Using the Monge-Kantorovitch Transform in Chromagenic Color Constancy for Pathophysiology

Ghalia Hemrit[1(✉)], Futa Matsushita[2], Mihiro Uchida[2], Javier Vazquez-Corral[1], Han Gong[1], Norimichi Tsumura[2], and Graham D. Finlayson[1]

[1] University of East Anglia, Norwich NR4 7TJ, UK
ghalia.hemrit@gmail.com
https://www.uea.ac.uk/computing/colour-and-imaging
[2] Chiba University, Chiba 263-8522, Japan
http://www.mi.tj.chiba-u.jp/

Abstract. The Chromagenic color constancy algorithm estimates the light color given two images of the same scene, one filtered and one unfiltered. The key insight underpinning the chromagenic method is that the filtered and unfiltered images are linearly related and that this linear relationship correlates strongly with the illuminant color. In the original method the best linear relationship was found based on the assumption that the filtered and unfiltered images were registered. Generally, this is not the case and implies an expensive image registration step.

This paper makes three contributions. First, we use the Monge-Kantorovich (MK) method to find the best linear transform without the need for image registration. Second, we apply this method on chromagenic pairs of facial images (used for Kampo pathophysiology diagnosis). Lastly, we show that the MK method supports better color correction compared with solving for a 3×3 correction matrix using the least squares linear regression method when the images are not registered.

Keywords: Color constancy · Chromagenic computer vision

1 Introduction

Color constancy is the vision property that allows humans to identify the color of an object independently of the color of the light source. For example, we are able to perceive a banana as yellow both in a room illuminated with a tungsten bulb – i.e. under reddish light – and outside in a cloudy day i.e. under bluish light. This property means that solving for color constancy – in other words, removing the color of the light – is a fundamental step in digital color image processing.

The chromagenic method for color constancy [3,4] solves for the illuminant color given 2 images of the same scene, one captured with a color filter and another without. This method can be decomposed in two different steps. In the first step, a set of linear transform matrices are calculated using a set of pairs of filtered and unfiltered images. In particular, each of these matrices relates a particular unfiltered image to its filtered counterpart. In the second step, given a pair of new chromagenic images, the method estimates the light color (the illuminant estimation step) by finding the best transform among the ones calculated in the pre-processing step.

The chromagenic color constancy approach can deliver good estimates of the illuminant [5]. However, the filtered and the unfiltered images need to be registered [7]. This is a limitation for some real-life applications as image registration is usually time-consuming and computationally expensive. In fact, image registration is still an important field of research on its own [6] and cannot always be solved reliably.

In this paper, we present an approach that aims at avoiding the need for image registration in the first step of the chromagenic color constancy algorithm. In particular, we propose to use the Monge-Kantorovitch (MK) transform for obtaining the linear relations between the filtered and unfiltered images.

To show the effectiveness of our new method, we introduce a new *pilot* database of 63 scenes of chromagenic facial images (to be used in Kampo diagnosis). Using this dataset we demonstrate that our new method supports better color correction compared with assuming registered images (when registration cannot be carried out or is insufficiently accurate).

While the focus of this paper is color correction – of a normal capture and a second image taken though a colored filter – without registration we have investigated using the discovered color corrections for illuminant estimation using the full chromagenic algorithm. However, we found that the dataset is too small to conclude much about estimation performance. Indeed, for this small dataset, we found the modified chromagenic algorithm can work almost perfectly (and conversely chromagenic working with unregistered images can fail). But, in order to study algorithm performance in depth we will need to capture a much large corpus of images. We plan to compile a large set of chromagenic face images in the near future.

This paper is organized as follows. We start by recalling the background of our research: color constancy, the chromagenic color constancy, and an overview of what is the image-based Kampo diagnosis system. Then, we introduce our new dataset of facial images. Section 4 presents our approach. This is followed by the experiments and results. Finally, the paper is summed up in the conclusions.

