Mathematical Problems of Computer Science 57, 18–29, 2022.
doi: 10.51408/1963-0083

UDC 004.932

A New Image Decolorization Evaluation Quality Metric

Hrach Y. Ayunts[1] and Sos S. Agaian[2]

[1]Yerevan State University, Yerevan, Armenia
[2]College of Staten Island (CSI), City University of New York, New York, USA
e-mail: hrach.ayunc@gmail.com, sos.agaian@csi.cuny.edu

Abstract

Image decolorization, the process of color-to-gray conversion, plays a crucial role in single-channel processing, computer vision, digital printing, and monochrome visualization. This process induces new artifacts, the impact of which on visual quality has to be identified. While visual quality assessment has been the subject of many studies, there are still some open questions regarding new color-to-gray conversion quality metrics. For example, computer simulations show that the commonly used grayscale conversion quality metrics such as CCPR, CCFR, and E-score depend on parameters and may pick different best decolorization methods by changing the parameters.

This paper proposes a new quality metric to evaluate image decolorization methods. It uses the human visual properties information and regression method. Experimental results also show (i) strong correlations between the presented image decolorization quality metric and the Mean Opinion Score (MOS), (ii) more robust than the existing quality metrics, and (iii) help to choose the best state-of-the-art decolorization methods using the presented metric and existing quality metrics.

Keywords: Color-to-gray conversion, Decolorization, Grayscale, Regression, Quality metric.

Article info: Received 18 April 2022; accepted 20 May 2022.

Acknowledgement: This work was supported by the professors of the Department of Informatics and Applied Mathematics of Yerevan State University (S. Sargsyan, E. Danoyan, Yu. Hakopian) and the Armenian Engineers and Scientists of America.

The authors would like to thank the Picsart employees for participating in the user study. The authors would also like to thank Sowmya et al. [13] and Lu et al. [4] for sharing source codes and experiment materials.

1. Introduction

Image decolorization aims to convert a color image into a grayscale image to improve the image's visual appearance or provide a "better" gray-level representation for the future automated image. The overall purpose of image decolorization is to preserve the visible color contrast, which usually suffers from information loss. It plays an essential role in single-channel image processing (analysis, detection, segmentation, and recognition), computer vision, monochrome printing, e-ink display, etc. [1]. The analysis of the existing image decolorization techniques shows the common problems that need to be solved because such methods introduce certain artifacts. It isn't easy to evaluate decolorization methods and select their optimal parameters. There are various quality metrics for color images [2, 3]. Thus, these types of metrics are not suitable for the evaluation of color-to-gray conversion. There is also no efficient measure that can be served as a building criterion for image decolorization.

Practically, all quality metrics for image decolorization are based on the fact that the human visual system cannot perceive color differences smaller than a certain threshold [4, 5, 6]. Extensive computer simulations show that (i) commonly used grayscale conversion quality metrics such as CCPR, CCFR and E-score depend on the color difference parameter, and (ii) by changing the parameter, we pick a different decolorization method. Thus, one needs to develop a new robust threshold-independent quality metric that does not require a reference image.

This paper makes several key contributions:

1. Propose a non-parametric, robust, monotonic, and non-reference quality metric for image decolorization.
2. Present extensive computer simulation results.
3. Present qualitative and perceptual evaluation of state-of-the-art decolorization methods.

The structure of this paper is organized as follows. Section 2 discusses the existing image decolorization methods and quality metrics. Section 3 presents a new non-parametric quality metric. Section 4 provides the results of extensive computer simulation. Section 5 validates the new metric using preference scores from the user study. Finally, Section 6 concludes the work.

2. Background

This section presents the existing color-to-gray conversion methods and quality metrics. Traditional color-to-gray conversion methods usually use a linear combination of R (red), G(green), B (blue) channels of a color image. It is based on the theory of T. Young (1802), which states that any color can be created by combining three primary colors: R, G, and B. $Gray = aR + bG + cB$ [7], were a, b, c coefficients are calculated as

(i) Lightness method: $Gray = \frac{max(R,G,B)+min(R,G,B)}{2}$
(ii) Average method: $a = b = c = 1/3$, or $Gray = \frac{R+G+B}{3}$
(iii) Luminosity method: $a = 0.21, b = 0.72, c = 0.07$, or $Gray = 0.21R + 0.72G + 0.07B$.

