are not able to distinguish such weak chromatic differences.

2) Evaluation on CSDD Dataset

Our CSDD dataset includes 22 different images with abundant colors and patterns. Thus, preserving all details of these images are challenging to the decolorization task. We have compared the proposed method with 4 state-of-the-art methods on this dataset. The source codes or executable files of these methods are obtained either from the authors' websites or provided by the authors.

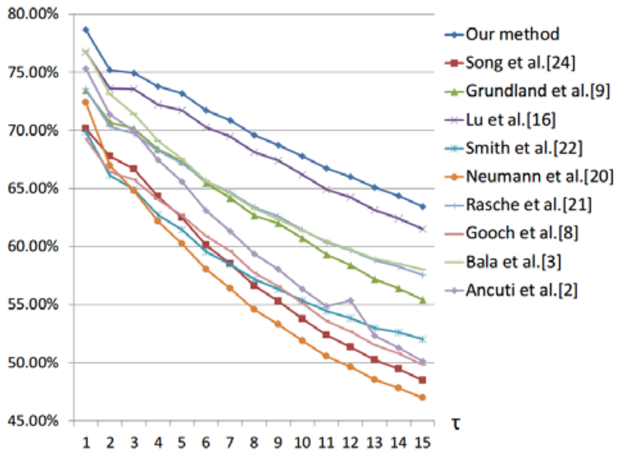


Fig. 9. CCPR comparison on Cadik's dataset. The grayscale images are either produced from the codes provided by the authors or from Cadik's dataset. Our method performs well in preserving details.

The qualitative comparison on 7 representative images from the CSDD dataset is shown in Fig. 8. Results show that the proposed method consistently outperforms the other methods in these complex scenes. In the first row, we can see that the details in (b) and (e) are missing, due to their global mapping strategies. In addition, the method in [9] as shown in Fig. 8(b) tends to produce darker grayscale values in some regions of most of the images, causing higher local contrasts than the original images. Loss of contrast is also observed in Fig. 8(d) (rows 2 – 5), as global optimization may not be able to capture all local details. Although the method in [22] as shown in Fig. 8(c) preserves the details well in row 1, it is not able to preserve the original visual perception well as the output image exhibits low contrast overall. Objects highlighted by red rectangles in the last two rows of Fig. 8(c) also show that the corresponding method cannot preserve the local contrast well. On the contrary, the proposed method produces better results in terms of perceptual accuracy and color orders.

B. Quantitative Evaluation

We have also quantitatively evaluated the proposed method on the two datasets using the color contrast preserving ratio (CCPR) [16] defined as follows:

$$CCPR = \frac{\#\{(x,y)|(x,y)\in\Omega,|g_x-g_y|>\tau\}}{||\Omega||}, \qquad (8)$$

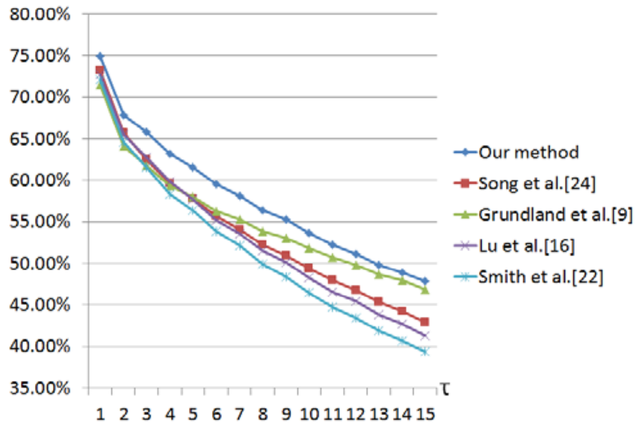


Fig. 10. CCPR comparison on the CSDD dataset. The grayscale images are all produced from the codes provided by the authors.

where Ω is the set that contains all neighboring pixels pairs, $||\Omega||$ is the total number of pixel pairs in Ω. τ is a threshold indicating that if the color/grayscale difference is smaller than τ, the difference will not be visible. $\#\{(x,y)|(x,y)\in\Omega,|g_x-g_y|>\tau\}$ indicates the number of pixel pairs with the difference higher than τ. The main idea of CCPR is to evaluate the percentage of distinct pixels in the original image remaining distinctive after decolorization. We vary τ from 1 to 15 in the experiment.

