Colour-to-Greyscale Image Conversion by Linear Anisotropic Diffusion of Perceptual Colour Metrics

Ivar Farup*, Marius Pedersen*, and Ali Alsam‡
Department of Computer Science
Norwegian University of Science and Technology
*Gjøvik and ‡Trondheim, Norway
{ivar.farup, marius.pedersen, ali.alsam}@ntnu.no

***Abstract*—We present an algorithm for conversion of colour images to greyscale. The underlying idea is that local perceptual colour differences in the colour image should translate into local differences in greylevel in the greyscale image. This is obtained by constructing a gradient for the greyscale image from the eigenvalues and eigenvectors of the structure tensor of the colour image, which, in turn, is computed by means of perceptual colour difference metrics. The greyscale image is then constructed from the gradient by means of linear anisotropic diffusion, where the diffusion tensor is constructed from the same structure tensor. By means of psychometric experiments, it is found that the algorithm gives the most accurate image reproduction when used with the ΔE_{99} colour metric, and that it performs at the level of, or better than, other state-of-the-art spatial algorithms. Surprisingly, the only algorithm that can compete in terms of accuracy is a simple luminance map computed as the L^* channel of the image represented in the CIELAB colour space.**

***Index Terms*—colour-to-greyscale conversion, anisotropic diffusion, perceptual colour metrics, psychometric evaluation**

I. Introduction

For many image reproduction applications, it is necessary to convert colour images to greyscale. The most common way to achieve this, is to replace the colours with a correlate of their luminance values. These can be computed, e.g., as the L^* channel of CIELAB or as a weighted sum of the RGB colour channels. Although this approach works well for many applications, there is a significant risk of loosing image information in the conversion, since all chrominance information is lost.

Bala and Braun [1] proposed a method to convert business graphics from colour to grayscale to preserve detail visibility. Their method depended on a discrete and finite colour palette and made a global mapping from that to the greyscale values. Rasche et al. [2] followed a similar approach, but formulated it as a discrete optimization problem, optimising for global contrast between colours. Grundland and Dodgson [3] focused instead on the ordering of colours, and construced an algorithm that produced a continuous mapping with global consistency, greyscale preservation, and a consistent ordering of luminance, saturation and hue. Kim et al. [4] improved on the ideas of Rasche et al. [2] by using a perceptual colour difference equation based on CIELAB ΔE_{ab}, and optimising for a global mapping that was even temporally coherent for video conversion. Cui et al. [5] found the greyscale image by fitting the parameters of a prescribed linear global mapping in order to minimize the loss of local gradients. Very recently, Alsam and Rivertz [6] developed an image dependent global luminance mapping method based on defining the grey axis as a linear, quadratic, or root polynomial of the RGB channels with weights optimized for preserving the structure tensor of the original image.

Gooch et al. [7] introduced the idea to preserve local colour differences, resulting in a discrete optimization problem that was solved by conjugated gradient descent. Alsam and Drew [8] followed a similar approach, but local colour differences were computed in the RGB domain, and the optimization problem was formulated as a Poisson equation that was solved by Jacobi iteration.

A combination of global and local approaches were followed by Du et al. [9]. The local solution was found by optimising a quadratic cost function. Their method was found to outperform all previous methods in a perceptual experiment. Similarly, Smith et al. [10] followed a similar approach, first ordering the greys globally taking the Helmholtz–Kohlrausch effect into account, and secondly improving the resulting image locally in order to preserve the local contrast. In a thorough perceptual evaluation of colour-to-greyscale conversion algorithms performed by Čadík [11], the algorithms by Smith et al. [10] and Grundland et al. [3] were found to have the best overall performance in terms of both accuracy and preference.

In the present paper, we build on the ideas of Alsam and Drew [8] with two major improvements. Firstly, we use perceptual Riemannian or 'Riemannized' colour differences [12] instead of Euclidean RGB colour differences. Secondly, we change the Poisson equation to anisotropic diffusion [13] based on diffusion tensors computed from the image structure tensor [14]. The resulting algorithm is tested against the methods of Du et al. [9] and Smith et al. [10] as well as standard luminance maps for various parameters and perceptual colour metrics.