These are the most popular and straightforward conversions used in electronic displays, printers, computer vision, image processing, and many other algorithms as a preprocessing step.

However, Fig. 1 shows that the grayscale conversion suffers from information loss (many details didn't preserve, and the color contrast was lost in the grayscale images). It is natural to ask.

(a) Can we have a better decolorization algorithm?

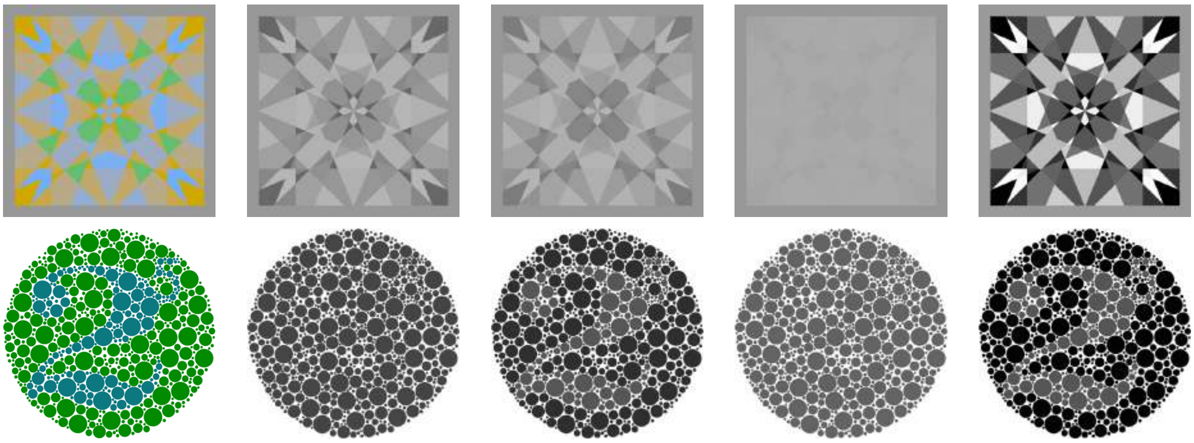


Fig. 1. Comparison of traditional grayscale conversion methods. Decolorized images can lose can lose the contrast and become hardly visible.

(b) How to quantitatively evaluate the performance of different methods or choose the parameters such as a, b, and c?

(c) How do you improve the quality of a color image using decolorized images?

More advanced decolorization methods use the values of other pixels to specify color orders to preserve the color contrast. Local methods rely on local chrominance edges to enhance the contrast [8, 9]. Most recent notable decolorization methods are based on the parametric decolorization model and its modification [5, 4, 10].

Parametric Decolorization Model(PDM). The basic idea here is to convert a color image into gray using a combination of a polynomial of R, G, and B components: $\{R, G, B, RG, RB, GB, R^2, G^2, B^2\}$. It generalizes commonly used linear and nonlinear color-to-gray conversion/mapping systems. More details on this method one can find in [5].

There are also neural network solutions to this problem [11].

Decolorization needs quantitative evaluation to understand the performance of different methods.

Exiting decolorization quality metrics. The most commonly used decolorization quality metrics are based on the fact that the human visual system cannot perceive color difference δ smaller than a certain threshold. For example, the Color Contrast Preserving Ratio (CCPR) (suggested by Lu et al. [4]), defined as

$$\mathrm{CCPR} = \frac{\#\{(x,y)|(x,y) \in \Omega, |g_x - g_y| \geq \tau\}}{||\Omega||}, \tag{1}$$

where Ω is the set of all pixel pairs with $\delta_{x,y} \geq \tau$, and g_x, is the value of the x pixel after decolorization.

CCPR shows the percentage of distinctive pixel pairs after the conversion, but it does not necessarily indicate if the grayscale image was "distorted" after conversion. To complement CCPR, Lu et al. [4] suggested Color Content Fidelity Ratio (CCFR). It is defined as

$$\mathrm{CCFR} = 1 - \frac{\#\{(x,y)|(x,y) \in \Theta, \delta_{x,y} \leq \tau\}}{||\Theta||}, \tag{2}$$

where Θ is the set of all pixel pairs with $|g_x - g_y| > \tau$. This metric shows how much the converted image has changed in terms of structure.

