Robust Color-to-gray via Nonlinear Global Mapping

Yongjin Kim Cheolhun Jang Julien Demouth Seungyong Lee
Pohang University of Science and Technology (POSTECH)

Abstract

This paper presents a fast color-to-gray conversion algorithm which robustly reproduces the visual appearance of a color image in grayscale. The conversion preserves feature discriminability and reasonable color ordering, while respecting the original lightness of colors, by simple optimization of a nonlinear global mapping. Experimental results show that our method produces convincing results for a variety of color images. We further extend the method to temporally coherent color-to-gray video conversion.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms

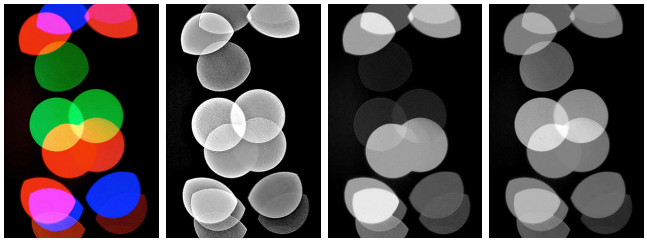
Keywords: color-to-grayscale, video decolorization

1 Introduction

Color-to-gray conversion is required in many applications, including black-and-white printing, single channel image processing, and non-photorealistic rendering with black-and-white media. The conversion should preserve the appearance of original color images in the resulting grayscale images. Unfortunately, a traditional method using the intensity channel (e.g., CIE Y) may fail to achieve this goal by losing feature discriminability in iso-luminant regions.

To retain feature discriminability in color-to-gray conversion, color differences in the input image should be reflected as much as possible onto the converted grayscale values. Several color-to-gray conversion methods have been proposed, which can be divided into two main categories: *local mapping* and *global mapping*. In a local mapping method, the color-to-gray mapping of pixel values is spatially varying, depending on the local distributions of colors. Although a local mapping has advantages in accurately preserving local features, constant color regions could be converted inhomogeneously if the mapping changes in the regions (Fig. 1(b)). In a global mapping method, the same color-to-gray mapping is used for all pixels in the input. It consistently maps the same colors to the same grayscale values over an image, guaranteeing homogenous conversion of constant color regions. However, it would be more challenging to determine such a global mapping that preserves local features at different locations at the same time (Fig. 1(c)).

In this paper, we present a fast and robust color-to-gray conversion algorithm using a nonlinear global mapping. Our global mapping is a nonlinear function of the lightness, chroma, and hue of colors. When a color image is given, the parameters of the function are optimized to preserve feature discriminability and color ordering in the color-to-gray conversion. In addition, the nonlinear function is designed to respect the original lightness of colors in the conversion. As a result, our nonlinear global mapping method can successfully reproduce the visual appearance of color images (Fig. 1(d)). We show the robustness of our method by testing it with diverse images used in recent user experiments [Čadík 2008], performed for the state-of-the-art color-to-gray conversion algorithms.



(a) Input (b) Apparent (c) Decolorize (d) Our result

Figure 1: *Comparison of color-to-gray image conversion results. (a) Input image. (b) Local mapping by Apparent Greyscale [Smith et al. 2008]. (c) Linear global mapping by Decolorize [Grundland and Dodgson 2007]. (d) Our nonlinear global mapping. Apparent Greyscale and Decolorize are respectively the best ranked local and global mapping algorithms in the user experiments by Čadík [2008]. The color image is courtesy of Laura O'Halloran (http://www.flickr.com/photos/mrsenil/3074871806/).*

We also present a temporally coherent video stream conversion method by extending our image method. Temporally coherent video conversion has only been addressed in a local mapping method by Smith et al. [2008], where each frame is processed with common fixed parameters. However, when the scene changes, their method may require user intervention to adjust the parameters. To maintain temporal coherency while providing adaptive conversion at each frame, our simple but novel idea is to make the temporal variation of grayscale values similar to that of colors. Experimental results show that this idea works well both with static and dynamic videos, running at interactive frame rates.