II. Proposed Algorithm

Let $\mathbf{u} : \Omega \to C$ denote the original colour image, where $\Omega \subset \mathbb{R}^2$ is the image domain and $C \subset \mathbb{R}^3$ is the colour space.

2018 Colour and Visual Computing Symposium (CVCS)
978-1-5386-5645-7/18/$31.0

The contravariant components[1] of $\mathbf{u}$ are denoted $u^\mu(x^i)$ where μ indexes the three colour components, and i the two spatial image dimensions. It was demonstrated by Pant and Farup [12] that the most common perceptual colour difference metrics (ΔE_{ab} etc.) either are Riemannian, or can be suitably 'Riemannized' by local linearization. Letting $g_{\mu\nu}(\mathbf{u})$ denote the components of the Riemannian metric tensor $\boldsymbol{g}$ (derived from the colour metric in accordance with [12]), the components of the structure tensor, $\boldsymbol{S}$ [14], can be generalized as

$$S_{ij} = g_{\mu\nu}(\mathbf{u}) \frac{\partial u^\mu}{\partial x^i} \frac{\partial u^\nu}{\partial x^j}\,, \tag{1}$$

with eigenvalues λ^+ and λ^- and corresponding normalized eigenvectors $\mathbf{e}^+$ and $\mathbf{e}^-$ that are stored as columns in the eigenvector matrix $\boldsymbol{E}$.

A one-channel estimate measure of the gradient of the image can be taken as one of the following

$$\mathbf{v} = \mathbf{e}^+ \sqrt{\lambda^+}\,, \tag{2}$$

$$\mathbf{v} = \mathbf{e}^+ \sqrt{\lambda^+ - \lambda^-}\,. \tag{3}$$

Since there is an ambiguity in this definition in that both $\mathbf{e}^\pm$ and $-\mathbf{e}^\pm$ will be valid eigenvectors corresponding to the eigenvalues $\lambda^\pm$, we select the sign such that $\mathbf{e}^+ \cdot \nabla u^L \geq 0$, where u^L denotes the luminance channel of the original image (in our implementation we use CIELAB L^* for this). We can then find a greyscale image c corresponding to this prescribed image gradient by solving the Poisson equation

$$\nabla^2 c = \nabla \cdot \mathbf{v} \tag{4}$$

by means of gradient descent,

$$\frac{\partial c}{\partial t} = \nabla^2 c - \nabla \cdot \mathbf{v} = \nabla \cdot (\nabla c - \mathbf{v}), \tag{5}$$

subject to the constraint $c \in [0, 1]$.

The main drawback of this approach is the potential production of halos and other artefacts. This can be strongly reduced by using anisotropic diffusion instead. The diffusion tensor can, in agreement with Sapiro and Ringach [13], be defined as

$$\boldsymbol{D} = \boldsymbol{E}^T \operatorname{diag}(d(\lambda^+), d(\lambda^-))\boldsymbol{E}, \tag{6}$$

where $d(\lambda)$ is a nonlinear diffusion coefficient function,

$$d(\lambda) = \frac{1}{1 + \kappa\lambda^2}\,, \tag{7}$$

and κ is a suitably chosen numeric constant. The anisotropic diffusion equation can then be written

$$\frac{\partial c}{\partial t} = \nabla \cdot (\boldsymbol{D}(\nabla c - \mathbf{v})), \tag{8}$$

subject to the constraint $c \in [0, 1]$.

[1]We do not require the colour space to be Euclidean, and thus we have to distinguish between contravariant and covariant vector and tensor components. These are indicated using raised and lowered greek indices, respectively. For the spatial coordinates with latin indices, this distinction is not important, since the image plane is taken to be Euclidean, and the coordinates Cartesian. Einstein's summation convention is used throughout.