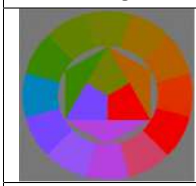
Finally, the combination of CCPR and CCFR, E-score [4], is defined as

$$\text{E-score} = \frac{2 \cdot \text{CCPR} \cdot \text{CCFR}}{\text{CCPR} + \text{CCFR}}. \tag{3}$$

3. Proposed Quality Metric

This section shows the shortcomings of the existing decolorization metrics and suggests a better quality metric for quantitative evaluations.

Table 1: E-score metric for some threshold values for different decolorization methods

Image	Method	$\tau = 3$	$\tau = 5$	$\tau = 7$	$\tau = 9$	$\tau = 15$	$\tau = 25$
	PDM	**0.9934**	0.9776	0.9759	0.9737	**0.9644**	**0.9551**
	LUM	0.9613	0.9174	0.8526	0.7990	0.5956	0.3997
	SPD	0.9862	0.9769	0.9751	0.9713	0.9475	0.9192
	SVD	0.9896	**0.9821**	**0.9765**	**0.9744**	0.9279	0.8514
	PDM	0.9726	0.9502	0.9272	**0.9035**	**0.8298**	**0.6647**
	LUM	0.9646	0.9356	0.9046	0.8704	0.7447	0.4993
	SPD	**0.9777**	**0.9550**	0.9275	0.8956	0.7823	0.5662
	SVD	0.9745	0.9525	**0.9279**	0.9003	0.7987	0.5965

The commonly used grayscale conversion quality metrics such as CCPR, CCFR, and E-score depend on the color difference parameter τ. Computer simulations show that by changing the parameter τ, we pick a different decolorization method.

To verify this statement, we compare different decolorization methods on a couple of images from Ĉadík's dataset [12].

We calculate the E-score quality metric for different values of threshold. We use three state-of-the-art methods (Lu et al. [5], Sowmya et al. [13], Liu et al. [10]) and the Luminosity method for comparison. The results are listed in Table 1. Obviously, the best method differs depending on the threshold value. For example, we can pick three different best methods by changing parameter τ in the case of the second image. The visual results of decolorization on these images are shown in Fig. 4.

In the previous work, the quantitative evaluation of color-to-gray conversion was performed using E-score for fixed values of threshold [4, 5]) or the average of several threshold values [10]. Therefore, there is a need for more independent metrics to investigate the conversion process for each image.

We introduce a new quality metric called Threshold-Independent Slope (TIS), which shows the decreasing speed of the E-score as the threshold value grows. We calculate the E-score metric for different τ values ($\tau = 1, 2, ..15$) and choose the slope of the linear regression of this data as a new metric. The main advantage of the new metric is that it is not dependent on the τ parameter.

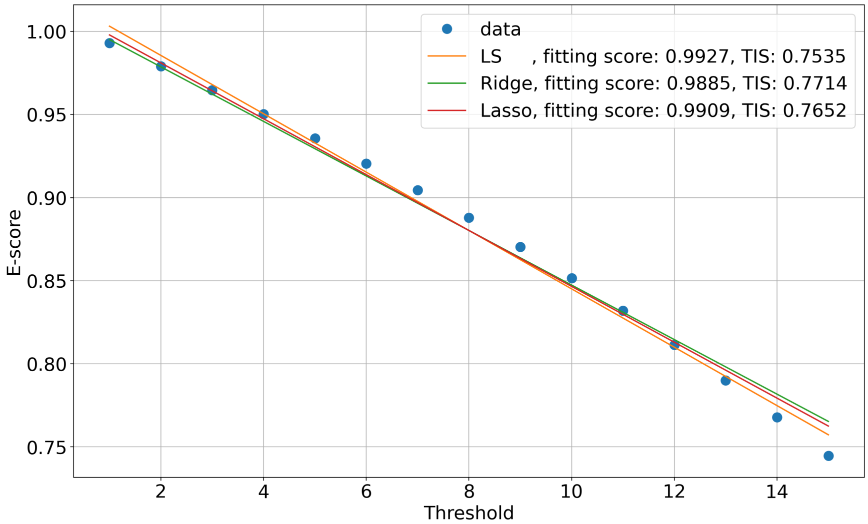


Fig. 2. Simple linear function estimation using the Least Squares method, Ridge regression, and Lasso regression.

