Høgskolen i Gjøviks rapportserie, 2009 nr. 4

Proceedings from Gjøvik Color Imaging Symposium 2009

Gabriele Simone, Alessandro Rizzi
and Jon Yngve Hardeberg (eds.)

Gjøvik 2009

ISSN: 1890-520X ISBN: 978-82-91313-18-4

CONTENTS

A Hybrid Image Quality Measure for Automatic Image Quality Assessment

Atif Mansoor
National University of Sciences and Technology
Islamabad, Pakistan
atif.mansoor@gmail.com

The aim of image quality assessment is to provide a quantitative metric that can automatically and reliably predict how an image will be perceived by humans. However, human visual system is a complex entity, and despite all advancements in the vision sciences, the phenomenon of image perception by humans is not clearly understood. Understanding the human visual perception is a challenging task, encompassing the complex areas of biology, psychology, vision etc. Likewise, developing an automatic quantitative measure that accurately correlates with the human perception of images is a challenging assignment.

Automatic image quality assessment has many diverse applications including image quality control systems, benchmarking etc. In the absence of a well defined mathematical model, researchers have attempted to find a quantitative metric based upon various heuristics to model the human image perception. The existing quality measures are not accurate representatives of the human perception, but they certainly give acceptable results in certain conditions.

In our research, instead of forming an altogether new image quality measure, we consider that investigating an intelligent combination of non-redundant existing image quality measures is a worthwhile effort. In this talk, a hybrid image quality (HIQ) measure, which is a combination of four existing measures using an 'n' degree polynomial to accurately model the human image perception, is presented. First we undertook time consuming human experiments to subjectively evaluate a given set of training images corrupted with fast fading distortion and resultantly formed a Human Perception Curve (HPC). Next we define a HIQ measure that closely follows the HPC using curve fitting techniques. The HIQ measure is then validated on a separate set of images by similar human subjective experiments and is compared to the HPC. The coefficients and degree of the polynomial are estimated using regression on training data obtained from human subjects. Validation of the resultant HIQ was performed on a separate validation data. Our results show that HIQ gives an RMS error of 5.1 compared to the best RMS error of 5.8 by a second degree polynomial of an individual measure HVS (Human Visual System) absolute norm (H1) amongst the four considered metrics. Our data contains subjective quality assessment of 174 images with various degrees of fast fading distortion. Each image was evaluated by 50 different human subjects using double stimulus quality scale, resulting in an overall 8,700 judgments.

From Perceptual Features to a Quality Metric

Chaker Larabi*
XLim-SIC, University of Poitiers
Poitiers, France
chaker.larabi@ieee.org

Quality assessment is becoming an important issue in the framework of image processing. This need is expressed by the fact that the quality threshold of end-users has been shifted up because of the large availability of high fidelity sensors at very affordable price. This observation has been made for different application domains such as printing, compression, transmission, and so on. Starting from this, it becomes very important to manufacturers and producers to provide products of high quality to attract the consumer. This high interest on quality means that tools to measure it have to be available. For this two types of measurements are possible: Subjective assessment that takes the end-observer in the loop of assessment and objective assessment based on the use of mathematical tools that can use properties of the Human Visual System (HVS). The objective tools are separated in three categories: Full-reference metrics using the original image and its impaired version to compute the fidelity, Reduced-reference metrics that embed some characteristics of the original image in order to compare them with impaired ones and no-reference metrics based on the detection of artifacts such as bluriness, blockiness, flickering, ringing... Objective tools are very interesting because they represent implementable tools that can be more or less complex and can be run when needed without any protocol or specific preparation.

However, some tasks or some artifacts cannot be handled by a mathematical tool because there is a subjectivity dimension that it cannot capture. For example, the cinematographic studios still need golden-eyes for the evaluation of the final work to be projected. Moreover, images and their associated processing (compression, halftoning, ...) are produced for the enjoyment or education of human observers so their opinion of the quality is very important. Subjective measurements have always been, and will continue to be, used to evaluate system performance from the design lab to the operational environment.

The aim of this study was to construct a perceptual metric dedicated to the evaluation of color reproduction devices. In other words, the idea was to learn from the human observer, what are the important perceptual features allowing quality judgment and then integrate them in perceptual metric trying to mimic the HVS. Thereby, a subjective study was performed using a wall of LCD displays from different manufacturers. Participants were asked to assess the images shown on these displays and to answer a detailed questionnaire. The perceptual features retained for this experiment are: Naturalness, Hue, Contrast, Saturation and Texture. Quality was the global criterion asked to the observer in order to study the correlation with the perceptual features.

A strong statistical analysis, composed by multivariate analysis of variance (MANOVA), repeated contrast, correlation coefficient,... has been run in order to determine the contribution of each perceptual feature in quality assessment. The first results demonstrated that all the quality factors used in this experiment (hue, saturation, contrast, texture, naturalness) allowed to discriminate between the displays in a qualitative way. This means that each of them can be used to assess the quality of a display. However, Pearson's correlation coefficient showed that hue, naturalness and saturation correlated most closely with overall quality. Nevertheless, the correlations of the other factors were high enough to estimate the quality ($\geq 87\%$). Lastly, the repeated-measures contrast study showed hue to be the most appropriate factor for estimating the quality of a display, as it defined the same display categories as the overall quality factor.

*This work has been done with Ludovic Quintard from the LNE - France

111 Full-Reference Image Quality Metrics and Still Not Good Enough?

Marius Pedersen
Gjøvik University College
Gjøvik, Norway
Océ Print Logic Technologies S.A.
Créteil, France
marius.pedersen@hig.no

We aim to give a survey of full-reference image quality metrics, including metrics specifically designed to evaluate image quality, but also metrics for image difference, image fidelity, and more. These metrics have in common that they try to predict the perceived difference between an original image and a modified version of it, this modification can typically be compression, halftoning and blurring. They output one numerical value and/or an image difference map. More than 100 image quality metrics have been reviewed and categorized, and short descriptions and analyses of all metrics and their relationships are given. This should prove valuable to researchers in various fields, to find the most appropriate metric for their application, and to give a better understanding of the state of the art of the field of image quality metrics.

Image Quality Measures and Individualized Conjoint Analysis for Evaluating Gamut Mapping Algorithms

Zofia Barańczuk
EMPA, Swiss Federal Laboratories for Materials Testing and Research
Duebendorf, Switzerland
zofia.baranczuk@empa.ch
Peter Zolliker
EMPA, Swiss Federal Laboratories for Materials Testing and Research
Duebendorf, Switzerland
peter.zolliker@empa.ch
Joachim Giesen
University of Jena
Jena, Germany
joachim.giesen@uni-jena.de

In the present study we compare different methods of predicting psycho-visual choice data. The accuracy of prediction is considered to be a good measure for an assessment of those methods. The first class of methods uses image quality measures for the prediction of psycho-visual choices. The second class of methods is the evaluation of part of the data using psycho-metric scaling models such as Thurston's Law of Comparative Judgment or conjoint analysis.

Accuracy of predictions is measured in terms of hit rates, i.e. the percentage of correctly predicted choices in a pair comparison test. Hit rates were always obtained from data not involved in the evaluation, i.e. for the psycho-metric scaling models cross validation was used.

We used four data sets, all using pair comparison data of gamut mapping algorithms. Three of those each compared a set of eight algorithms, the fourth one used parameterized gamut mapping algorithms. Each test set was based on sets between 70 and 100 different images.

For each data set we computed choice predictions using several image quality measure from the literature, such as Structural Similarity (SSIM), Laplacian Mean Square Error (LMSE) and ΔE. Also a new image quality measure was included. It consists in an optimized linear combination of ΔE and measure for the local constrast, ΔLC, basically an averaged difference in Michelson contrast.

For each data set we computed global scale values for the tested algorithms using either Thurstone's Law of Comparative Judgment or, for the parameterized algorithms, conjoint analysis. On the basis of those scale values an expected preference can be calulated for each pair. In a further step scale values were calculated also indiviually for each image. Choice predictions can than be made, based on a linear combination of the global and the individual scale values.

The results obtained for the best image quality measures and Thurstone's method or conjoint analysis were close to each other. Better predictions can be achieved by computing individualized Thurstone's scale values, but only if enough test data is available. For the test of parametrized gamut mappings the optimized linear combination of ΔE and ΔLC had the highest percentage of correct predictions. The pixel-wise image quality measures correlated weakly with real preferneces apart from the test with parametrized gamut mapping algorithms. But there were large differences in ΔE between compared images in this test and that is probably the reason for good results of those measures.

Image quality measures like SSIM can be used for predicting choices in psycho-visual tests concerning the evaluation of gamut mapping algorithms. Evaluating gamut mapping algorithms automatically by using image quality measures can be an attractive alternative or complement to psycho-visual test whenever the latter are too expensive or difficult to carry out.

Towards a Comprehensive Image Quality Metric

Sabine Süsstrunk
School of Computer and Communication Sciences (IC),
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
sabine.susstrunk@epfl.ch

Evaluating quality is a necessary task for many imaging applications. Consequently, the color science and image processing communities have developed several image quality models and image metrics. But what is image quality? While a human observer can quite easily judge if an image is ”good” or ”bad,” formulating which image attributes influence the overall judgment is non-trivial. Even when image attributes (noise, sharpness, colorfulness, etc.) can be isolated, the individual contribution to image quality is not always evident.

Especially in color imaging, there are fundamental gaps in our knowledge of what color - or a change in color - contributes to the perception of an image. Most color imaging models and algorithms are based on single pixel transformations and measures, even though it is a well-known fact that the appearance of color is directly influenced by its size, surround, and viewing conditions. To develop meaningful color image quality models and algorithms based on visual perception needs the collaboration of several fields, including photography, computer vision, image processing, psychology, and physiology.

In this talk, the most commonly applied image quality metrics will be reviewed and their use (and abuse) illustrated. A framework of (well understood) visual characteristics will be presented, and how they can be integrated into an image quality metric. The areas where we still lack sufficient knowledge and models will be discussed.

From Contrast to Sensitivity: Vision and Image Processing

Eli Peli
Schepens Eye Research Institute, Harvard Medical School
Boston, MA, U.S.A.
eli.peli@schepens.harvard.edu

When image processing is performed in the context of displayed images presented to observers, the properties of the displays and the observers' visual system have to be considered. Contrast sensitivity is the main feature of human image perception, thus understanding image contrast is essential for effective image processing. Various contrast measures adequately describe the perception of simple targets but fail to capture the perception of even slightly more complex patterns (Peli 1997), while our measure of contrast in complex images contrast measure (Peli 1990) holds. Many applications, in particular image quality assessment, relied on simulating human vision. Our contrast metric and its associated vision model lends itself to simulations of image appearance (Peli et al. 1991). Bovik and colleagues (Damera-Venkata et al. 2000) and others used our model to develop computational image quality measures. Our simulations are parameter free except for the use of measured contrast sensitivity function (CSF), which varies by factor of 10 across studies. The spatial and temporal characteristics of the stimuli can account for all of the differences (Peli et al. 1993). This leaves open the question: which of these CSF data sets should be considered in image processing? Studies testing the simulations also determine the type of CSF function that should be implemented in such models (Peli 2001, Peli Geri 2001). Sensitivity to contrast represents an adaptation of the visual system. However, the visual system adapts also to changes in the contrast levels it faces (Wilson et al. 1993). Webster et al. (2002) demonstrated strong adaptation to image blur and to sharpening. When observers adapted to a blurred (or sharpened) image for just 3 minutes (or even less), they reported the image and subsequent images to be perceived relatively sharpened (or blurred). This phenomenon has significant implications for the perception of processed images. To test the hypothesis that the blur (sharp) adaptation is simply a manifestation of the contrast adaptation we attempted to simulate the effect of blur adaptation using our vision model. Contrast adaptation that causes an increase in threshold indeed simulated adaptation to sharpened images. However, there is no data that supports lowering the thresholds employed in the model in response to reduced contrast (blur). This led us to realize that the CSF data used by us and any others are obtained following "adaptation" to a blank gray screen. Clearly, the CSF that is relevant to seeing natural images is the one that exists following adaptation to the natural images previously seen. These thresholds should already be elevated due to the adaptation and therefore may decline when presented with a lower contrast image such as a blurred image. Thus, the relevant CSF (for modeling or any other application) needs to be tested following adaptation to natural images and not following a blank screen. Peter Bex et al. (2009) have recently measured the effects of adaptation following adaptation to movies and found a modest increase in threshold at low spatial frequencies.

An Alternative Color Difference Formula for Computing Image Difference

Gabriele Simone
Gjovik University College, Norway
gabriele.simone@hig.no

Claudio Oleari
University of Parma, Italy
claudio.oleari@fis.unipr.it

Ivar Farup
Gjovik University College, Norway
ivar.farup@hig.no

Abstract

In this paper, we approach color-image-difference metrics by a Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space. We start from previous image-difference metrics by replacing the CIE color-difference formulae with the new one. Tests are performed on th TID database and current results show improvements in the actual state of art, making this formula the future key for image- difference metrics.

1 Introduction

The CIE published the CIELAB color space [3], with the idea of a perceptually uniform color space. In this color space it is straightforward computing the distance between two colors, by using the Euclidean distance ΔE^*_{ab}. This metrics formula has been used also for computing the difference between color images as color difference of all the pixels and averaged. The inadequateness of the original CIELAB formula was the origin of other important proposals. The British Colour Measurement Committee proposed the ΔE_{CMC} formula [6], defined on the CIELAB system. The CMC formula is today used as standard formula in industrial color control [16]. The BFD [9] formula, introduced in 1987 by Luo and Rigg, is a modification of the ΔE_{CMC} formula. It provided a correction for the CMC in the blue region [6]. In 1994, CIE proposed the ΔE_{94} [2] formula with the main intention to reduce the high complexity of the CMC formula. All these formulas (CMC, BFD and CIE94) are based on the BFD color-difference data [10] and none of them resulted completely satisfactory. The last CIE formula for small-medium color differences is the ΔE_{00} one [8], known as CIEDE2000 and based on a wider set of empirical data, known as COM10 dataset. Very recently, in 2009, a Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space, termed ΔE_E, has been published [11, 5]. The BFD empirical color difference data represented in the OSA-UCS space show a regularity not existing in the CIELAB space. First, this induced the authors of this formula to represent the small-medium color differences by a simple ellipsoidal equation [5], termed ΔE_{GP}, and finally to propose a proper logarithmic compression of the OSA-UCS space with a consequent reduction of the parameters and new formula, that is Euclidean and termed as ΔE_E [11]. So far in the years many different color-image-difference metrics have been proposed [12], some with the intent of measuring general image quality and some for detecting specific distortions. However, at the moment, a universal color-image-difference metric does not exist. A spatial extension to the CIELAB color-difference formula (S-CIELAB) was proposed by Zhang and Wandell [17] in 1997, introducing a spatial filter, which simulates the human visual system, as spatial pre-processing to the CIELAB color difference formula [3]. Johnson and Fairchild [7] followed the same approach but the spatial filter is implemented in the frequency domain, allowing for more precise control of the filter. In 2002 Hong and Luo [4] proposed the hue angle algorithm still based on the CIELAB color difference correcting some of the drawbacks. However, not including a spatial filtering of the image, it is unsuitable for halftone images [13, 15]. In 2008 Pedersen et al. [14] proposed two new image-difference metrics with spatial filtering simulating the human visual system. These two metrics (called SHAME and SHAME-II) apply a spatial filtering of the images similar to that used by Zhang and Wandell [17] and Johnson and Fairchild [7], before applying the hue angle measure to the filtered images.

2 A Proposal of Two New Metrics

The first metric that we propose and analyze is the simple pixel value difference but instead of using ΔE^*_{ab} formula we use ΔE_E in the Log-Compressed OSA-UCS space. The second metrics that we consider is based on the S-CIELAB developed by Johnson et al. [7]. This metric works with the following steps:

- The original and the reproduced image are converted to the opponent color space
- Afterwards they are spatially filtered
- Then they are converted to CIELAB color space
- In final a pixelwise difference is done using ΔE^*_{ab} formula, obtaining an image difference representation generally called S-CIELAB representation.

We have modified the last step changing again the ΔE^*_{ab} with the ΔE_E obtaining a different image difference representation that we call S-DEE.

3 Preliminary Results

For the evaluation of the proposed metrics we used the TID2008 database [1], which is composed by 25 original images. These images have been altered and divided into seven categories: Noise, Noise2, Safe, Hard, Simple, Exotic, Exotic2. Each category represents different kind of distortions. These two new metrics have been tested on 1700 images. Three types of correlation are computed for the results, the Pearson-product-moment-correlation coefficient, the Spearman-rank-correlation coefficient and the Kendall-tau-rank-correlation coefficient20.

Table 1: ΔE_E correlations compared to $DeltaE^*_{ab}$ ones on each category of the TID2008 database.

Dataset	Pearson correlation		Spearman correlation		Kendall correlation	
	ΔE^*_{ab}	ΔE_E	ΔE^*_{ab}	ΔE_E	ΔE^*_{ab}	ΔE_E
Noise	0.294	0.203	0.333	0.238	0.223	0.158
Noise2	0.243	0.338	0.297	0.412	0.213	0.285
Safe	0.336	0.405	0.338	0.461	0.221	0.303
Hard	0.492	0.643	0.466	0.665	0.324	0.481
Simple	0.418	0.585	0.434	0.608	0.309	0.433
Exotic	0.252	0.311	0.201	0.26	0.087	0.133
Exotic2	0.019	0.049	0.041	0.053	0.007	0.017
All	0.174	0.212	0.173	0.248	0.121	0.166

As shown in Table 1, ΔE_E performs better than ΔE^*_{ab}, excluding the noise dataset, with the same computational complexity and computational time. However either ΔE^*_{ab} and ΔE_E show a low performance considering all database set; only in the category ”hard” and ”simple” ΔE_E shows a reasonable result. A T-test at 5% confidence level on Spearman correlation values confirms the performance of the metric.

The simple pixelwise difference using ΔE_E performs better than the ΔE^*_{ab}, hue angle metrics but it is still worse of some others metrics previously developed. The S-DEE metric performs better than ΔE^*_{ab}, ΔE_E and hue angle metric. It performs slightly worse than S-CIELAB by Zhang et al. and S-CIELAB by Johnson et al., while it is still not efficient like SHAME-II, SSIM and UIQ.

Table 2: ΔE_E and S-DEE compared against other metrics considering all TID2008 database set.

METRICS	Pearson correlation	Spearman correlation	Kendall correlation
ΔE^*_{ab}	0.174	0.173	0.121
Hue angle	0.179	0.161	0.113
ΔE_E	0.212	0.248	0.166
S-DEE	0.443	0.456	0.335
S-CIELAB	0.476	0.482	0.354
S-CIELAB (Johnson)	0.542	0.538	0.4
SHAME	0.544	0.55	0.414
SSIM	0.547	0.653	0.437

4 Conclusion

The ΔE_E color difference formula makes improvements to the previously developed image-difference metrics and, at the moment, seems promising, but more studies must be done. Future studies will encapsulate the ΔE_E in other image difference metrics and applied to other spatial filters and evaluated on other different kind of dataset.

4.1 Acknowledgments

The authors would like to thank Marius Pedersen and Jon Yngve Hardeberg for their advice, suggestions and feedback regarding this project.

References

[1] E. Bando, J. Y. Hardeberg, and D. Connah. Can gamut mapping quality be predicted by color image difference formulae. In *Human Vision and Electronic Imaging X, ed. B. Rogowitz, T. Pappas, S. Daly, Proc. of SPIE - IST Electronic Imaging,SPIE*, volume 5666, pages 180 – 191, 2005.

[2] CIE. Industrial colour-difference evaluation. CIE Technical Report 116, Central Bureau of the CIE, Vienna, Austria, 1995.

[3] *Colorimetry*, volume 15:2004 of *CIE Publications*. Central Bureau of the CIE, Vienna, Austria, 2004.

[4] G. Hong and M.R. Luo. Perceptually based colour difference for complex images. In R. Chung and A. Rodrigues, editors, *Proceedings of SPIE: 9th Congress of the International Colour Association*, volume 4421, pages 618–621, 2002.

[5] R. Huertas, M. Melgosa, and C. Oleari. Performance of a color-difference formula based on OSA-UCS space using small-medium color differences. *Journal of the Optical Society of America*, 23(9):2077–2084, September 2006.

[6] F.H. Imai, N. Tsumura, and Y. Miyake. Perceptual color difference metric for complex images based on mahalanobis distance. *Journal of Electronic Imaging*, 10:385–393, April 2001.

[7] Garrett M. Johnson and Mark D. Fairchild. Darwinism of color image difference models. In *The 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, 2001.

[8] M.R. Luo, G. Cui, and B. Rigg. The development of the cie 2000 colour-difference formula: Ciede2000. *Color Research and Application*, 26(5):340–350, 2001.

[9] M.R. Luo and B. Rigg. Bfd(l:c) colour-difference formula: Part 1 - development of the formula. *Journal of the Society of Dyers and Colourists*, 103:86–94, 1987.

[10] B. Rigg M. R. Luo. Chromaticity-discrimination ellipses for surface colours. *Color Research & Application*, 11:25–42, 1986.

[11] Claudio Oleari, Manuel Melgosa, and Rafael Huertas. Euclidean color-difference formula for small-medium color differences in log-compressed osa-ucs space. *J. Opt. Soc. Am. A*, 26(1):121–134, 2009.

[12] M. Pedersen and J. Y. Hardeberg. Survey of full-reference image quality metrics. *ACM Computing Surveys*, 2008. Submitted.

[13] Marius Pedersen and Jon Yngve Hardeberg. Rank order and image difference metrics. Jun 2008.

[14] Marius Pedersen and Jon Yngve Hardeberg. A new spatial hue angle metric for perceptual image difference. In *2009 Computational Color Imaging Workshop*, Saint Etienne, France, Mar 2009. Submitted.

[15] Marius Pedersen, Jon Yngve Hardeberg, and Peter Nussbaum. Using gaze information to improve image difference metrics. In Bernice Rogowitz and Thrasyvoulos Pappas, editors, *Human Vision and Electronic Imaging VIII (HVEI-08)*, volume 6806 of *SPIE proceedings*, pages 680611–1–680611–12, San Jose, USA, Jan 2008. SPIE.

[16] Gaurav Sharma. *Digital Color Imaging Handbook*. CRC Press, Inc., 2002. ISBN: 084930900X.

[17] X. Zhang and B. A. Wandell. A spatial extension of cielab for digital color image reproduction. In *Soc. Inform. Display 96 Digest*, pages 731–734, San Diego, 1996.

Face Saliency in Human Visual Saliency Models

Puneet Sharma, Faouzi Alaya Cheikh and Jon Yngve Hardeberg
Faculty of Computer Science and Media Technology
Gjøvik University College,Gjøvik, Norway
er.puneetsharma@gmail.com

Abstract

Under natural viewing conditions humans tend to fixate on specific parts of image that interests them naturally. Saliency map is the map of regions which are more prominent than other regions in terms of low level image properties such as intensity, color and orientation. With some modifications it can be used to simulate the natural human fixation also known as gaze. There are numerous applications in the field of engineering, marketing and art that can benefit from understanding the human visual fixation such as image quality evaluation, label design etc. The objective of this research is to understand the factors that influence the saliency map and gaze map and to modify the saliency map in order to make it similar to the gaze map. Eye movements of 20 test subjects were captured using eye tracking equipment available in the lab. Three different algorithms Saliency [17], GAFFE [12], GBVS [4] were modified by addition of face saliency module to generate modified maps corresponding to each of the algorithms. Experimental gaze maps were analyzed and compared with modified saliency map, modified GAFFE map and modified GBVS map.

1 Introduction

When processing images for a human observer, it is important to consider how images are converted into information by the viewer. Understanding visual perception helps during image and video algorithm development. Image data represents physical quantities such as chromaticity and luminance. Chromaticity is the color quality of light defined by its wavelength. Luminance is the amount of light. To the viewer, these physical qualities may be perceived by such attributes as color and brightness. For the viewer, image perception begins at the eyeball. There are at least three reasons why gaze control is an important topic in scene perception [5]. First, vision is an active process in which the viewer seeks out task-relevant visual information. Secondly attention plays an important role in visual and cognitive processing. Thirdly eye movements provide unobstructive, sensitive, real-time behavioral index of ongoing visual and cognitive processing. Visual attention is the ability of a vision system, biological or artificial, to rapidly detect potentially relevant parts of a visual scene, on which higher level vision tasks, such as object recognition, can focus. It is generally agreed nowadays that under normal circumstances human eye movements are tightly coupled to visual attention [8]. Attention implements an information-processing bottleneck that allows only a small part of incoming sensory information to reach short term memory thus understanding a complex scene is a series of computationally less demanding, local visual analysis problems [6, 7]. These regions are valuable for understanding the dynamics of human visual system and the development of applications in various fields. The selection of these regions depends on stimulus and goal driven objectives. The subjects selectively direct attention to objects in a scene using both bottom-up, image-based cues and top-down, task-dependent cues [6]. Human visual perception is task specific to objects in a scene for e.g. in a busy restaurant when we are looking for an empty table then we ignore all the other details like people, decorations, background etc. However if we are free to look we will pay attention to all the details like lights, decorations etc. of the same restaurant. So human vision is the perception of the scene depending on stimulus (saliency) and task assigned visual attention is the ability of a vision system, biological or artificial, to rapidly detect potentially relevant parts of a visual scene, on which higher level vision tasks, such as object recognition, can focus. It is generally accepted nowadays that under normal circumstances human eye movements are tightly coupled to visual attention [8].

1.1 Saliency Map

Saliency is the quality of an object or item to stand out from rest of the objects or items [1]. There are many different physical qualities that can make an object more salient than other objects in the scene such as its color, orientation, size, shape, movement or unique onset .The analysis of the saliency maps shows the role played by low level image features. Intensity, color and orientation maps contribute to the saliency of the regions in an image. As saliency map emphasizes the object which is differentiated from the other objects in terms of its higher intensity, chromacity and orientation. The order in which the scene should be inspected is determined by saliency map [1]. During human scene perception, high quality visual information is acquired only from a limited spatial region surrounding the center of gaze [6] .Visual quality falls off rapidly and continuously from the center of gaze into a low resolution center surround. We move our eyes three times a second to reorient the fovea through the scene. Pattern information is only acquired during fixations.

Moran Cerf states that a combined model of face detection and low-level saliency outperforms a low-level model in predicting locations humans fixate on [1]. Viola & Jones(VJ) [16] feature-based template matching algorithm combined with bottom-up saliency map model of Itti [6] was used. Seven subjects viewed a set of 250 images in a three phase experiment. Overall in both the experimental conditions i.e."free viewing" and "search" faces were powerful attractors of attention, accounting for a strong majority of early fixations when present. The new saliency model combined "bottom-up" feature channels of color, orientation and intensity with a clear special face detection channel based on VJ algorithm [16]. The combination was linear in nature with uniform weight distribution for maximum simplicity. In attempting to predict the fixations of human subjects, additional face channel improved the performance of both a standard and a more recent graph-based saliency model in images with faces. In the few images without faces, false positives represented in the face-detection channel did not significantly alter the performance of the saliency maps - although in a preliminary follow-up on a larger image pool they boost mean performance. These findings pointed towards a specialized "face channel" in our vision system, which is subject to current debate in the attention literature. Inspired by biological understanding of human attention allocation to meaningful objects - faces - a new model for computing an improved saliency map which is more consistent with gaze deployment in natural images containing faces than previously studied models was developed. It suggested that faces always attract attention and gaze, relatively independent of the task. They should therefore be considered as part of the bottom-up saliency pathway [1].

Humans have a collection of passive mechanisms reducing the amount of incoming visual information. For instance, the signal stemming from the photo receptors is assumed to be compressed by a factor of about 130:1, before it is transmitted to the visual cortex [3]. Nevertheless, the visual system is still faced with too much information. To deal with the still overwhelming amount of input, an active selection, involving eye movement, is required to allocate processing resources to some parts of our visual field. Oculomotor mechanisms involve different types of eye movements. A saccade is a rapid eye movement allowing jump from one location to another. The purpose of this type of eye movement, occurring up to three times per second, is to direct a small part of our visual field into the fovea in order to achieve a closer inspection. This last step corresponds to a fixation. Saccades are therefore a major instrument of the selective visual attention. This active selection is assumed to be controlled by two major mechanisms called bottom-up and top-down controls. The former, the bottom-up attentional selection, is linked to involuntary attention. This mechanism is fast, involuntary, and stimulus-driven. Observers unconsciously tend to select central locations of the image in order to catch the potentially most important visual information [3]. Human vision relies extensively on the ability to make saccadic

eye movements to orient the high-acuity fovea region of the eye over the targets of interest in the visual scene [12, 15].