The algorithm has been implemented in Python[2] within the framework by Farup [15] using FDM with explicit Euler time integration, forward differences for the gradient and structure tensor and backward differences for the divergence in order to ensure an overall well balanced numerical scheme. Although this is an efficient numerical method, it is still significantly computationally heavier than global methods.

III. Experimental Evaluation

A. Preliminary Studies

The algorithm was run on all the images of the Kodak[3] (24 images), CSDD [9] (22 images), and CSIQ [16] (30 images) data sets with both alternatives (2) and (3) for the gradient, both isotropic (5) and anisotropic (8) diffusion (with $\kappa = 10^4$), and the ΔE_{ab} [17], ΔE_{uv} [17], ΔE_{00} [18], ΔE_E [19], $\Delta E_{99\{\cdot,b,c,d\}}$ [5], and the hyperbolic $\Delta E_{99c,hyp}$ [20] colour metrics for computing the structure tensor (1), resulting in a total of $(24 + 22 + 30) \times 2 \times 2 \times 9 = 2736$ greyscale image reproductions. The following observations were made:

- There were no visible and barely any measurable differences between the two strategies (2) and (3) for computing the gradient of the colour image (average PSNR of 50.6 dB with a standard deviation of 2.9 dB). Thus, we decided to use (3) for the following experiments in agreement with Sapiro and Ringach [13].
- Whenever there was a visible difference between the isotropic (5) and anisotropic (8) methods, it was clearly in favour of the anisotropic method, being without the halo artefacts typically produced by the isotropic method. Thus, we decided to use only the anisotropic method for the following experiments.
- Despite performing better as a colour difference metric, there were no visible and barely any measurable difference between the Euclidean ΔE_{99c} and its hyperbolic counterpart $\Delta E_{99c,hyp}$ (average PSNR of 52.0 dB with a standard deviation of 6.1 dB). The few cases where there was a significant difference between the images were examined visually, but no clear trend was observed in favour of one over the other. Thus, we decided to use only the original Euclidean version for the following experiments.
- A small psychometric pair comparison experiment with two observers evaluating 9 different images reproduced with the different parameters was conducted to select the best parameters for the algorithm. The results indicated that the Euclidean ΔE_{99} overall was more accurate.

B. Psychometric Experiment

The preliminary experiment showed that the anisotropic Euclidean ΔE_{99} method was the overall most accurate, and therefore this is selected for the comparison against other methods. We also included the ΔE_{00}, which is widely assumed to be the best currenct colour metric. We evaluate this

[2]Available from https://github.com/ifarup/colourlab since v0.0.3 as `colourlab.image.c2g_diffusion`.

[3]Downloaded from http://r0k.us/graphics/kodak/ January 2018.

TABLE I
ALGORITHMS INCLUDED IN THE PSYCHOMETRIC EXPERIMENT.

Algorithm	Source	Parameters
CIELAB L^*	–	–
ΔE_{00}	proposed	$\kappa = 10^4$
ΔE_{99}	proposed	$\kappa = 10^4$
Du	[9]	–
Smith	[10]	$r = 5, a = 0.15, \gamma = 1$

TABLE II
UNCORRECTED p-VALUE RESULTS OF THE BINOMIAL TEST FOR THE ACCURACY EXPERIMENT. THE GREEN COLOUR DENOTES REJECTION OF THE NULL HYPOTHESIS AT THE 5% SIGNIFICANCE LEVEL.

	ΔE_{00}	ΔE_{99}	Du	Smith
L^*	0.0114	0.6831	0.0000	0.0275
ΔE_{00}		0.0000	0.0179	0.1208
ΔE_{99}			0.0000	0.5675
Du				0.0000

method against CIELAB L^*, being one of the most commonly used methods for colour to gray. We also include the method from Smith et al. [10], as this was found to best the most accurate colour to gray method in the study by Ĉadík [11]. At last we include the method from Du et al., which was found by the authors to be superior in a psychometric experiment [9]. An overview of the algorithms and parameters included in the experiment is listed in Table I.