Linear Regression can be solved using several linear models. A simple linear model function is defined as

$$y = \alpha + \beta x, \tag{4}$$

which describes a line with a slope β and y-intercept α. One of the easiest ways to estimate the slope is to use the **Least Squares** method:

$$\hat{\beta}_{ls} = \arg\min_{\beta} ||y - \beta x||_2^2. \tag{5}$$

Another method for coefficient estimation of (4) is **Ridge regression** [14]. It is most suitable when data contains a higher number of predictor variables than the number of observations. The ridge regression estimator solves the regression problem using l_2 penalized least squares:

$$\hat{\beta}_{ridge} = \arg\min_{\beta} ||y - \beta x||_2^2 + \lambda ||\beta||_2^2, \tag{6}$$

where $\lambda > 0$ is a tuning parameter that controls the strength of the penalty term. Similar to ridge regression, **Lasso regression** can be used for slope estimation [15]. The lasso estimator uses l_1 penalized least squares for solving the following optimization problem with λ tuning parameter:

$$\hat{\beta}_{lasso} = \arg\min_{\beta} ||y - \beta x||_2^2 + \lambda ||\beta||_1. \tag{7}$$

Fig. 2 compares these three regression models on a sample image from Ĉadík's dataset [12]. Each of these models is used to calculate the TIS metric. To find the best model for our case, we calculate the fitting scores of each model on every image from the dataset. The Least Squares method has the best average fitting score: thus, we use it for further evaluations. Therefore, our TIS metric is defined as

$$\mathrm{TIS} = \max(1 - |\alpha\beta|, 0), \tag{8}$$

where α and β are coefficients of a simple linear function (4) estimated with the Least Squares method (5). TIS ranges in $[0, 1]$, and higher values mean a lower decreasing speed of the E-score metric when the threshold is increased.

TIS	0.53565	0.72862	0.82799	0.87882

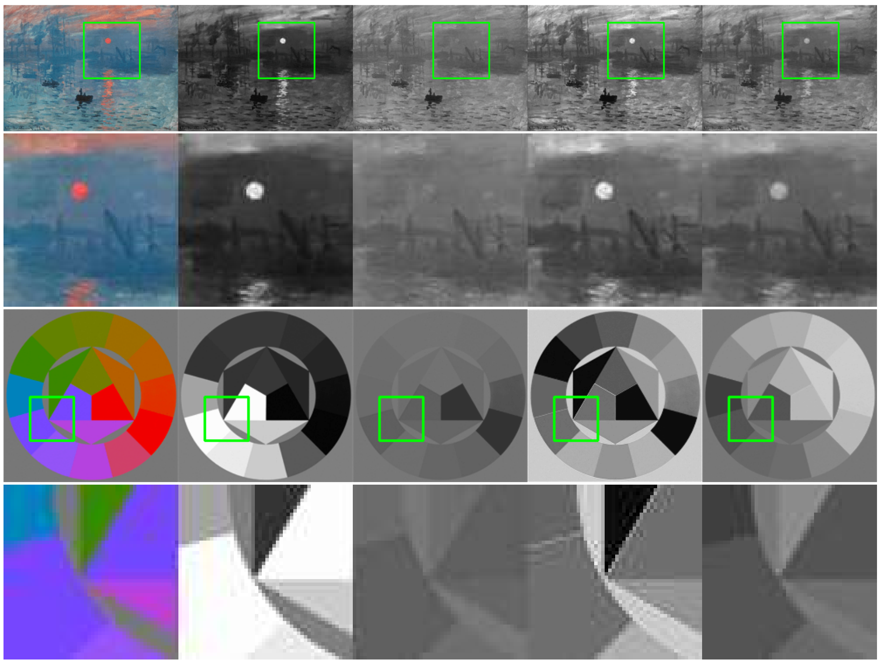
adftgsdfg

Fig. 3. The TIS metric grows with a contrast and visibility increase.

Fig. 3 shows the decolorization result on a sample image with four different levels of visibility. The value of our TIS metric grows with better visibility and contrast in the result. Therefore, the TIS is also a monotonic metric.