A new bottom-up visual saliency model known as Graph-Based Visual Saliency (GBVS) was proposed in [4]. It consisted of two steps: first forming activation maps on certain feature channels, and then normalizing them in a way which highlighted the conspicuity and admitted combination with other maps. This model powerfully predicted human fixations on 749 variations of 108 natural images, achieving 98% of the Receiver Operating Characteristics (ROC) area of a human-based control, whereas the classical algorithms of Itti & Koch [7] achieved only 84%.

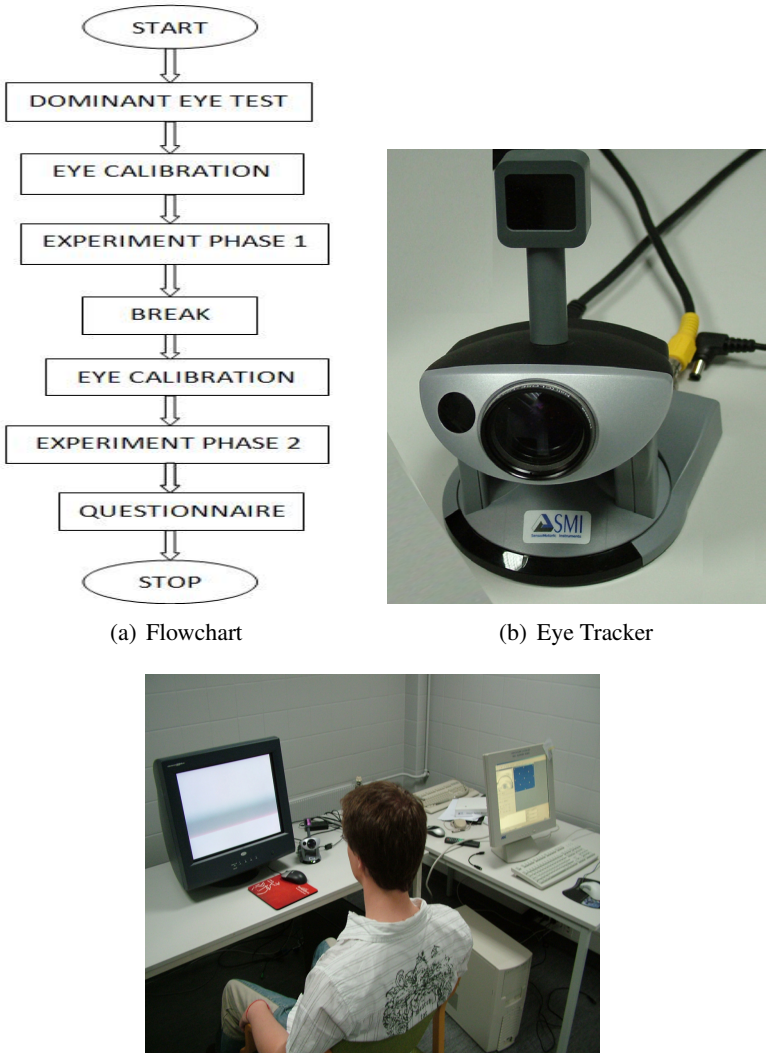
The authors of [12] used four low-level local image features: luminance, contrast and bandpass outputs of both luminance and contrasts to conclude that image patches around human fixations had, on average, higher values of each of these features than image patches selected at random. Foveated framework was used for analysis of the fixations. Circular patches of diameters 32, 64, 96, 160, 192 pixels centered at each fixation were extracted. Each image was displayed for 5 sec. in a fixed order for all observers. The image patches around observers' fixation points were then analyzed to determine if the statistics of the four image features: luminance, contrast, luminance-bandpass, and contrast-bandpass were statistically different from image patches that were picked randomly. Contrast-bandpass showed the greatest difference between the human and random fixations, followed by luminance bandpass, RMS contrast and luminance. Using these measurements GAFFE (Gaze-Attentive Fixation Finding Engine) is proposed by the authors [12].

The first objective of this paper is to study the experimentally obtained gaze maps and for this purpose an experiment was carefully designed in order to analyze the human fixation under natural conditions. The details of the experiment are discussed in the section 3. Second objective is to compare the three algorithms (Saliency [17], GAFFE [12] and GBVS [4]) and modify them in order to improve their performance in prediction of the gaze maps.

2 Experiment Setup

Subjects were shown a set of 190 images and data was recorded as they viewed each image. Each image was shown for 2 sec only, similar to the experiment performed in [1]. The experiment was divided into two phases in order to provide adequate rest to the test subjects. As shown in figure 1(a) the experiment was performed in 6 steps. First of all the dominant eye of the test subject was found by using Porta test [13]. After finding the dominant eye of the test subject the eye tracker was calibrated to it. Performing the calibration on the dominant eye yields accurate result which is an important issue in this project [[13, 11, 10]. After calibration experiment phase 1 was started. In this phase each subject was shown a set of 90 images for a total period 180 seconds. Each image was shown for a period of 2 seconds only. After this phase subject was given a break to rest his or her eyes. The eye tracker is recalibrated before entering the experiment phase 2. In experiment phase 2 each subject was shown a set of 100 images for a total period of 200 seconds. Each image was shown for 2 seconds only. After completing this step subject was given a questionnaire to fill. The distance between the eye of the subject and the monitor was kept at 70 cm giving a viewing angle of 30x23 degrees.The intensity of light at the front of display, back of the display was 159 lux and 146 lux respectively as measured from i1Display device. The subjects were not given any instructions to judge the images or look for something in particular. The trails were performed under free view conditions on the dominant eye of the user using eye tracker equipment available in the lab. The eye tracker shown in figure 1(b) was used to measure the fixation

of the user's eye at important parts of the image.Figure 1(c) shows the setup of the experiment with the subject seated on the chair. Subject looking on the display pc (on the left) and the eye movements being recorded by the pc connected to the eye tracker (on the right). Remote Eye Tracking Device was used for the experiment. Observers felt worse with Head Mounted Eye Tracking Device (HED) than Remote Eye Tracking Device (RED) because of the size and weight of HED. The difference of precision between HED and RED is about 10-16 pixels for printed images [9]. These reasons favored the use of RED for this experiment. It can be observed that none of the two images have the similar background but same people are appearing in different images. Some random images were added to break the sequence feeling during viewing of similar images and to maintain the level of interest among the subjects.



(a) Flowchart (b) Eye Tracker

(c) Experiment Setup

Figure 1: Experiment Flow and Setup

3 Results

The gaze maps obtained for 20 users were summed and normalized to get 190 gaze maps corresponding to 190 images. There were 190 images with 212 faces in them, 17 grayscale images and 39 images with

no faces. Objects including faces at center of the image were fixated more as compared to faces or objects at other parts of the image. But in a few images faces at the other parts of images were fixated more than faces at the center. This was possibly because subjects fixated more on the changes i.e. new faces or new objects appearing in the images. The area of the face regions in the image was also a factor as larger face (as a result of person standing near to the camera) was fixated more as compared to smaller face. After the faces the most prominent regions for subjects were objects like toy car, toy banana, mobile phones, magic cube, books etc. Numbers, alphabets on posters were also strong in attracting attention of subjects.

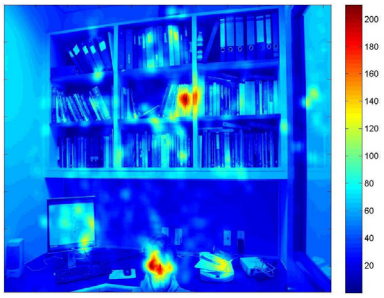
Figure 2(a) shows the averaged gaze map. We can see that there is large variability in the viewing patterns of the observers. But some regions are fixated more than others. In this picture the face region and the book at center of the book shelf are the two most fixated regions which can be judged by the redness in these regions of the heatmap. Scanner and the PC monitor can also be classified as other less fixated regions. However there are also other regions near the top of book shelf and corner of the book shelf that have been fixated which may or may not contain any information. This may be due to the randomness in the viewing pattern of eyes.

Figure 2(b) shows the Modified Saliency map (MSM) which is obtained by combination of saliency map and face detection module. MSM is superimposed over the original image for better analysis of the results. Figure 6 shows the Modified GAFFE map (MGAFFE) which is obtained by combination of GAFFE map and face detection module. Figure 2(d) shows the Modified GBVS map (MGBVS) which is obtained by combination of Graph based Visual Saliency map and face detection module. In the Face detection module the feature selection and classifier training was done using a variant of Ada Boost algorithm. Each weak classifier depends on a single Modified Census Transform (MCT) feature. They were combined into a cascade to make a strong classifier, so called MCT cascade, which is a weighted combination of weak classifiers [14]. All the three maps MSM, MGAFFE and MGBVS are calculated for 7 most salient regions so that their performance can be compared with Gaze map. In this manner the maps were obtained for all the 190 images used in the experiment. Area under the ROC curve was used as performance metric for the three algorithms. Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1.0. It is the measure of the performance of a classifier. Ideal classifier gives an AUC of 1.0 where as random guessing produces an AUC of less than 0.5 [2].

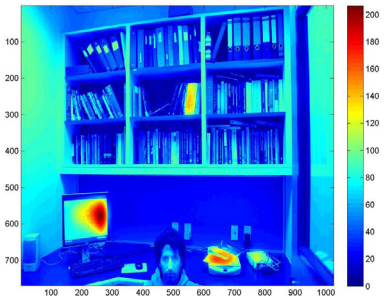
The histogram for AUC of Saliency map has a mean value of 0.5281. It means that its performance is marginally better than random classifier. The histogram for AUC of Modified Saliency map has a mean value of 0.7035. It shows an improvement in the performance of the classifier by more than 33 percent. The histogram for AUC of GAFFE map has a mean value of 0.4587. It means that its performance is worse than random classifier. The histogram for AUC of Modified GAFFE map has a mean value of 0.7035. The histogram for AUC of GBVS map has a mean value of 0.5579. Its performance is marginally better than random classifier and mean AUC of Saliency map. The histogram for AUC of Modified GBVS map has a mean value of 0.7450. It has the highest value among the modified classifiers of the above mentioned algorithms. Table 1 summarizes the performance of all the algorithms.

Table 1: Performance of Saliency Algorithms

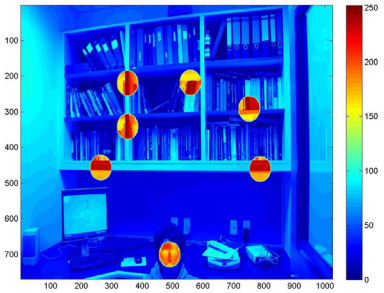
Algorithm	Mean AUC
Saliency Map	0.5281
Modified Saliency Map	0.7035
GAFFE	0.4587
Modified GAFFE	0.7035
GBVS	0.5579
Modified GBVS	0.7450



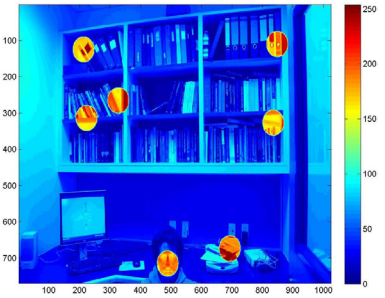
(a) Gaze Map



(b) Modified Saliency Map



(c) Modified GAFFE Map



(d) Modified GBVS Map

Figure 2: Comparison of Gaze with Modified Saliency Maps

4 Conclusion

The analysis of saliency maps shows the role played by low level image features. Intensity, color and orientation maps play a key role in the selection of the salient regions in an image. As saliency map selects the object which is differentiated from the other objects in terms of its higher intensity, color and

orientation. Additionally familiar objects tend to attract our attention, especially human faces. Therefore, we opted to investigate the performance improvement of different saliency algorithms by including a top-down face saliency module. Assuming that attention is drawn to the changes in the environment of the image; the order in which the scene should be inspected is determined by the saliency map. Since the mean AUC for GBVS is the highest (i.e. 0.7450) , It concludes that Modified GBVS is able to predict the gaze map more accurately as compared to Modified GAFFE Map and Modified Saliency Map. The addition of the face detection improved significantly the performance of all the three algorithms i.e. Saliency Map, GAFFE and GBVS. In can be further improved by addition of familiar object detection.

References

[1] Moran Cerf, Jonathan Harel, Wolfgang Einhauser, and Christof Koch. Predicting human gaze using low-level saliency combined with face detection. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems (NIPS 2007)*, pages 241–248, Cambridge, MA, 2007.

[2] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters, Elsevier*, (27):861–874, 2006.

[3] Jillian H. Fecteau and Douglas P. Munoz. Salience, relevence, and firing: a priority map for target selection. *Trends in Cognitive Sciences, Elsevier*, 10:382–390, 2006.

[4] Jonathen Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in Neural Information Processing Systems (NIPS 2006)*, pages 545–552, 2006.

[5] John M. Henderson. Human gaze control during real-world scene perception. *TRENDS in Cognitive Sciences, Elsevier*, 7(11):498–504, November 2003.

[6] Laurent Itti. *Models of Bottom-Up and Top-Down Visual Attention*. PhD thesis, California Institute of Technology, Pasadena, California, 2000.

[7] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Neuroscience 2001*, 2(3):194–203, 2001.

[8] Timothee Jost, Nabil Ouerhani, Roman Von Wartburg, Rene Muri, and Heinz Hugli. Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding, Elsevier*, 100:107–123, 2005.

[9] Barbora Kominkova, Marius Pedersen, Jon Y. Hardeberg, and Marie Kaplanova. Comparison of eye tracking devices used on printed images. In *Human Vision and Electronic Imaging XIII.Proceedings of the SPIE*, volume 6806, pages 68061I–68061I–12, 2008.

[10] Ayame Oishia, Shozo Tobimatsub, Kenji Arakawaa, Takayuki Taniwakia, and Jun ichi Kira. Ocular dominancy in conjugate eye movements at reading distance. *Neuroscience Research, Elsevier*, 52(3):263–268, July 2005.

[11] Marius Pedersen. Importance of region-of-interest on image difference metrics. Master's thesis, Department of Computer Science and Media Technology, Gjovik University College, 2007.

[12] Umesh Rajashekar, Ian van der Linde, Alan C. Bovik, and Lawrence K. Cormack. Gaffe: A gaze-attentive fixation finding engine. *IEEE Transcations on Image Processing*, 17(4):564–573, April 2008.

[13] Heidi L. Roth, Andrea N. Lora, and Kenneth L. Heilman. Effects of monocular viewing and eye dominance on spatial attention. *Brain*, 125(9):2023–2035, 2002.

[14] Tiffany Sauquet, Yann Rodriguez, and Sebastien Marcel. Multiview face detection. IDIAP-RR 49, IDIAP, 2005.

[15] Benjamin W. Tatler, Roland J. Baddeley, and Iain D. Gilchrist. Visual correlates of fixation selection: effects of scale and time. *Vision Reserach, Elsevier*, (45):643–659, 2005.

[16] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*, 2001.

[17] Dirk Walther. *Interactions of Visual attention and Object Recognition : Computational Modeling, Algorithms and Psychophysics*. PhD thesis, California Institute of Technology, Pasadena, California, 2006.

Correlation between Visual and Instrumental Color Differences of Metallic Paints

Alain Trémeau
Laboratoire Hubert Curien UMR CNRS 5516
Université Jean Monnet, Saint Etienne, France
alain.tremeau@univ-st-etienne.fr

Christoph Godau
Master CIMET
Université Jean Monnet, Saint Etienne, France

Abstract

The objective of this paper is to propose a new dataset more reliable than the previous datasets and to evaluate from this dataset more accurate just noticeable differences. In our study we have considered not only surface specimens (reflective objects) but also color pairs assessed using visual displays (e.g., self-luminous LCD). The aim is to use this dataset to develop/test new color spaces with Euclidean color-difference formulas more robust than the CIEDE2000 color-difference formula.

1 Introduction

The Technical Committee 1-55 of the International Commission on Illumination on ”Uniform color space for industrial color difference evaluation” requested in 2006 reliable experimental datasets (that is, color pairs assessed by a considerable number of observers with non-defective color vision, under well controlled experimental conditions and using proper experimental methods). The aim was to use these datasets to develop/test new color spaces with Euclidean color-difference formulas more robust than the CIEDE2000 color-difference formula. The objective of this paper is to propose a new dataset more reliable than the previous datasets and to evaluate from this dataset more accurate just noticeable differences. In our study we have considered not only surface specimens (reflective objects), but also color pairs assessed using visual displays (e.g., self-luminous LCD). Our experimental results have been obtained under illuminating/viewing conditions close to the ”reference conditions” suggested for the CIEDE2000 color-difference formula (CIE Publication 142-2001) and by the ISO standard 3668:2003 (ISO standard 3668:2003 Paints and varnishes - Visual comparison of the colour of paints). In our study we have controlled all viewing conditions including the spectral power distribution of the viewing environment and the spectral reflectance factors of surface specimens. Detailed description of the experimental conditions and method employed are given below. Thirty three color standards (paint metallic surface specimens) had been used (see Table 1), for each color (center) 26 samples had been considered (see example given in Table 4). The standards had been chosen such as (see Figure 1):

- Some of them are closed to the color centers investigated by other studies, such as those published by Melgosa in 1997 [3], in order to compare our results with those previously published.
- Most of them are closed to the axis of hue discontinuity defined by Sharma in 2005 [4]. We have split the L*a*b* color space in eighteen areas delimited by six hue domains (see Figure 1) and three lightness domains (L*<30, 30<L*<60, 60<L*) in order to study the computation errors due to color difference formulas.
- Others were chosen to cover the entire color space. That is most of these standards have a lightness 30<L*<60.
- Their saturation is low enough to be reproduced correctly on the LCD display without problems of color gamut mapping.

The samples had been defined such as:

- The color difference (ΔE76) between each sample and the corresponding color center is lower than 1.5.

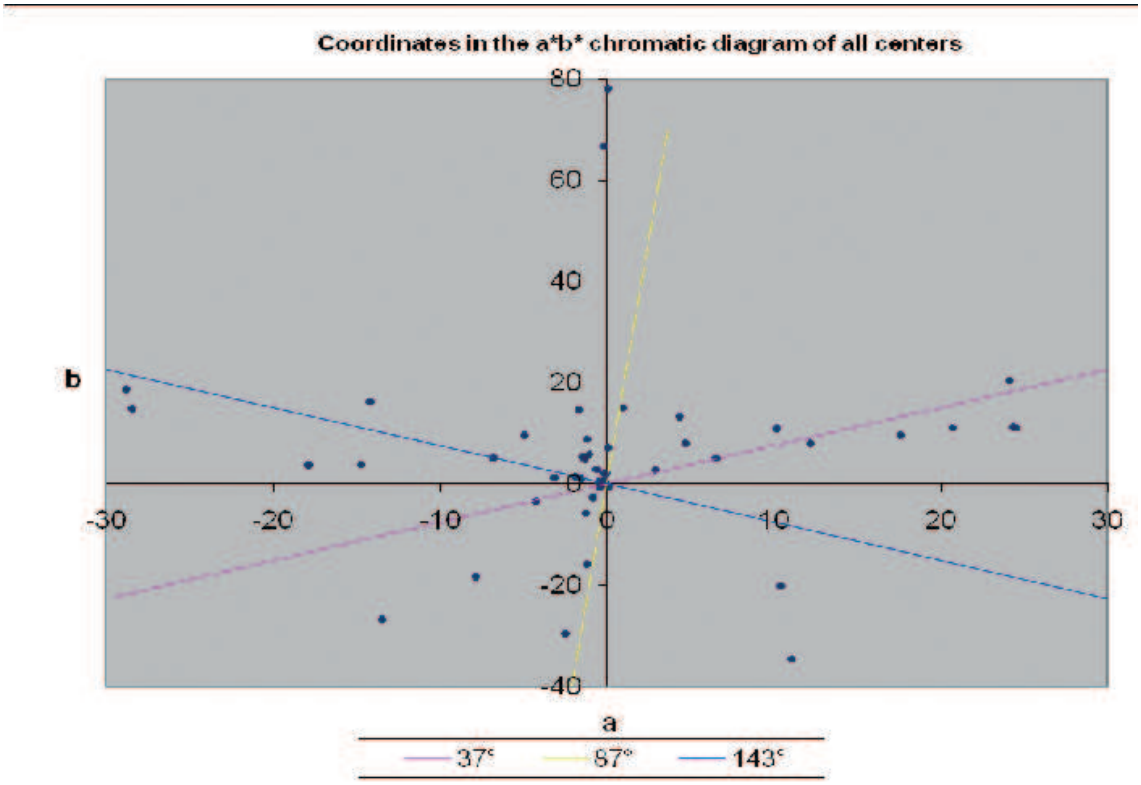


Figure 1: Color coordinates in a*b* of surfaces specimens. Color axes for which the CIEDE2000 color-difference formula introduces computation errors due to various hue discontinuities.

- The color difference are defined in six directions (+a*, -a*, +b*, -b*, +L*, -L*) in order to obtain assumed color matching points.
- At least 10 samples (among the 26 assessed) can be fitted by the ellipsoids obtained earlier (e.g. [3]).
  - The data had been firstly adjusted in order to take care about hue discontinuities (see [4]) and inconsistencies between ellipsoid patterns underlined by previous studies (see [1]).
  - Next the data had been tested by five observers then adjusted (after this first campaign of evaluation) in order to increase the number of samples for which the difference to the corresponding color center is just noticeable.

2 Experimental Procedures

2.1 Observers

Twenty ”standard” observers took part in the experiment: sixteen males and four females, eleven from European countries and nine from outside Europe countries, 60% of observers aged between 24 and 27 years, 10% observers aged between 31 to 35 years, and 30% aged between 36 and 45 years. All were color science postgraduate students or researchers experienced in making color judgments and in performing psychophysical tasks. The observers were screened for color vision deficiencies by Ishihara

Table 1: Color coordinates in L*a*b* of surfaces specimens and of color displayed on the LCD display used (after calibration of the LCD). Average=0,461. These values had been measured with the Minolta CS1000 tele-spectro-radiometer.

Center	L *	a*	b*	Le (cd/m2)	L*	a*	b*	Δ E76
1	79,15	-2,34	14,52	82,96	79,21	-1,89	14,22	0,544
2	21,70	-1,55	-8,66	5,77	21,95	-1,502	-8,99	0,413
3	23,38	19,81	-45,70	6,05	23,32	20,2	-45,98	0,483
4	49,52	-0,35	-18,29	28,47	49,68	0,263	-18,11	0,662
5	33,48	-8,02	13,01	11,67	33,35	-7,466	12,79	0,607
6	50,33	-17,38	17,83	27,44	50,22	-17,54	17,97	0,239
7	21,26	-25,54	3,48	4,86	21,21	-25,6	3,439	0,086
8	25,08	-1,17	-4,11	6,56	24,94	-1,13	-4,61	0,524
9	27,51	-0,58	2,89	7,64	27,13	-0,92	2,779	0,522
10	70,88	-1,51	9,47	61,49	70,87	-1,457	9,048	0,423
11	79,08	-1,57	0,07	80,35	79,03	-1,034	-0,38	0,705
12	71,82	-1,69	4,50	63,31	71,78	-1,68	4,297	0,210
13	25,84	14,40	18,17	6,97	25,51	14,18	17,58	0,710
14	22,21	6,80	14,57	5,35	22,07	6,743	14,44	0,199
15	31,72	30,53	17,86	11,28	31,63	30,93	17,22	0,760
16	18,83	4,64	4,64	4,48	19,03	4,705	4,212	0,476
17	88,37	0,21	7,39	107,5	88,32	0,0182	7,128	0,327
18	85,02	-1,56	4,88	97,48	85,02	-1,15	4,709	0,439
19	66,31	-0,50	-1,81	52,88	66,2	-0,457	-1,99	0,219
20	47,90	-0,04	1,55	24,18	47,84	-0,74	1,83	0,754
21	76,38	-1,34	5,12	74,41	76,22	-1,075	4,982	0,336
22	86,32	-1,88	0,76	101,7	86,47	-1,853	0,641	0,195
23	85,40	0,57	15,51	97,27	85,26	0,487	14,82	0,708
24	28,09	27,80	19,21	8,77	28,2	27,8	19,11	0,148
25	89,82	-1,39	4,53	111,2	89,88	-1,64	4,106	0,497
26	38,23	27,85	29,02	15,78	38,27	28,35	28,78	0,556
27	29,56	-25,15	2,79	8,93	29,36	-25,04	2,27	0,566
28	36,49	53,71	36,08	13,99	36,53	53,01	36,4	0,770
29	61,88	4,22	14,06	43,49	61,76	4,705	13,86	0,541
30	28,55	1,69	-38,28	8,02	28,59	1,76	-38,42	0,161
31	6,94	0,78	-1,61	1,13	6,766	0,945	-1,334	0,364
32	25,32	24,02	18,47	6,7	25,18	23,73	17,98	0,586
33	38,23	27,85	29,02	82,96	79,21	-1,89	14,22	0,544

pseudo-iso-chromatic plates and by Farnsworth-Munsell 100 Hue Test. The experiments were also done by ten professional colorists. Each observer made two repeated comparisons of each of the 20 pairs of stimuli using a two-stage procedure:

- In the first stage, the observers established an asymmetric comparison between the standard patch (center) and the sample patch (target) in the cabinet.
- In the second stage, the observers established an asymmetric comparison between the standard patch (center) and the sample patch (target) on the monitor.

The observers were instructed to begin the matching with the light booth cabinet; this choice is arbitrary, with the sole purpose of standardizing the conditions for all observers.

The color of the patches displayed on the LCD was controlled using a two-stage procedure:

- In the first stage, the LCD was calibrated by a colorimeter Gretag Macbeth i1Pro.
- In the second stage, all colors were adjusted individually (using the Minolta CS1000 tele-spectro-radiometer) in order to match color values for the color patches and the LCD color patches. We carried out that color patches vary according to the viewing conditions (see [1]).

2.2 Experimental Procedures

All the pairs of samples considered in this study had been compared from surface specimens viewed in a light booth cabinet (Verivide cabinet) and from color patches displayed on a LCD screen (Eizo ColorEdge CG241 TFT monitor). The viewing conditions used for the light booth cabinet was: geometry 45°/0 (i.e. perpendicularly to their surface), D65, neutral background (Munsell N5: Ye = 0.36 W/sr.m2, L* = 77, a* = -0.77, b* = 0.55). The luminance of the light booth cabinet was controlled such as: Lw the luminance of the white patch set to 120 cd/m2 (Ye = 0.614 W/sr.m2, L* = 100, a* = 0.011, b* = 0.025). The viewing conditions used for the LCD screen was set to: D65, neutral background (Ye = 0.36 W/sr.m2, L* = 77, a* = -0.77, b* = 0.55), Lw =120 cd/m2 (Ye = 0.614 W/sr.m2, L* = 100, a* = 0.011, b* = 0.025). The observers were instructed to face the cabinet or monitor directly, so that the observational angle with all stimuli was normal. The laboratory was dark except for the LCD display and the viewing cabinet. The size of surface specimens and of color patches displayed on the LCD screen is identical: 170x80 mm (± 3mn). For each color, the gloss of surface specimens is identical(± 3). The two samples are compared side by side (at about 500 mm from the eyes), separated by a black area of 2mm wide.

The instructions given to the observers are in both cases identical. If the observer see a noticeable color difference (rating value 2) between the two samples he/she have to indicate if there is a difference of hue, of chroma or of lightness between the two sample considered (see Table 2).

The difference of Lightness (DL) is assessed according a rating scale graded from -2 to 2. As example if the lightness of the sample at right is just noticeably lower than the lightness of the sample at left DL = -1. The difference of chroma (DC) is assessed according a rating scale graded from -2 to 2. The differences of hue (DH) are assessed according four rating scales (yellow, green, red, and blue) graded from 0 to +2. As example if the sample at right is noticeably more yellowish than the sample at left DH = +2.

3 Results

In this paper we report only results computed for some color sample values. The result of our experimentations showed that after spectral calibration of the LCD display its accuracy was enough good to

Table 2: Rating Values Used to Assess Color Differences.

Rating value	Degre of difference
0	No noticeable difference
1	Just noticeable color difference
2	Noticeable color difference

Table 3: Color coordinates (in L*a*b*) of color pairs tested to compute the parameter of the ellipsoid corresponding to standard N° 30. To automatically fit ellipses to our test data we have used the method provided in [3], according to the authors at least 6 data points are needed for fitting. For this target the ellipsoid was computed from 10 points (those for which the average rating value is lower than 1).