Five images from the CSIQ [16] and five images from Kodak data set were selected for the experiment (Figure 1). The images were chosen based on covering different attributes (different levels of lightness, different levels of saturation, larger areas of the same color, fine details, memory colors, and different hues) following the recommendations from Field [21].

The first evaluation was done as a paired comparison experiment using QuickEval [22]. 15 observers, 2 female and 13 male, with an average age of 35 years. All observers had normal or corrected-to-normal vision. The viewing distance was set to approximately 50 cm, and the ambient illumination was dark. The experiment was done on a BenQ SVV2700-B, calibrated to sRGB (80 cd/m^2 and a color temperature of 6500K). The original color image was shown in the middle with two reproductions on each side. The observers were asked to 'Select the most accurate reproduction of the color image'. This experiment is hereby referred to as the accuracy experiment.

The second evaluation was also done as a paired comparison experiment using QuickEval [22]. 16 observers participated in an uncontrolled online experiment rating the same images as in the accuracy experiment. Two reproductions were shown (without the original) and the observers were asked to 'select the image you prefer'. This experiment is referred to as the preference experiment.

The raw data from the psychometric experiment is processed into z-scores [23] with 95% confidence intervals. In addition, we analyze the raw data with a binomial test at 5% significance level with Bonferroni [24] correction to counteract the problem of multiple comparison. For completeness, we tested also with the Sidák [25] correction, and came to exactly the same conclusions.

IV. RESULTS AND DISCUSSION

A. Behaviour of the proposed algorithm

Resulting images from running the algorithms on the test images are shown in Figure 2 and 3 for the Kodak and CSIQ images, respectively. Most of the algorithms provide stronger local contrasts than the classic CIELAB L^*, but this also makes the L^* images smooth and pleasing. The local behaviour of the proposed algorithm can in some cases amplify textures and noise in the original images to an unwanted degree. See, e.g., the sky in the third image of the CSIQ data set. The Du algorithm in some cases makes surprising choices in the reproduction of saturated colours, in particular red, see, e.g, the hat in the first image of the Kodak data set. For some of the images, the Smith algorithm produces images that are somewhat darker than what the other algorithms do. All in all, the differences between the different algorithms are small but visible.

B. Results of the psychometric experiment

The resulting Z-scores from the psychometric accuracy experiment is shown in Figure 4. The non-overlapping confidence intervals indicate that L^*, ΔE_{99} and Smith performs statistically significantly better than ΔE_{00} and Du, and that ΔE_{00} performs statistically significantly better than Du.

However, the Z-scores are computed in agreement with Thurstone's case V [23], where the assumptions are that all the measured variables are normally distributed with the same variance. That is not necessarily the case here, and has not been tested. Thus, we also computed the binomial test for all the pairs of algorithms, for which no such assumptions are made. The resulting p-values (before Bonferroni correction) listed in Table II show that L^*, ΔE_{99} and Smith perform statistically significantly better than Du, and that ΔE_{99} performs statistically significantly better than ΔE_{00}.

Figure 5 shows the z-score for the preference experiment. CIELAB L*, ΔE_{99} and Smith perform significantly better than ΔE_{99} and Du. Compared to Figure 4 the results are similar, but for the preference experiment Du is not significantly different than ΔE_{00}, indicating that Du performs better in the preference experiment than in the accuracy experiment. Table III shows the results of the binomial test for the preference experiment. We can notice that the Du algorithm performs better compared to the accuracy experiment.

Despite their clearly well-founded constructions, one of the proposed algorithms perform significantly better than L^*, neither in terms of accuracy nor preference. This is contradicting the results of Čadík [11], who found that the Smith algorithm outperformed CIE L^* both in terms of accuracy and preference. However, Čadík used a mix of artificial and photographic images, whereas only photographic images were used in this study. This can influence the results significantly. Secondly,

Fig. 1. Images used for the psychometric experiment. The images in the top row are from from CSIQ and the images in the bottom row are from the Kodak data set.