In summary, the specific contributions of this paper are

- Fast and robust color-to-gray conversion using a nonlinear global mapping to preserve visual appearance of color images.
- Temporally coherent conversion of streaming videos at interactive frame rates.

2 Previous Work

Several local and global mapping algorithms have been proposed for color-to-gray conversion. In local mapping approaches, Bala and Eschbach [2004] presented a method which adds high-frequency chromatic components to luminance. Neumann et al. [2007] reconstructed the grayscale image from the gradients of a color image. Smith et al. [2008] decomposed the image into several frequency components and adjusted combination weights using chromatic channels. These local mapping algorithms effectively preserve local features but may distort appearances of constant color regions (Fig. 1(b)).

In global mapping approaches, Gooch et al. [2005] tried to preserve color differences between all pairs of pixels. Rasche et al. [2005b] also tried to preserve color differences but considered all pairs of existing colors, instead of pixels. Despite the technical soundness, these methods require heavy computation.

To reduce the computation, a global mapping can be constrained to have a simple functional form. Rasche et al. [2005a] projected

colors onto a linear axis in a 3D color space, where a linear mapping was optimized for local feature preservation. However, due to the limited flexibility of a linear mapping, the lightness of colors can be changed in the projection with possible inversion of desired color ordering (e.g., luminance ordering). Grundland and Dodgson [2007] proposed a fast linear mapping algorithm that adds a fixed amount of chrominance to the luminance, where the original luminance and color ordering can be better preserved by restraining the added amount of chrominance. However, in the final addition of the luminance and chrominance, their contrasts may cancel out, causing feature losses (green circles in Fig. 1(c) and tulips in Fig. 3). In addition, a linear global mapping cannot discriminate colors aligned perpendicular to the projection axis (blue and red circles in the bottom right of Fig. 1(c)). Our method reduces these artifacts by optimizing a *nonlinear* global mapping for a given color image.

3 Nonlinear Global Mapping

Our color-to-gray conversion algorithm considers the following criteria to preserve the visual appearance of a color image:

- *Mapping consistency:* Same colors should be mapped to the same grayscale values.
- *Feature preservation:* Features in the color image should remain discriminable in the grayscale image.
- *Ordering preservation:* Reasonable ordering of colors should be respected in their converted grayscale values.
- *Lightness fidelity:* Color and grayscale images should have similar lightness stimuli.

For *mapping consistency*, our method uses a global mapping for color-to-gray conversion. This approach prevents inhomogeneous conversions of constant color regions, which could disrupt the appearance of images, e.g., diagrams and illustrations.

In this paper, a global mapping is specified by a simple nonlinear functional form,

$$g(x,y) \quad = \quad L + f(\theta)C, \tag{1}$$

where L, C, and θ are the lightness, chroma, and hue angle of the color value at a pixel (x,y) in the CIE LCH color space, respectively. Because the hue angle θ has 2π periodicity, we model the function $f(\theta)$ as a trigonometric polynomial by

$$f(\theta) \quad = \quad \sum_{k=1}^{n} (A_k \cos k\theta + B_k \sin k\theta) + A_0, \tag{2}$$

where A_k, B_k, and A_0 are unknown parameters that are optimized to preserve the visual appearance in color-to-gray conversion.

To retain *feature discriminability*, local differences among color pixels should be similarly reproduced in the corresponding grayscale values. We use image gradients to represent local pixel differences, and minimize the difference between color and grayscale image gradients. Because a color space is 3D and grayscale is 1D, we have to define 1D measurements of the color differences of pixels. In this paper, we use a normalized distance of two colors in the CIE $L^*a^*b^*$ color space for the measurement, similar to [Gooch et al. 2005]. However, color-to-gray conversion is a dimensionality reduction process, and the normalized distances cannot be preserved by a global mapping for all pairs of neighboring color pixels. To reproduce the features as much as possible in the conversion, we determine the unknown parameters of $f(\theta)$ so that the differences of neighboring grayscale pixels are made as similar as possible to the normalized distances of corresponding color pixels. Consequently, the global mapping g adaptively changes with the distribution of color values in a given image. Specific formulations for preserving feature discriminability are described in Sec. 4.