Sample			Targets								
L*	a*	b*	L*	a*	b*	Δ L*	Δ a*	Δ b*	DE76	Average Rating value	Standard deviation
8,64	0,34	2,64	9,64	0,34	2,64	1	0	0	1	0,54	0,69
8,64	0,34	2,64	9,54	1,24	2,64	0,9	0,9	0	1,27	1	0,77
8,64	0,34	2,64	9,54	0,34	3,54	0,9	0	0,9	1,27	0,63	0,67
8,64	0,34	2,64	9,24	1,04	3,34	0,6	0,7	0,7	1,16	1,73	0,64
8,64	0,34	2,64	7,74	0,34	2,64	-0,9	0 0	0,9		1,27	1,01
8,64	0,34	2,64	7,84	-0,46	2,64	-0,8	-0,8	0	1,13	1,54	0,69
8,64	0,34	2,64	7,84	0,34	1,84	-0,8	0	-0,8	1,13	1,63	0,67
8,64	0,34	2,64	8,04	-0,26	2,04	-0,6	-0,6	-0,6	1,04	1,36	0,92
8,64	0,34	2,64	7,84	1,14	2,64	-0,8	0,8	0	1,13	1,54	0,69
8,64	0,34	2,64	8,64	1,24	2,64	0	0,9	0	0,9	1,45	0,82
8,64	0,34	2,64	8,64	1,14	1,84	0	0,8	-0,8	1,13	1,09	0,83
8,64	0,34	2,64	8,64	1,14	3,44	0	0,8	0,8	1,13	1,54	0,82
8,64	0,34	2,64	8,64	0,34	3,54	0	0	0,9	0,9	0,54	0,69
8,64	0,34	2,64	8,64	-0,56	2,64	0	-0,9	0	0,9	0,54	0,69
8,64	0,34	2,64	8,64	-0,46	1,84	0	-0,8	-0,8	1,13	1,63	0,67
8,64	0,34	2,64	8,64	-0,46	3,44	0	-0,8	0,8	1,13	1	0,89
8,64	0,34	2,64	8,64	0,34	1,74	0	0	-0,9	0,9	1,18	0,75
8,64	0,34	2,64	7,84	0,34	3,44	-0,8	0	0,8	1,13	1,27	0,90
8,64	0,34	2,64	9,44	0,34	1,84	0,8	0	-0,8	1,13	1,18	0,98
8,64	0,34	2,64	9,44	-0,46	2,64	0,8	-0,8	0	1,13	1,27	0,78
8,64	0,34	2,64	9,24	-0,26	2,04	0,6	-0,6	-0,6	1,04	0,91	0,83
8,64	0,34	2,64	8,04	0,94	3,24	-0,6	0,6	0,6	1,04	1,45	0,82
8,64	0,34	2,64	9,24	0,94	2,04	0,6	0,6	-0,6	1,04	0,82	0,87
8,64	0,34	2,64	8,04	-0,26	3,24	-0,6	-0,6	0,6	1,04	1,45	0,69
8,64	0,34	2,64	8,04	0,94	2,04	-0,6	0,6	-0,6	1,04	1,27	0,90

Table 4: Color differences (in L*a*b*) of color pairs tested to compute the parameter of the ellipsoid corresponding to standard N° 23. At right: the average rating values corresponding to the experimental campaign based only on LCD display assessments with 26 pairs of samples. The results confirm that the size of the ellipsoid was correctly estimated and that we can't increase the size of the ellipsoid for this target even if the Δ E is quite small!

Δ L*	Δ a*	Δ b*	ΔE76	ΔECIE2000	Average Rating value	Standard deviation
0,6	0	0	0,6	0,79	1	0,77
0,5	0,6	0	0,78	0,80	1,45	0,82
0,5	0	0,8	0,94	0,72	1,91	0,30
0,4	0,5	-0,8	1,02	0,88	1,63	0,67
-0,6	0	0	0,6	0,38	1,45	0,82
-0,5	-0,6	0	0,78	0,79	2	0
-0,5	0	-0,8	0,94	0,74	1,91	0,30
-0,4	-0,5	-0,8	1,02	0,98	2	0
-0,5	0,6	0	0,78	0,80	2	0
0	0,7	0	0,7	0,86	2	0
0	0,6	-0,8	1	0,93	1,72	0,47
0	0,6	0,8	1	1,05	2	0
0	0	0,8	0,8	0,65	1,91	0,30
0	-0,7	0	0,7	0,83	1,91	0,30
0	-0,6	-0,8	1	1,04	1,91	0,30
0	-0,6	0,8	1	0,89	1,72	0,64
0	0	-0,8	0,8	0,67	1,6	0,67
-0,5	0	0,8	0,94	0,73	1,91	0,30
0,5	0	-0,8	0,9	0,74	1,82	0,40
0,4	-0,6	0	0,72	0,75	1,63	0,67
0,4	-0,5	-0,8	1,02	0,98	1,91	0,30
-0,4	0,5	-0,8	1,02	0,88	1,55	0,82
0,4	0,5	-0,8	1,02	0,88	1,18	0,98
0,4	-0,5	0,8	1,02	0,85	1,36	0,50
-0,4	-0,5	0,8	1,02	0,85	2	0
-0,4	0,5	-0,8	1,02	0,88	1,6	0,81

provide the same visual stimulation to the user's eye as the reflected light form the hard copy proof. Over 30 color discrimination ellipsoids were fitted from our datasets:

- 10 from visual assessments done by "standard" observers (from assessments done in the cabinet).
- 5 from visual assessments done by professional colorists (from assessments done in the cabinet).
- 10 from visual assessments done by "standard" observers (from assessments done on the LCD display).
- 5 from visual assessments done by professional colorists (from assessments done on the LCD display).

33 color discrimination ellipsoids were computed from our datasets from the ellipsoids coefficients investigated by Melgosa [1]. It was found that:

- 40% of the color samples used in our datasets had been rated by ”standard” observers with a score equal to 1 (1 = Just noticeable color difference), which is sufficient to establish an ellipsoid model. 75% of these color samples are consistent with an ellipsoid model. The sampling method that has been used to define these color samples seems therefore to give good results to compute a new ellipsoid model which better fit our dataset than well known ellipsoid models.

- Most of the ellipsoids computed from our datasets are consistent with the ellipsoids computed from the RIT-Dupont coefficients (see [3]) than with the ellipsoids computed from the DE2000 od DE94 formulas. They are in general consistent in terms of shape and orientation, but not always in terms of size. However, few are not consistent in term of shape and orientation. This was caused by different psychophysical methods and different anchoring pairs being used in various perceptibility experiments. The less consistent ellipsoids are either closed to the hue axis shown in Figure 1 or closed to the achromatic axis.

- The shape and the orientation of these ellipsoids are consistent between standard observers and professional colorists, but not in size. The size of ellipsoids is always over-estimated by standard observers in regards to the size estimated by professional colorists (as example see Table 3).

- The shape and the orientation of these ellipsoids are consistent for visual assessments done in the cabinet and visual assessments done on the LCD display, but not in size. For most of centers the size of ellipsoids is over-estimated by visual assessments done in the cabinet in regards to the size estimated on the LCD display (as examples see Tables 4 and 5).

- Most of the color differences (CIELAB based, CIE94, and CIEDE2000) computed from our datasets are not consistent with visual assessment because color-difference formulas introduce computation errors due to some discontinuities for various values of hue and of chroma (see [2]).

All these results showed that LCD displays can be used to compute color differences of metallic surfaces by ”virtual proofing” (see Table 6).

References

[1] H. Komatsubara, S. Kobayashia, N. Nasuno, Y. Nakajima, and S. Kumada. *Visual Color Matching under Various Viewing Conditions* . Color research and application, 2002. Vol. 27, no 6, pp 399-420.

[2] R.G. Kuehni. *Analysis of five sets of color difference data*. Color Research and Application, 2001. Vol. 26, no. 2, pp 141-450.

[3] M. Melgosa, E. Hita, A. J. Poza, David H. Alman, and Roy S. Berns. *Suprathreshold color-difference ellipsoids for surface colors* . Color Research and Application, 1997. Vol. 22, no. 3, pp 148-155.

[4] G. Sharma, W. Wu, and E. N. Dalal. *The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations*. Color Research and Application, 2005. Vol. 30. no. 1.

Table 5: Color differences (in L*a*b*) of color pairs tested to compute the parameter of the ellipsoid corresponding to standard N° 26. At right: the average rating values corresponding to the experimental campaign based only on LCD display assessments with 26 pairs of samples. The results confirm that the ellipsoid is now better adjusted and that we can't increase the size of the ellipsoid for this target even if the Δ E is quite small!

Δ L*	Δ a*	Δ b*	ΔE76	ΔECIE2000	Average Rating value	Standard deviation
0,5	0	0	0,5	0,45	0,63	0,81
0,4	0,4	0	0,56	0,45	0,82	0,98
0,4	0	0,7	0,80	0,51	0,73	0,90
0,3	0,3	0,4	0,58	0,33	1,09	0,70
-0,5	0	0	0,5	0,44	1,09	0,94
-0,4	-0,4	0	0,56	0,45	1,45	0,82
-0,4	0	-0,6	0,72	0,48	1,18	0,87
-0,3	-0,3	-0,4	0,58	0,33	1,54	0,69
-0,4	0,4	0	0,56	0,44	1,09	0,83
0	0,5	0	0,5	0,34	0,82	0,87
0	0,4	-0,6	0,72	0,53	0,82	0,87
0	0,4	0,6	0,72	0,29	0,82	0,75
0	0	0,7	0,7	0,37	0,36	0,67
0	-0,5	0	0,5	0,35	1,73	0,47
0	-0,4	-0,6	0,72	0,29	1,73	0,44
0	-0,4	0,6	0,72	0,52	1,27	0,78
0	0	-0,7	0,7	0,38	1,09	0,83
-0,4	0	0,6	0,72	0,47	0,63	0,67
0,4	0	-0,6	0,72	0,48	0,82	0,87
0,4	-0,4	0	0,56	0,45	1,18	0,60
0,3	-0,3	-0,4	0,58	0,33	1,09	0,83
-0,3	0,3	0,4	0,58	0,33	1,27	0,78
0,3	0,3	-0,4	0,58	0,45	1,36	0,81
0,3	-0,3	0,4	0,58	0,45	1,54	0,82
-0,3	-0,3	0,4	0,58	0,45	1,45	0,69

Table 6: Rates of Good Detection (RGD) and of Bad Detection (RBD) of the LCD display assessment done by experts when the overall rating value = 2. Most of the errors of assessment done by experts are due to the lightness component, next to the a* and to the b* component. When two samples are perceptually different the experts attribute in general this difference to the lightness rather than to the hue and never to the chroma. These rates of Good Detection are similar to those obtained for the assessment of surface samples.

L*	a*	b*
RBD = 30%	RBD = 25%	RBD = 10%
RGD = 70%	RGD = 75%	RGD = 90%

Performance of the Difference of Gaussians Model in Image Difference Metrics

Sebastien A. Ajagamelle
Grenoble-INP Pagora and Gjøvik University College
Grenoble, France - Gjøvik, Norway
Sebastien.Ajagamelle@bvra.etu.grenoble-inp.fr

Gabriele Simone
Gjøvik University College
Gjøvik, Norway
gabriele.simone@hig.no

Marius Pedersen
Gjøvik University College
Gjøvik, Norway
marius.pedersen@hig.no

Abstract

We propose two novel image difference metrics using an extension of the S-CIELAB framework, which are based on the Difference of Gaussians model. The first metric uses the Difference of Gaussians model as a basis for the spatial filtering with the ΔE^*_{ab} as a color difference formula, while the second uses the same model in association with the ΔE_E color difference formula. A dataset with 20 gamut mapped images was selected for the evaluation of the metrics. The performance in correlation of the metrics are significantly better than the S-CIELAB difference metric for this set of images.

1 Introduction

Image quality is a key issue in the graphical industries and it can be defined as the truthfulness to the original image, as well as the naturalness or the degree of excellence exhibited by an image [5]. Quality is influenced by a considerable number of parameters, and a metric is therefore required to get an absolute and accurate assessment of this feature. So far many different image quality metrics have been proposed, some with the intent of measuring general image quality and some for detecting specific distortions. Determining image quality is difficult and complex, because of this many researches have chosen a less complicated approach by calculating image difference. However, at the moment, an universal image-difference metric does not exist. Color difference formulas were primarily implemented as global measures to quantify the difference between two patches. These formulas have also been used to calculate image difference. In these cases a pre-processing based on characteristics of the human visual system is usually performed before computing the pixelwise color difference over the image.

An example of an image difference metric is the spatial extension of the CIELAB color-difference formula [1] (S-CIELAB) suggested in 1997 by Zhang and Wandell [14]. It introduces a spatial filter which simulates the human visual system. Johnson and Fairchild [6] followed the S-CIELAB approach, but the spatial filter is implemented in the frequency domain, allowing for more precise control of the filter. In 2002 Hong and Luo [4] proposed the hue angle algorithm still based on the CIELAB color difference and correcting some of its drawbacks. However, not including a spatial filtering of the image, it is unsuitable for halftone images [10]. Because of this Pedersen et al. [10] proposed two new image-difference metrics based on the hue angle algorithm including a spatial filtering similar to that used by Zhang and Wandell [14] and Johnson and Fairchild [6].

In general these metrics compute the difference using the CIELAB color-difference formula. Very recently a new one called ΔE_E has been published. It is an Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space [9]. The BFD color difference data represented in the OSA-UCS space revealed a regularity not existing in the CIELAB space. This induced the authors to represent the small-medium color differences by a simple ellipsoidal equation, termed ΔE_{GP}, and finally to suggest a proper logarithmic compression of the OSA-UCS space with a consequent reduction of the parameters and new formula.

The Contrast Sensitivity Function (CSF) is commonly used in image difference metrics as a spatial filtering to simulate the human visual system. The achromatic sensitivity is equivalent to a band-pass spatial filter, while the chromatic CSF demonstrates low-pass behavior [6]. The Movshon CSF [8] is among the most common functions, and is an exponential function that behaves as a band-pass filter and it peaks around four cycles per degree. One shortcoming of this simple model is that the function does not take into consideration the viewing conditions [7]. By comparison, the Daly model [2] is a function of several parameters and is closer to the behavior of the human visual system. Luminance CSFs are still an area of research. Unfortunately, chrominance CSFs have not been investigated enough, leading to a fragile part of image difference metrics.

We present here a method based on the Difference of Gaussians (DOG) model, which has also been used as contrast measure by Tadmor and Tolhurst [13]. In the conventional model, a bi-dimensional Gaussian describes the spatial sensitivity in the center of receptive fields (central component) and the surround component is represented by a Gaussian curve as well, with a larger radius. The response of a neurone is assumed to depend solely on the local luminance difference between the center and the surround, and the model response depends on the contrast stimulus [11]. Different criteria for the measure of contrast were proposed. The incorporation of the DOG model has shown positive results for predicting perceived contrast, mostly due to its ability to consider edges and gradients in the image.

2 A Proposal of Two New Metrics

We propose two new image difference metrics referred as $S_{DOG} - CIELAB$ and $S_{DOG} - DEE$ based on the framework of S-CIELAB, which has been extended by Simone et al. [12]. Simone et al. proposed to exchange the color difference formula in S-CIELAB, with the new formula ΔE_E. We further refine the work by Simone et al. by focusing on the spatial filtering. The novelty of our technique is substituting the previous spatial filtering with a DOG calculation. The DOG model was chosen as a surrogate to the CSF filtering because it has revealed to be beneficial in the identification of edges while the CSFs are mainly used to modulate the less perceptible frequencies. We propose two different metrics, where the first ($S_{DOG} - CIELAB$) uses the DOG model as a basis for the spatial filtering with the ΔE^*_{ab} as a color difference formula, the other ($S_{DOG} - DEE$) uses the DOG model and ΔE_E as color difference formula.

3 Preliminary Results

For the evaluation of the proposed metrics we used a dataset with gamut mapped images from [3]. 20 different images were gamut mapped with 5 different algorithms. The 20 different images were evaluated by 20 observers in a pair comparison experiment. This is a more complex task for the observers since many artifacts must be considered, and also a demanding task for the image difference metrics. Because of that this data set is suitable for testing these two new image difference metrics.

In Figure 1 we can see the results from our new difference metrics and the S-CIELAB [14]. The $S_{DOG} - DEE$ correlation is clearly better than S-CIELAB for all the images except for one (Image 10) and the $S_{DOG} - CIELAB$ shows better correlation than S-CIELAB for 15 images.

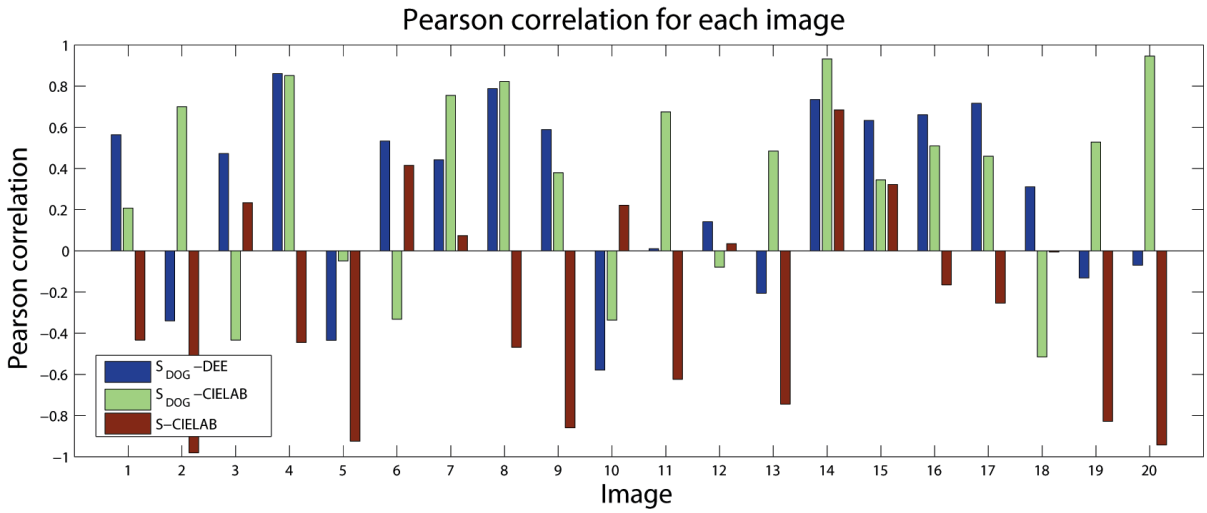


Figure 1: Pearson correlation for $S_{DOG} - DEE$, $S_{DOG} - CIELAB$, and S-CIELAB.

4 Conclusion

We presented here two new metrics based on the S-CIELAB framework. We proposed to perform the spatial filtering in S-CIELAB with the Difference of Gaussians model. The preliminary results of the two algorithms perform better than S-CIELAB on a data set of gamut mapped images. Further work includes in-depth testing of the parameters on the DOG model, as well as extensive evaluations on other data sets.

References

[1] CIE. Colorimetry. Technical Report 15, 2004.

[2] S. Daly. Subroutine for the generation of a two dimensional human visual contrast sensitivity function. Technical Report 233203y, Eastman Kodak, 1987.

[3] F. Dugay, I. Farup, and J. Y. Hardeberg. Perceptual evaluation of color gamut mapping algorithms. *Color Research & Application*, 33(6):470–476, Dec 2008.

[4] G. . Hong and M. R. Luo. New algorithm for calculating perceived colour difference of images. *Imaging Science Journal*, 54(2):86–91, 2006.

[5] R.E. Jacobsen. An evaluation of image quality metrics. *J. Photograph. Sci.*, 43:7–16, 1995.

[6] Garrett M. Johnson and Mark D. Fairchild. Darwinism of color image difference models. In *The 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, 2001.

[7] Garrett M. Johnson and Mark D. Fairchild. On contrast sensitivity in an image difference model. In *PICS*, pages 18–23, 2002.

[8] J. A. Movshon and L. Kiorpes. Analysis of the development of spatial sensitivity in monkey and human infants. *J. Opt. Soc. Am. A*, 5:2166–2172, Dec 1988.

[9] C. Oleari, M. Melgosa, and R. Huertas. Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space. *Journal of the Optical Society of America*, 26(1):121–134, 2009.

[10] Marius Pedersen and Jon Yngve Hardeberg. A new spatial hue angle metric for perceptual image difference. In *2009 Computational Color Imaging Workshop*, Saint Etienne, France, Mar 2009. Submitted.

[11] Alessandro Rizzi, Thomas Algeri, Giuseppe Medeghini, and Daniele Marini. A proposal for contrast measure in digital images. In *CGIV 2004 – Second European Conference on Color in Graphics, Imaging and Vision*, pages 187–192, 2004.

[12] G. Simone, C. Oleari, and I. Farup. Performance of the euclidean color-difference formula in log-compressed OSA-UCS space applied to modified-image-difference metrics. In *11th Congress of the International Colour Association (AIC)*, Sydney, Australia, Oct 2009. Accepted.

[13] Y. Tadmor and D.J. Tolhurst. Calculating the contrasts that retinal ganglion cells and lgn neurones encounter in natural scenes. *Vision Research*, 40:3145–3157, 2000.

[14] X. Zhang and B.A. Wandell. A spatial extension of CIELAB for digital color image reproduction. In *Soc. Inform. Display 96 Digest*, pages 731–734, San Diego, 1996.

Optimal Color Mapping of Multispectral Images

Bjorn Skovlund Dissing, Jens Michael Carstensen
(bdi@imm.dtu.dk)
Informatics and Mathematical Modelling
Technical University of Denmark, Building 321
DK-2800 Kgs. Lyngby, Denmark

Abstract

In this study we propose a method to map a multi-spectral image into a trichromatic color model, CIE XYZ and RGB.
The image is acquired with a relatively new camera technology called VideometerLab, which rapidly acquires 1 MegaPixel images in 18 distinct wavelengths in under 10 seconds. Eleven of the 18 recorded wavelengths resides in the visible spectrum, and the last 8 wavelengths are near infra red channels below 1000 nanometer.
To obtain this mapping we estimate an optimal projection into a subspace resembling a trichromatic color model. The projection is estimated by calibrating a modified ridge regression method with a set of physical standard colors, also known as the x-rite colorchecker standard. The true value of the physical colors in color spaces RGB and CIELAB is known, which makes it possible to regress an acquired image into either of these color spaces. Because of the sparse amount of data, we use a leave-one-out cross validation to validate the estimated model.
We see that by penalizing the curvature of the ridge coefficients we get slightly better results than when we penalize the slope of the ridge coefficients, and significantly better results than by using ordinary least square regression.

1 Introduction

The majority of today's color images are acquired with a CCD or CMOS chip equipped with some sort of band-filter, such as e.g. a Bayer filter which works as a mosaic filter. The filter is responsible for binning incoming photons into three primary channels representing colors; Red, Green and Blue.
This binning is based on an approximation to the CIE color matching functions(CMF) which again is created as a very good approximation to the human visual system. As also indicated by the CMF, the human eye is only sensitive in the area between 400 and 700 nanometer(nm) in the electromagnetic spectrum - also called the visible area. The CMF bands peaks around 600nm, 555nm and 445nm corresponding to just Red, Green and Blue colors.
This way of capturing color has definitely proven itself usable, but unfortunately it also has some drawbacks. The binning of the photons has the drawback that the intensity recorded in each channel is an integration over a range of wavelengths, passed through the band-filter, thus hiding the spectral radiant power distribution of the scene. This can lead to metameric failure which basically is what happens when color stimuli with the same tristimulus values has different spectral radiant power distribution; so basically due to a bad sampling. A drawback of the CMF approximation performed in traditional digital cameras is that the colors recorded are device dependent. This means that two different cameras records the scene slightly different from each other and different from the independent CIE tristimulus values - i.e. it is not colorimetric. Exactly these problems were investigated in [5] which made thorough investigations of these differences, and modeled them using polynomial regression.
One way to overcome these problems is by using multi-spectral imaging. In multi-spectral imaging, an image is recorded for each member in a set of wavelengths, where the set contains more than the normal three channels. There are different ways of creating multi-spectral images. One approach is to use a set of narrowband filters which basically makes a more delicate grouping of the reflected light from the

scene, and thereby makes a better approximation to the spectral radiant power distribution as suggested in [2]. Another approach, which is the one used in this paper is commercialized by Videometer[1] in the camera VideometerLab. Here a set of Light Emitting Diodes(LEDs) is strobing successively, and an image is recorded for each LED. This means we can get exact information about reflection of specific wavelengths, and thereby sample the spectral radiant power distribution at specific points, corresponding to the properties of the LEDs. Further description of this camera can be found in next section.
This type of multi-spectral imaging device works somewhat like a spectrometric device, and therefore have many industrial applications where it is advantageous to be able to gain reflection information of specific wavelengths as well as getting spatial information. The specific wavelength information might reveal information about the actual chemical composition of the subject, because it is known which chemical bonds reflects/absorbs which wavelengths. Furthermore, such a sampling of the spectral radiant power distribution can also overcome the problems with metameric failure, and together with a proper mapping routine, also estimate device independent colors.
Creating this mapping routine is however not a trivial task, and we need some way to map the multi-spectral image into the device independent tristimulus values based on the CMF. Such a mapping routine is definitely not a trivial task and is the motivation for this paper. [1] studied the accuracy of transformations between RGB and CIELAB color spaces and developed alternative approximations, because existing transformations could be time consuming. A study very similar to this was carried out by [6], who used linear and polynomial regression to map multi-spectral images into standard color spaces, CIELAB and CIEXYZ.
Applying a color reconstruction using regression from a multi-spectral space into RGB or CIEXYZ space is a matter of finding a set of weights for the recorded channels which make out an optimal projection. We will test different regression themes, primarily by regularizing an ordinary least squares fit[4]. A set of training data is needed in order to calibrate the model properly. Different standards could be used for this, such as the NCS colorsystem which covers a vast amount of colors. We have limited ourselves to the use of the *X-rite color checker standard* [2], containing n=24 squares of different colors which represents the gamut of CIE XYZ quite well.

2 Image Acquisition and Data

The acquisition of data is done using VideometerLab which acquires multi-spectral images in 18 different wavelengths ranging from 430 to 970 nm. The camera setup is seen in Figure 1. The object, in this paper an xrite colorchecker, is placed inside an integrating or Ulbricht sphere which has its interior coated to obtain high diffuse reflectivity for optimal light conditions. In the top of the sphere a camera is located with the sensitivity spectrum seen in Figure 3. Naturally the sensitivity decays towards the near-infrared area, which means that the illuminating diodes, which has the spectral radiant power distributions seen in Figure 2 just needs higher intensity in the near infrared area. The LEDs are strobing successive, and for each LED an image of dimensionality 1280x960 is acquired, which in the end yields a multi-spectral image.
An image of the X-rite ColorChecker standard, seen in Figure 4, has been recorded in this manner. In theory, the first p=11 bands, residing in the visible spectrum are the only ones containing color information, which is why the rest are discarded. Finally we have an multi-spectral image of the xrite colorchecker with 11 distinct wavelengths; 435, 450, 470, 505, 525, 570, 590, 630, 645, 660 and 700 nanometer.