2   Research Background

2.1   Color Constancy

Color constancy is the ability of a visual system to see objects with same colors regardless of the lighting conditions. In Fig. 1, we can see that the gray ball color varies with the color of the light, this happens when color constancy is not performed, here in the case of a digital camera.

While the human visual system is designed to achieve color constancy, machines – in particular modern digital cameras – need algorithms to accomplish this function (also known as white-balancing in digital photography). In computer vision, color constancy is achieved by first determining the color of the light under which the image scene was captured. Once the light color is estimated, it can be "divided out". Illuminant estimation is a core component of modern digital cameras reproduction pipelines.
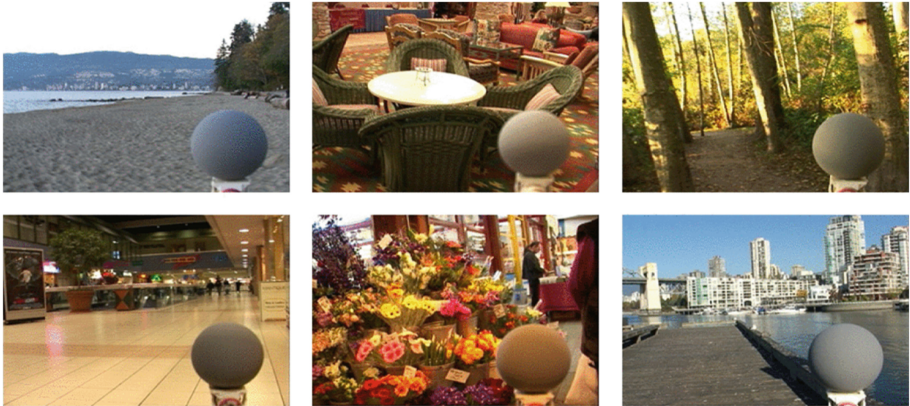
Fig. 1. The rendering of a gray ball under various lights. The image is from the SFU gray-ball dataset [8], appeared in [9].

Illuminant estimation algorithms can be split into two broad classes: algorithms that estimate the illuminant via a 'bag of pixels' statistical approach [14–17], and learning-based methods [18,19] (including deep learning [20,21]). There are also less commonly used methods that look for physical insights to drive the light estimation. For example, in the specular highlight method [22], highlights are sought in the scene. It is then assumed that the highlight color is the same as the illuminant color (true for dielectic materials). Another example is the blackbody-model-based algorithm [30,31] that uses the sensors responses to form an illuminant invariant color space and estimate the power spectrum of the illuminant. Another physics based method is the eponymous chromagenic algorithm [3,4,7], see Sect. 2.2, below.

Color constancy – the ability to estimate and then remove the color bias due to illumination – is important in several applications including, object tracking [12], facial recognition [11] and scene understanding [13]. In this paper we focus on a medical application requiring color constancy. Matsushita et al. [2] developed a pathophysiology system to reproduce a Kampo medical diagnosis for number of diseases based on facial images. The method only works when face color in an image is directly related to the physical reflectance properties of a face. This condition is only accomplished when the light illuminating the scene is equienergetic (i.e. achromatic), meaning that color constancy should be applied.

2.2   Chromagenic Color Constancy

In the chromagenic color constancy approach two images are taken of each scene. The first image is a normal capture and the second is an image taken through a specially chosen chromagenic filter. Given reasonable assumptions about the dimensionality of lights and surfaces it was shown in [7] that the filtered and unfiltered responses are related by a linear transform and that this relationship varies with (is intrinsic to) the illuminant color. Put another way, the relationship between filtered and unfiltered RGBs indexes – and so identifies – the illumination.