4. Computer Simulation

This section evaluates four decolorization methods using our TIS metric and the existing quality metrics. We also show the usefulness of our metric in picking the best parameters for grayscale conversion.



sdfasdfasd

Fig. 4. Visual results of different decolorization algorithms
(from left to right: source image, PDM, LUM, SPD, SVD)

Evaluation of decolorization methods. We chose one traditional conversion method: the Luminosity method (denoted as LUM in tables) is the most popular conversion used in many image processing algorithms and electronic devices. In many cases, it fails to preserve the contrast because the conversion considers only current pixel information. We also chose three state-of-the-art contrast preserving decolorization methods for evaluation. These methods are suggested by Lu et al. [5], Liu et al. [10], and Sowmya et al. [13] (we use PDM, SPD, and SVD acronyms in the tables, respectively). These methods consider global pixel information and color differences in the image for better conversion.

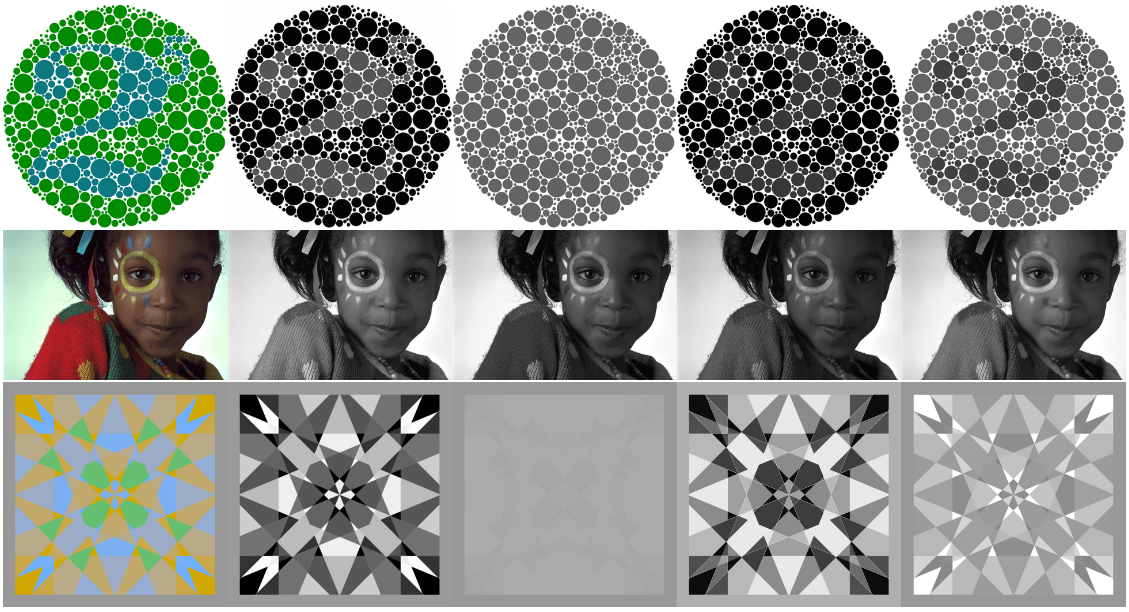


Fig. 5. Visual results of different decolorization algorithms (from left to right: source image, PDM, LUM, SPD, SVD)

Figs. 4 and 5 show the visual results of four decolorization methods on several images. The simple Luminosity method usually fails to preserve the color contrast, while the other three methods produce better visual outputs.

We use Ĉadík's dataset [12] for performance evaluation. It contains 24 PNG images and mainly consists of synthetically generated images and some colorful real-life photos. Most of these images are challenging for traditional color-to-gray conversion methods. That's why Ĉadík's dataset is the most popular in this field and can be beneficial for the evaluation of decolorization methods.

Table 2: Average TIS and E-score for different thresholds on Ĉadík's dataset.

Method	$\tau = 3$	$\tau = 5$	$\tau = 7$	$\tau = 9$	$\tau = 15$	$\tau = 25$	TIS
PDM	0.98222	0.97009	0.95866	0.94697	**0.90971**	**0.84409**	**0.91635**
LUM	0.96340	0.93755	0.91399	0.89556	0.83167	0.71761	0.84992
SPD	**0.98241**	**0.97060**	**0.95922**	**0.94810**	0.90966	0.83835	0.91560
SVD	0.98045	0.96651	0.95334	0.94040	0.89324	0.81638	0.90121

The quantitative evaluation of four decolorization methods using the E-score metric for different thresholds and our new TIS metric on Ĉadík's dataset are presented in Table 2. It presents the performance of each metric (average value) of all images from Ĉadík's dataset.