Fig. 2. Resulting greyscale images for the Kodak test images shown in Figure 1. Left to right: CIELAB L^*, ΔE_{00}, ΔE_{99}, Du, and Smith.

Fig. 3. Resulting greyscale images for the CSIQ test images shown in Figure 1. Left to right: CIELAB L^*, ΔE_{00}, ΔE_{99}, Du, and Smith.

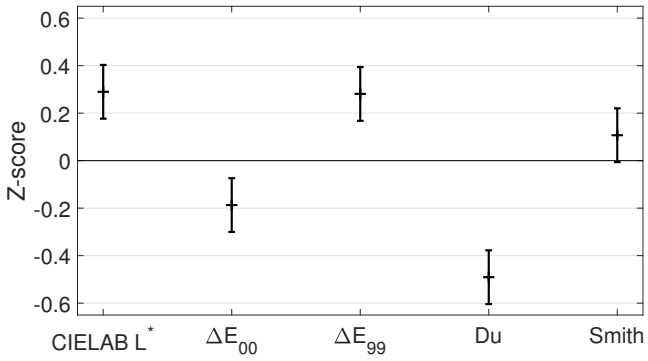

Fig. 4. Z-score with 95% confidence intervals from the psychometric experiment for accuracy.

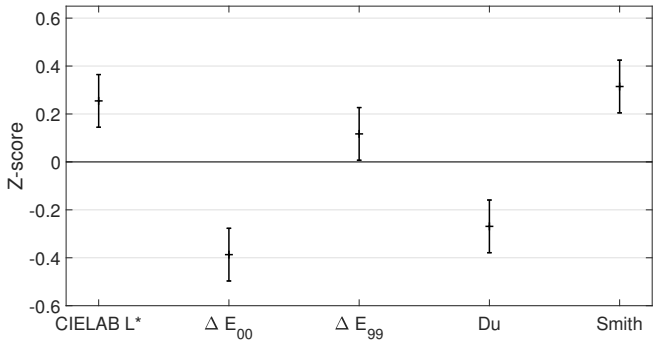

Fig. 5. Z-score with 95% confidence intervals from the psychometric preference experiment.

far more observers were used in Čadík's experiments, making it easier to obtain statistically significant results.

V. Conclusion

In terms of accuracy, the proposed algorithm for conversion of colour images to greyscale based on linear anisotropic diffusion of local perceptual colour difference metrics performs better than some of the state-of-the-art spatial algorithms and at the same level as others, including simple L^* luminance maps. For the anisotropic diffusion algorithm, the best results are obtained with the ΔE_{99} colour metric. The same trends are found also when the observers are asked to rate their preference rather than the accuracy of the reproduction.

TABLE III
Uncorrected p-value results of the binomial test for the preference experiment. The green colour denotes rejection of the null hypothesis at the 5% significance level.

	ΔE_{00}	ΔE_{99}	Du	Smith
L^*	0.0001	0.0409	0.2205	0.6832
ΔE_{00}		0.0000	0.3692	0.0000
ΔE_{99}			0.0069	0.0409
Du				0.0000

References

[1] R. Bala and K. M. Braun, "Color-to-grayscale conversion to maintain discriminability," in *Color Imaging IX: Processing, Hardcopy, and Applications*, vol. 5293. International Society for Optics and Photonics, 2003, pp. 196–203.

[2] K. Rasche, R. Geist, and J. Westall, "Re-coloring images for gamuts of lower dimension," *Computer Graphics Forum*, vol. 24, no. 3, pp. 423–432, 2005.

[3] M. Grundland and N. A. Dodgson, "Decolorize: Fast, contrast enhancing, color to grayscale conversion," *Pattern Recognition*, vol. 40, no. 11, pp. 2891–2896, 2007.