Mathematically, by adopting the Lambertian model of image formation, if we denote as $\underline{\rho}$ the normal captured image, and $\underline{\rho}_F$, the image captured by placing a color filter in front of the camera, we can write:

$$\begin{aligned} \underline{\rho}_k &= \int_\omega E(\lambda)S(\lambda)Q_k(\lambda)d\lambda \\ \underline{\rho}_{F,k} &= \int_\omega E(\lambda)S(\lambda)F(\lambda)Q_k(\lambda)d\lambda \end{aligned} \tag{1}$$

where λ denotes a particular wavelength, ω the visual spectrum (normally from 380 to 740 nm), E is the illuminant, S is the set of scene objects reflectances, k corresponds to R, G or B, the color channels of the digital camera, Q is the camera sensitivity function, and F is the spectral response of the selected filter.

As stated above, it was shown in [7] that under reasonable assumptions about the dimensionality of lights and surfaces, the unfiltered and filtered responses should be related by a 3×3 linear transform:

$$\underline{\rho}_F \approx T_E^F \underline{\rho} \tag{2}$$

The chromagenic algorithm works in two steps. First, in pre-processing, we calculate a range of illuminant transform matrices T_i (for $i = 1, .., N$ illuminants) using a least squares approach. In a second step, given a chromagenic pair of images $\underline{I}(x, y)$ and $\underline{I}_F(x, y)$, we determine the illuminant color by minimizing:

$$argmin_i(\Sigma_x \Sigma_y ||T_i \underline{I}(x, y) - \underline{I}_F(x, y)||) \tag{3}$$

where (x, y) represents a particular pixel of the image.

A limitation of chromagenic color constancy is that images need to be registered. Image registration is required both in the least squares minimization of the first step and also in the selection of T_i in the second step. In this work, we present a method to avoid the need for registration in the first step.

2.3   Kampo Medical Diagnosis

Kampo medicine is the traditional Japanese medicine used in Japan and, in alternate forms, across Asia. A Kampo medical diagnosis [25] requires a visual observation, an olfactory examination, an inquiry and a palpation. A face-only diagnosis is, however, possible for various diseases: blood stagnation (due to a poor blood circulation), blood deficiency (resulting from the lack of blood, in other terms when the blood is not regenerated in normal proportions) and yin deficiency (which is a sign of a lack of water at the face level).

Matsushita et al. [2] developed an image-based system for facial Kampo diagnosis. The system emits a diagnosis in the form of a score (from 1 to 5) where 1 indicates a non-disease state and 5 indicates a severe disease state. The Kampo system works as follows. First, given an image of an ill patient, the system generates a hemoglobin density image and a gloss image. The hemoglobin image is the result of a pigmentation component separation by independent component analysis (ICA) [23] and the gloss image is obtained by using a polarizer (the face is captured with and without a polarizing plate) [24]. Five regions of interest are extracted for each of the two images: one region from the forehead area in the image, 2 regions under the eyes and 2 other regions at the cheeks level. A final region is the sum of all these 5 regions. Five features values are calculated from the RGBs values of these 5 regions. In total 60 features are extracted from the images. The system emits a diagnosis by support vector regression (an optimization problem).

In [2], the system was evaluated and tested on a dataset of images generated from images of healthy patients (taken in a lab under a white light) by the modulations of gloss and hemoglobin. The results were compared to Kampo medical doctor diagnostic. In this paper, we also present a new dataset of images we collected for Kampo diagnosis, in order to allow the testing of the system under different lights. However, here we capture every scene with and without a colored filter.

3   A Chromagenic Face Image Dataset for Pathophysiology

We introduce in this section a new dataset of facial images for Kampo pathophysiology diagnosis. The dataset has 63 initial scenes (a set of three facial images of a healthy subject taken under a determined light). Every scene was captured 3 times: one time without a filter and 2 other times with a red and a yellow filter (respectively a Tiffen 85 and a Tiffen 81EF). The images were taken in Chiba University in Japan during the Summer 2018. Nine participants took part in the

data collection. All images were taken with a Nikon D5200 camera in a lighting room equipped with 2 Thouslite LED cubes (which allowed us to simulate a range of illuminant color temperatures).