When we define the 1D measurement of a color difference, we also introduce the sign of the measurement, which gives an ordering of two colors. With this approach, by minimizing the gradient differences between color and grayscale images, we can preserve a desired *color ordering* in color-to-gray conversion.

A global mapping g defined by Eq. (1) is well-suited for achieving *lightness fidelity*. When the chroma C is zero, a given color is gray and will not be changed by mapping g. Similarly, when C is small, a given color is less saturated and the lightness of the pixel would not change much in the conversion. If we have enough almost-gray pixels, these pixels will guide the conversion of other pixels through the image gradients, preventing abrupt changes of lightness. Even when the input color image does not contain gray pixels, the lightness L is a major factor in Eq. (1) and dominates the converted grayscale values.

By modeling $f(\theta)$ with Eq. (2), Eq. (1) contains $C\cos\theta$ and $C\sin\theta$, which are the same as a^* and b^* in the CIE $L^*a^*b^*$ color space, respectively. This means that our nonlinear mapping extends the linear combination of L, a^*, and b^*, which is used in [Rasche et al. 2005a], except that L is not scaled in our mapping. Because the YPQ space used in [Grundland and Dodgson 2007] is defined similarly to CIE $L^*a^*b^*$, our mapping could cover their linear model.

Eq. (1) has a similar form to a Helmholtz-Kohlrausch (H-K) lightness predictor of Nayatani [1997], $L^{HK} = L + Lf_1(\theta')S$, where saturation S and hue θ' are defined by the CIE LUV representation of a color. $f_1(\theta')$ has the same form as Eq. (2) with $n = 4$ but the parameters are fixed constants determined by fitting the function to an experimental data set of perceptual lightness of colors. In [Nayatani 1997], this functional form was used to model the influence of chromatic components on the perceptual lightness of an isolated color. In contrast, we use a similar functional form to respect local color differences in color-to-gray conversion, and the parameters are adaptively determined for a given image.

4 Optimization

4.1 Objective Function

To preserve features in color-to-gray conversion, we minimize the difference of image gradients between the input color and the resulting grayscale images. The energy function is defined by

$$E_s \quad = \quad \sum_{(x,y)\in\Omega} \|\nabla g(x,y) - \mathbf{G}(x,y)\|^2, \tag{3}$$

where (x,y) is a pixel in an image Ω. ∇g is the gradient of the grayscale image, defined by

$$\nabla g(x,y) = \left(g(x{+}1,y){-}g(x{-}1,y),\, g(x,y{+}1){-}g(x,y{-}1)\right).$$

$\mathbf{G}$ is the difference of input pixel colors $\mathbf{c}$, defined by

$$\mathbf{G}(x,y) = \begin{pmatrix} G^x(x,y) \\ G^y(x,y) \end{pmatrix}^T = \begin{pmatrix} \mathbf{c}(x+1,y) \ominus \mathbf{c}(x-1,y) \\ \mathbf{c}(x,y+1) \ominus \mathbf{c}(x,y-1) \end{pmatrix}^T.$$

The color difference operator $\ominus$ is defined by

$$\mathbf{c}_i \ominus \mathbf{c}_j = sign(\mathbf{c}_i, \mathbf{c}_j)\sqrt{{\Delta L_{ij}}^2 + \left(\alpha \frac{\sqrt{{\Delta a^*_{ij}}^2 + {\Delta b^*_{ij}}^2}}{\mathcal{R}}\right)^2}, \tag{4}$$

where L, a^*, and b^* are the CIE $L^*a^*b^*$ representation of color $\mathbf{c}$, $\Delta L_{ij} = L_i - L_j$, $\Delta a^*_{ij} = a^*_i - a^*_j$, and $\Delta b^*_{ij} = b^*_i - b^*_j$. $\mathcal{R}$ is the normalization constant, $\mathcal{R} = 2.54\sqrt{2}$, to equalize the ranges of the chromatic contrast, $\sqrt{{\Delta a^*_{ij}}^2 + {\Delta b^*_{ij}}^2}$, and the lightness contrast, ΔL_{ij}. α is a user-specified parameter to control the influence of the chromatic contrast on feature discriminability.