It also shows that the presented TIS metric is more stable and picks only the best method for this dataset. So it can be helpful in both individual and large-scale evaluations of the grayscale conversion methods.

Picking the best parameter for the simple grayscale conversion. Image decolorization quality metrics can not only be useful in method evaluation, but they can also help pick the best parameters for an algorithm. For example, in simple grayscale conversion, coefficients can be changed to get a "better" conversion.

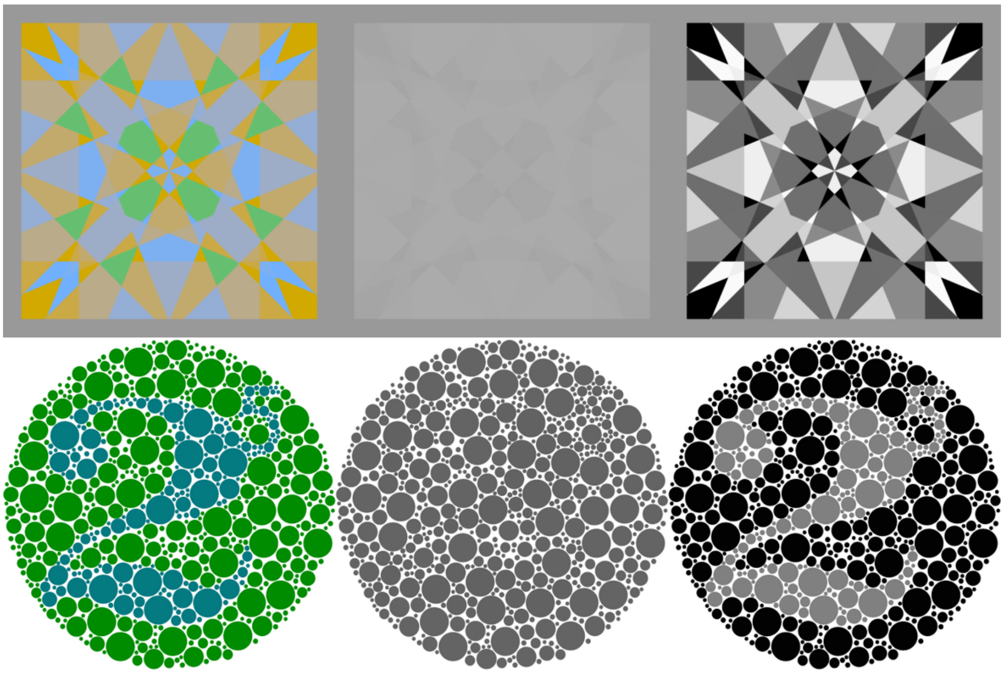


Fig. 6. Comparison of the "best" linear conversion with the luminosity method (from left to right: source image, luminosity, the best conversion).

We pick the best parameters of the linear grayscale conversion by maximizing the value of the quality metric for each image individually. Fig. 6 shows the results corresponding to the highest values of the TIS for two images ($a = 0.02, b = 0, c = 0.98$ for the first image, and $a = 0, b = 0.06, c = 0.94$ for the second one).

5. Perceptual Validation

This section validates our TIS metric using the preference scores.

We invited 20 users to participate in a survey to show the effectiveness and importance of our metric. After a small introduction to decolorization, they were asked to rate the color-to-gray conversion for ten random images from the Čadík's dataset on a scale of one to three. To facilitate the scoring process, we use the three-scale modification of the Mean Opinion Score (MOS). One means the conversion is bad, and it failed to preserve the contrast. Score two corresponds to mediocre conversion. Finally, three is for the best conversion with contrast preservation and the most visually pleasing result.