[1] URL: http://www.videometer.com
[2] URL: http://www.xrite.com

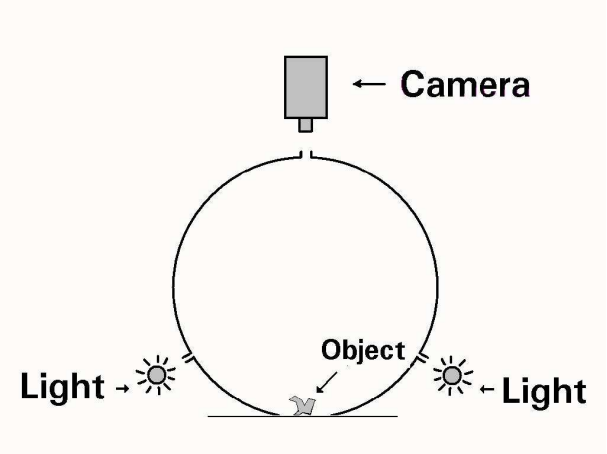


Figure 1: *Camera setup*

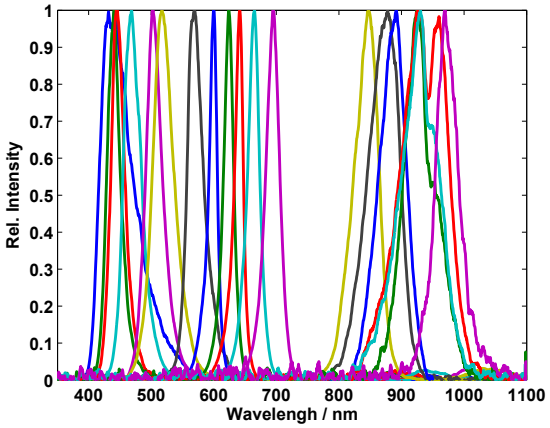


Figure 2: *Spectral power of LED illuminants*

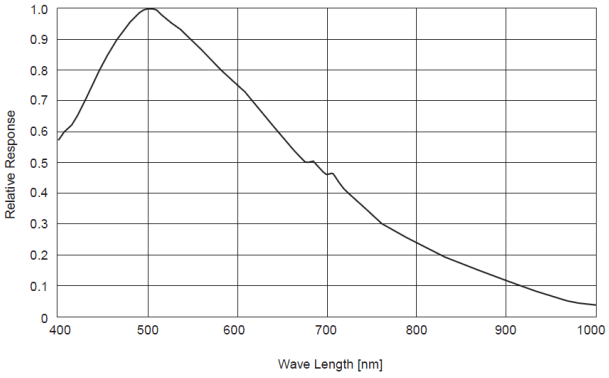


Figure 3: *Spectral sensitivity for camera*

Figure 4: *Xrite colorchecker*

As seen, the x-rite standard contains a matrix of colors, four rows and six columns, which sums to a total of 24 squares. These squares are extracted from the image manually, and a mean value is calculated for each square thus yielding a total of 24 samples. The real values of the colors in RGB and L*a*b* space are known, and can thus be used for calibration. Before calibration, the CIEL*a*b* values are transformed into CIEXYZ space using Eq. 1.

$$
\begin{aligned}
Y &= \begin{cases} Y_n f_y^3 & \text{for } f_y > \delta \\ f_y - \frac{16}{116} 3\delta^2 Y_n & \text{otherwise} \end{cases} \quad , \quad f_y \equiv \frac{L* + 16}{116} \\
X &= \begin{cases} X_n f_x^3 & \text{for } f_x > \delta \\ f_x - \frac{16}{116} 3\delta^2 X_n & \text{otherwise} \end{cases} \quad , \quad f_x \equiv f_y + \frac{a*}{500} \quad , \quad \delta = \frac{6}{29} \\
Z &= \begin{cases} Z_n f_z^3 & \text{for } f_z > \delta \\ f_z - \frac{16}{116} 3\delta^2 Z_n & \text{otherwise} \end{cases} \quad , \quad f_z \equiv f_y - \frac{b*}{200}
\end{aligned} \tag{1}
$$

3 Mapping the Multi-spectral Image into a Standardized Color Model

To map an 11-dimensional multi spectral image into a standardized color space such as CIERGB or CIEXYZ many different approaches could be used. A very obvious regression technique is a straight forward least squares, also called Ordinary Least Square(OLS) which is seen in Eq. 2, written as the

Residual Sum of Squares(RSS).

$$\begin{aligned}\widehat{\beta} &= \arg\min_{\beta}\left\{\sum_{i=1}^{N}(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j)^2\right\} \\ &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\end{aligned} \tag{2}$$

$\widehat{\beta}$ is the projection matrix which projects the n dimensional space into the three dimensional color space $\widehat{y} = \mathbf{X}\beta$, where $\mathbf{X}$ is the multispectral data and $\mathbf{y}$ is the matrix containing the true color values. When a full linear model like OLS is used, all variables are weighed equally in the model. This might be a problem, as the model can get to complex, and some variables might have a bad influence on the model. A well known solution to this problem is the Tikhonov regularized least square, also known as ridge regression, 3, proposed by Carl and Hoerl[4].

$$\begin{aligned}\widehat{\beta}^{\text{ ridge}} &= \arg\min_{\beta}\left\{\sum_{i=1}^{N}(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j)^2+\lambda\sum_{j=1}^{p}\beta_j^2\right\} \\ &= \left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{y}\end{aligned} \tag{3}$$

$\mathbf{I}$ is the $p \times p$ identity matrix. The penalty term of the Ridge Regression shrinks the coefficients towards zero, and thereby weighs down badly influencing variables more than other. [3] provides a good overview and discussion on regression methods and variable selection methods.
A modification to the ridge regression is here proposed as a penalization of slope and curvature in the coefficients which will reduce the variance. In other words we will be able to regularize the coefficients by smoothing, which should make the model more robust. We differentiate the penalization term of Eqn. 3 twice and obtain Eqn.4.

$$\lambda\sum_{j=1}^{p}(\beta_j-2\beta_{j-1}+\beta_{j-2})^2 \tag{4}$$

i.e. the second derivative of β. Some ordering of the elements of β is assumed, which naturally is an ordering in the wavelength of the spectrum. The final regression model for the curvature penalization is seen in eq. 5

$$\begin{aligned}\widehat{\beta}^{\text{ ridge}} &= \arg\min_{\beta}\left\{\sum_{i=1}^{N}(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j)^2+\lambda\sum_{j=1}^{p}(\beta_j-2\beta_{j-1}+\beta_{j-2})^2\right\} \\ &= \left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{X}^T\mathbf{y}\end{aligned} \tag{5}$$

where A is a $p \times p$ pentadiagonal matrix. Similarly the slope penalization is derived.
Due to the sparse amount of observations, Leave One Out Cross Validation(LOOCV) is used to select λ optimally with best possible generalization error. This means the model is trained with all observations except a single observation which is used to generate a test-error residual. This is repeated for all observations in the model, and finally a total testerror can be calculated.

4 Results

When the experiments were done in practice, a gamma correction of the response was needed in order to get suitable coefficients.

$$v_{out} = v_{in}^{\gamma}$$

For each channel, a unique γ and λ was chosen that minimizes the generalization Root Mean Square Error(RMSE), which as mentioned was approximated with LOOCV. A finite difference scheme was used to find the minimum testerror for the two variables. Optimal γ and λ values found for the four tested regression schemes are presented in table 1.

OLS				**Ridge Regression**			
λ	γ	rmse	df	λ	γ	rmse	df
0.00	2.18	4.18	11.00	79.48	2.19	3.87	7.16
0.00	1.67	7.15	11.00	403.79	1.89	5.00	5.60
0.00	1.45	16.46	11.00	809.87	1.75	10.82	4.96
0.00	0.85	1.85	11.00	378.72	0.89	1.45	5.66
0.00	0.80	2.98	11.00	1440.40	0.85	2.14	4.46
0.00	0.74	5.49	11.00	1628.50	0.76	3.75	4.35
Ridge Regression 1				**Ridge Regression 2**			
λ	γ	rmse	df	λ	γ	rmse	df
54.27	2.17	3.78	6.79	16.49	2.17	3.78	7.26
211.32	1.89	5.07	5.70	96.25	1.88	5.10	6.04
1453.30	1.74	9.96	4.33	4217.70	1.74	9.77	3.98
305.27	0.89	1.40	5.44	168.20	0.89	1.40	5.69
3406.10	0.85	2.09	3.82	17242.00	0.86	2.13	3.42
4547.10	0.76	3.70	3.67	38041.00	0.76	3.67	3.19

Table 1: *Key-values for different regression methods. λ and γ are the regression parameters, rmse is the root mean square of the residuals and df is the efficient degree of freedom.*

All methods seems to approximate the standard color values fairly well if looking at the rmse. The modified ridge regressions seems however to perform slightly better. In all cases it was seen that the CIE XYZ values were approximated significantly better than the RGB values. This might be caused by the constrained nature of the RGB space. The effective degree of freedom which of course is dependent on the regularization parameter λ, is also calculated, and for all the regularized regression methods, the coefficients seems to be shrunk slightly more for the approximation to XYZ values than for RGB values. This indicates that more information is actually needed to predict RGB than XYZ. More information of the calibration is to be gained by looking at plots of the estimated coefficients. Clearly it is desired to have coefficient curves corresponding to the tristimulus values of the estimated color model i.e. sRGB or XYZ curves, as this will have the same effect as the binning previously mentioned.
The resulting coefficients are seen in Figure 8. Each curve is plotted by itself in order for it not to ”drown” in the scaling. In the lower right corner of each group the curves are plotted together. In the top row the OLS fit is seen. Although the RMSE for OLS shows a rather good fit, the curves seems very messy. For the ridge regression we already see the tristimulus curves appear, which indicates that the regularization is working as desired. To gain smoother functions a regularization of OLS with the derivative of the Ridge regression was used as described in previous section. Using this, it is seen that the coefficient curves indeed seems slightly more smooth as seen from ridge to first derivative, and again to the slightly smoother second derivative.
To obtain CIE L*a*b* values, the estimation of XYZ space is used together with the nonlinear relation given in [7]. To visualize the result the sRGB coefficients are applied to a multi spectral image of the x-rite colorchecker. We now obtain the result shown in Figure 7. Comparing this to the neighboring

image, Figure 6, which was produced by just directly extracting the bands sampled closest to the primary RGB channels, i.e. the wavelengths $\lambda = \{470, 505, 645\}$ nm. Figure 5 shows colors from a webpicture. Showing the correct colors in an article such as this is impossible because colors are device dependant, so a picture like this might to some extent represent the real colors of the colorchecker. It seems that the red and brown colors are the colors hardest for the weights to estimate. However, compared to the direct extraction, it seems to be quite satisfying.

Figure 5: *"Real" colors of the colorchecker*

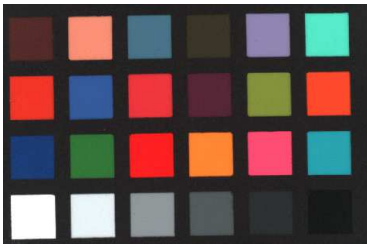


Figure 6: *Extraction of bands* $\lambda = \{470, 505, 645\}$

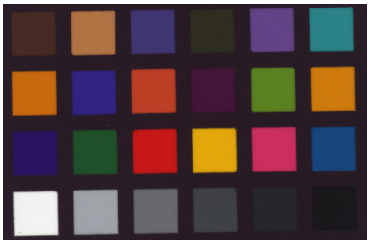


Figure 7: *Applying second derivative ridge weights*

5 Conclusion

We have presented a method to map multivariate images into a three dimensional space using regularized regression techniques. The mapping method was evaluated on a x-rite ColorChecker Chart in two different color spaces, yielding quite good and interpretable results as seen in the reconstruction of the x-rite ColorChecker Chart colors, using obtained weights in the RGB colors space. These were verified by looking at the estimated coefficient plots as well as test errors and parameter values.
In our future work the regression schemes needs to be refined. Furthermore we want to build a pipeline to model colorimetric data based on recorded reflectance spectrum, different sensitivity curves for different observers, and different lightsources.

References

[1] Christine Connolly and Thomas Fliess. A study of efficiency and accuracy in the transformation from rgb to cielab color space. *IEEE Transactions on Image Processing*, 6(7):1046–1048, 1997.

[2] J.Y Hardeberg. *Acquisition and Reproduction of Color Images: Colorimetric and Multispectral Approaches*. Dissertation.com, 2001.

[3] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *Elements of Statistical Learning: data mining, inference and prediction*. Springer-Verlag, 2001.

[4] A. E. Hoerl and R. W. Kennard. Ridge regression. biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[5] Guowei Hong, M. Ronnier Luo, and Peter A. Rhodes. A study of digital camera colorimetric characterization based on polynomial modeling. *Color Research & Application*, 26(1):76–84, 2001.

[6] Daniel Nystrom. Colorimetric and multispectral image acquisition using model-based and empirical device characterization. In Bjarne K. Ersboll and Kim Steenstrup Pedersen, editors, *SCIA*, volume 4522 of *Lecture Notes in Computer Science*, pages 798–807. Springer, 2007.

[7] Gunter Wyszecki and W.S. Stiles. *Color science: Concepts and Methods, Quantitative Data and Formulae, second edition*. Wiley & sons Inc.

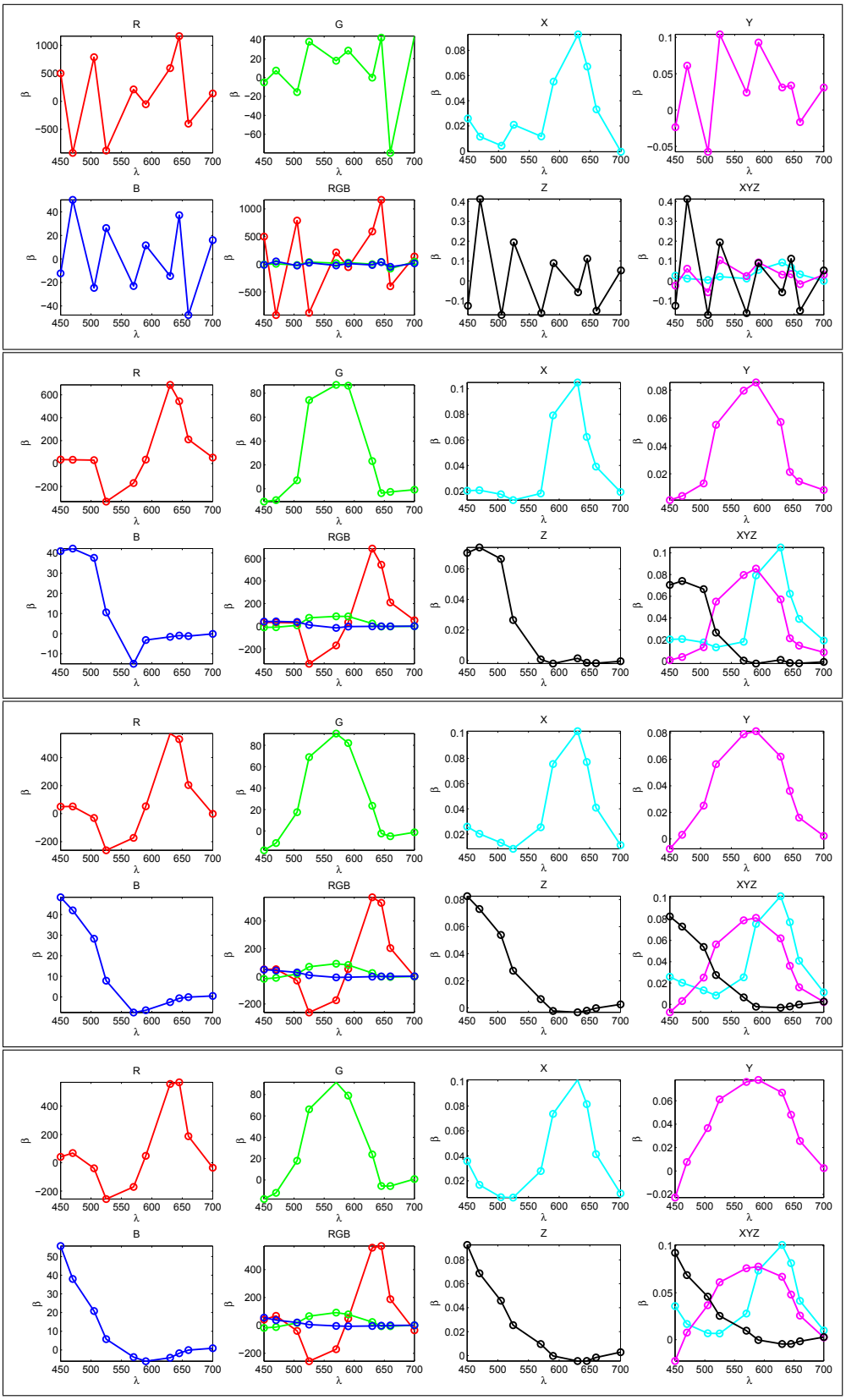


Figure 8: Koefficient plots for four regression schemes. Row 1: Ordinary Least Squares. Row 2: Ridge Regression. Row 3: Ridge Regression 1. derivative. Row 1: Ridge Regression 2. derivative. First column are coefficients for sRGB values, and second for XYZ values.

Full Reference Image Quality Metrics for Optimizing Example-based Total Variation Deblurring

Hyrum S. Anderson
University of Washington
Seattle, Washington, USA
hyrum@ee.washington.edu

Jon Y. Hardeberg
Gjøvik University College
Gjøvik, Norway
jon.hardeberg@hig.no

Maya R. Gupta
University of Washington
Seattle, Washington, USA
gupta@ee.washington.edu

Abstract

Researchers have recently explored increasing the resolution of a digital image by inferring high-resolution edge content from a corpus of sharp training images. We have shown that a similar example-based approach to deblurring can produce good results, but is prone to objectionable image artifacts. Here we consider example-based deblurring as a complement to traditional sharpening and deblurring techniques, and propose a deblurring algorithm with tunable parameters that trade-off between an example-based method and total-variation image deblurring. We investigate automatically optimizing algorithm parameters using a full-reference image quality metric during cross-validation. We report the efficacy of several different image quality metrics for optimizing algorithm parameters. We present preliminary findings for the automated method, where results are compared to a human-guided result using the proposed algorithm.

1 Introduction

Unlike traditional image sharpening and deblurring techniques that aim to improve image quality given only a blurred image, the goal of example-based image deblurring is to infer high-res detail about the blurred image using a set of high-res training images. The basic premise underlying example-based approaches is that edges at different resolution scales locally exhibit correlation, as noted for wavelet coefficients at different scales in [9]. Therefore, local edge detail in a low-res image can be matched to high-res edges in a training corpus and the low-res image can locally inherit the high-res edge with the best correlation. Freeman et al. used this approach to artificially magnify a low-res image by matching edges in the low-res image with down-sampled high-res training images on a block-by-block basis; the matched-blocks' high-res edges were then added to the low-res image [4]. To decrease the number of image artifacts, other researchers have investigated example-based methods that regularize for de-aliasing [5] and extended the idea in [4] to k nearest neighbor (k-NN) matches with outlier rejection [3].

In this work, we consider example-based deblurring, which is a more general problem than the single-frame super-resolution studied in [4]. Applying the simple super-resolution approach in [4] often produced surprisingly good results, but is also prone to objectionable image artifacts [1]. Recently, we demonstrated how example-based deblurring could be optimally fused with traditional Wiener deconvolution to produce superior results. However, that approach required knowledge or a reasonable estimate of the true image spectrum and reconstruction error spectrum.

In this paper, we investigate noisy image deblurring by fusing total variation (TV) image reconstruction with example-based image deblurring. The algorithm trades off between TV reconstruction and example-based reconstruction via three tunable parameters. The value for each parameter must be chosen to maximize the visual quality of the reconstructed image when compared to the true (unknown) image. Although the parameters may be tuned manually for each image, we investigate in this paper how full-reference image quality metrics (IQMs) may be used to automatically tune the parameters. In particular, consistent with our use of training images to estimate high-frequency detail, we cross-validate the

regularization parameters on a hold-out set of example images. One hold-out image is blurred, then reconstructed using the remaining images as examples. We repeat this process for several different choices of image deblurring parameters, and the deblurred image is compared to the true image using a full-reference image quality metric. The parameter set that achieves the best quality as judged by the metric is used to reconstruct the actual blurred image.

We address the following questions in this paper. Using a "perfect" IQM (a human observer), can we reliably choose parameters using cross-validation on training images that will result in a good reconstruction for the test image? If so, then to automate the process, how do several popular IQMs compare with the manually-guided reconstruction? Which one correlates best with manually-guided reconstruction?

2 Example-based Total Variation Deblurring Algorithm

First, we propose a reconstruction algorithm. Given a column-scanned $M \times N$ noisy blurred image $z \in \mathbb{R}^{MN}$ and a known blur matrix $H \in \mathbb{R}^{MN \times MN}$, we solve for a restored image $x \in \mathbb{R}^{MN}$ and an auxiliary image $v \in \mathbb{R}^{MN}$ and weights $\alpha_i \in \mathbb{R}^k$ that solve

$$\begin{aligned} \min_{x,v,\{\alpha_i\}} \quad & \|Hv - z\|^2 + \lambda_1 \|x - v\|^2 + \lambda_2 \|x - \sum_{i=1}^{K} T_i \alpha_i\|^2 + \lambda_3 \|x\|_{TV} \\ \text{s.t.} \quad & \alpha_i \in \text{simplex}(k), i = 1, \ldots, B, \end{aligned} \tag{1}$$

where the vector α_i provides convex weights on $T_i \in \mathbb{R}^{RM \times k}$ for ith $m \times m$ block, $\|\cdot\|_{TV}$ denotes total variation, and the columns of T_i contain the k "most similar" high-res training blocks to the ith low-res block of the input image z (with entries outside the block set to zero).

Equation (1) can be analyzed as follows. The $\|Hv - z\|^2$ term is *fidelity preserving*: it ensures that the reconstructed image matches the input if blurred by H. The $\|x - \sum_i T_i \alpha_i\|^2$ term is *detail enhancing*: high-res edges are inferred from low-res edges in z using training examples. The $\|x\|_{TV}$ term promotes *smoothness* of the reconstruction; TV is also known to be edge-preserving. The criteria *fidelity preserving*, *detail enhancing* and *smoothness* were proposed by Liu et al. in [7] for optimizing visual quality, though implemented differently. The $\|x - v\|^2$ term is for computational efficiency, and links the *smoothness* and *detail enhancing* criteria that operate on x to the *fidelity preserving* criterion that operates on v. Since example-based deblurring can induce objectionable image artifacts, we optimize over local weights $\{\alpha_i\}$ so training blocks in T_i that poorly match (locally) the *fidelity preserving* and *smoothness* criteria are given a small weight; this is analogous to the outlier rejection explored in [3].

To generate $\{T_i\}$, we pre-process training images as prescribed in [4] and as done in our previous work [1]: edges are extracted using a high-pass filter, and the images are (locally) contrast-normalized. Training image blocks are generated by extracting every $m \times m$ block from the corpus of training images. For each $m \times m$ block of the test image in raster-order, we select the $k = 5$ nearest neighbors from the corpus of training blocks that minimize Euclidean distance of the test block to a blurred version of each training block. To promote continuity across block boundaries, we allow the blocks to overlap, and the distance computation is regularized by the error norm between the sharp training block and an average of the neighbors in the overlap region of previously estimated blocks, analogous to [4, 1]. When a nearest neighbor is selected, it inherits the local contrast and low-frequency information of the test block.

We have found an efficient implementation of (1) by extending previous work [2] that efficiently solves the subproblem $\min_x \|x - v\|^2 + \gamma \|x\|_{TV}$, but we omit details here for brevity[1]. In experiments, we evaluate (1) on the luminance channel in YUV space.

[1] MATLAB code can be downloaded from `http://students.washington.edu/hyrum`.

Figure 1: Training images used for reconstructions. Images are 768×512 pixels. The second image has been rotated to fit in the figure.

3 Testing the Cross-Validation Hypothesis

Adjusting the values for parameters λ_1, λ_2 and λ_3 in (1) can lead to a variety of reconstructions. For $\lambda_1 = \lambda_2 = \lambda_3 = 0$, (1) results in a least-squares reconstruction; for $\lambda_2 = 0$, x is the TV reconstruction; for $\lambda_1 = \lambda_3 = 0$, we have blockwise k-NN reconstruction, analogous to [4]. In general, the optimal result may lie somewhere among these extrema, but manually finding the optimal choice of parameters for a particular image can be time-consuming.

Training images are shown in Fig. 1. Freeman et al. suggested that the training images need not be of the same scene as the test image, though a small drop in reconstruction quality may occur [4]. A set of similar training images could be manually selected or automatically assembled via web-based image similarity searches, such as those provided by Microsoft[2] and Google[3]. It has been noted in [6] that collecting a good training set, and properly aligning and scaling them during a pre-processing step can improve performance. In our experiments, we select training images from the Kodak benchmark dataset that contained high structural content, but do not require them to be of the same scene.

In order to develop our automated method for choosing parameters, we hypothesize that for a set of images with similar local edge content, an "optimal" choice of λ_1, λ_2 and λ_3 for one image will be nearly "optimal" for every image in the set. To test this hypothesis, optimality is measured by the authors' judgement of how well a reconstructed image matches a true image. We choose λ_1 from the set $\{0, 0.1, 0.5, 1\}$, and choose λ_2 and λ_3 from $\{0, 10^{-5}, 10^{-3}, 0.1\}$; these values were chosen because they produce a range of perceptual changes in preliminary experiments.

One image from the set was artificially blurred (Gaussian blur kernel with 2.25 pixel bandwidth), corrupted with noise (Gaussian white noise with 34 dB PSNR), then reconstructed using (1) with the other two images as training images. For each blurred image, we generated a reconstruction for each combination of λ_1, λ_2 and λ_3 (64 images). We grossly categorize the images into four groups: very good (score=5), good (score=4), fair (score=3), unsatisfactory (score=2) and objectionable (score=1). The process of reconstructing and scoring was performed on each of the training images, and the scores were averaged across the training set. The parameters corresponding to the top 10 average scores are shown in Table 1. The hypothesis is that these parameters will result in a good reconstruction on the test image.

The training images in Fig. 1 are then used to reconstruct the test image shown in Fig. 2. We grossly categorize the results as before and compare the best quality image to what was predicted via cross-validation. The deblurring estimates corresponding to the top 5 human-guided cross-validation choices are shown in Fig. 3. One of the images received a score of 5, one received a score of 4, and the other three received a score of 3.

[2] `http://www.live.com/`

[3] `http://similar-images.googlelabs.com/`

Table 1: Top 10 cross-validation parameters chosen by a human with those parameters' scores averaged across training images, shown for various IQMs. Larger numbers are better for all metrics except "95% ΔE" and "95% spatial ΔE", for which smaller numbers are better.

λ_1	λ_2	λ_3	human	PSNR (dB)	95% ΔE	95% spatial ΔE	SSIM	VIF
0.1	0.1	10^{-5}	4.7	19.30	13.80	9.24	0.947	0.611
0.5	0.1	10^{-5}	4.7	19.24	13.93	9.52	0.946	0.635
0.5	10^{-5}	10^{-5}	4.3	19.14	14.09	10.03	0.943	0.670
1.0	0.1	10^{-5}	4.3	19.06	14.23	9.96	0.942	0.611
0.1	0	10^{-3}	4.3	19.42	13.57	9.59	0.944	0.712
0.1	10^{-3}	10^{-3}	4.3	19.44	13.55	9.55	0.944	0.709
0.5	10^{-3}	10^{-3}	4.3	19.13	14.13	10.05	0.943	0.661
0.1	0	0.1	4.3	19.31	13.83	9.76	0.943	0.660
0.1	10^{-5}	0.1	4.3	19.31	13.83	9.76	0.943	0.660
1.0	0	10^{-5}	4.0	18.95	14.47	10.44	0.940	0.663

Figure 2: Blown up region of test image and ground truth.

Although the human scoring is error-prone and subjective, this small preliminary experiment does suggest the cross-validation method we have proposed may produce reasonable reconstructions.

4 Choosing Parameters with IQMs

We now wish to replace the human scoring with an IQM. We consider five generic IQMs: peak signal to noise ratio (PSNR), the 95% quantile of the error in CIELAB space when averaged over all the pixels (95% ΔE), the 95% quantile of the error using spatial CIELAB (spatial 95% ΔE) [11], structural similarity index (SSIM) [10], and visual information (VIF) [8].

First, we compute cross-validations scores using the IQMs. The score for each IQM averaged over the training images is shown for the top-10 human-guided cross-validation parameters in Table 1. Generally, the IQM scores are correlated to our gross category scores; however, the top IQM score was not a good predictor of the top human score: the top PSNR choice received a score of 2.3 by the human; 95% ΔE was 1.3; spatial 95% ΔE was 2.3; SSIM was 1.3; and VIF was 1. In experiments, we noticed

Figure 3: Images reconstructed using top five human-guided parameters, and ground truth for comparison.

that the IQMs generally gave a low score to objectionable images, but that they did not discriminate well between, e.g., good (4) and very good (5) images.

Although the IQMs do not perform as well as the human, it may be beneficial if the IQMs give self-consistent results. That is, does the parameter choice that a particular IQM selects during cross-validation result in a high score by the IQM on the test image? Indeed, the magnitude of the correlation coefficients between average cross-validation score and testing score exceeds 0.98 for each of the IQMs.

Cross-validation parameters chosen by the best IQM scores are used to reconstruct the test image. The results are shown in Fig. 4. The human scores for these images are 4, 3, 3, 1, and 1, respectively, for PSNR, 95% ΔE, 95% spatial ΔE, SSIM and VIF. SSIM- and VIF-optimal parameters resulted in images that were too grainy and too smooth, respectively.

5 Discussion of Preliminary Results and Future Work

Preliminary experiments in this paper have shown that with a human in the loop, reconstruction parameters can be chosen via cross-validation on a training set of images. We have taken the first steps to automate this process using an IQM.