The $sign$ function in the color difference operator $\ominus$ encodes relative ordering of two pixel colors. For perceptually meaningful ordering, we design a prioritized sign decision scheme using different lightness-based orderings. Since an H-K effect predictor offers more accurate lightness estimation than CIE L, we assign the highest priority to the sign of the H-K effect predictor difference ΔL^{HK} and use it whenever ΔL^{HK} is non-zero. To compute ΔL^{HK}, we use the VAC model of Nayatani [1997], which is also used in [Smith et al. 2008]. If ΔL^{HK} is zero, we then use the sign of ΔL. Finally, if the sign of ΔL is also zero, we use the sign of $\Delta L^3 + \Delta a^{*3} + \Delta b^{*3}$, similarly to [Neumann et al. 2007].

4.2 Optimization with Nonlinear Global Mapping

With the definition of g in Eq. (1), the objective function E_s can be reduced to a quadratic equation of the unknown parameters of g. For compact representation, we use a vector form of Eq. (2),

$$f(\theta) \quad = \quad \mathbf{t}^T\mathbf{x},$$

where $\mathbf{t} = (t_i)$ and $\mathbf{x} = (x_i)$ are $(2n+1) \times 1$ vectors. For $1 \le i \le n$, $t_i = \cos(i\theta)$ and $x_i = A_i$. For $n+1 \le i \le 2n$, $t_i = \sin((i-n)\theta)$ and $x_i = B_{i-n}$. For $i = 2n+1$, $t_i = 1$ and $x_i = A_0$. Then,

$$E_s \quad = \quad \mathbf{x}^T\mathbf{M}_s\mathbf{x} - 2\mathbf{b}_s^T\mathbf{x} + \mathcal{C}, \tag{5}$$

where $\mathbf{M}_s = \sum_\Omega(\mathbf{u}\mathbf{u}^T + \mathbf{v}\mathbf{v}^T)$, $\mathbf{b}_s = \sum_\Omega(p\mathbf{u} + q\mathbf{v})$, $\mathbf{u} = (C\cdot\mathbf{t})_x$, $\mathbf{v} = (C\cdot\mathbf{t})_y$, $p = G^x - L_x$, and $q = G^y - L_y$. Subscripts x and y denote partial derivatives with respect to x and y directions, respectively. The partial derivatives are computed using the central differences, similar to ∇g and $\mathbf{G}$.

The energy function E_s is minimized by $\hat{\mathbf{x}}_s = \mathbf{M}_s^{-1}\mathbf{b}_s$, where $\mathbf{M}_s^{-1}$ is a $(2n+1) \times (2n+1)$ matrix and $\mathbf{b}_s$ is a $(2n+1) \times 1$ vector. n corresponds to the number of cos and sin bases in Eq. (2), which determines the degrees of freedom of mapping function g. We experimented with different values of n, and $n = 4$ proved to be a good choice. Higher values of n do not lead to significant improvements of conversion results in most cases. For $n = 4$, the matrix $\mathbf{M}_s$ is 9×9 and $\mathbf{M}_s^{-1}$ can be computed immediately.

However, $\mathbf{M}_s$ can be singular in rare cases. For example, when all image pixels have the same chroma and hue with different lightness, $\mathbf{u}$ and $\mathbf{v}$ become null vectors, making $\mathbf{M}_s$ a null matrix. Also, the elements of $\mathbf{x}$ could have large values, causing g to go out of the visible range. We resolve these problems by adding a regularization term, $E_r = \mathbf{x}^T I\mathbf{x}$, to our objective function with a weight λ. E_r increases with the magnitude of $\mathbf{x}$ and, therefore, an appropriate choice of λ would make $\mathbf{x}$ small, keeping g in the visual range. To reflect the number of terms in the summation of Eq. (3), we use the number of pixels N for λ, and obtained good results in our experiments.