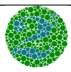
To validate our metric, we use the Kendall rank correlation coefficient [16]. It is defined as

$$R = \frac{\#\{\text{concordant pair}\} - \#\{\text{disconcordant pair}\}}{\frac{1}{2}n(n-1)}, \tag{9}$$

where $n = 4$ denotes the number of methods. Let s_i be the score for the result produced by the ith method, and p_i be the preference score for the same result. If two pairs (s_i, s_j) and (p_i, p_j) are with the same order (i.e., $(s_i - s_j)(p_i - p_j) > 0$), the pair (i, j) is concordant. Otherwise, it is discordant. R ranges in $[-1, 1]$. We get $R > 0$ if the two rankings agree with each other and $R < 0$ otherwise.

We calculate the Kendall rank correlation coefficient (R) for the existing metrics and our TIS metric. The TIS metric has a high correlation with the user preference scores and can easily replace the existing quality metrics for quantitative evaluations of decolorization. The ranks for several images presented in the survey are listed in Table 3.

Table 3: Kendall rank correlation coefficient for the E-score metric with different thresholds, and our TIS metric

Image	$\tau = 5$	$\tau = 7$	$\tau = 9$	$\tau = 15$	$\tau = 25$	TIS
	0.333	0.333	0.333	0.667	0.667	0.667
	0.333	0.333	0.333	1	1	1
	0.667	0.333	0.333	0.333	0	0.333

6. Conclusion

This paper proposes a new TIS image quality metric for accurately evaluating image decolorization methods. The TIS quality metric is a blind, robust, monotonic, non-parametric metric and correlates with subjective preference scores. The quantitative and qualitative computer simulations on the Ĉadík's dataset demonstrate that the proposed metric outperforms the current state-of-the-art metrics. The TIS metric is also helpful in picking the best parameters of the grayscale algorithm.

Our future work will extend the proposed work to other types of distortion, generate new decolorization methods, and evaluate them on other databases.

References

[1] C. Saravanan, "Color image to grayscale image conversion", *Second Inter. Conference on Computer Engineering and Applications, IEEE*, vol. 2, pp. 196–199, March 2010.

[2] K. Panetta, C. Gao, and S. Agaian, "No reference color image contrast and quality measures", *IEEE trans. on Consumer Electronics*, vol. 59, no. 3, pp. 643–651, 2013.

[3] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures", *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2015.

[4] C. Lu, L. Xu, and J. Jia, "Contrast preserving decolorization with perception-based quality metrics", *Inter. Journal of computer vision*, vol. 110, no. 2, pp. 222–239, 2014.

[5] C. Lu, L. Xu, and J. Jia, "Contrast preserving decolorization", *Proc. of IEEE inter. conference on computational photography (ICCP)*, pp. 1–7, 2012.

[6] S. Agaian, "Visual Morphology", *Proc. of SPIE, Nonlinear Image Processing X*, San Jose CA, vol. 3646, pp. 139–150, 1999.

[7] J. Cook, (2009) Three algorithms for converting color to grayscale. [Online]. Available: https://www.johndcook.com/blog/2009/08/24/algorithms-convert-color-grayscale/

[8] R. Bala, and R. Eschbach, "Spatial color-to-grayscale transform preserving chrominance edge information", *Color and Imaging Conference, Society for Imaging Science and Technology*, vol. 2004, no. 1, pp. 82–86, 2004.

[9] L. Neumann, M. Ĉadík, and A. Nemcsics, "An efficient perception-based adaptive color to gray transformation", *Proc. of the Third Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging* pp. 73–80, 2007.

[10] Q. Liu, P. Liu, Y. Wang, and H. Leung, "Semiparametric decolorization with Laplacian-based perceptual quality metric", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 27, no. 9, pp. 1856–1868, 2016.

[11] Q. Liu, and H. Leung, "Variable augmented neural network for decolorization and multi-exposure fusion", *Information Fusion*, vol. 46, pp. 114–127, 2019.

[12] M. Ĉadík, "Perceptual Evaluation of Color-to-Grayscale Image Conversions", *Computer Graphics Forum*, vol. 27, no. 7, Wiley Online Library, pp. 1745–54, 2008.

[13] V. Sowmya, D. Govind, and K. Soman, "Significance of incorporating chrominance information for effective color-to-grayscale image conversion", *Signal, Image and Video Processing*, vol. 11, no. 1, pp. 129–136, 2017.