The quest for an IQM that correlates well with human vision is elusive. In this paper, we explored only a few generic IQMs for optimizing a example-based total variation deblurring algorithm. Compared to parameter selection via human-in-the-loop cross-validation, the IQMs we tested gave did not give satisfying results: three IQMs (PSNR, 95% ΔE, and 95% spatial ΔE) resulted in good or fair images, but two (SSIM and VIF) resulted in objectionable images. We do believe that using an IQM for automatic parameter selection may be feasible, but will require a more complete study than what we have presented in this paper.

References

[1] H. S. Anderson and M. R. Gupta. Joint deconvolution and imaging. *Proc. SPIE Conf. on Computational Imaging*, 2009.

[2] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging and Vision*, 20:89–97, 2004.

[3] D. Datsenko and M. Elad. Example-based single document image super-resolution: a global MAP approach with outlier rejection. *Multidim. Syst. Sig. Process.*, 18:103–121, 2007.

[4] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Comp. Graph. and Appl.*, 22(2):56–65, Mar/Apr 2002.

[5] C. V. Jiji, S. Chaudhuri, and P. Chatterjee. Single frame image super-resolution: should we process locally or globally? *Multidim. Syst. Sig. Process.*, 18:123–152, 2007.

[6] S. Kondo, H. Amirshahi, T. Toma, and T. Aoki. Example-based suoer-resolution using internet photo collection. In *SIGGRAPH*, Aug.

[7] F. Liu, J. Wang, S. Zhu, N. Gleicher, and Y. Gong. Visual-quality optimizing super resolution. 28(1):127–140, 2009.

[8] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Trans. Image Processing*, 15(2):430–444, Feb 2006.

[9] E. Simoncelli. Modeling the joint statistics of images in the wavelet domain. *Proc. SPIE*, pages 188–195, 1999.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, Apr 2004.

[11] X. Zhang and B. A. Wandell. A spatial extension of CIELAB for digital color image reproduction. In *SID Symposium Technical Digest*, volume 27, pages 731–734, 1996.

Figure 4: Images reconstructed by top choice for human, PSNR, 95% ΔE, 95% spatial ΔE, SSIM and VIF.

Common Assumptions in Color Characterization of Projectors

Arne Magnus Bakke
Gjøvik university College
The Norwegian color research laboratory
Gjøvik, Norway

Jean-Baptiste Thomas
Université de Bourgogne
Le2i, Dijon, France
and
Gjøvik university College
The Norwegian color research laboratory
Gjøvik, Norway

Jérémie Gerhardt
Fraunhofer FIRST
Berlin, Germany

Abstract

This work focuses on the evaluation of assumptions commonly made in projector color characterization. We investigate how these assumptions are valid along the spatial dimension. Features studied include normalized response curve, chromaticity constancy of primaries and channel independence along the display. We provide qualitative and quantitative data analysis from different projectors to support our discussion. Some features seem to vary noticeably spatially, such as the normalized response curve. Some others appear to be quite invariant, such as the channel independence.

1 Introduction

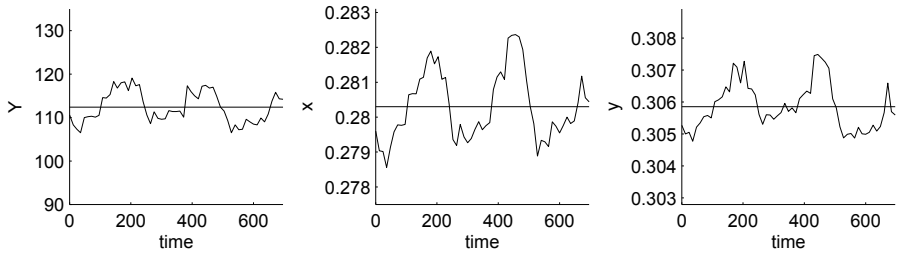
There are a variety of models that are used for colorimetric characterization of projection displays. These models make different assumptions about the devices, usually based on preexisting knowledge about the technologies utilized in the displays, but also sometimes determined by empirically investigating the output of the devices. Problems arise when a model is used without verifying whether these assumptions are true for a specific display device. We have previously shown[10] that there can be a strong spatial color shift for some displays. In this study, we check out some of the major hypothesis involved in color characterization models related with the spatial dimension.

One issue with some models is that they originally were introduced to characterize devices using a different technology to that which is currently in use today, e.g., models for CRT monitors are used for LCD projector displays. In physical color characterization models, assumptions are made considering the display in order to establish the most simple and as fast as possible model. These assumptions are mainly: spatial color uniformity (or only a luminance shift), temporal stability, chromaticity constancy of primaries, independence between channels, gamma or s-curve response curve, etc. Many of these assumptions has been shown to be reasonably correct for a CRT monitor [2, 3, 9].

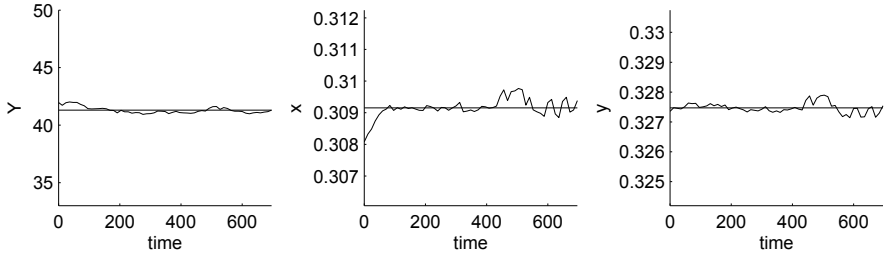
Many studies have investigated LCD monitors [9, 11, 5], and only a few studies have performed verification of these hypothesis on projectors [1, 6, 5, 8, 7]. With the exception of Bastani, these studies investigate mostly projector features as defined by the IEC draft [4].

In this work, we extend previous works by analyzing the characteristics of several projector displays along the spatial dimension. We focus on checking the validity of the most common assumptions.

In the following, we present our experimental setup, then we analyze the normalized response curves of the displays, the chromaticity constancy of primaries, and the independence between channels. We use a method described by Bastani et al.[1] in order to analyze the cross-channel interaction of the displays. By keeping the input of two channels at either full or no intensity, and varying the input of the third channel, the amount of channel interaction can be found.



(a) Y, x and y versus time in minute for display DLP



(b) Y, x and y versus time in minute for display LCD2

Figure 1: Visualizations of the temporal shift for the DLP and one of the LCD tested projectors. One can notice that the DLP is less stable than the LCD. However, both devices appear to be stable enough to be used in normal applications. We can notice that for the LCD projector, there is an optimal time between the warming up time and a overheat time.

2 Experimental Setup

We have used measurements of three projection displays as the basis for our analysis, two of them are LCD projectors and one is using DLP technology. The LCD projectors are both manufactured by Sony and are of the same model (VPL-AW 15), from now on referred to as LCD1 and LCD2. The DLP projector is the Projection Design Action One, which is named DLP throughout the rest of this paper.

All measurements were done using a CS-1000 spectroradiometer in a dark room, after the devices were allowed to warm up for a period of one hour and fifteen minutes in order to achieve a satisfactory temporal stability. Measurements of the RGB cube using 5 subdivisions were taken at 9 locations corresponding to the center and extreme horizontal and vertical positions of a 5 x 5 spatial grid, using a geometrical setup similar to [6]. The decision of reducing the number of positions from 25 to 9 was made due to time constraints, after having verified the spatial caracteristics of the projectors by measuring white patches at all 25 locations.

2.1 Temporal Stability

In order to ensure that our measurements at different locations are significant compared with the normal drift of the equipment, we performed a temporal stability check of the projectors we used. We started by performing an evaluation close to what is proposed in the IEC draft [4]. We measured a white full screen patch (full intensity) at regular intervals of 12 minutes, for about 700 minutes (11h40min). The Y, x and y coordinates are plotted for projectors DLP and LCD2 in Figure 1. We used another range for x and y than the one proposed in the IEC draft since we could not see any information while plotting between 0.25 and 0.35 chromaticity diagram unit.

It appears that the LCD projector is really stable after one hour warming up, and before approximately

Table 1: Temporal stability estimation

	DLP						LCD2				
	R	G	B	W	All		R	G	B	W	All
ΔE^*_{ab} Mean	1.29	1.21	0.78	1.17	1.11		0.60	0.33	0.58	0.22	0.43
ΔE^*_{ab} Max	2.79	2.73	1.64	2.41	2.79		4.74	1.32	1.83	0.64	4.74
ΔE^*_{ab} STD DEV	0.72	0.66	0.37	0.56	X		0.86	0.25	0.46	0.11	X

7 hours of use. The DLP projector however vary in intensity from 106 to 118 $cd.m^{-2}$ in a regular way. The chromaticity values are following the same pattern.

To complete this evaluation, and to have a better idea of the global temporal stability in normal use, we measured the primaries and the graylevel at full intensity at 12 minutes interval for the same time, and compute the differences compared with the average in $L^*a^*b^*$ for each colors after one hour warming up. Results are presented in Table 1.

We notice the same thing as can be seen in the graphs, that the LCD is pretty stable, and the DLP is slightly less stable. However, there is a big maximum shift of the red channel for the LCD that appears at around 8h10 minutes after switching on.

In overall, the stability of these devices is pretty good for normal use, and should be good enough for our experiment.

3 Normalized Response Curves

A pretty common assumption in display characterization is to consider the normalized response curve of each channel to have the same shape. By extension, each channel may have the same shape as the graylevel response curve. In many common methods this assumption can reduce the number of intensity measurements or evaluations that have to be taken or done. This assumption has been shown to be valid for CRT monitors but not for LCD ones [9]. For projectors, if we look at the works of Seime and Hardeberg [8, 7] or of Kwak and MacDonald [6], the LCD projector does not show to fit this assumption, however the DLP in [8, 7] seems to show approximately equivalent normalized response curves for ech channel. Note that in [5], one LCD projector they tested seems to fit the hypothesis. However, no quantitative data is given in these studies to assess this. The purpose of this section is to evaluate this with quantitative data, and to extend the investigation to the spatial dimension.

In Figure 2 we show the response curve of a normalized graylevel intensity ramp at the reference location of the DLP we tested, and a normalized sRGB response curve sampled as the first curve. We propose a simple indicator of similarity that consists in the absolute difference between the integrals (i.e. the surface between both curves). We multiplied the surface found by 1000 to avoid king sized numbers. We compared the sRGB and the response curve of our three projectors and found a δ of 4.31, 4.29 and 5.26 for LCD1, LCD2 and DLP. That enables us to relate the following results to something known.

Based on this indicator, we perform three experiments:

First we compute the average and maximum mismatch δ_{mean} between the intensity response curve of each channel and the graylevel response curve at each position. If there is no mismatch it can be enough to measure only the graylevel response curve at each spatial location. Results are reported in Table 2. We observe that the centers of the displays are among the locations with the largest shift between curves for each display. The mismatch does not appear as negligible for many colorimetric accurate applications, while a *sRGB accuracy* could be reached considering normalized response curves equivalent at each location whatever the channel.

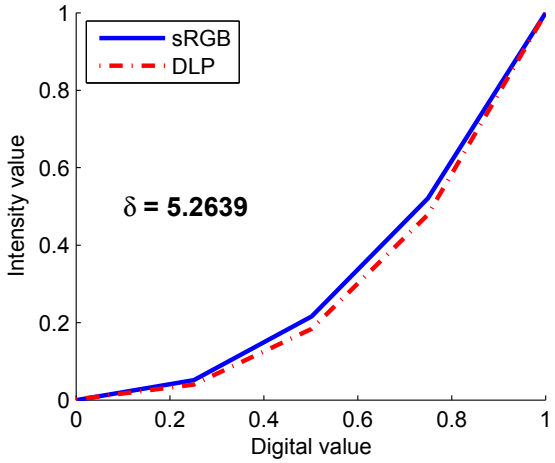


Figure 2: Visualizations of the normalized response curves of the DLP projector compared with the normalized sRGB response curve. The indicator δ is the surface between both curves $\times$ 1000.

Table 2: Mismatch between the intensity response curves of each channel and the graylevel curve, depending on the location on the screen. The maximum and average mismatches are reported.

Average mismatch					Maximum mismatch			
LCD1								
	1	3	5			1	3	5
1	0.96	2.59	2.61		1	2.63	3.88	4.54
3	1.15	3.00	2.43		3	2.57	4.85	3.71
5	2.01	2.64	1.87		5	2.87	3.69	3.42
LCD2								
	1	3	5			1	3	5
1	1.21	2.01	2.24		1	2.05	3.30	3.68
3	1.29	2.15	1.72		3	2.76	3.30	3.01
5	2.05	1.53	1.31		5	3.31	3.27	3.14
DLP								
	1	3	5			1	3	5
1	1.43	1.24	0.98		1	3.44	3.16	1.46
3	1.34	2.05	1.01		3	2.49	4.03	2.64
5	2.62	1.37	0.84		5	4.38	2.82	1.68

Our second experiment consists in computing the mismatch between each primary at different locations, and the same primary at the reference location. If there is no mismatch, we could consider that measuring the response curves at one random location is enough for each primary. Results are reported in Table 3. It seems to be a valid assumption for DLPs. However, for the LCDs it is approximately as different as supposing an sRGB answer (that can be a correct hypothesis depending on the accuracy one want to reach).

Our last experiment testing this assumption is to compare response curves at all locations and for all channels with the reference location graylevel normalized response curve (as it can be measured in some

Table 3: Mismatch between channels for each primary and the channel response curve at the reference location. The graylevel response curve mismatch is shown as well.

Average mismatch					Maximum mismatch			
LCD1								
Red	Green	Blue	Gray		Red	Green	Blue	Gray
3.02	1.49	1.26	1.71		5.15	3.71	3.40	3.87
LCD2								
Red	Green	Blue	Gray		Red	Green	Blue	Gray
1.94	2.36	2.02	2.38		3.85	5.50	3.43	4.99
DLP								
Red	Green	Blue	Gray		Red	Green	Blue	Gray
0.48	0.24	0.97	0.57		1.83	0.80	1.67	0.95

case for applying a classic physical color characterization model). If there is no mismatch, it is enough to measure only one ramp at a given location. We found an average mismatch of 2.13, 2.48 and 1.10, and a maximum of 6.29, 8.30 and 3.85 for LCD1, LCD2 and DLP. In average, the difference is not as big as the difference compared with an sRGB curve, especially for the DLP. However, the maximum error found in LCDs shows that for this technology (or at least for these projectors) one can introduce a critical error through this approximation.

More analysis should be performed, especially to find a not noticeable difference. As a first conclusion, we would not use this assumption for projectors for accurate color rendering. However, it seems that within DLP technology, one can consider the normalized response curve of a given channel as invariant along the spatial dimension. If a *sRGB accuracy* is enough for a given application, then it seems that measuring only one ramp for one projector could be a compromise to do.

4 Chromaticity Constancy

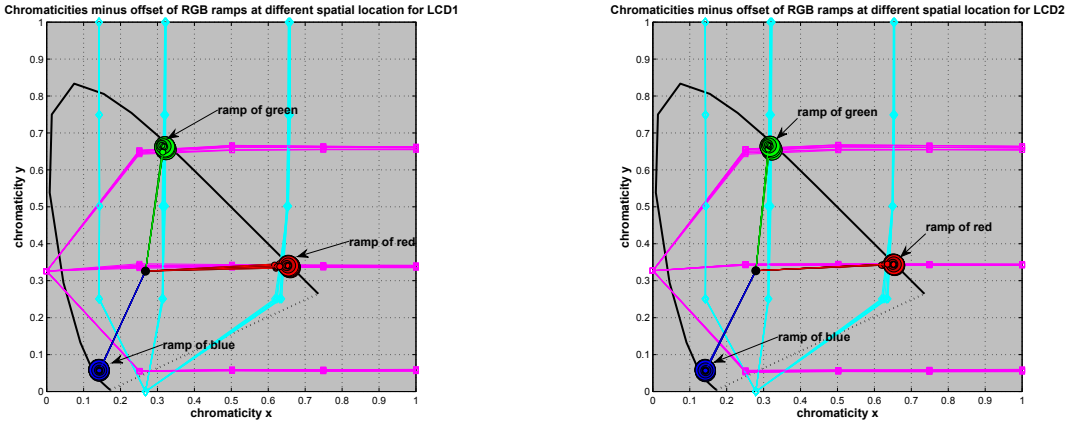
We present in Figure 3 the chromaticity values of the ramps of red, green and blue for each projector and a different locations. In these figures the offset has been removed using the local offset for each location. We can observe a slightly better chromaticity constancy for the DLP projector (Fig. 3 (c)) than for the two LCD projectors, each chromaticity of each ramp at the various level and location being almost identicals.

We still miss a good quantitative indicator, but it does not seem that the primaries spatial shift of chromaticity is a big issue for projector color characterization.

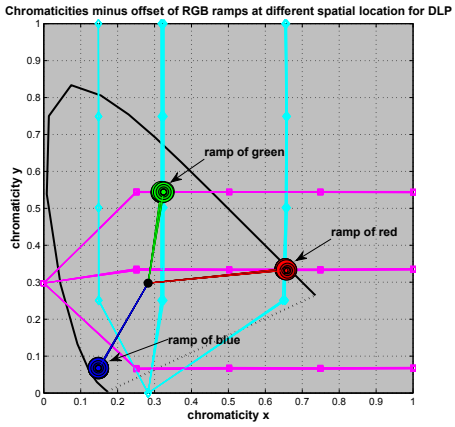
5 Channel Independence

An assumption made by several models is that of channel additivity, e.g., that the output of a gray ramp is equal to the sum of the three R, G, and B ramps. For each projector, we have plotted the measured gray ramp and compared it with the computed sum of the individual ramps in Figure 5.

The lack of additivity we can observe in Figure 5 is due to the existence of channel interaction. Bastani et al. suggested that the amount of interaction for a channel at a given intensity can be calculated using the formula in Equation 1, where $L(r,g,b)$ represents the luminance that is measured for a specified RGB input. a and b are constant values for two of the channels, while v is the varying input of the third

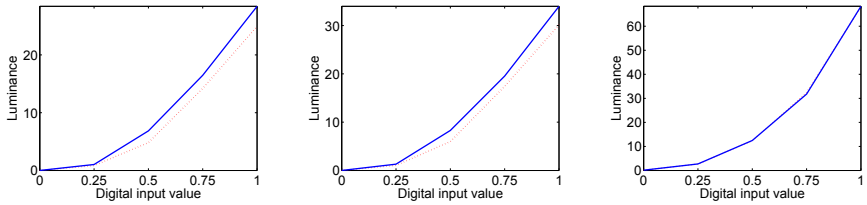


(a) LCD1 (b) LCD2



(c) DLP

Figure 3: Illustration of the chromaticity constancy for the projectors at different locations, in (a) for projector LCD1, in (b) for projector LCD2 and in (c) for the DLP projector. In each figure the x and y chromaticity values are shown twice: once as a regular chromaticity diagram and secondly in the background of the figure in line and square versus the ramp digital steps in horizontal axis for the chromaticity x and in line and diamond versus the ramp digital steps in vertical axis for the chromaticity y. For clarity the chromaticity of the average offset over the various locations has been used in the graphs.



(a) LCD1 (b) LCD2 (c) DLP

Figure 4: The luminance of the gray ramp (solid line) compared with the sum of the individual ramps (dashed line) for the three projectors.

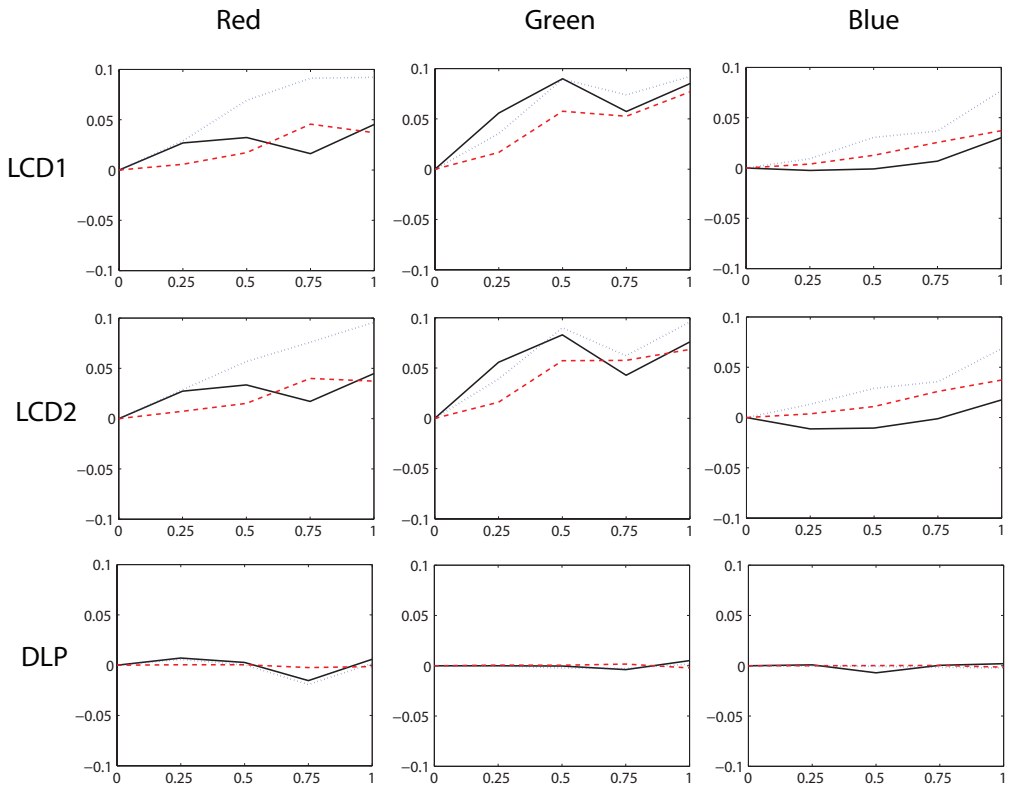


Figure 5: Channel interaction for three displays. The horizontal axis represents the input value of the denoted channel, while the vertical axis represents the calculated interaction value. The solid black line is the interaction found when the two other channels are kept at maximum input value, while the dashed lines are when the a or b, respectively, is set to 0 when computing the interaction metric.

channel. Equation 1 defines the interaction for the red channel. The interaction for the other channels are found in a similar manner. We preferred this method to the more complete, but more complex, method proposed in the IEC draft [4] for visualization purpose.

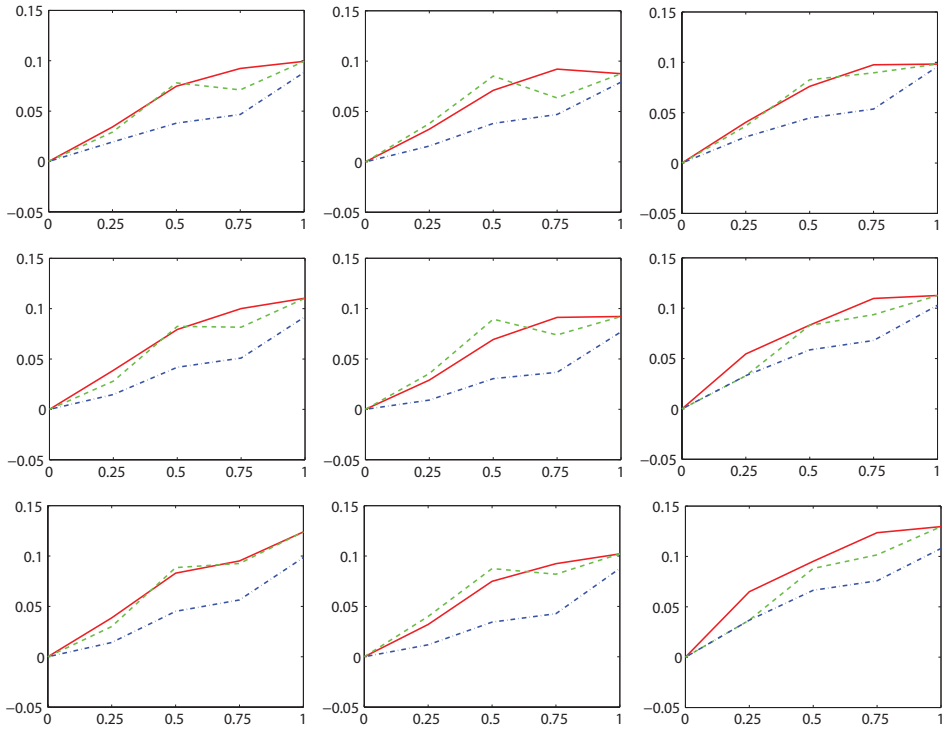
$$CI_{RED}(v,a,b) = \frac{(L(v,a,b) - L(0,a,b)) - (L(v,0,0) - L(0,0,0))}{L(255,255,255) - L(0,0,0)} \quad (1)$$

Figure 5 shows the interaction between the channels for the three projectors. We can clearly see that the LCDs have much more interaction than the DLP. The LCDs feature quite similar interaction characteristics, which is unsurprising given that they are of the same manufacturer and model.

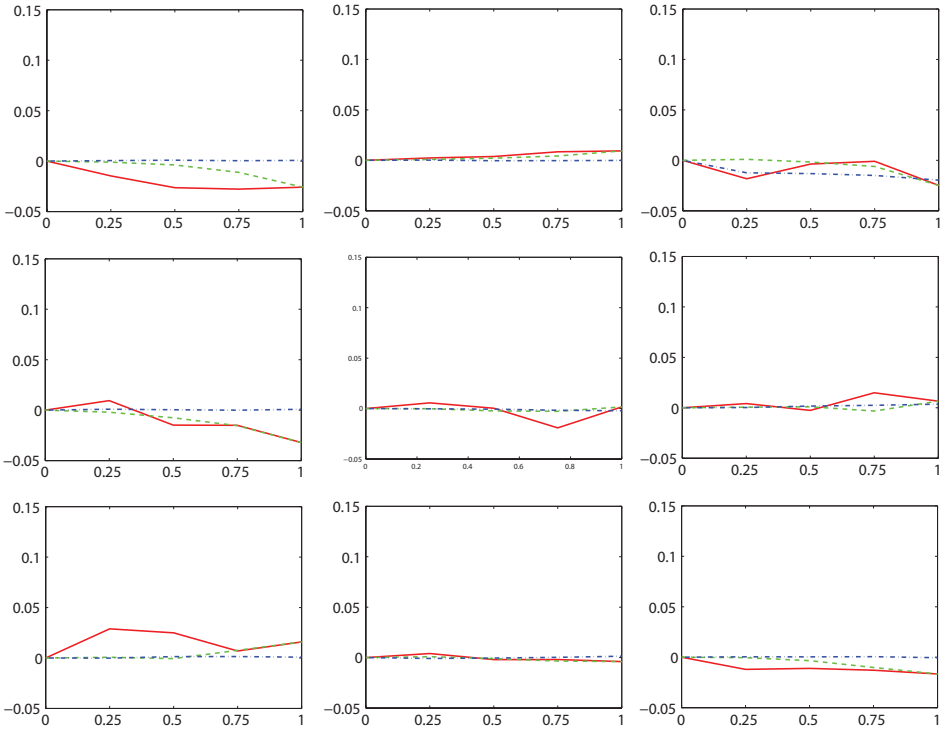
The spatial effect on the interaction is shown in Figure 6(a) and 6(b) for LCD1 and DLP, respectively. We noticed more interaction in the corners of the image in DLP technology. That could be due either to the motion of the color wheel that is less synchronized with the micro-mirrors motion at the corner or to some lens diffraction effect.

6 Conclusion

To summarize, we can say that the normalized response curves vary enough with the spatial location to influence strongly the accuracy of the characterization, except for the tested DLP where the spatial normalized response curve seems to be consistent by channel. We confirm previous studies, saying that



(a) LCD1



(b) DLP

Figure 6: Spatial channel interaction for two of the projectors, where each graph represents the interaction at the corresponding spatial location. The combination of *a* and *b* that give the highest interaction are chosen for each channel.

LCDs projectors have a lot of interaction, and their channel additivity is bad meanwhile, DLP technology shows more independence, and a good additivity. However, LCDs show a more consistent behavior along the spatial dimension than DLP for this feature. The study of the chromaticity constancy shows as well better performances for the DLP.

If we relate these results with the color shift measured in our previous study [10], we can say that,beside a small temporal shift and a lens effect, a strong contribution to the spatial color shift could come from the difference in response curve for LCDs technology. That induces a spatial different mix of primaries for the same input. For the DLP tested the spatial difference considering channel interaction can play a role as well as a more important temporal variation.