Fig. 2. Two images of the same scene from the dataset showing the ColorChecker chart, the left image is a normal capture and the right image was captured through a red filter, note that these 2 images are the camera pipeline outputs.

The left of Fig. 2 shows one normal image from our dataset, the right shows the same capture but through the red filter. The right image is, of course, redder in appearance. The images are not registered but the experiment conditions were very well controlled, for this reason the difference in the images alignment is not easily noticeable in this case. Notice that there is a ColorChecker in the scene and this is true for all our images. Placing a ColorChecker in every scene is useful for two reasons.

First, we can use it to measure the white point (the RGB of the color of the light). In line with [10] this is defined to be the RGB taken from the brightest unsaturated gray patch of the ColorChecker. In Fig. 4 we show the ground-truth chromaticities for the illuminants in the 63 scenes of the dataset. It is clear that our dataset has a range of illuminant colors.

Second, given a chromagenic pair of images, with the Macbeth ColorChecker in each image, we can solve for the best possible 3×3 matrix relating the colors of the two color charts (without requiring the pixel-wise registration of the images in this case). We will consider this ColorChecker-based 3×3 matrix transform as our reference when solving for the linear transform in the case of non-registered images (see Sect. 5). Of course, usually there is no ColorChecker in the scene and solving for the best possible transform is not possible.

While Fig. 2 shows the capture environment, in the Kampo diagnosis we need only the face image. In Fig. 3 we show two sets of 3 images for two of our subjects. From left to right we show the original image. Then there is the same person imaged through a yellow and a red filter.

In the future, we will generate new images with various diseases states from this dataset. These images will be obtained by modulations of gloss and hemoglobin [2].

Fig. 3. Two sets of filtered and unfiltered images from the dataset. The 2 sets represent 2 scenes (2 subjects), from left to right: normal capture, image with yellow filter and image with red filter. (Color figure online)
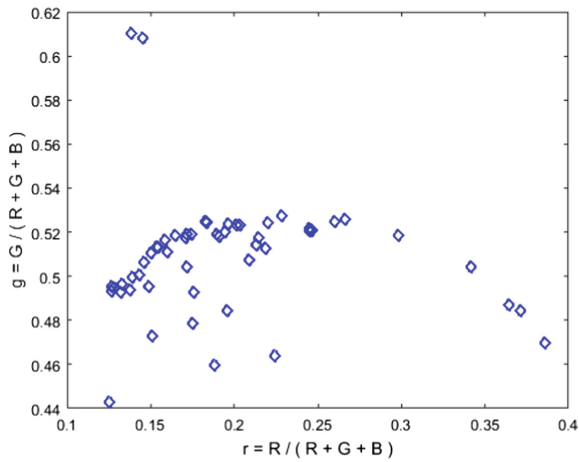
Fig. 4. Ground-truth chromaticities of the Kampo images dataset.

4   Color Correction Without Registration: The Monge-Kantorovitch Linear Transform

The chromagenic color constancy algorithm works in two steps. In the pre-processing step we solve for the best 3×3 matrices relating unfiltered to filtered RGBs for a large range of scenes and illuminants. Each transform by construction is associated with a light color. Then in the second step, when we have a chromagenic image pair for an unknown illuminant, we test each of the pre-computed transforms in turn to see which when applied to the unfiltered RGBs best predicts the filtered counterparts. The color of the light is then defined to be the associated light color.

Both the pre-processing step and the application part of the chromagenic algorithm (estimating the illuminant using a pair of chromagenic images) requires registered pairs of images. Unfortunately, even after decades of investigation, image registration remains a hard problem and even when it works it delivers imperfect results [6]. Further, even images that are only slightly out of registration can result in significantly different transforms (3×3 matrices) that best relate the unfiltered to filtered RGBs. This is the case for our dataset where the differences in the alignment of the unfiltered and filtered images of the same scene is not significantly visible but this difference still impacts on the transforms (see results in Sect. 5).