[14] A. Hoerl, and R. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[15] R. Tibshirani, "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[16] H. Abdi, "The Kendall rank correlation coefficient", *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, pp. 508–510, 2007.

Պատկերի գունազրկման որակի գնահատման նոր չափորոշիչ

Հրաչ Յու. Այունց[1] և Սոս Ս. Աղայան[2]

[1]Երևանի պետական համալսարան, Երևան, Հայաստան
[2]Սթեյթեն Այլենդի քոլեջ, Նյու Յորքի քաղաքային համալսարան, Նյու Յորք, ԱՄՆ
e-mail: hrach.ayunc@gmail.com, sos.agaian@csi.cuny.edu

Ամփոփում

Պատկերի գունազրկումը՝ գունավոր պատկերից մոնոխրոմ պատկերի փոխակերպման գործընթացը, կարևոր դեր է խաղում մեկ բաղադրիչով պատկերների մշակման, համակարգչային տեսողության, թվային տպագրության և մոնոխրոմ վիզուալիզացիայի մեջ: Այս գործընթացը առաջացնում է նոր աղավաղումներ, որոնց ազդեցությունը տեսողական որակի վրա պետք է բացահայտվի: Թեև տեսողական որակի գնահատումը եղել է բազմաթիվ ուսումնասիրությունների առարկա, դեռևս կան որոշ բաց հարցեր կապված փոխակերպման որակի նոր մետրիկաների հետ: Օրինակ՝ համակարգչային մոդելավորումը ցույց է տալիս, որ հաճախ օգտագործվող որակի չափորոշիչները, ինչպիսիք են՝ CCPR-ը, CCFR-ը և E-score-ը, կախված են պարամետրերից և կարող են ընտրել տարբեր լավագույն մեթոդներ՝ փոփոխելով պարամետրերը:

Հոդվածում առաջարկվում է պատկերների գունազրկման որակի գնահատման նոր չափորոշիչ, որը հիմնված է մարդու տեսողական հատկությունների վրա և հաշվվում է ռեգրեսիայի մեթոդի միջոցով: Փորձարարական արդյունքները ցույց են տալիս, որ առաջարկվող մետրիկան ավելի կայուն է, քան գոյություն ունեցողները, այն նաև ունի բարձր կորելացիա միջին կարծիքի գնահատականի (MOS) հետ, և դրա օգնությամբ հնարավոր է ընտրել լավագույն գունազերծման մեթոդները:

Բանալի բառեր՝ գունավոր պատկերների փոխակերպում մոնոխրոմ պատկերների, գունազրկում, մոնոխրոմ պատկեր, ռեգրեսիա, որակի չափորոշիչ:

Новая метрика качества оценки обесцвечивания изображения

Грач Ю. Аюнц[1] и Сос С. Агаян[2]

[1]Ереванский государственный университет, Ереван, Армения
[2]Колледж Статен-Айленда, Городской университет Нью-Йорка, Нью-Йорк, США
e-mail: hrach.ayunc@gmail.com, sos.agaian@csi.cuny.edu

Аннотация

Обесцвечивание изображения, процесс преобразования цветного изображения в монохромное, играет решающую роль в одноканальной обработке, компьютерном зрении, цифровой печати и монохромной визуализации. Этот процесс вызывает новые артефакты, влияние которых на визуальное качество должно быть определено. Несмотря на то, что оценка визуального качества

была предметом многих исследований, все еще остается несколько открытых вопросов, касающихся новых показателей качества преобразования цветного изображения в серый. Например, компьютерное моделирование показывает, что обычно используемые показатели качества преобразовании, такие как CCPR, CCFR и E-score, зависят от параметров и могут выбирать различные наилучшие методы путем изменения параметров.

В этой статье предлагается новая метрика качества для оценки методов обесцвечивания изображения. Она использует информацию о зрительных свойствах человека и метод регрессии. Экспериментальные результаты также показывают сильную корреляцию между представленной метрикой качества обесцвечивания изображения и средней оценкой мнений (MOS), более надежную, чем существующие метрики качества, и помогают выбрать лучший из современных методов обесцвечивания с использованием представленной метрики и существующих метрик качества.

Ключевые слова: преобразования цветного изображения в монохромное, обесцвечивание, монохромное изображение, регрессия, метрика качества.