The average CCPR results on Cadik's dataset and the CSDD dataset are shown in Fig. 9 and 10, respectively. We can see that the proposed method outperforms all the state-of-the-art methods tested on both datasets. This is mainly because the proposed method aims at preserving visual perception based on local information, rather than ensuring global consistence.

C. User Experiment

We have further conducted a user experiment to compare the proposed method with existing methods, in terms of preference and accuracy. There were a total of 20 randomly selected color images, 10 from Cadik's dataset [4] and 10 from the CSDD dataset. We invited 30 participants (15 males and 15 females) at the age of 19 to 40, with no eye-sight deficiency and engaging in different jobs, to participate in the study. The existing methods for comparison are [24, 9, 22, 16]. The images are

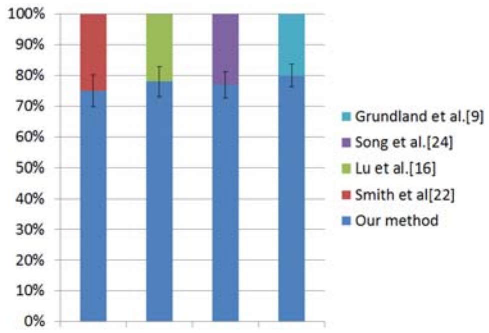


Fig. 11. Results of the preference experiment. Each time, users are asked to choose a grayscale image that they prefer from two images, one from our method and the other from one of the four methods [9, 24, 16, 22]. The error bars are the 95% confidential level.

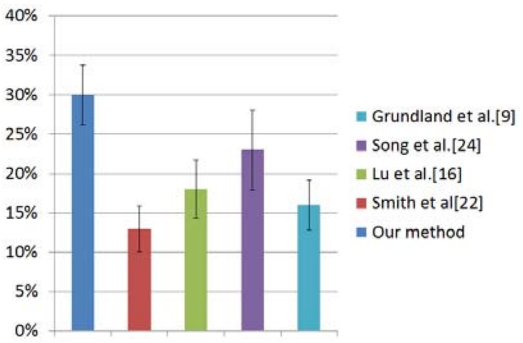


Fig. 12. Results of the accuracy experiment. Each time, users are asked to choose a grayscale image that best represents the original image, from five candidate grayscale images, one from our method and the others from the four methods [9, 24, 16, 22]. The error bars are the 95% confidential level.

presented on an iPad2 with a 9.7-inch screen of resolution 1024 ×768, in a well illuminated room with a light level of 300 lux.

The user experiment consists of two stages, preference experiment and accuracy experiment. In the first stage of the user experiment, each participant is asked to choose one image from two grayscale images to indicate his/her preference. The two images are the results of our method and another randomly selected method. Each participant is asked to choose 80 image pairs. In the second stage, we present 6 images to each participant each time, including the original color image and the grayscale images from all the methods. The participant is asked to choose a grayscale image that best representative the original image. Each participant is asked to choose 20 sets of images.

Results of the user experiment are shown in Fig. 11 and 12. They show that the proposed method performs better both in terms of preference and accuracy. Repeated measures analysis shows a significant main effect between different methods (F(4, 116) = 4.29, p = 0.0034). The error bars show that the proposed method consistently outperforms the other methods. After the experiments, we asked the participants for their comments on their selections. They said that it was difficult to choose results between Song et al. [24], Lu et al. [16] and our method for some images in Cadik's dataset. However, the methods of Song et al. [24] and Lu et al. [16] produced some obvious artifacts in the CSDD dataset.

D. Temporal Properties

Although the proposed method is designed primarily for images, it is robust enough to be able to handle video decolorization to certain extent. However, the proposed method currently does not consider temporal coherency and the output grayscale videos may suffer from such a problem.

Fig. 13 shows an example of video decolorization. The color blocks of the input color video in Fig. 13(a) are gradually disappearing. From the output grayscale video frames shown in Fig. 13(b), we can see that some frames may be able to preserve the temporal coherency (e.g., columns 1 and 2), while some frames may not (e.g., columns 2 and 3). This is due to the change in the detected saliency. One possible solution to this problem is to add spatio-temporal cues to the saliency model (e.g., [7, 27]) to prevent the saliency from abruptly changing. Another possible solution is to add some temporal coherence constratints to the optimization process. This may be an interesting future work.