In what follows, we only focus on the pre-processing step of the chromagenic method. In particular, we propose using the Monge-Kantorovitch (MK) linear transform to replace the 3×3 matrix relating the images in the chromagenic pair without registration. Note that the MK transform is a 3-D similarity transform. MK has its roots in the Earth Movers Distance (EMD) [26] (or Wasserstein Metric [28]) which has proven to be a useful tool in image recognition [27]. Imagine we have a few piles of earth. Equal to the volume of all the earth we have several holes to deposit the earth. Clearly if we wish to move the earth into the holes to minimize the energy expelled, we wish to move each shovel full of earth as little as possible. The minimum distance we have to move all the earth is exactly the earth movers distance. It can be efficiently solved using linear programming [29].

Rather usefully, there is a simple and closed form linear restriction to EMD. Given $M \times N$ data matrices A and B (where $N > M$) the classic linear least squares minimization solves:

$$\min_T ||TA - B|| \tag{4}$$

Of course to solve the above then we exploit the fact the columns of A and B are in correspondence (not the case for non-registered images). In the linear restriction EMD, we seek to find a transform T such that the correlation structure of AT and B matches and that TA is as close to A as possible (that is the colors in A move as little as possible). Specifically, we minimize

$$\min_T ||TA - A|| \; s.t. \; TAA^t T^t = BB^t \tag{5}$$

Pitié et al. [1] have shown that MK (or linear restriction to EMD) can be used in color grading (to map the colors of an input image to match the look and feel of a target image).

Here we use Eq. 5 to find the transform relating an RGB image to its filtered counterpart. In Eq. 5, A would contain the pixels from the unfiltered image and B pixels for the same scene captured through a colored filter. The pixels in A and B may not be in correspondence. Indeed, there is no constraint that the number of pixels in the filtered and unfiltered images need to be the same.

5   Experiments and Results

In order to evaluate the effectiveness of our approach, we compare the results given by our method (Eq. 5; MK) to those obtained by the usual least squares procedure (Eq. 4; LST) on our Kampo dataset.

As a reference point, we use the fact that a Macbeth ColorChecker chart is included in our images. This allows us to compute the best possible linear transform between the two images (T_{CC}) by computing the least squares regression only considering those colors of the color charts. Thanks to this, we can now evaluate the difference between the best linear transform and the two other solutions as

$$\epsilon_m = ||T_{CC}A - f_m(A)||$$

where A is the unfiltered image and f_m states the method being computed where $m = \{LS, MK\}$.

Figure 5 plots the individual errors for all images: the upper graph is the result for the red filter and the lower graph is the result for the yellow filter. These results are summarized in Table 1, where we present the mean and RMS errors for the dataset. Our method improves the usual procedure by at least 75%. Note all image values are in the interval [0,1] so a mean error of 0.01 corresponds to a 1% error.

Visual examples of our results are presented in Fig. 6 where we show from left to right: an unfiltered image, the image corrected by using the ColorChecker (i.e., the best possible result or reference), the image corrected using MK (the approach proposed in this paper), and the image corrected with the LST approach. The upper example was generated with the red filter and the lower with the yellow filter. The images were linearly converted from RAW format and demosaiced. We can clearly see that our approach generates colors that are very close to the best possible solution.