To construct a spatial color characterization model, performing measurements at many spatial locations on the displayed area might be required. However, the number of measurements could be reduced depending on the display characteristics. For instance, considering the DLP we tested, it could be enough to evaluate each channel normalized response curve at one location. Or considering the interaction between channels stable along the spatial dimension, it could be enough to take some model's parameters at one location.

Further work includes performing a more in-depth statistical analysis of the results, and testing more projectors to improve the significance of the experiment.

References

[1] Behnam Bastani, Bill Cressman, and Brian Funt. Calibrated color mapping between LCD and CRT displays: A case study. *Color Research & Application*, 30(6):438–447, 2005.

[2] D. H. Brainard. Calibration of a computer-controlled color monitor. *Color Research & Application*, 14:23–34, 1989.

[3] W.B. Cowan and N. Rowell. On the gun independency and phosphor constancy of color video monitor. *Color Research & Application*, 11:S34–S38, 1986.

[4] IEC:61966-6. *Color measurement and management in multimedia systems and equipment, part 3: Equipment used for digital image projection, committee Draft*. August 1998.

[5] Y. Kwak, C. Li, and L. MacDonald. Controling color of liquid-crystal displays. *Journal of the Society for Information Display*, 11(2):341–348, 2003.

[6] Y. Kwak and L. MacDonald. Characterisation of a desktop LCD projector. *Displays*, 21(5):179–194, 2000.

[7] Lars Seime and Jon Y. Hardeberg. Colorimetric characterization of LCD and DLP projection displays. *Journal of the Society for Information Display*, 11(2):349–358, 2003.

[8] Lars Seime and Jon Yngve Hardeberg. Characterisation of lcd and dlp projection displays. In *Color Imaging Conference*, pages 277–282. IS&T - The Society for Imaging Science and Technology, 2002.

[9] G. Sharma. LCDs versus CRTs: Color-calibration and gamut considerations. *Proc. IEEE*, 90(4):605–622, Apr. 2002. special issue on Flat Panel Display Technologies.

[10] Jean-Baptiste Thomas and Arne Magnus Bakke. A colorimetric study of spatial uniformity in projection displays. In *Lecture Notes in Computer Science*, number 5646, pages 160–169. Springer, 2009.

[11] Yasuhiro Yoshida and Yoichi Yamamoto. Color calibration of LCDs. pages 305–311. IS&T - The Society for Imaging Science and Technology, 2002. Scottsdale, Arizona, USA.

Color Management Using Device Models and Look-Up Tables

Satyam Srivastava, Thanh H. Ha, Jan P. Allebach, Edward J. Delp
Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA

Abstract

Color management for digital media has become increasingly important because of the huge variety of systems used to create and view content. Traditionally, this has been achieved using color profiles of various capture and display devices and is used mainly for still images. In this paper, we describe some of the issues related to a profile-based color management system. We then construct an alternative system for managing color by visually matching a given display device to a reference device. This is achieved by constructing analytical models of the devices and using these to develop a transformation between their color spaces. Optimal 3D look-up tables (LUT) are used to realize this transformation in an accurate and computationally efficient mannner. We describe experiments using two LCD displays and assess the accuracy of the LUT-based method for tables comparable in size to typical ICC (International Color Consortium) profiles.

1 Introduction

A digital image is a representation of two-dimensional visual data. A display device recreates a likeness of the visual data using this representation. In most cases the visual output of two display devices will be different when the same digital image is used as the input. This is because the commonly used RGB space is device dependent. We ignore compression artifacts and assume that the digital image is a 2D array of RGB values, the image when displayed is essentially uncompressed. Figure 1 illustrates a typical scenario in which visual data is delivered to a group of users as a digital image. The content designer would like to ensure that each user obtains the "correct" visual representation independent of their display device.

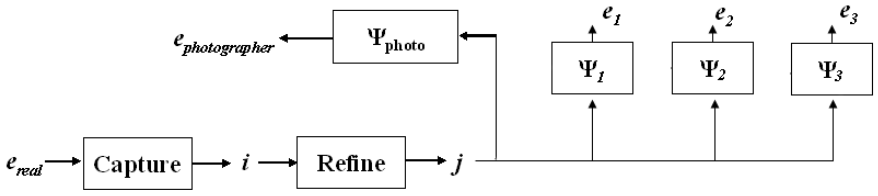
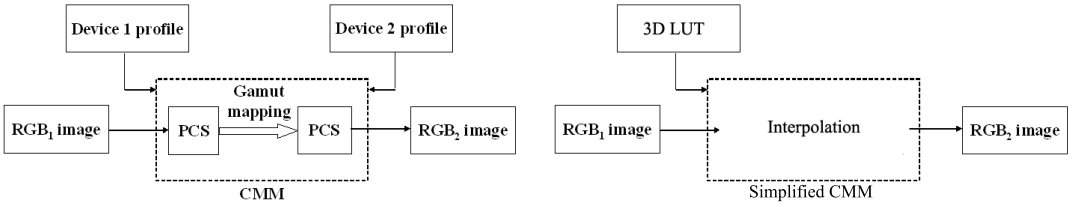


Figure 1: A framework for delivery of visual data.

Let us define a set $\Re$ consisting of all still and moving visual data and a set $\Im$ containing all digital image and video data. Then, using the notation in Figure 1, $e_{real}, e_{photograher}, e_1, e_2, \ldots, e_n \in \Re$ and $i, j \in \Im$. The goal is to achieve $e_1 = e_2 = \ldots = e_n = e_{photographer}$. It should be noted that our reference for visual matching is the visual output as obtained by the content designer. This is important because a

This work was partially supported by a grant from the Indiana 21st Century Research and Technology Fund and by the endowment of the Charles William Harrison Distinguished Professorship at Purdue University. Address all correspondence to E. J. Delp (ace@ecn.purdue.edu).



(a) A profile-based system (b) The proposed system (see Sections 2 and 3)

Figure 2: Block diagrams of the conventional and the proposed methods for color management.

digital image (or video) often undergoes a series of enhancements before being delivered to a user and hence, may not resemble e_{real}. Finally, $\Psi : \Im \rightarrow \Re$ represents the basic functionality of a display device.

Color management is the general collection of techniques which are used to achieve visual similarity across display devices when the same visual data (image or video) is viewed on them [1]. Typically, this is performed using *color profiles* consisting of information about the devices. Profiles provide a way of transforming an image from the color space of a device to an independent color space which is known as the profile connection space (PCS). An example of a profile-based color management is shown in Figure 2(a). The various transformations between the device spaces and the PCS are done in the color management module (CMM) which is also responsible for any gamut mapping that may be required.

While this method is very generic and has been widely used, there are issues which limit its suitability:

- A color management system must be able to understand the profile format. Although the International Color Consortium (ICC) profile format [2] is widely used, several vendors now use their own color profiles [3]. This leads to problems in cross-platform usage.
- Several applications used to open digital image files do not support color profiles [4].
- This is a computationally complex process.
- This approach is not used on digital video [5].

We approach this problem by examining the case of motion pictures. More specifically, we consider features compatible with the Digital Cinema (DCI) specification [6] and have the goal of achieving $e_{user} = e_{artist}$ where e_{artist} is the visual output obtained during cinema *post production*. During post production, the motion picture undergoes several transformations aimed at improving the audio-visual appeal as well as operations such as editing and addition of special effects. To justify the efforts of artists in obtaining the "correct" visual appeal for the picture, it is important that the visual output obtained by the artists be as close as possible to that obtained when the motion picture is screened in a cinema theatre.

From a color management point of view, this is the problem of visually matching two display devices - the cinema theatre screen and the displays used in the post production operations. However, the traditional solution for the cinema industry has been the use of custom displays for the post production. These displays closely match the visual attributes of the cinema theatre and help the artists preview a faithful rendering of the motion picture. However, this is an expensive solution and not very flexible. We also note that a profile based color management may not be useful in this situation. This is primarily because the equipment in the cinema theatre projection system usually does not have the general purpose computing power to utilize profiles for color correcting a full length motion picture.

Our solution is based on the premise that for most professional visual data, there exists a reference display device and we can achieve visual similarity by matching all other devices to this reference. In

the case of motion pictures, the DCI Reference Projector serves as the reference. Using our approach, we aim to visually match the cinema theatre screen and the artists' displays without having to custom manufacture the latter. In the following sections, we describe an analytical method for perceptual matching two displays [7] and a computationally efficient way to realize the method using 3D look-up tables. Note that we can construct a more specific solution to match a user device to a reference device rather than match it to an unknown device using profiles.

2 Analytical Transformation Using Device Models

We need to develop a function $f : \Im \rightarrow \Im$ such that for given display devices Ψ_{artist} and Ψ_{user} and a digital image (or video) $i \in \Im$, $\Psi_{artist}(i) \approx \Psi_{user}(f(i))$ where $\approx$ represents visual similarity. In this paper, we achieve visual similarity in a perceptual sense [7] by accounting for differences in viewing conditions by means of chromatic adaptation ability of the human visual system [8]. We say that two colors are perceptually similar if they transform to the same values in the CIE LAB space [9]. Moreover, Euclidean distance between two points in the LAB space describes the deviation in ΔE units which is a widely accepted measure of color difference. We design our experimental framework using two LCD monitors by taking one as the reference (or target) device and the other as the user device. These are denoted by subscripts 1 and 2, respectively.

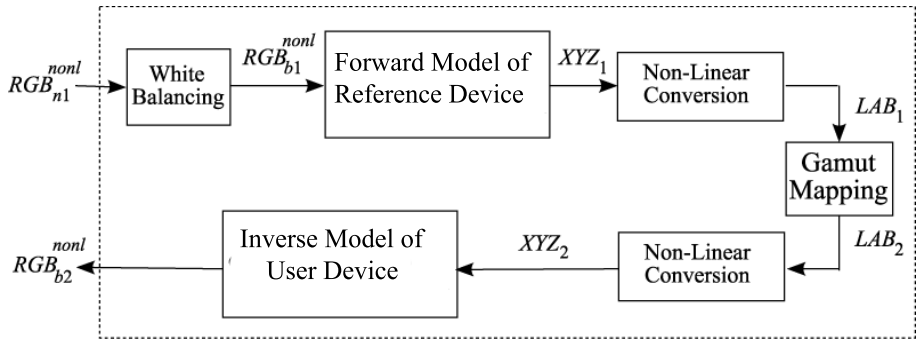


Figure 3: Block diagram of the non-linear transformation for perceptual matching (NLXWP).

Various methods have been proposed to obtain such a function, we use the scheme we proposed in [10] and refer to as NLXWP (Non-Linear Transformation with White Point correction). A block diagram of the scheme is shown in Figure 3 and we briefly describe the basic components below.

The white balancing step is used to ensure that the two devices achieve the chromaticity of a common white point. We choose $D65$ [11] as the desired common white in our experiments. In the first step, we obtain the highest luminance white on the devices with the desired chromaticity (R_{dw}, G_{dw}, B_{dw}). Then, this value is used to "white balance" other RGB values in the absolute linear RGB space as follows:

$$R^{lin}_{balanced} = \frac{R^{lin}_{dw}}{R^{lin}_{nw}} \cdot R^{lin}$$

where *nw* implies native monitor white (the color produced when an input of (255,255,255) is applied to one of the monitors) and the other channels are treated similarly.

The quantitiy commonly referred to as RGB is more specifically denoted as *non-linear* RGB in this paper to emphasize the use of "gamma correction" [12]. Linear RGB values are obtained from the non-linear RGB by the gray balancing process which is achieved by using $1D$ look-up tables for each color channel.

This operation simplifies the inclusion of chromatic adaptation into our model. Let the reference device and the user device be viewed under viewing conditions 1 and 2, respectively. Then, a chromatic adaptation transformation (CAT) based on Von Kries model [13] can be constructed as follows:

$$\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} = H^{-1} \begin{bmatrix} \frac{L_{w2}}{L_{w1}} & 0 & 0 \\ 0 & \frac{M_{w2}}{M_{w1}} & 0 \\ 0 & 0 & \frac{S_{w2}}{S_{w1}} \end{bmatrix} H \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix};$$

where subscripts $i = 1,2$ imply the viewing conditions of the reference device and user device, respectively; LMS_{wi} is the LMS cone responses [14] to the reference white w_i in viewing condition i; H is a non-singular 3×3 matrix which transforms the XYZ values to the LMS cone responses.

However, under common white chromaticity assumption, this transformation reduces to:

$$\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} = \left(\frac{Y_{w2}}{Y_{w1}}\right) \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix};$$

where, Y_{wi}, $i = 1,2$ is the luminance of the reference white under viewing condition i.

The white balanced (non-linear) RGB values are transformed into XYZ values using the forward model of the respective device. This consists of two steps:

- Absolute linear RGB values are obtained using a gray balancing process as described above. For each channel, we use 64 entry LUTs and spline interpolation. Normalized linear RGB is then obtained with respect to the absolute linear RGB values of the white.
- CIE XYZ values are obtained from the linear RGB by a linear transformation represented by a 3×3 matrix M:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \quad M_{3\times3} \quad \end{bmatrix} \begin{bmatrix} R_{nor}^{lin} \\ G_{nor}^{lin} \\ B_{nor}^{lin} \end{bmatrix}$$

The conversions between XYZ and LAB spaces use the standard transformations [9]. Note that these conversions are performed with respect to a white point XYZ_w. Therefore, we use XYZ_{w1} for device 1 and XYZ_{w2} for device 2. Also note that non-linear RGB values corresponding to device 2 can be obtained from the LAB values by using the inverse steps performed for device 1. These details are not described here.

Gamut mapping addresses the gamut mismatch problem between the two monitors. In this paper, we use the straight chroma clipping technique (SCC) [15]. SCC is simple and has been demonstrated to perform better than some other popular clipping approaches [15] as well as some compression techniques [16].

In [15], SCC is implemented in the LAB space by first mapping the L* (lightness) extremes of the two gamuts to each other, and then clipping the out-of-gamut colors to the destination gamut's boundary while still preserving L* and h* (hue). In our approach, the L* extremes of the two gamuts are already equal to each other (the maximum is 100 and the minimum is 0 units). Therefore, the L* mapping step for SCC is not necessary. This is desirable since the sequential L* mapping approach used in [15] has some potential problems as discussed in [17]. The SSC technique is depicted in Figure 4. To clip an out-of-gamut color described by L*, h*, and c* (chroma), we fix L* and h*, and then use the inverse monitor model of the user device to determine its maximum achievable chroma (clipped c*).

The information about the displays, needed to develop the NLXWP function, is obtained by measuring colors displayed on the monitors (in CIE XYZ units) corresponding to RGB inputs by using a spectro-radiometer (Photo Research Model PR-705).

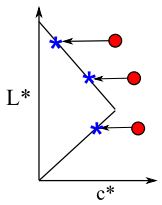


Figure 4: The SSC technique for gamut mapping: red circles are out-of-gamut colors and blue stars are the mapped colors.

3 Efficient Realization Using Look-Up Tables

The NLXWP function is mathematically complex and may not be easy to implement for all displays. Therefore, we use 3D look-up tables (3D LUT) to realize the function in a computationally efficient manner. In [18] we describe the applicability of a 3D LUT for achieving complex transformations and also obtain optimal parameters for such a system. This is briefly described below. The desired LUT is constructed using the NLXWP function and also cross-validated against it.

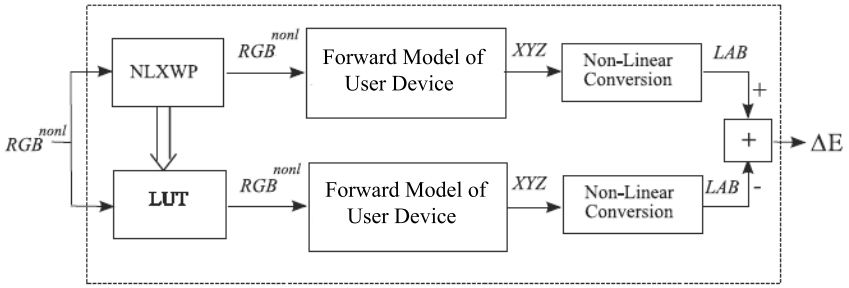


Figure 5: Proposed system for constructing and evaluating a 3D LUT.

Let $f()$ represent the non-linear transformation (NLXWP) using the device models. Since our transformations map non-linear RGB space to non-linear RGB space, let $[r,g,b]^{in}$ and $[r,g,b]^{out}$ indicate the input and output vectors respectively. Next, we let the table look-up be described by a function $\Phi()$. As in the previous sections, a subscript 1 stands for the first monitor (reference/target device) and subscript 2 stands for the user device or the second monitor. Then, for the same input $[r,g,b]^{in}$ to the two functions:

$$[r,g,b]^{out}_{LUT} = \Phi([r,g,b]^{in})$$

$$[r,g,b]^{out}_{NL} = f([r,g,b]^{in})$$

Using the forward monitor models, we can obtain the corresponding values in the LAB space and compute the difference between the colors:

$$\Delta E = \|[LAB]^{out}_{NL} - [LAB]^{out}_{LUT}\|$$

This is illustrated in Figure 5.

The transformation f requires a three dimensional (3D) LUT with a vector $[r,g,b]$ as input. Each entry in the LUT is also a vector $[r,g,b]$ corresponding to the corrected input to the device. For 8-bits per

channel, the input space is of size $256 \times 256 \times 256$ but practical LUTs would be much smaller. Consider a LUT of size $n_R \times n_G \times n_B$ and sample points selected according to the set

$$\Omega = \{r_0, r_1, \ldots, r_{n_R-1}\} \times \{g_0, g_1, \ldots, g_{n_G-1}\} \times \{b_0, b_1, \ldots, b_{n_B-1}\}$$

where $\times$ denotes a Cartesian product. Then, the function $\Phi()$ can be specified as:

$$\Phi([r,g,b]) = \begin{cases} f([r,g,b]), & \forall [r,g,b] \in \Omega \\ \pounds([r,g,b]), & \text{otherwise} \end{cases}$$

Here $\pounds()$ represents an interpolation operation which is used to transform input values not aligned with any of the table entries.

The accuracy of a LUT based transformation is measured in terms of the deviation of the colors in ΔE units. Using the results presented in [18], it can be shown that an optimal LUT function (Φ) would use all nine neighbors of the desired point in $3D$ space and weigh their values according to the square of the inverse Euclidean distances. Further, for a specified size, the optimal selection of sample points represented by the set Ω also affects the transformation accuracy. This is formulated as a constrained optimization problem as follows:

Let S_T be a training set of N RGB values taken randomly from the RGB space. The difference between the colors produced by the NLXWP model (f) and the LUT (Φ) is given by:

$$\Delta E_i = \| [L,A,B]_{NL}^{out,i} - [L,A,B]_{LUT}^{out,i} \|$$

An average of all the differences is referred to as the transformation error:

$$\overline{\Delta E} = \frac{1}{N} \cdot \sum_{i=0}^{N-1} \Delta E_i$$

Then the minimization problem can be represented as:

$$\Omega_{opt} = \arg\min_{\Omega} \overline{\Delta E}$$

subject to the constraints that

$$x_i \in 0, 1, \ldots, 255 \qquad \text{and}$$

$$x_i \leq x_j \qquad \forall i < j$$

where $x = \{r,g,b\}$, $i, j \in \{0, 1, \ldots, n_X - 1\}$ and $X = \{R, G, B\}$.

Let the optimal LUT be represented by $\Phi : \Im \rightarrow \Im$ such that $\Psi_{reference}(i) \approx \Psi_{user}(\Phi(i))$ for any $i \in \Im$. We would like to compare the visual similarity achieved by a LUT based approach with that achieved by a profile-based approach.

Typical ICC profiles can range from 500 bytes to more than 20 kilobytes [19] and a profile-based color management system would use two such profiles as shonw in Figure 2(a). Therefore, for a fair comparison we assume that our LUT can be as large as two ICC profiles. Consider a platform using color profiles of size 15 kilobytes, each. Hence, we can use roughly 30 kilobytes to construct the LUT. Since each entry of the table stores an (R,G,B) triple using 3 bytes, there can be approximately $10,000$ entries in the LUT. Assuming an equal number of samples are collected from each color channel this then corresponds to a maximum size of $21 \times 21 \times 21$.

4 Experimental Results

In this paper we present results of evaluating the performance of smaller LUTs. Given a size specification, optimal parameters for these LUTs are obtained using the methods described in [18]. For the purpose of evaluation, we assume that the analytical models of the display devices (used in the NLXWP function) are accurate and hence, can be used to predict the visual output. This precludes the need for actual measurement of color on the monitors.

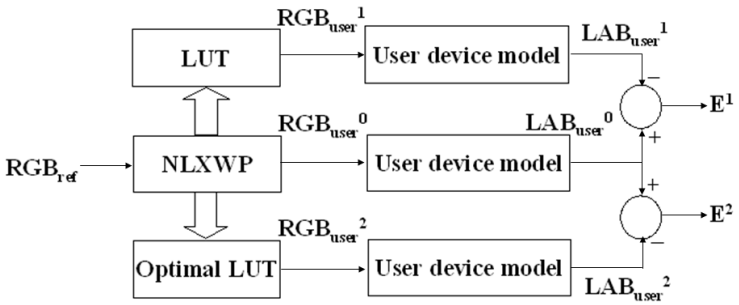


Figure 6: System used testing the LUT-based approach.

A block diagram of the test system is shown in Figure 6. "LUT" represents a non-optimized look-up table which uses uniform sampling of the RGB space and trilinear interpolation while the "Optimal LUT" uses optimal parameters. The testing set consists of 1000 points in the RGB space which are randomly selected. Some statistics of the transformation errors are presented in Table 1.

Table 1: Testing error statistics (ΔE units) for LUTs

Type\Statistic	Mean	Median	Max	1-Std. Dev.
$6 \times 6 \times 6$	7.1495	6.7516	19.0061	3.4082
$6 \times 6 \times 6$ Optimal	4.0886	3.6881	14.3982	2.3822
$9 \times 9 \times 9$	4.6013	4.2622	13.0551	2.2819
$9 \times 9 \times 9$ Optimal	2.5357	2.2743	9.5911	1.5266
$12 \times 12 \times 12$	3.3975	3.1558	10.7637	1.7393
$12 \times 12 \times 12$ Optimal	1.8550	1.6820	6.2326	1.0904

5 Conclusions and Future Work

We observe that an optimal $6 \times 6 \times 6$ LUT achieves an average error of approximately $4\Delta E$ units by using slightly more than 600 bytes whereas an optimal $12 \times 12 \times 12$ LUT limits the average error to well below $2\Delta E$ while using less than 20% of our 30 kilobyte upper bound. Based on our assumption that the NLXWP function provides a highly accurate perceptual matching, we can conclude that a LUT-based implementation retains the accuracy while reducing the associated complexity.

Two other aspects of a LUT-based color management system deserve mention:

- This approach can be extended to a (finite) collection of LUTs corresponding to different reference devices similar to the DCI speification, the sRGB specification or the HDTV specification. Depending on the display device, optimal LUTs of appropriate sizes can be generated and stored. At the time of displaying content, a LUT corresponding to the desired reference can be selected.
- Once the device selects a LUT from the collection, it can generate a much larger LUT (say $64 \times 64 \times 64$). This may be useful because for large LUTs, sample points and interpolation techniques do not greatly affect the accuracy of transformation. Therefore, the device can perform simpler (and faster) linear interpolations on a regular grid of data points.

References

[1] B. Fraser, C. Murphy, and F. Bunting, *Real World Color Management*. Berkeley, CA: Peachpit Press, 2004.

[2] International Color Consortium, "File formats for color profiles," ICC.1:2001-04, Reston, VA, 2001.

[3] Microsoft Corporation, "Windows color system: The next generation color management system," Microsoft White Paper, 2005.

[4] C. Smith and J. Kabili, *How to Wow: Photoshop CS2 for the Web*. Berkeley, CA: Peachpit Press, 2005.

[5] M. D. Fairchild, "A color scientist looks at video," *International Workshop on Video Processing and Quality Metrics*, 2007.

[6] Digital Cinema Initiative, LLC, "Digital cinema system specification," Adopted and released document, version 1.2, Hollywood, CA, 2008.

[7] R. W. G. Hunt, *The Reproduction of Color*. USA: Wiley, 2004.

[8] M. D. Fairchild, *Color Appearance Models*. USA: Wiley, 2005.

[9] International Organization for Standardization, "CIE colorimetry part 4: 1976 l*a*b* colour space," ISO 11664-4:2008(E)/CIE S 014-4/E:2007, Geneva, Switzerland, 2007.

[10] T. H. Ha *et al.*, "Model based methods for developing color transformation between two display devices," Submitted to IEEE International Conference on Image Processing, 2009.

[11] N. Ohta and A. R. Robertson, *Colorimetry*. USA: Wiley, 2005.

[12] C. Poynton, *Digital Video and HDTV: Algorithms and Interfaces*. San Fransisco, CA: Morgan Kaufmann, 2003.

[13] J. V. Kries, *Chromatic Adaptation*. Friboug, Switzerland: Festschrift der Albrecht-Ludwig-Universitat, 1902.

[14] R. W. G. Hunt and L. M. Winter, "Colour adaptation in picture-viewing situations," *Journal of Photographic Science*, vol. 23, 1975.

[15] E. D. Montag and M. D. Fairchild, "Evaluation of chroma clipping techniques for three destination gamuts," *Proceedings of IS&T/SID 6th Color Imaging Conference: Color Science, Systems and Applications*, 1998.

[16] R. S. Gentile, E. Walowit, and J. P. Allebach, "A comparison of techniques for color gamut mismatch compensation," *Journal of Imaging Technology*, vol. 16, 1990.

[17] J. Morovik, *To Develop a Universal Gamut Mapping Algorithm*. Derby, UK: Doctoral dissertation, University of Derby, 1998.

[18] S. Srivastava *et al.*, "Generating optimal look-up tables to achieve complex color space transformations," Submitted to IEEE International Conference on Image Processing, 2009.

[19] T. Newman, "Improved color for the world wide web: A case study in color management for distributed digital media," `http://www.color.org/wpaper2.xalter`.

Analysis of Omnidirectional Spectral Images in Natural Scenes

Shoji Tominaga
Chiba University
Chiba, Japan
shoji@faculty.chiba-u.jp

Atsushi Matsuura
Chiba University
Chiba, Japan

Takahiko Horiuchi
Chiba University
Chiba, Japan
horiuchi@faculty.chiba-u.jp

Abstract

The present paper describes a method for analyzing omnidirectional spectral images in a natural scene. A multiband imaging system with a fisheye lens is used for acquiring omnidirectional images with six spectral bands. We first recover the spectral-power distributions of omnidirectional illuminants from the captured multiband images. Second, we estimate the spectral composition of the omnidirectional illuminants in a natural scene by using the Principal Component Analysis (PCA). Moreover, we analyze the omnidirectional illuminants at a fixed location over different seasons. We show that the overall omnidirectional illuminants can be described in a linear combination of only three principal components. Finally, we examine the reliability of the analysis method in an experiment using the omnidirectional spectral images captured in different seasons in an outdoor scene.

1 Introduction

The analysis of spectral images captured in a natural scene is definitely one of the most interesting topics in the recent field of color imaging science and technology. Some representative works on the spectral analysis are listed in Refs. [1, 2]. These works are mainly classified into two types of analysis of (1) surface-spectral reflectances as physical properties of materials and (2) daylight spectra as effects of natural light. We note that the previous works had a goal to analyze radiance or reflectance spectra at specific areas in a scene of interest. However, it should be noted that the light sources forming color images in a natural scene are not only daylight and sunlight, but also all lights reflected from the surrounding object surfaces.

The present paper describes a method for analyzing omnidirectional spectral images in a natural scene, containing direct illuminants and indirect illuminants. Our goal is to recover spatial distribution of light sources by omnidirectional observations at a particular point in space, and estimate the spectral composition of the omnidirectional scene illuminations over different seasons of a year. In a previous study [3], we proposed a multiband omnidirectional imaging system, which used a high-resolution trichromatic digital camera, a fisheye lens, and two color filters.

In this paper, we first recover the spectral-power distributions of omnidirectional illuminants from the captured multiband images. Second, we estimate the spectral composition of the omnidirectional illuminants in a natural scene by using the PCA. Moreover, we analyze the omnidirectional illuminants at a fixed location over different seasons. Then, it is shown that all omnidirectional illuminants are expressed in a linear combination of a small number of principal components. We show the experimental results for the omnidirectional images captured in an outdoor scene on campus.