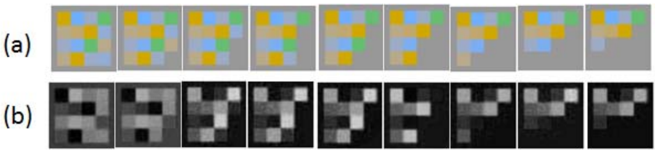


Fig. 13. Video conversion results. (a) Input video. (b) Our results.

V. Conclusions

This paper presents a saliency-preserving decolorization method based on a region-based contrast optimization algorithm, which optimizes a parametric color-to-gray mapping function according to the local contrast. The proposed method applies location-dependent contrast in both the color-to-gray mapping function and the saliency model to preserve the contrast and details. Since the color-to-gray mapping function considers not only the color information of the target pixels but also the surrounding color information, the same color in the original image can be mapped to different grayscale values. In addition, the proposed regional saliency model also helps improve the perceptual accuracy of the conversion. Experimental results show that the proposed method is able to produce grayscale images that are close to the visual perception of the original color images.

VI. References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, SLIC Superpixels Compared to State-of-the-art Superpixel Methods, *IEEE TPAMI*, **34**(11):2274 -2282, 2012.

[2] C. Ancuti, C. Ancuti, and P. Bekaert, Enhancing by Saliency-guided Decolorization, *Proc. IEEE CVPR*, 2011.

[3] R. Bala and R. Eschbach, Spatial Color-to-grayscale Transform Preserving Chrominance Edge Information, *Proc. Color Imaging Conference*, pp. 82–86, 2004.

[4] M. Cadik, Perceptual Evaluation of Color-to-grayscale Image Conversions, *Computer Graphics Forum*, **27**(7):1745–1754, 2008.

[5] D. Corney, J. Haynes, G. Rees, R. Lotto, and O. Sporns, The Brightness of Color, *PLoS ONE*, **4**(3), e5091, 2009.

[6] R. de Queiroz and K. Braun, Color to Gray and Back: Color Embedding into Textured Gray Images, *IEEE TIP*, 15(6):1464–1470, 2006.

[7] S. Goferman, L. Zelnik-Manor, and A. Tal, Context-aware Saliency Detection, *Proc. IEEE CVPR*, 2010.

[8] A. Gooch, S. Olsen, J. Tumblin, and B. Gooch, Color2gray: Salience-preserving Color Removal, *ACM TOG*, **24**(3):634–639, 2005.

[9] M. Grundland and N. Dodgson, Decolorize: Fast, Contrast Enhancing, Color to Grayscale Conversion, *Pattern Recognition*, **40**(11):2891–2896, 2007.

[10] L. Itti, C. Koch, and E. Niebur, A Model of Saliency-based Visual Attention for Rapid Scene Analysis, *IEEE TPAMI*, **20**(11):1254–1259, 1998.

[11] Y. Kim, C. Jang, J. Demouth, and S. Lee, Robust Color-to-gray via Nonlinear Global Mapping, *ACM TOG*, **28**(5), 2009.

[12] C. Koch and S. Ullman, Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry, *Human Neurobiology*, **4:**219–227, 1985.

[13] J. Kuk, J. Ahn, and N. Cho, A Color to Grayscale Conversion Considering Local and Global Contrast, *Proc. ACCV*, LNCS 6495, pp. 513–524, 2011.

[14] R. Lotto and D. Purves, A Rationale for the Structure of Color Space, *Trends in Neurosciences*, **25**(2):84–89, 2002.

[15] C. Lu, L. Xu, and J. Jia, Real-time Contrast Preserving Decolorization, *Proc. ACM SIGGRAPH ASIA Technical Briefs*, 2012.

[16] C. Lu , L. Xu, and J. Jia, Contrast Preserving Decolorization, *Proc. IEEE ICCP*, 2012.