Table 1. Mean and RMS error for MK vs LST approaches with red and yellow filters

Filter	Mean		RMS	
	MK	LST	MK	LST
Red	**0.0024**	0.0140	**0.0033**	0.0236
Yellow	**0.0025**	0.0123	**0.0031**	0.0207
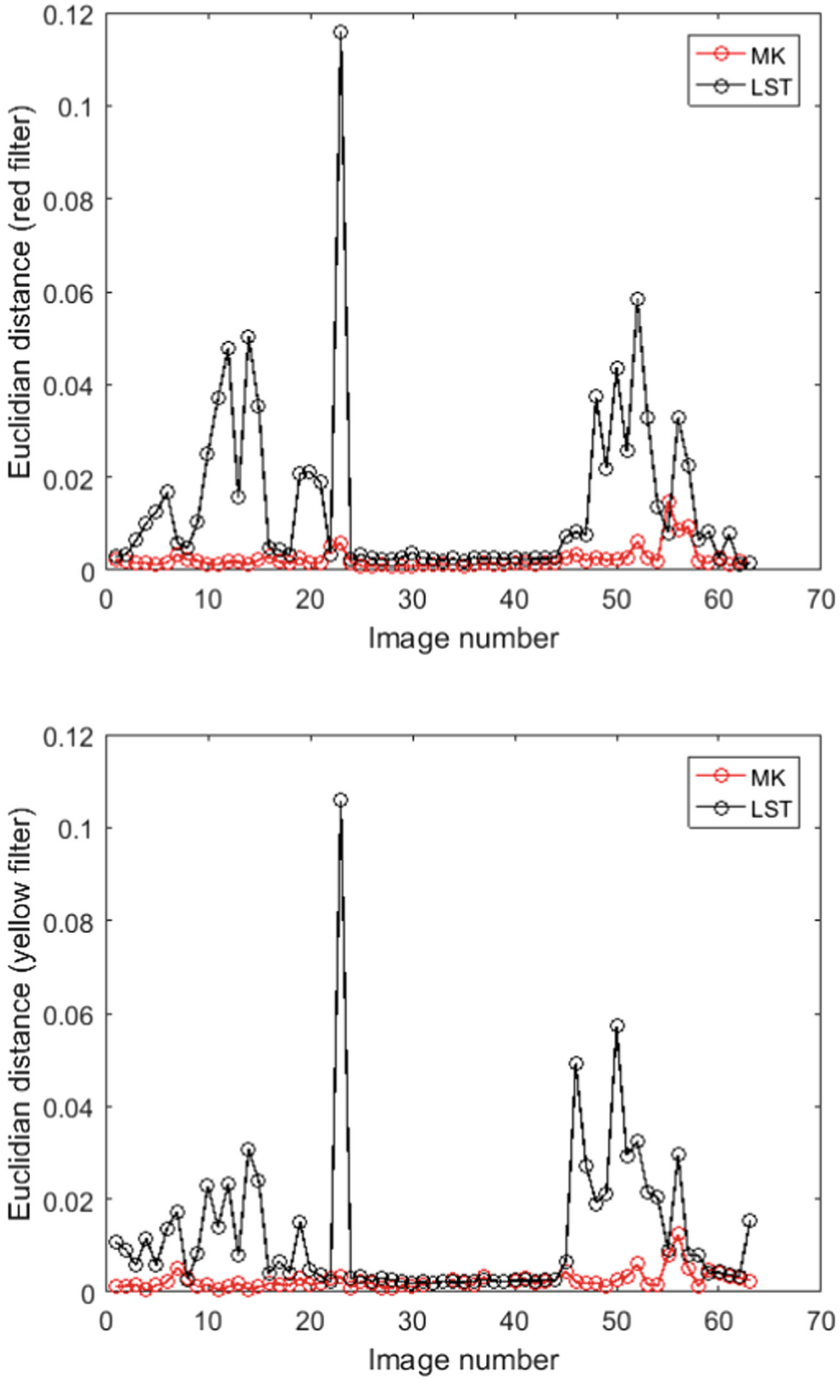
Fig. 5. Chromagenic distance when using MK vs the least squares transform LST. Top: red filter. Bottom: yellow filter. (Color figure online)