After incorporating the regularization term, we have

$$E_{image} \quad = \quad E_s + \lambda E_r, \tag{6}$$

which is minimized by

$$\hat{\mathbf{x}}_{image} \quad = \quad (\mathbf{M}_s + \lambda\mathbf{I})^{-1}\mathbf{b}_s. \tag{7}$$

Notice that the size of $\mathbf{M}_s + \lambda\mathbf{I}$ remains the same as $\mathbf{M}_s$, and that $\mathbf{b}_s$ is unchanged. The vector $\hat{\mathbf{x}}_{image}$ uniquely determines $f(\theta)$ and, consequently, the global color-to-gray mapping g.

5 Temporally Coherent Video Conversion

Our image conversion method can easily be extended for temporally coherent video conversion by including the following temporal energy term,

$$E_t \;=\; \sum_{(x,y)\in\Omega} \left\{(g_c(x,y) - g_p(x,y)) - (\mathbf{c}_c(x,y) \ominus \mathbf{c}_p(x,y))\right\}^2,$$

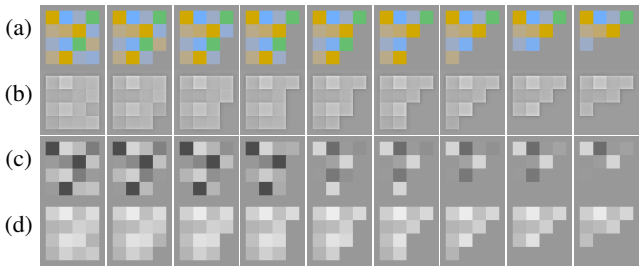


Figure 2: *Comparison of video conversion results. (a) Input video. (b) Local-mapping-based conversion [Smith et al. 2008]. (c) Frame-by-frame conversion using a global-mapping-based image method [Grundland and Dodgson 2007]. (d) Our method.*

where the subscripts c and p denote the current and previous frames, respectively. This energy term does not make consecutive grayscale values similar among frames, because such an approach would fail with dynamic scenes. Instead, it makes the grayscale variations similar to the color variations among frames in the temporal domain. Then, the resulting grayscale video would be as temporally coherent as the original color video, converting both static and dynamic scenes appropriately.

The energy function E_{video} for video conversion is defined as the combination of spatial, temporal, and regularization terms,

$$E_{video} \quad = \quad ((1-\beta)E_s + \beta E_t) + \lambda E_r, \tag{8}$$

where β is a user-specified parameter to control the relative weights of spatial and temporal terms. The minimizer of E_{video} is

$$\hat{\mathbf{x}}_{video} \quad = \quad (\mathbf{M}_v + \lambda\mathbf{I})^{-1}\mathbf{b}_v, \tag{9}$$

where $\mathbf{M}_v = (1-\beta)\mathbf{M}_s + \beta\mathbf{M}_t$, $\mathbf{b}_v = (1-\beta)\mathbf{b}_s + \beta\mathbf{b}_t$, $\mathbf{M}_t = \sum_\Omega({C_c}^2\mathbf{t}_c{\mathbf{t}_c}^T)$, and $\mathbf{b}_t = \sum_\Omega C_c(g_p + (\mathbf{c}_c \ominus \mathbf{c}_p) - L_c)\mathbf{t}_c$.

6 Results

We implemented our method in C++ using OpenCV 1.1pre1 and OpenMP 2. All experiments were performed on a PC with a 2.66 Ghz Intel Core i7 CPU and 6 GB memory. We tested our implementation both using single and quad cores. For a 320×240 video, single core implementation took 0.102 seconds, and the quad core took 0.058 seconds to process each frame. The computation time is almost linearly proportional to the number of pixels in an image or video.