[4] Y. Kim, C. Jang, J. Demouth, and S. Lee, "Robust color-to-gray via nonlinear global mapping," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5, p. 161, 2009.

[5] M. Cui, J. Hu, A. Razdan, and P. Wonka, "Color-to-gray conversion using isomap," *The Visual Computer*, vol. 26, no. 11, pp. 1349–1360, 2010.

[6] A. Alsam and H. J. Rivertz, "A mathematical approach to best luminance maps," *JOSA A*, vol. 35, no. 4, pp. B239–B243, 2018.

[7] A. A. Gooch, S. C. Olsen, J. Tumblin, and B. Gooch, "Color2gray: salience-preserving color removal," *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, pp. 634–639, 2005.

[8] A. Alsam and M. S. Drew, "Fast colour2grey," in *Color and Imaging Conference*, vol. 2008. Society for Imaging Science and Technology, 2008, pp. 342–346.

[9] H. Du, S. He, B. Sheng, L. Ma, and R. W. Lau, "Saliency-guided color-to-gray conversion using region-based optimization," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 434–443, 2015.

[10] K. Smith, P.-E. Landes, J. Thollot, and K. Myszkowski, "Apparent greyscale: A simple and fast conversion to perceptually accurate images and video," in *Computer Graphics Forum*, vol. 27, no. 2. Wiley Online Library, 2008, pp. 193–200.

[11] M. Ĉadík, "Perceptual evaluation of color-to-grayscale image conversions," in *Computer Graphics Forum*, vol. 27, no. 7. Wiley Online Library, 2008, pp. 1745–1754.

[12] D. Raj Pant and I. Farup, "Riemannian formulation and comparison of color difference formulas," *Color Research & Application*, vol. 37, no. 6, pp. 429–440, 2012.

[13] G. Sapiro and D. L. Ringach, "Anisotropic diffusion of multivalued images with applications to color filtering," *IEEE Transactions on image processing*, vol. 5, no. 11, pp. 1582–1586, 1996.

[14] S. Di Zenzo, "A note on the gradient of a multi-image," *Computer vision, graphics, and image processing*, vol. 33, no. 1, pp. 116–125, 1986.

[15] I. Farup, "A computational framework for colour metrics and colour space transforms," *PeerJ Comp. Sci.*, vol. 2, p. e48, 2016.

[16] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, 2010.

[17] CIE, *Colorimetry*, 2nd ed., ser. CIE Publications. Vienna, Austria: Central Bureau of the CIE, 1986, vol. 15.2.

[18] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Res. Appl.*, vol. 26, no. 5, pp. 340–350, 2001.

[19] C. Oleari, M. Melgosa, and R. Huertas, "Euclidean color-difference formula for small–medium color differences in log-compressed OSA-UCS space," *J. Opt. Soc. Am. A*, vol. 26, no. 1, pp. 121–134, 2009.

[20] I. Farup, "Hyperbolic geometry for colour metrics," *Optics Express*, vol. 22, no. 10, pp. 12 369–12 378, 2014.

[21] G. G. Field, "Test image design guidelines for color quality evaluation," in *Color and Imaging Conference*, vol. 1999, no. 1. Society for Imaging Science and Technology, 1999, pp. 194–196.

[22] K. Van Ngo, J. J. Storvik, C. A. Dokkeberg, I. Farup, and M. Pedersen, "Quickeval: a web application for psychometric scaling experiments," in *Image Quality and System Performance XII*, vol. 9396. International Society for Optics and Photonics, 2015, p. 93960O.

[23] P. G. Engeldrum, *Psychometric scaling: a toolkit for imaging systems development*. Imcotek, 2000.

[24] C. Bonferroni, "Teoria statistica delle classi e calcolo delle probabilita," *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, vol. 8, pp. 3–62, 1936.

[25] Z. Šidák, "Rectangular confidence regions for the means of multivariate normal distributions," *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 626–633, 1967.