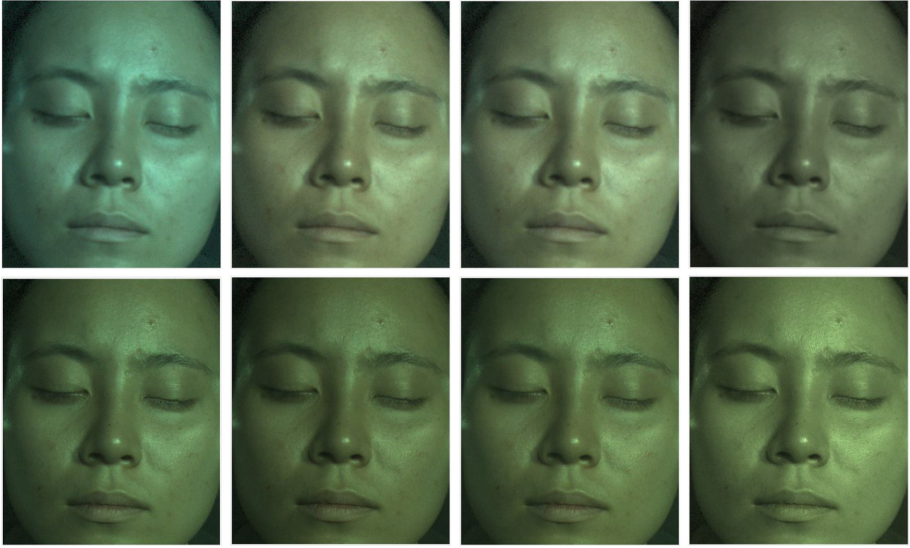
Fig. 6. Two scenes from the dataset (upper with red filter and lower with yellow filter), from left to right: the original unfiltered image, the color corrected image with CCT (the ColorChecker-based transform), the color corrected image with MK and the color corrected image with LST. (Color figure online)