2 Multiband Omnidirectional Imaging System

Our system for multi-band omnidirectional imaging is composed of three parts of a digital camera, a fisheye lens, and color filters. The camera is a Canon EOS camera with the image size of 4082×2718 pixels and the bit depth of 12 bits. The fisheye lens is a SIGMA Circular Fisheye lens based on the orthographic projection. For multi-spectral image acquisition, we use two types of additional color filters. Each filter is placed between the lens and the camera body. Combining these color filters to the

original camera spectral sensitivities leads to different sets of trichromatic spectral sensitivity functions. Therefore, two sets of the modified trichromatic spectral sensitivities result in an imaging system with six spectral bands. Figure 1 shows the overall spectral-sensitivity functions of the present multi-spectral imaging system.

Since the present fisheye system takes pictures of a scene in a hemisphere, we need at least two sets of images in opposite viewing directions for completing an omnidirectional image. To eliminate a certain distortion at the edge of the image plane, we can combine three sets of images observed at rotation angle intervals of 120 degrees.

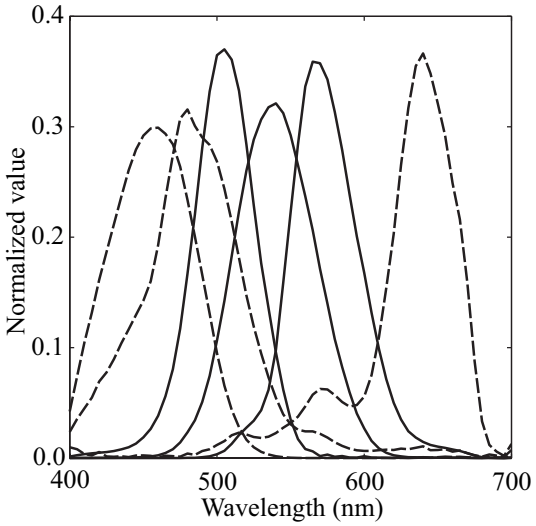


Figure 1: Spectral sensitivity functions of the imaging system.

3 Spectral Image Analysis

3.1 Estimation of Illuminant Spectra

Omnidirectional illuminants to be estimated are not only direct illuminations from light sources, but also indirect illuminations from light reflected from object surfaces in a scene. The observed images can include various noise on sensors and optical process. The image sensor outputs are modeled as a linear system

$$\rho_i = \int_{400}^{700} E(\lambda) R_i(\lambda) d\lambda + n_i, \quad (i = 1, 2, \cdots, 6), \tag{1}$$

where $E(\lambda)$ is the illuminant spectrum, $R_i(\lambda)$ is the spectral sensitivity function of the i-th sensor, and n_i is the noise component with zero mean. We sample each spectral function at n points with an equal interval $\Delta\lambda$ in $[400, 700\text{nm}]$. Let $\mathbf{e}$ be an n-dimensional column vector representing the illuminant spectrum $E(\lambda)$ and $\mathbf{R}$ be a $6 \times n$ matrix with the element $r_{ij} = R_i(\lambda_j)\Delta\lambda$. Moreover, define a six-dimensional column vector $\boldsymbol{\rho}$ representing a set of the sensor outputs ρ_i. Then the above imaging relationships are summarized in a linear matrix equation,

$$\boldsymbol{\rho} = \mathbf{R}\mathbf{e} + \mathbf{n}. \tag{2}$$

When the signal $\mathbf{e}$ and the noise $\mathbf{n}$ are uncorrelated, a linear estimate of $\mathbf{e}$ is given as

$$\hat{\mathbf{e}} = \mathbf{C}_{ss}\mathbf{R}^t(\mathbf{R}\mathbf{C}_{ss}\mathbf{R}^t + \boldsymbol{\Sigma})^{-1}\boldsymbol{\rho}, \tag{3}$$

where $\mathbf{C}_{ss}$ is the correlation matrix of illuminant signals $\mathbf{C}_{ss} = \mathbf{E}[\mathbf{e}\mathbf{e}^t]$ and $\mathbf{\Sigma}$ is the covariance matrix of noises $\mathbf{\Sigma} = \mathbf{E}[\mathbf{n}\mathbf{n}^t]$. We can assume that the noises in each spectral channel are statistically independent. In this case, the covariance matrix is reduced to be diagonal as $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_6^2)$.

3.2 PCA of Illuminants and Data Compression

The PCA analysis is applied to the whole set of illuminant spectra observed for an omnidirectional scene. This analysis is done in the original observation coordinate system of orthographic projection. The omnidirectional image is represented with the azimuth angle θ and the rotational angle ϕ. The spectral distribution of illuminant $\mathbf{e}$ is a function of θ and ϕ. A singular value decomposition (SVD) of the set of $\hat{\mathbf{e}}(\theta, \phi)$ provides that each illuminant spectrum can uniquely be expressed in a linear combination of the n orthogonal vectors as

$$\hat{\mathbf{e}}(\theta, \phi) = c_1(\theta, \phi)\mathbf{u}_1 + c_2(\theta, \phi)\mathbf{u}_2 + \cdots + c_n(\theta, \phi)\mathbf{u}_n, \tag{4}$$

where $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ are the left singular values. Because the observation space is six-dimension, n is limited to six as a practical matter. Consider an approximate representation of the illuminant spectra in terms of some component vectors chosen from $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$. The performance index of the chosen principle components is given by the percent variance $P(K) = \sum_{i=1}^{K} \mu_i^2 / \sum_{i=1}^{n} \mu_i^2$, where $\{\mu_1, \mu_2, \cdots, \mu_n(\mu_i > \mu_{i+1})\}$ are the singular values of the matrix of all $\hat{\mathbf{e}}(\theta, \phi)$. For a typical example, the percent variance $P_1 = 0.95$ for the first component only and $P_3 = 0.99$ for the first three components. When we choose the first K components, all illuminants are represented as

$$\hat{\mathbf{e}}_K(\theta, \phi) = c_1(\theta, \phi)\mathbf{u}_1 + c_2(\theta, \phi)\mathbf{u}_2 + \cdots + c_K(\theta, \phi)\mathbf{u}_K. \tag{5}$$

Because the basis functions of principal components are common to all the illuminants, the set of illuminant spectral data represented by the high-dimensional vectors are reduced to the set of a small number of coefficients $\{c_i\}$. Thus the spectral image data are compressed.

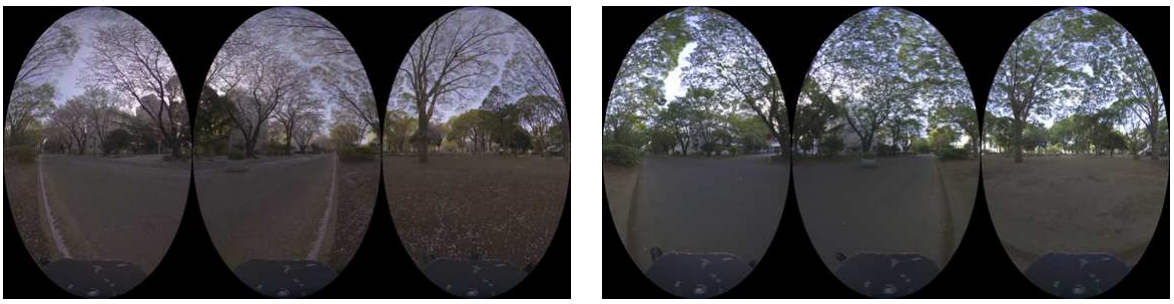
To achieve a high compression rate, we build the linear model for individual omnidirectional scenes. It should be noted that an appropriate model dimension can depends on spectral contents of the objective scene. For the performance index, we compute the sum of the squared error $\| \hat{\mathbf{e}}_K - \mathbf{e}_m \|^2$ or the root-mean squared error $RMSE(\hat{\mathbf{e}}_K, \mathbf{e}_m)$ in illuminant approximation. Here, $\mathbf{e}_m$ means the direct measurement of illuminant $\mathbf{e}$.

4 Experimental Results

4.1 Measurement

There are four seasons in Japan. Each of the four seasons has its characteristics in natural scenes. We suppose that the spectral compositions of omnidirectional light sources and the chromaticity distributions are quite different from each other.

We captured the multi-band omnidirectional images with six-spectral channels for an outdoor scene on campus in Chiba University. The image acquisition was repeated five times at the same location at a fixed time in the mornings in four seasons of April (spring), August (summer), October (autumn), December (winter), and February (winter). All the days had clear weather. Figure 2 shows the captured images in April and August, which were taken with the 120-degree visual field at three different viewing directions. The outdoor scene contains direct illuminations from the sky and indirect illuminations from the surrounding object surfaces. The image intensities vary greatly on pixels.



(a) April 4, 2008 (b) August 6, 2008

Figure 2: Omnidirectional images in two seasons.

Fifteen images of the same scene were taken with the camera at shutter speeds of 2, 1, 1/2, $\cdots$, 1/8000, respectively, and then these images were combined into a high dynamic range images with 22 bits. Second, three high dynamic range images were combined into three directions which are separated 120 degrees as shown in Figure 2.

We can see season features in the natural scene. The cherry blossom is characteristic of the first image in April. The second in August is in the season of green leaves. The images in winter are characterized by dry vegetation.

4.2 Illuminant Estimation

An omnidirectional image was created as a latitude/longitude image in a polar coordinate system in order to represent the spatial distribution of ambient lighting. Figure 3 shows the omnidirectional image of August in the polar coordinate system.

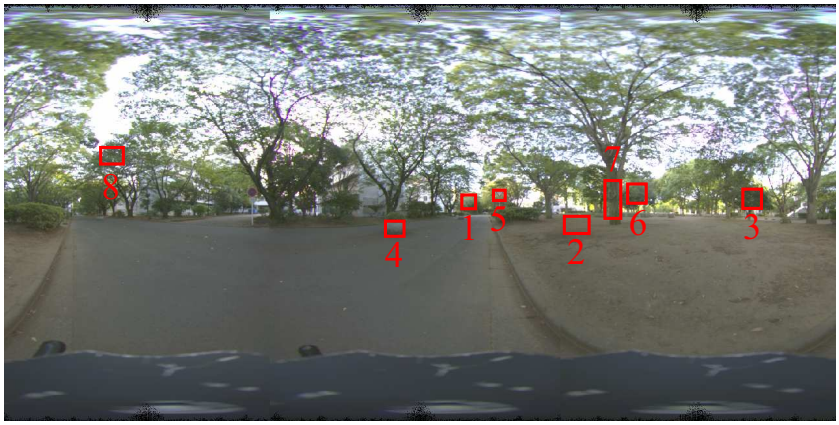


Figure 3: Omnidirectional image of August in the polar coordinates.

Each spectral-power distribution was sampled at 61 equally spaced wavelength points as $n = 61$. The illuminant vector was estimated from the camera data at each pixel of the omnidirectional image by using the algorithm of Eq. (3). To determine the correlation matrix $\mathbf{C}_{ss}$, we used two kinds of database for surface-spectral reflectances and illuminant spectra. One is a set of 507 surface-spectral reflectances for natural objects and artificial objects. Another is an illuminant database consisting of nine light sources, which are the CIE standard spectral-power distributions of daylight of different correlated color temperatures from 5000K to 10000K [4] and the measured spectral-power distribution of daylight by using a

spectro-radiometer. We created a large database of about 5000 indirect illuminants by multiplying the surface-spectral reflectances and the illuminant spectra. Moreover, the set of direct illuminants were added to the indirect illuminant database to complete the overall illuminant database.

We examined the accuracy of the estimated illuminant spectra for the rectangular areas in Figure 3. Figure 4 shows the estimation results of illuminant spectra from Area 2 and Area 5 in the scene. We measured directly the illuminant spectral-power distribution of each area by using the spectro-radiometer. The comparisons between the estimates and the measurements suggest that the reliability of the present estimation method. The second column in Table 1 lists the root-mean squared errors (RMSE).

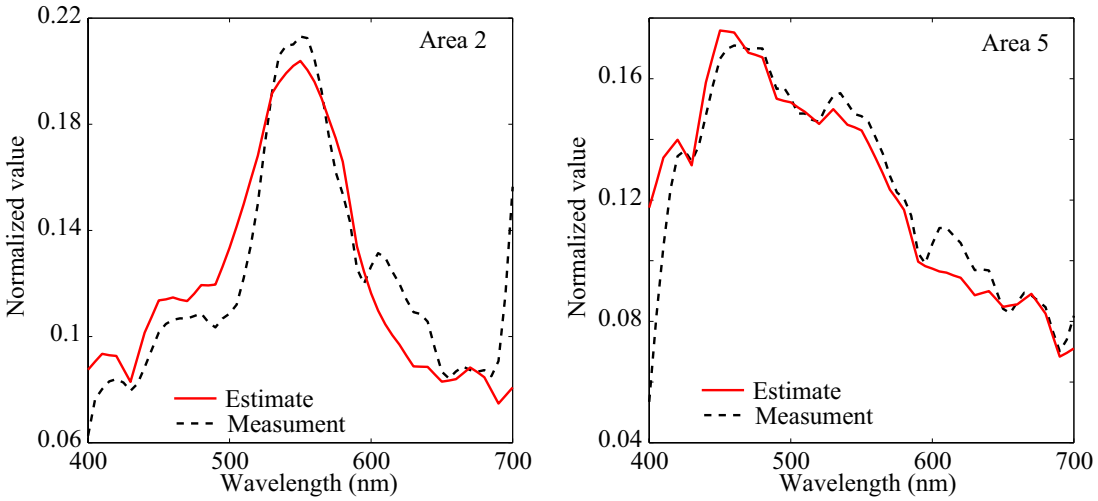


Figure 4: Estimation results of illuminant spectra from Area 2 and Area 5.

Table 1: Estimation errors in RMSE for some areas in Figure 3.

Area	$RMSE(\hat{\mathbf{e}}, \mathbf{e}_m)$	$RMSE(\hat{\mathbf{e}}_3, \mathbf{e}_m)$
1	0.020	0.022
2	0.017	0.018
3	0.022	0.022
4	0.009	0.009
5	0.012	0.013
6	0.017	0.019
7	0.006	0.006
8	0.026	0.025

4.3 Principal Component Analysis

The principal components of illuminant spectrum for the respective omnidirectional scenes were extracted from SVD of the data set of the estimated illuminant vectors . Four figures in Figure 5 show the spectral curves of the first four principal components. Each figure contains five spectral curves belonging to the same rank of principal component in different seasons. We calculated the performance indices $P(K)$ for use of the first K components for the respective data sets. Figure 6 depicts the performance

indices as a parameter of the number K of principal components. We note that the performance indices are $P(3) > 0.995$ for all seasons. Therefore the omnidirectional illuminants can be described using only three principal components.

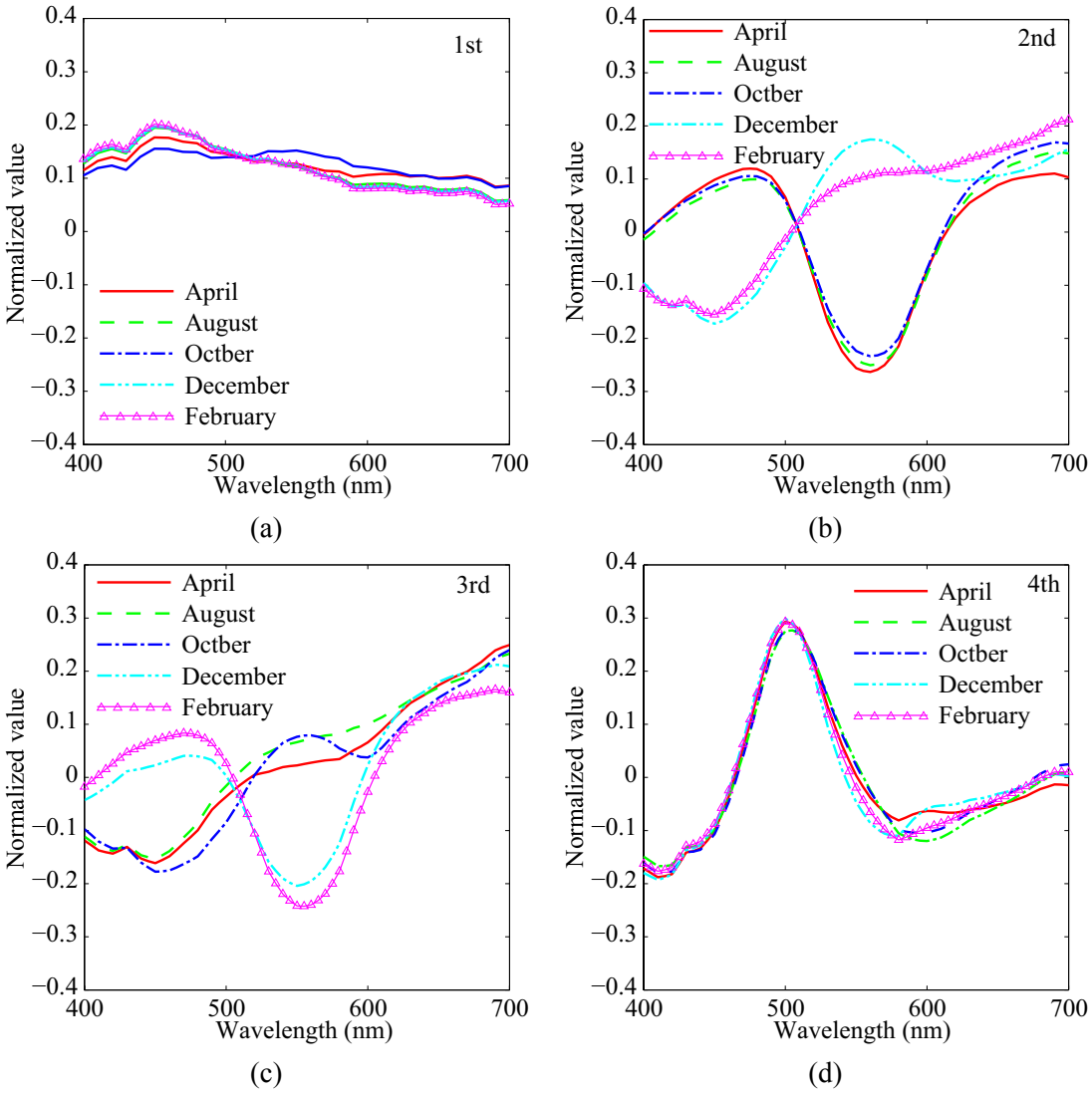


Figure 5: Spectral curves of the first four principal components.

The detailed comparison between Figure 5 (b) and (c) suggests that the second component curve and the third one of December and February are different from the components of the other seasons. In the winter, dry leaves fall to the ground and so the visible area of the sky increases, while in the other seasons the sky is mostly covered with green leaves and cherry flowers. Moreover, we should note that the spectral curves in winter in Figure 5 (b) are close to the curves in the other seasons in Figure 5 (c), and in the same way, the spectral curves in winter in Figure 5 (c) are close to the ones in the other seasons in Figure 5 (b). Therefore, we switch the spectral curves between the second and the third components in winter. As a result, the omnidirectional illuminants for all seasons can be represented using three common principal components. Figure 6 also contains the performance indices $P(K)$ for the overall omnidirectional illuminant data of all seasons. Figure 7 shows the spectral curves of the three principal components for all omnidirectional scenes. With these spectral curves, the illuminant can be estimated

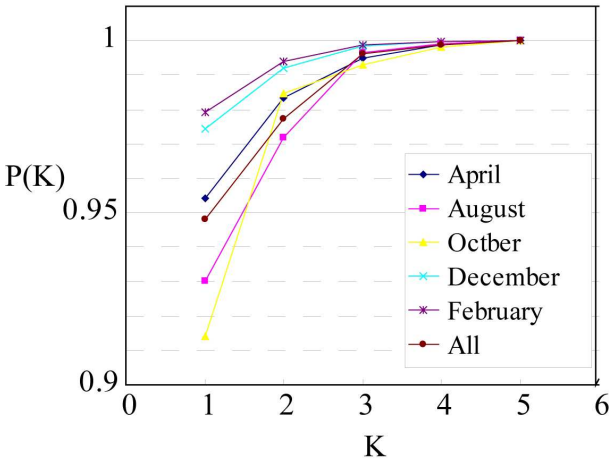


Figure 6: Performance indices $P(K)$ for approximation by principal components.

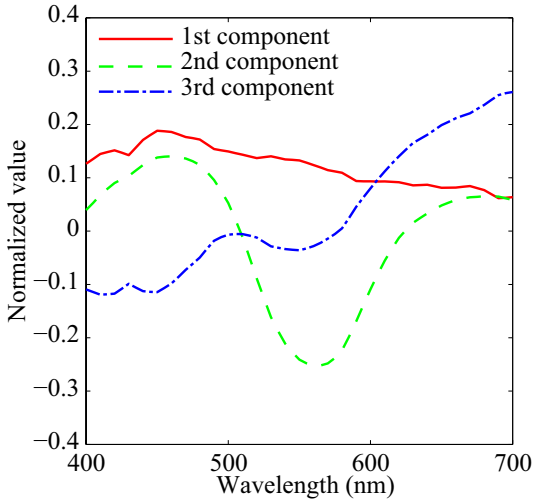


Figure 7: Three principal components for all omnidirectional scenes.

Figure 8: Omnidirectional color image rendered by only three components.

as

$$\hat{\mathbf{e}}_3(\theta,\phi) = c_1(\theta,\phi)\mathbf{u}_1 + c_2(\theta,\phi)\mathbf{u}_2 + c_3(\theta,\phi)\mathbf{u}_3. \quad (6)$$

In order to confirm the reliability of the proposed method, we have estimated the illuminant spectral-power distributions in a linear combination of the three spectral bases. The estimation accuracy was examined for the eight areas shown in Figure 3. We compared the estimation results to the direct measurements. The third column in Table 1 lists the RMSE between the illuminant estimate by the three principal components and the direct measurement. These comparisons suggest that the reliability of the proposed method based on the omnidirectional multiband imaging. Finally, Figure 8 shows the omnidirectional color image rendered by using only the three principal components. The image is almost completely coincident with the observed image in Figure 3.

5 Conclusions

This paper has described a method for analyzing omnidirectional spectral images in a natural scene. A multiband imaging system with a fisheye lens based on the orthographic projection was used for acquiring omnidirectional images with six spectral bands. We first recovered the spectral-power distributions of omnidirectional illuminants from the captured multiband images. Second, we estimated the spectral composition of the omnidirectional illuminants in a natural scene by using the PCA. Moreover, we analyzed the omnidirectional illuminants at a particular location over different seasons. It was shown that the overall omnidirectional illuminant spectral distributions could be described in a linear combination of three principal components. Finally the reliability of the proposed analysis method was confirmed in an experiment using the omnidirectional spectral images captured in different seasons in an outdoor scene.

References

[1] T.W. Cronin C.C. Chiao and D. Osorio. Color signals in natural scenes: characteristics of reflectance spectra and effects of natural illuminants. 2000.

[2] D.L. MacAdam D.B. Judd and G.W. Wyszecki. Spectral distribution of typical daylight as a function of correlated color temperature. 1964.

[3] T. Fukuda S. Tominaga and A. Kimachi. A high-resolution imaging system for omnidirectional illuminant estimation. 2008.

[4] G. Wyszecki and W.S. Stiles. *Color Science: Concepts and Methods*. 1982.

Least-Square Technique for Color Reproduction of Semi-Transparent Material

Dibakar Raj Pant
Gjovik University College, Norway
dibakar.pant@hig.no

Abstract

Industrial coloring process depends upon physical behavior of the colorants (dyes and pigments).To reproduce color to match the standard sample, it is necessary to calculate the absolute value of the scattering and absorption coefficients of individual pigments , denoted by 'K' and 'S'. Industries associated with the product development are being concerned how to mix pigments correctly. Coloring process with the pigments is complex especially in the case of semi-transparent material because of their non-linear behavior. The reasons are: first, pigments do not respond to the ideal theory exactly. Second, the semi- transparent materials exhibit two physical properties known as reflectance and transmittance. The above phenomenon states that calculation of the scattering and absorption coefficients can be done efficiently only by using the approximation theory like, The Least Square. In this method, many useful information as the inputs can be applied to evaluate the absolute 'K' and 'S'. In this experiment, first, reflectance of the mixture color sample panels (panels colored with more than two pigments) were measured and then applied the Least Square method to calculate the absolute value of the scattering and absorption coefficients of individual pigments. The K and S value of the pure white and black sample panels were also used in this method as the dependent variables . The final results are interesting and shown in graphical format .

1 Introduction

Color reproduction is becoming more important now a days in industries especially working on product development.Coloring process associated with the dyes and pigments(colorants) is complex due to their non linear nature[4].Stated in another way, many common materials colored with pigments and dyes have capability to absorb, transmit and scatter the light. Thus , it is necessary to apply appropriate theory to know what color the mixture of colorants give in order to match the color of materials [5],[9] . The authors in this paper deal color, in viewing industrial application as an objective matter which depends on the following three factors[3].

- The nature of the prevailing illumination under which the colored surface is viewed.
- The interaction of the illuminating radiation with the colored species in the surface layers.
- The ability of the radiation that is transmitted,reflected and scattered from the colored surface to induce the sensation of color.

Coloring industries main problem is how to predict the proportion of the three or more colorants required to produce a color match for an object showing the same color as a given refrence object or standard.This complexity occurs when the colorants scatter and absorb light because the light scattering properties of colorant layers deposited on the substrate such as plastic materials(semi-transparent mateials) are difficult to describe exactly in quantitative terms [1], [2],[8].In general, the prediction of color of semi-transparent materials necessitate characterization of colorants according to two coefficients K and S, the absorption and scattering coefficients respectively[11].The coefficient K describes absorption of diffuse flux by the colored material,while the coefficient S describes the net scattering of flux from the surface layer.These coefficients are known as the most fundamental optical properties of pigment colored materials.Therefore, to predict the reflectance spectra of mixtures of colorants, it is necessary to determine unit scattering and absorption coefficents also known as absolute coefficients. There exists various theories to formulate colorants for matching and the most conventional one is the Kubelka-Munk theory.In many

literatures, it is described that the Kubelka-Munk theory alone can not give true value of the absorption and scattering coefficients[3],[5],[11],[6].This fact arises because of the restrictions and assumption imposed during its derivation.One of the major restriction of this theory is its assumption of uniform light traveling media which in parctical cases do not occur because of change of refractive index in air to layer and vice versa[3],[7].Further, real dyes and pigments do not respond ideally with theory.There are also practical problems with very bright colors such as yellow, orange and red having low scattering power[2]. From the above discussion, it is clear that color reproduction of semi-transparaent mateial associated with pigments is not an exact science but it is an approximation. Least-square technique is an approximation technique in which the values are adjusted automatically untill the sum of squares is as small as possible.Walowit et al. also described the least- squares method is an effective approach for evaluating K and S of semitransparent materials[3],[1]. The authors have sucessfully implemented Least square technique to determine absolute value of K and S of real industrial sample which will be fruther discussed in the section experimental result.