In our experiments, we used $\gamma = 2.4$ for gamma correction to linearize the input sRGB color image before processing. In RGB to CIE $L^*a^*b^*$ conversion, we assumed D65 reference white. For user parameters, we used $\alpha = 1.0$ and $\lambda = N$ (the number of pixels) for image conversion. For videos, we used $\beta = 0.5$. These fixed values of parameters worked well for all test images and videos.

We compared the results of our method with previous techniques, using the test images chosen by Čadík [2008] to compare the state-of-the-art color-to-gray algorithms. The values of parameters to run previous techniques were taken from [Čadík 2008]. Fig. 3 shows a partial comparison table with five color images. Comparisons using other test images can be found in the supplementary material. From the comparisons, we can see that our method preserves the visual appearance more robustly than previous techniques in most test images.

In some cases, more exaggeration of chromatic features would be desirable. Then, we can use a larger α to put more emphasis on the chromatic contrasts in Eq. (4). The supplementary material contains conversion results with different values of α.

Fig. 2 shows our color-to-gray video conversion result and comparison with other methods. In the input video, iso-luminant color patches are erased one by one. While the local mapping method of Smith et al. [2008] produces a temporally coherent output, it introduces non-homogeneity artifacts near region boundaries (Fig. 2(b)).

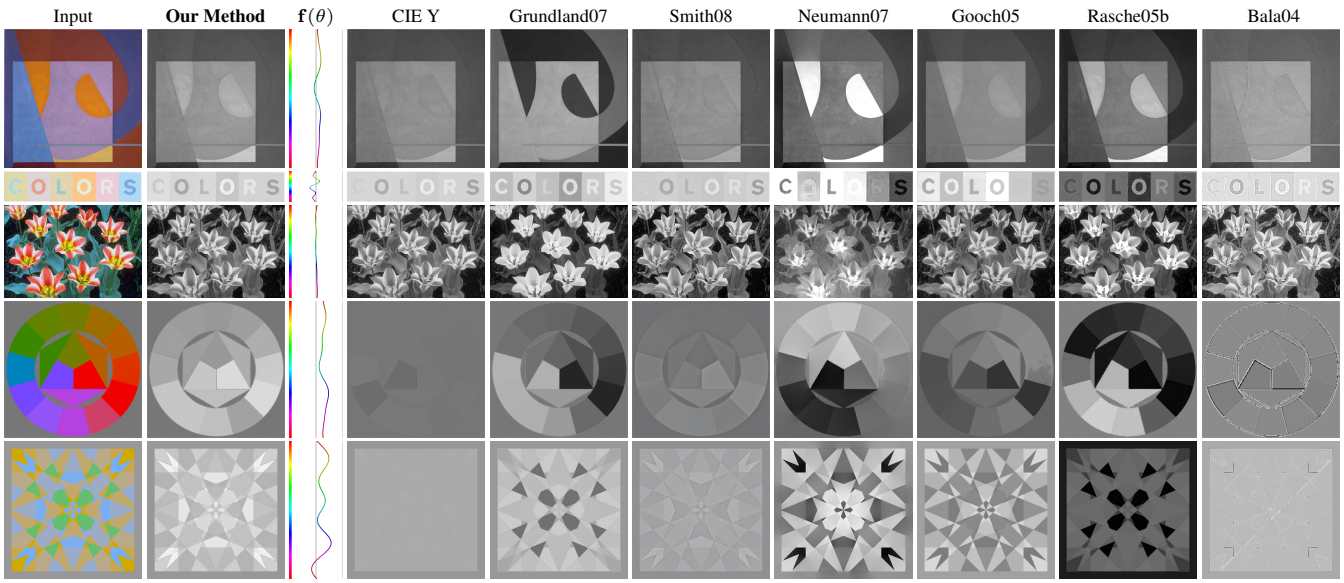


Figure 3: *Comparison with other methods. For more results and comparison, refer to the supplementary material. Result images of other methods are courtesy of Čadík [2008].*

To avoid such artifacts, a global mapping method for image conversion can be used for a video by converting each frame independently. We experimented with this approach using the method of Grundland and Dodgson [2007]. However, since an image method does not specifically consider temporal coherency, the result shows inconsistent conversions among consecutive frames (Fig. 2(c)). In contrast, our result show temporally coherent grayscale frames while discriminating different color regions (Fig. 2(d)).