6   Conclusion

This paper introduces a new image dataset comprising 63 scenes of facial images taken under a variety of lights and, novelly, with and without a color filter. Given a pair of filtered and unfiltered images it will be possible to use the chromagenic approach to illuminant estimation. The chromagenic algorithm has two parts: first we need to relate the unfiltered to filtered image using a linear transform. Second, we need to identify the illuminant by searching for the best transform. This paper focuses on the first question only.

We show, using the Monge-Kantorovitch (MK) transform, how we can solve for the linear map without the need to register the images. This is of significant practical importance. Not only is registration a hard problem it cannot always be solved in a pixel-wise manner. Here we remove the need for registration altogether. Moreover, we show that the MK method outperforms direct least squares (where we assume good registration when this not the case) by a factor of about 4:1.

Looking to the future our plan is to capture a larger set of facial images so we can test the second part of the chromagenic algorithm. That is, we will investigate whether MK suffices to allow the chromagenic algorithm to estimate the illuminant for face images.

References

1. Pitie, F., Kokaram, A.: The linear Monge-Kantorovitch linear colour mapping for example-based colour transfer. In: IET 4th European Conference on Visual Media Production, pp. 23–23 (2007)
2. Matsushita, F., Kiyomitsu, K., Ogawa, K., Tsumura, N.: System for evaluating pathophysiology using facial image. In: Color and Imaging Conference, pp. 274–279 (2017)
3. Finlayson, G.D., Fredembach, C., Drew, M.S.: Detecting illumination in images. In: IEEE 11th International Conference on Computer Vision (2007)
4. Finlayson, G.D., Hordley, S.D., Morovic, P.: Colour constancy using the chromagenic constraint. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1079–1086 (2005)
5. Fredembach, C., Finlayson, G.D.: The bright-chromagenic algorithm for illuminant estimation. J. Imaging Sci. Technol. **52**, 137–142 (2008)
6. Zitova, B., Flusser, J.: Image registration methods: a survey. Image Vis. Comput. **21**, 977–1000 (2003)
7. Finlayson, G.D., Hordley, S., Morovic, P.: Chromagenic filter design. In: 10th Annual Congress of the International Colour Association, pp. 1023–1026 (2005)
8. Ciurea, F., Funt, B.V.: A large image database for color constancy research. In: Color and Imaging Conference, pp. 160–164 (2003)
9. Gijsenij, A., Gevers, T., Van De Weijer, J.: Computational color constancy: survey and experiments. IEEE Trans. Image Process. **20**, 2475–2489 (2011)
10. Hemrit, G., et al.: Rehabilitating the ColorChecker dataset for illuminant estimation. In: Color and Imaging Conference, pp. 350–353 (2018)
11. Samal, A., Iyengar, P.A.: Automatic recognition and analysis of human faces and facial expressions: a survey. Patt. Recognit. **25**, 65–77 (1992)
12. Yilmaz, A., Javed, O., Shah, M.: Object tracking: a survey. ACM Comput. Surv. **38**, 1–45 (2006)
13. Li, L.J., Socher, R., Fei-Fei, L.: Towards total scene understanding: classification, annotation and segmentation in an automatic framework. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2036–2043 (2009)
14. Buchsbaum, G.: A Spatial processor model for object colour perception. J. Franklin Inst. **310**, 1–26 (1980)
15. Land, E.H., McCann, J.J.: Lightness and retinex theory. J. Opt. Soc. Am. **61**, 1–11 (1971)
16. Finlayson, G.D., Trezzi, E.: Shades of gray and colour constancy. In: Color and Imaging Conference, pp. 37–41 (2004)
17. Vazquez-Corral, J., Vanrell, M., Baldrich, R., Tous, F.: Color constancy by category correlation. IEEE Trans. Image Process. **21**, 1997–2007 (2012)
18. Gehler, P.V., Rother, C., Blake, A., Minka, T., Sharp, T.: Bayesian color constancy revisited. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
19. Gijsenij, A., Gevers, T.: Color constancy using natural image statistics. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
20. Bianco, S., Cusano, C., Schettini, R.: Color constancy using CNNs. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 81–89 (2015)
21. Barron, J.T., Tsai, Y.-T.: Fast fourier color constancy. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)

22. Tan, R.T., Nishino, K., Ikeuchi, K.: Color constancy through inverse-intensity chromaticity space. J. Opt. Soc. Am. A. **21**, 321–334 (2004)
23. Tsumura, N., et al.: Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. ACM Trans. Graph. **22**, 770–779 (2003)
24. Ojima, N., Minami, T., Kawai, M.: Transmittance measurement of cosmetic layer applied on skin by using processing. In: 3rd Scientific Conference of the Asian Societies of Cosmetic Scientists, p. 114 (1997)
25. Sato, Y., Hanawa, T., Arai, M., Cyong, J.C., Fukuzawa, M., M.K.: Introduction to Kampo: Japanese traditional medicine. Japan Soc. Orient. Med. (2005)
26. Rubner, Y., Tomasi, C.: The Earth mover's distance. Percept, Metrics Image Database Navig (2001)
27. Rubner, Y., Tomasi, C., Guibas, L.J.: Earth mover's distance as a metric for image retrieval. Int. J. Comput. Vis. **40**, 99–121 (2000)
28. Vasershtein, L.N.: Probl. Pered. Inform. **5**, 64 (1969)
29. Dantzig, G.B., Orden, A., Wolfe, P.: Generalized simplex method for minimizing a linear form under linear inequality restraints. Pac. J. Math. **5**, 183–195 (1955)
30. Ratnasingam, S., Hernández-Andrés, J.: Illuminant spectrum estimation at a pixel. J. Opt. Soc. Am. A. **28**, 696–703 (2011)
31. Ratnasingam, S., Collins, S., Hernández-Andrés, J.: Optimum sensors for color constancy in scenes illuminated by daylight. J. Opt. Soc. Am. A. Opt. Image Sci. Vis. **27**, 2198–2207 (2010)