[17] Y. Ma and H. Zhang, Contrast-based Image Attention Analysis by Using Fuzzy Growing, *Proc. ACM Multimedia*, pp. 374–381, 2003.

[18] Y. Nayatani, Simple Estimation Methods for the Helmholtz-Kohlrausch Effect, *Color Research & Application*, **22**(6):385-401, 1997.

[19] Y. Nayatani, Relations between the Two kinds of Representation Methods in the Helmholtz-Kohlrausch Effect, *Color Research & Application*, **23**(5):288–301, 1998.

[20] L. Neumann, M. Cad´ık, and A. Nemcsics, An Efficient Perception-based Adaptive Color to Gray Transformation, *Proc. Computational Aesthetics*, pp. 73–80, 2007.

[21] K. Rasche, R. Geist, and J. Westall. Re-coloring images for gamuts of lower dimension. *Proc. Eurographics*, 423–432, 2005.

[22] K. Smith, P. Landes, J. Thollot, and K. Myszkowski, Apparent greyscale: A Simple and Fast Conversion to Perceptually Accurate Images and Video, *Computer Graphics Forum*, **27**(2):193–200, 2008.

[23] Y. Song, L. Bao, X. Xu, Q. Yang, Decolorization: Is rgb2gray() Out?, *Proc. SIGGRAPH Asia Technical Briefs*, 2013.

[24] Y. Song, L. Bao, and Q. Yang, Real-time Video Decolorization Using Bilateral Filtering, *Proc. IEEE WACV*, 2014.

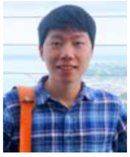
[25] A. Treisman and G. Elade, A Feature-integration Theory of Attention, *Cognitive Psychology*, **12**(1):97–136, 1980.

[26] B. Wong, Points of view: Color Coding, *Nature Methods*, **7**:573, 2010.

[27] Y. Zhai and M. Shah, Visual Attention Detection in Video Sequences Using Spatiotemporal Cues, *Proc. ACM Multimedia*, pp. 815–824, 2006.

Author Biographies

Hao Du received his B.Eng degree in Dept. of Computer Science and Technology at Nanjing University of Aeronautics and Astronautics (NUAA). He is now a graduate student in the Department of Computer Science and Engineering at Shanghai Jiao Tong University. His research interests include image/video processing and visual salient-region detection.



Shengfeng He received his B.Sc. and M.Sc. degrees in the Faculty of Information Technology, Macau University of Science and Technology, in 2009 and 2011, respectively. He is now a Ph.D. student in the Department of Computer Science, City University of Hong Kong. His research interests include image processing, computer vision, computer graphics, physically-based animation and machine learning.

Bin Sheng received his BA degree in English and BE degree in computer science from Huazhong University of Science and Technology in 2004, and MS degree in software engineering from University of Macau in 2007, and PhD Degree in computer science from The Chinese University of Hong Kong in 2011. He is currently an associate professor in Department of Computer Science and Engineering at Shanghai Jiao Tong University. His research interests include virtual reality, computer graphics and image based techniques.

Lizhuang Ma received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1985 and 1991, respectively. He is a Professor and the Head of the Digital Media and Data Reconstruction Laboratory, Shanghai Jiao Tong University, Shanghai, China. His research interests include digital media technology, computer graphics, digital image, and video processing.

Rynson W.H. Lau received his Ph.D. degree from University of Cambridge. He was on the faculty of Durham University and Hong Kong Polytechnic University. He is now with City University of Hong Kong. Rynson serves on the Editorial Board of Computer Animation and Virtual Worlds, and IEEE Trans. on Learning Technologies. He has served as the Guest Editor of a number of journal special issues, including ACM Trans. on Internet Technology, IEEE Multimedia, IEEE Trans. on Multimedia, IEEE Trans. on Visualization and Computer Graphics, and IEEE Computer Graphics & Applications. In addition, he has also served in the committee of a number of conferences, including Program Co-chair of ACM VRST 2004, ACM MTDL 2009, IEEE U-Media 2010, and Conference Co-chair of CASA 2005, ACM VRST 2005, ICWL 2007, ACM MDI 2009, ACM VRST 2010, ACM VRST 2014. His research interests include computer graphics and image processing.