7 Discussion and Future Work

Our color-to-gray conversion does not explicitly consider features from global contrasts but focuses more on local features. As a result, different colors of distant pixels can be mapped to similar grayscale values. Incorporating the differences among distant pixels into the energy function can be a solution for this problem. However, as shown in [Gooch et al. 2005], this approach may lose small feature details. A multi-resolution framework, similar to [Mantiuk et al. 2006], could be used for providing a conversion method to preserve both local and global features simultaneously.

Our prioritized sign decision scheme, which determines the color ordering, results in perceptually consistent grayscale orderings in most cases (Fig. 3). Our scheme is based on the observation that more-saturated colors are perceived to be brighter than their luminance [Nayatani 1997]. However, in some cases, users may prefer to map less-saturated colors to brighter grayscale values. Neumann et al. [2007] modeled this effect but did not give a specific formula. Color ordering is still an active research topic in color theory, and our sign decision scheme has room for improvement.

Color appearance phenomena related to spatial interactions, such as filtering effects and simultaneous contrasts [Fairchild 2005], are worth considering in color-to-gray conversion. We expect that globally mapped grayscale values reproduce such phenomena in some degree by spatially interacting with their neighborhoods, in a similar way to the corresponding color values. Still, more-accurate reproduction of color appearance phenomena could be achieved by explicitly considering them in the conversion process.

Acknowledgements

The authors thank anonymous reviewers for their valuable comments. We also thank Sunghyun Cho for implementation support, and Kaleigh Smith for providing the implementation of Apparent Greyscale [Smith et al. 2008]. This work was supported by the IT R&D program of MKE/MCST/KEIT (2008-F-031-01, Development of Computational Photography Technologies for Image and Video Contents) and the ERC program of MEST/NRF (R11-2008-007-01002-3).

References

BALA, R., AND ESCHBACH, R. 2004. Spatial color-to-grayscale transform preserving chrominance edge information. In *Proc. Color Imaging Conference 2004*, 82–86.

ČADÍK, M. 2008. Perceptual evaluation of color-to-grayscale image conversions. *Computer Graphics Forum (Proc. Pacific Graphics 2008) 27*, 7, 1745–1754.

FAIRCHILD, M. D. 2005. *Color Appearance Models*. Wiley.

GOOCH, A. A., OLSEN, S. C., TUMBLIN, J., AND GOOCH, B. 2005. Color2gray: salience-preserving color removal. *ACM Trans. Graphics (Proc. SIGGRAPH 2005) 24*, 3, 634–639.

GRUNDLAND, M., AND DODGSON, N. A. 2007. Decolorize: Fast, contrast enhancing, color to grayscale conversion. *Pattern Recognition 40*, 11, 2891–2896.

MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Applied Perception 3*, 3, 286–308.

NAYATANI, Y. 1997. Simple estimation methods for the helmholtz-kohlrausch effect. *Color Research and App. 22*, 6, 385–401.

NEUMANN, L., ČADÍK, M., AND NEMCSICS, A. 2007. An efficient perception-based adaptive color to gray transformation. In *Proc. Computational Aesthetics 2007*, 73– 80.

RASCHE, K., GEIST, R., AND WESTALL, J. 2005. Detail preserving reproduction of color images for monochromats and dichromats. *IEEE Computer Graphics and Applications 25*, 3, 22–30.

RASCHE, K., GEIST, R., AND WESTALL, J. 2005. Re-coloring images for gamuts of lower dimension. *Computer Graphics Forum (Proc. Eurographics 2005) 24*, 3, 423–432.

SMITH, K., LANDES, P.-E., THOLLOT, J., AND MYSZKOWSKI, K. 2008. Apparent greyscale: A simple and fast conversion to perceptually accurate images and video. *Computer Graphics Forum (Proc. Eurographics 2008) 27*, 2, 193–200.