2 Formulation of Least-Square Approach

To estimate the solution by the Least Square method, it is necessary to prepare a set of colorant panels for dependent and independent varibles . In this experiment, the authors have considerd sample panel containing colorant only as an independent variable and the other panels white only,black only,white and black,colorant and white and colorant and black as dependent vaiables.The absorption and scattering ratio of each individual panel can be expressed in terms of their reflectance value as follows:

$$\frac{K}{S} = \frac{(1-R)^2}{2R} \tag{1}$$

Again, the absorption and scattering ratio of mixture panel can be expressed according to additive laws as follows:

$$\frac{K_M}{S_M} = \frac{K_a + K_b + K_W}{S_a + S_b + S_W} \tag{2}$$

where,

- $\frac{K_M}{S_M}$ is the ratio of coefficients of mixtures
- $\frac{K_a}{S_b}$ is the ratio of coefficients of colorant only
- $\frac{K_b}{S_b}$ is the ratio of coefficients of black only
- $\frac{K_W}{S_W}$ is the ratio of coefficients of white only

Equation 2 can also be expressed in terms of relative to white and black standard reference.

$$\omega_M = \frac{K_M}{S_M} = \frac{K_a W + K_b W + K_W W}{S_a W + S_b W + S_W W} \tag{3}$$

For the fitting process, the Equation 3 must be expressed in the linear form which is accomplished by defining the relative reference coefficients by their concentration multiplied with per unit relative coefficient.So, the individual terms of the Equation 3 turn into:

$$\left.\begin{aligned} K_{aW} &= C_a \cdot k_{aW} \\ K_{bW} &= C_b \cdot k_{bW} \\ K_{WW} &= C_W \cdot k_{WW} = C_W \cdot \omega_W \end{aligned}\right\} \tag{4}$$

where,
K_{aW} and $k_{aW}$is the relative and unit absorption coefficient of colorant with standard white respectively
K_{bW} and k_{bW} is the relative and unit absorption coefficient of black with standard white respectively
K_{WW} is the relative absorption coefficient of white with standard white
ω_W is the ratio of coefficients of white only
C_a,$C_b$and C_W are the concentration of colorant, black and white in the mixture panel.
In the same fashion, relative scattering coefficients are expressed as:

$$\left.\begin{aligned} S_{aW} &= C_a \cdot s_{aW} \\ S_{bW} &= C_b \cdot s_{bW} \\ S_{WW} &= C_W \cdot S_{WW} = C_W \end{aligned}\right\} \tag{5}$$

where,
S_{aW} and $s_{aW}$is the relative and unit scattering coefficient of colorant with standard white respectively
S_{bW} and s_{bW} is the relative and unit scattering coefficient of black with standard white respectively
S_{WW} is the relative absorption coefficient of white with standard white
C_a,$C_b$and C_W are the concentration of colorant, black and white in the mixture panel.
Now, Substituting Equations 4 and 5 into the Equation 3 , the final structure in linear form can be written as

$$C_W(\omega_M - \omega_W) = C_a \cdot k_{aW} - C_a \cdot s_{aW} \cdot \omega_M + C_b \cdot k_{bW} - C_b \cdot s_{bW} \cdot \omega_M \tag{6}$$

Again, The value of ω_M is obtained by the reflectance R of the calibration panel using the Equation 1 . Here, the goal is to find unknown quantities k_{aW}, s_{aW}, k_{bW} and s_{bW} from the known quantities C_W,C_a,C_b, ω_M and ω_W from least square approximation.To carry out this, it is required to formulate a general equation and then compared with the Equation 6. In the standard formulation,two variables X and Y is used for n data set to find a function giving the value of the dependent variable Y from the values of an independent variable X.Considering a model function $Y = f(X,a)$,it can be expressed as:

$$Y = a_1 \cdot f_1 + a_2 \cdot f_2 + a_3 \cdot f_3 + a_4 \cdot f_4 \tag{7}$$

For being specific to our problem, we consider four constants a_1, a_2, a_3 and a_4 and the value of the expression for the i set of data is written as

$$Y_i = a_1 \cdot f_{1,i} + a_2 \cdot f_{2,i} + a_3 \cdot f_{3,i} + a_4 \cdot f_{4,i} \tag{8}$$

The weighted version of the Equation 8 is

$$WY_i = Wa_1 \cdot f_{1,i} + Wa_2 \cdot f_{2,i} + Wa_3 \cdot f_{3,i} + Wa_4 \cdot f_{4,i} \tag{9}$$

Now,this method calculates the constants a_1, a_2, a_3 and a_4 which minimize the sum of the squares between the measurements and the model. mathematically, the sum of the squares of the deviation E over the n data set is

$$E = \sum_{i=1}^{n} \varepsilon^2 \tag{10}$$

where, $\varepsilon = Y_i - X_i$ the deviation of i^{th} value of Y from the value of X . The weighted version of the Equation (10) is

$$W_i E = \sum_{i=1}^{n} W_i \varepsilon^2 \tag{11}$$

The set of n weighted equations can be expressed in matrix form as:

$$[WY] = [Wf] \cdot [a] \tag{12}$$

Finally, the solution of the Equation 12 in the matrix form is [3], [1],[10]

$$[a] = [Wf]^T \cdot [Wf]^{-1} \cdot [Wf]^T \cdot [WY] \tag{13}$$

Comparing the Equation 6 and Equation 9 , the following relationship is obtained

$$\left.\begin{array}{ll} WY_i = C_W(\omega_M - \omega_W) & \\ f_{1,i} = C_a \cdot W & f_{2,i} = -C_a \cdot \omega_M \cdot W \\ f_{3,i} = C_b \cdot W & f_{4,i} = -C_b \cdot \omega_M \cdot W \end{array}\right\} \tag{14}$$

The optical coefficients whose values have to be determined are related as follows

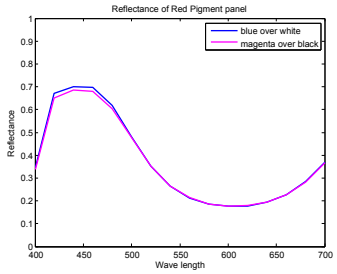
$$\left.\begin{array}{ll} a_1 = k_{aW} & a_2 = s_{aW} \\ a_3 = k_{bW} & a_4 = s_{bW} \end{array}\right\} \tag{15}$$

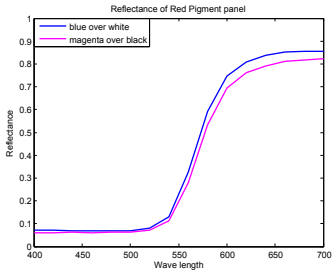
3 Experimental Result

The main aim of least square technique is to prepare optical data file of standard colorant samples. In this experiment, the samples are semi-transparent plastic materials colored by pigments.In the language of coloring industry, the samples are pigments suspended in the plastic material.The input to least square model are the ratio of coefficient of mixture(sample panel) defined by ω in equation 3, the ratio of coefficient of white(white sample panel) and the amount of concentrations of colorant only, white only and black only.To obtain the input data the first step is the measurement of reflectance of the sample panel and the white panel . From this reflectance data the ratio of coefficient can easily obtained by using the Equation 1. The author has measured reflectance of the sample panels having different concentrations as well as the white panel. The measurement is done by using spectrophotometer at 20 nm bandwidth.Two types of reflectance data is used , one is measured under the black claibration standard and other is measured under the white standard. These two data are necesary to know the translucent of a material. In this particular case, light trap method is used for the black calibaration and the standard ceramic plaque is used for the white calibaration.

The measured reflectance of blue and red sample panels are as shown in the Figure 1

The author measured sample of different color of different concentrations.Finally, with known values of concentrations and $\frac{K_M}{S_M}$ data of sample panel , the absolute value is calculted as shown in the Figure 2.





(a) Reflectance of a blue translucent sample under black and white calibration

(b) Reflectance of a blue translucent sample under black and white calibration

Figure 1: Example of Reflectance Measurement

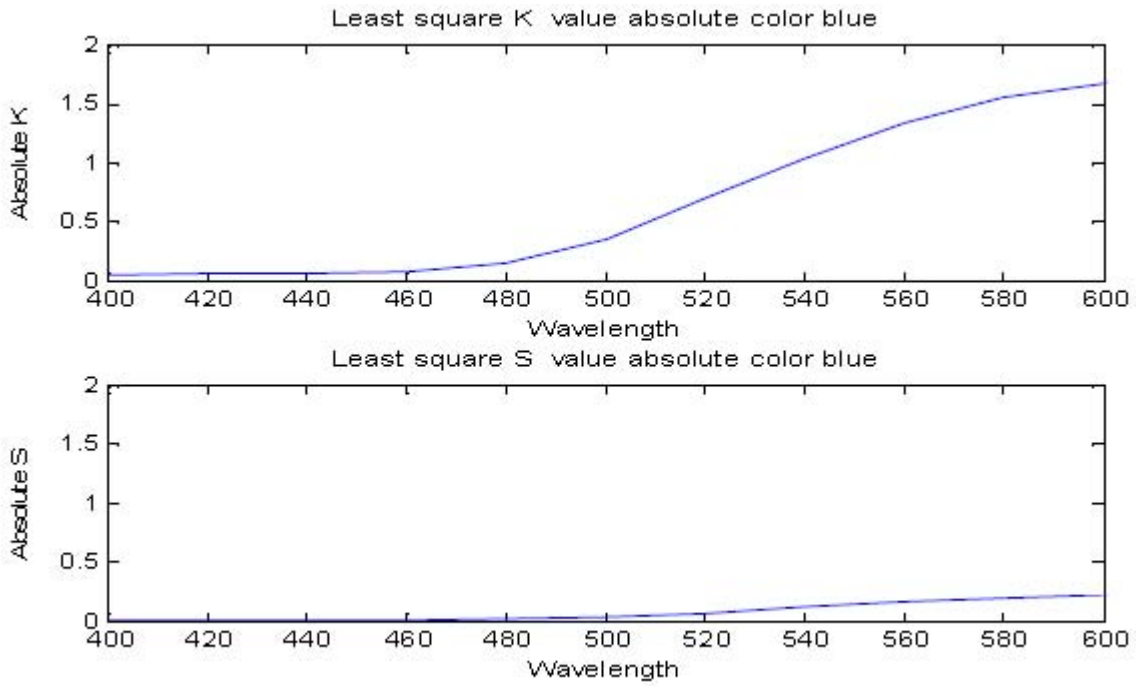


Figure 2: Absolute value of $K$and S of blue translucent sample.

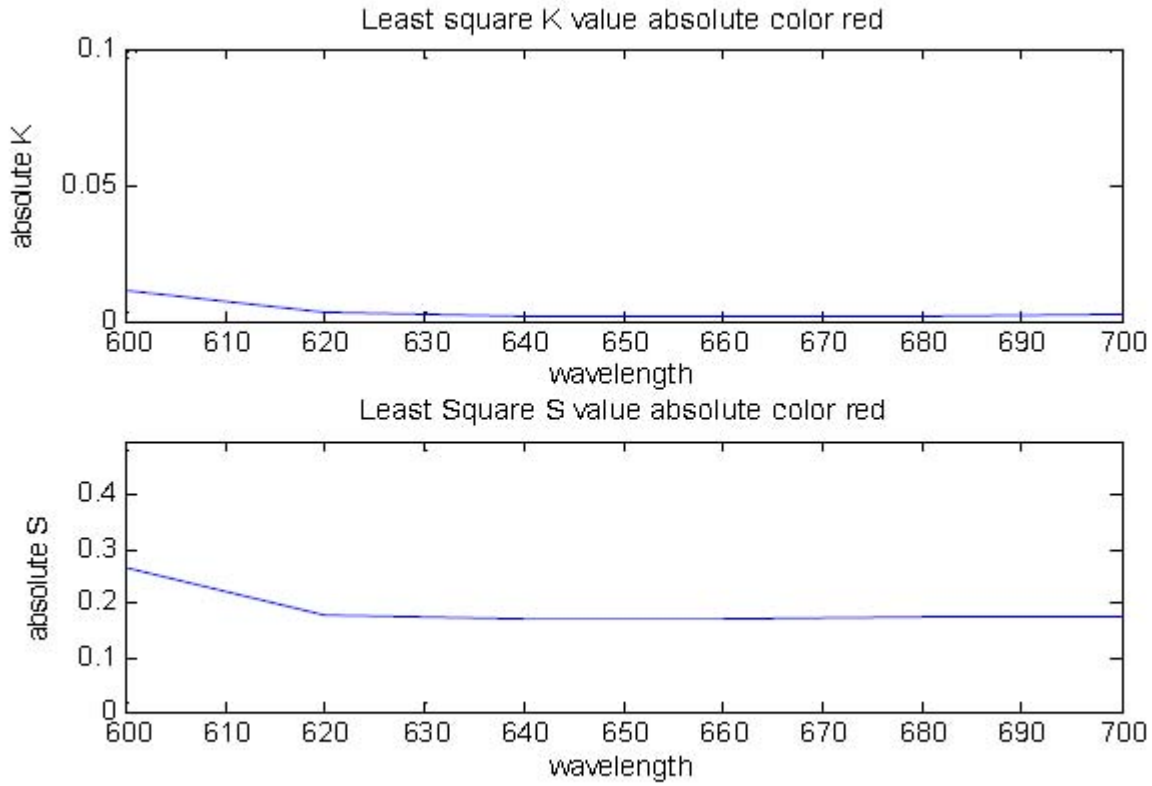


Figure 3: Absolute value of *K*and *S* of red translucent sample.

4 Conclusion

From the experimental result, it can be concluded that the least square method is simple and efficient approach to calculate the absolute value of absorption and scattering coefficient for reproduction of color according to the standard sample panel. The beauty of this method is that it takes the information about different pigment concentrations of the sample panels as well as concentration of the white and black panels simulatneously while doing analysis.This means it takes care about the tranlucent of those samples having low concentration of pigments which is the most important feature of the least square method.

4.1 Acknowledgments

The author would like to thank Marku Hauta Kasri, Department of Computer Science , University of Joensuu, Tuomas Backman and Janne Karila for theirsuggestion and support to perform this experiment. Similarly, I would like to give my sincere thanks to Ivar Farup for his valuable help to write this article.Finally, I am graetful to Prof.Jon Yngve Hardeberg for his continous encouragement to write about my work.

References

[1] Wyszecki & Stiles, 2000, Color Science:Concepts and Methods, Quantirative Data and Formula" *John Wiley and Sons,New york*

[2] Roy S.Berns,2000, Principles of Color Technology, *John Wiley and Sons,New york*

[3] Roderick McDonald, 1997, Color Physics for Industry *Society of Dyers and Colorist,England*

[4] Edward, J.Giorgianni & Thomas, E.Madden ,1997, Digital Color Management:encoding solutions *Addison Wesley Longman, Massachusetts*

[5] William E. Vargas & Gunnar A. Niklasson,1997 Applicability conditions of the Kubelka-Munk theory *J. optical Society of America, Volume 36 , Number 22*

[6] William E. Vargas,2002 Inversion Methods from Kubelka - Munk Analysis *J. optical Society of America, Volume 36 , Number 22*

[7] Li Yang & Bjorn Kruse , 2004 Revised Kubelka Munk Theory,2002 *J. opitical Society of America Volume 29, Number 10*

[8] Shah H. S. & Thaker A.A , 1990 Theoretical aids for the determination of colorimetric curves of binary powder mixtures *J. opitical Society of America Volume 29, Number 7*

[9] Frederic Cortat , 2004 The Kubelka Munk Theory , application and modification *Graduate course Report, Linkoping University, Sweden*

[10] Lewis-Beck. M & Bryman. A & Futing T.,2003 Least Squares:Encyclopedia of Social Sciences Research Methods. *The University of Texas ,Dallas*

[11] Sunderson J.L. , 1942 Calculation of the Color of pigmented plastics *J. opitical Society of America Volume 32, pp 727-736*

Highlight Removal Method for Spectral Image

Pesal Koirala
University of Joensuu
Joensuu, Finland
pkoirala@cs.joensuu.fi

Markku Hauta-Kasari
University of Joensuu
Joensuu, Finland
mhk@cs.joensuu.fi

Jussi Parkkinen
University of Joensuu
Joensuu, Finland
jussi.parkkinen@cs.joensuu.fi

Abstract

Highlight removal method for spectral image has been presented. Orthogonal subspace projection (OSP) has been applied to get the specular free spectral image. By setting the threshold value in the difference between original spectral image and specular free spectral image, the highlight affected part in the image is detected. The basis function calculated by transformation from principle component (PC) of specular free image to original image of highlight free part is applied in whole specular free image so that spectral image without highlight with original color information is achieved.

1 Introduction

In this research we have studied the highlight removal method for single spectral image. Different types of spectral image of single colored surface to multicolored textured surface rendered in different light sources have been tested. The spectral images within the visible range of 400nm to 700nm with 10 nm sampling have been used . As a first step of method the specular free image is achieved by projecting the given spectral image in the orthogonal subspace of light source since the highlight region is the result of scaling of light source [7]. The specular spectral free image preserves the shades of the image but when we visualize it in RGB space , the shift in the original color of the spectral image has been found and some radians values may change to negative values. The original color of image without highlight is calculated by using the linear transformation of specular free image. The basis function for transformation is calculated by linear transformation from PC of specular free image to original image by employing only the highlight free part of the spectral image. The highlight affected part and highlight free part of the original spectral image is detected by setting the threshold value in the difference between original spectral image and specular free spectral image in each pixel value for all wavelengths.
There are a lot of previous works regarding highlight removal from RGB images [9][3][1] [5] but in our knowledge there are only some works regarding highlight removal in spectral image [2] [7]. The method in reference [2] uses mixture model of Probabilistic PCA to detect highlight affected part and diffuse part in the image. Finally the highlight detected part mapped across the first eigen vector of diffused part was used to remove highlight during reconstruction process by PCA. However the accuracy of the result depends on the information the first principle component carries. And the segmentation of the highlight may suffer in multicolored image with background. However this method does not need the information of light source. Fu et al [7] proposes specular free spectral image using Orthogonal Subspace Projection. In that method the projector maps the radiance spectrum to the subspace orthogonal to the illumination spectrum as a result specular free spectral image is obtained. This method requires the prior information of light source. The specular free image accurately preserves the shades of color however shift of color value was detected when specular free spectral image was rendered in RGB space. As a result of orthogonal subspace projection, the resulted radiance of specular free image may give the negative value. However this method is computationally fast and can be applied to spectral image rendered with multiple light source. The method [7] is based on the dichromatic reflection model [6].

2 Specular Free Spectral Image

Shafer et al. [6] described dichromatic reflection model for modeling the reflectance of dielectric objects. The model suggests that the reflection can be represented by the linear combination of the diffuse and specular components. Based on this model the response of the receptor for each wavelength λ at geometric position x is defined as in eq. (1).

$$I(x,\lambda) = K_d(x)R(\lambda,x)S(\lambda)q(\lambda) + K_s(x)S(\lambda)q(\lambda) \tag{1}$$

Where $I(x,\lambda)$ is the radiance value or response of receptor at geometric position x at wavelength λ. q is the camera sensor sensitivity at each wavelength. K_d and K_s are the weighting factors for diffuse and specular reflections respectively. The weighting factors are dependent on the geometric structure of the surface. R(λ,x) is the diffuse reflectance or surface albedo at pixel position (x) and at wavelength λ. S(λ) is the spectral power distribution of the illuminant which is independent of geometry of the surface. λ represents each wavelength which lies within visible range. In this experiment we have used the wavelength between 400nm to 700 nm with 10 nm resolution. In the specular component the specular reflection is not considered since it is assumed that specular reflection is equal to the spectral power distribution of light source. Even if the sensitivity is not flat, it can be omitted by merging it into the illumination term $S(\lambda)$ for the same image sensor used [7]. eq.(1) is rewritten in simple matrix form as in eq.(2).

$$I = K_d I_d + K_s S \tag{2}$$

Where $I_d = R.S$ is the radiance value for diffuse reflectance and sign . shows the element wise product. Similarly S is the element wise product between spectral power distribution of illuminant and camera sensitivity function. Since the specular effect in the image is due to the scaling of illumination. Specular effect in the image is eliminated by by mapping the radiance spectrum of original image I to the subspace orthogonal to the illumination spectrum S. The orthogonal subspace projector of S is designed as $P = U - S(S^TS)^{-1}S^T$. Here U is identity matrix. The specular free image (SF) is obtained by multiplying the radiance image by orthogonal subspace projector. The specular free image is given in

$$SF = PI = K_d PI_d + K_s PS \approx K_d PI_d \tag{3}$$

Here SF gives the specular free image since specular component is eliminated. However some radiance value of specular free image may be negative. The color shift has been detected when the specular free image has been rendered in RGB space but it preserves the shading accurately.

3 Highlight detection and Highlight removal

The highlight detected area in the given image is calculated on the single pixel level, based on the difference between the original RGB image and the modified specular free image (MSF)[4]. The same method has been used in the spectral image based on the difference between original spectral image and the specular free spectral image. Threshold value has been set in the difference between original and SF image to classify each pixel in the group of highlight and highlight free part according to following rule.

$$pixel = \begin{cases} \text{highlight} & \text{if } d(\lambda,x) > th \text{ for all } \lambda \\ \text{diffuse} & \text{otherwise} \end{cases}$$

Where $d(\lambda,x) = I(\lambda,x) - SF(\lambda,x)$
However to set the accurate threshold value is the challenging task and the same threshold value may not be valid for all images. In this experiment threshold value has been set after visual assessment

(a) (b)

Figure 1: Highlight detection (a) Spectral Image represented in RGB (b) Black is diffuse part and white is highlight part

of classification results of each image. The highlight detected part and highlight free detected part of spectral image is shown in Figure.1. Threshold value used for Figure. 1 is 0.5. As we mentioned before SF image correctly preserves shades but there is change in color while rendering in RGB form. To get original radiance back but without highlight we employ the linear principle component analysis [8]. In our analysis we found that only three principle component of Spectral image carries about 99.9% fidelity ratio. At first we detect the highlight free area of image using above mentioned method and calculate the principle component (inner product image) as shown in eq. (4) using three eigen vectors according to first three largest eigen values from the correlation matrix of SF image .

$$P_d = V^T F_d \tag{4}$$

Where F represents SF image arranged in the dimension of $n * N$, here n is number of wavelengths used and N is number of pixels. Matrix V is eigen vectors sorted in the descending order of eigen values. In our experiment the size of V is $n * 3$. $[]^T$ denotes the transpose of matrix. As a result P is principle component of SF image. Subscript $_d$ represents only highlight free part of image. The new basis function W is calculated as shown in eq. (5) from the linear least square transformation from the principle component of SF image of highlight free area to the original image of highlight free area.

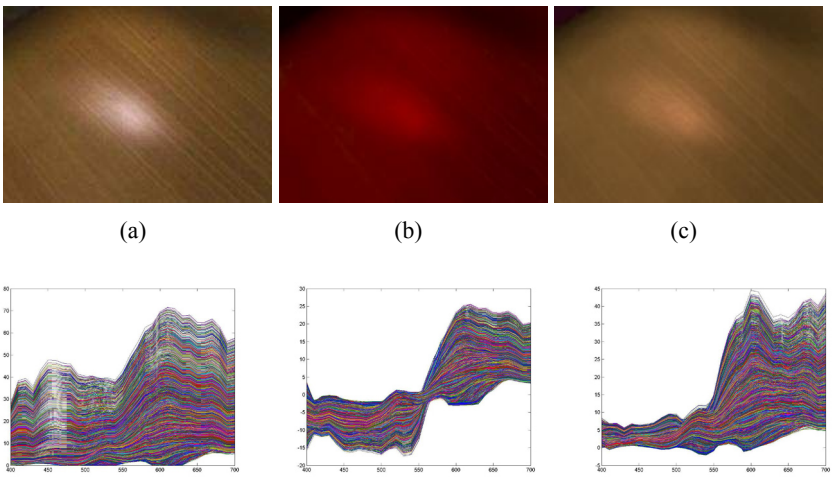
$$W = I_d({P_d}^T P_d)^{-1} {P_d}^T \tag{5}$$

Here $[]^{-1}$ and $[]^T$ represent inverse and transpose of matrix respectively. The spectral image without highlight$\tilde{I}$ is estimated by applying basis function W in the whole principle component of SF image as shown in eq.(6).

$$\tilde{I} \approx WP \tag{6}$$

4 Results and Discussions

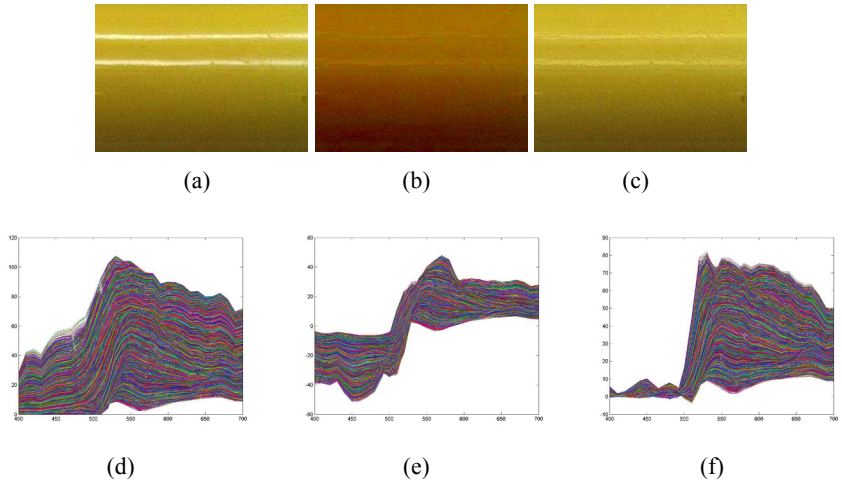
Dichromatic reflection model describes the reflection of the surface as a linear combination of specular and diffuse reflection components [6]. The scaling factor of the diffuse and specular reflection depends on the geometric properties of the surface therefore each component should be separated pixel wise. The specular component of the image is removed by projecting radiance image in orthogonal subspace of illuminant and original spectral image without highlight was obtained by linear transformation of PCA of specular free image to the original image of highlight free area. The result obtained is promising even for texture image as shown in Figure. 2, here we have shown original image, specular free image and final highlight removed image and corresponding radiance spectra. Similarly Figure. 2 shows the result of spectral image of uniform color surface with round shape. The described method needs the information of light source, we have shown result under day light source, the method works equally well for other light sources and multiple light sources. However we need the prior information of lightsource before estimating SF image. As a future work, the highlight removal for single spectral image without knowing prior information of light sources can be considered. The method presented can be extended easily to multiple light sources.



(a) (b) (c)

(d) (e) (f)

Figure 2: Highlight removal from textured image (a) Original image (b) Specular free image (c) Highlight removed image (d)(e)(f) Radiance spectra of original, specular free and highlight removed images respectively



(a) (b) (c)

(d) (e) (f)

Figure 3: Highlight removal from uniform surface image (a) Original image (b) Specular free image (c) Highlight removed image (d)(e)(f) Radiance spectra of original, specular free and highlight removed images respectively

5 Conclusions

Computationally fast highlight removal method for single spectral image has been described. At first the specular free spectral image was calculated by projecting the radiance spectral image in the orthogonal subspace of light source. However there is shift of color and negative radiance value in specular free spectral image. By setting the threshold value in the difference in each pixel between original image and specular free image, the highlight affected part is detected quite accurately. The basis function calculated from the linear relation ship between the principle component of specular free image and original image considering only highlight free area was employed to get spectral image without highlight. The method was tested for different types of spectral image under different light source, the result shows the feasibility

of the method.

References

[1] Robby T.Tan, Katsushi Ikeuchi : Seperating Reflection Components of Textured Surfaces Using a Single Image. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 178–193.

[2] Vladmir Bochko, Jussi Parkkinen: Highlight Analysis Using a Mixture Model of Probabilistic PCA. Proceedings of the 4th WSEAS International Conference on Signal Processing, Robotics and Automation, Salzburg, Austria Article No. 15 (2005) ISBN:960-8457-09-2.

[3] Method for computing the scene illuminant chromaticity from specular highlights. J. Opt. Soc. Am **3** (1986) 1694–1699

[4] Hui-Liang Shen, Hang-Gang Zhang, Si-Jie Shao, John H.Xin: Chromaticity-based Seperation of reflection components in a single image. Pattern Recognition **41** (2008), 2461-2469.

[5] S.K. Nayar, X.S. Fang, T. Boult: Separation of reflection components using color and polarization, Int. J. Computer Vision, Vol. 21, 1997, 163–186.

[6] S.A. Shafer: Using color to seperate reflection components, Color Res. App., Vol. 10, 1985, 210–218.

[7] Zhouyu Fu, Robby T. Tan Terry Caelli: Specular Free Spectral Imaging Using Orthogonal Subspace Projection, Proceedings of the 18th International Conference on Pattern Recognition, 2006

[8] Jolliffe I.T.: Principle Component Analysis, Springer series of statistics, Springer, 2002.

[9] Ping Tan, Stephen Lin, Long Quan, Heung-Yeung Shum: Highlight Removal by Illumination-Constrained Inpainting, Proceeding of the ninth IEEE International Conference on Computer Vision (ICCV 2003), 2-Volume.

A fully Automatic Redeye Correction Algorithm with Multilevel Eye Confirmation

Sony George
International School of Photonics
Cochin University
of Science and Technology, India
sonytgeorge@gmail.com

Jon Y Hardeberg
The Norwegian Color Research Laboratory
Gjøvik University College
Norway
jon.hardeberg@hig.no

Tomson G George
Wipro Technologies
Cochin, India
tomsongg@gmail.com

Abstract

Redeye is a well-known problem in portrait photography. This effect is caused by the light entering the subject's eye through the pupil and reflecting from the retina blood cells back to the sensor. Many of the existing redeye correction methods sometimes fail to detect the actual redeye artifacts and incorrectly change other red colored areas. The proposed algorithm describes a fully automatic redeye correction system with multilevel eye confirmation stages. The algorithm first identifies the skin and locates the redeye region using color information. The detected region is then confirmed as redeye by its redness variation, glint, eyes-lips triangle and comparison with surrounding regions. Finally the defect is corrected by desaturating the red pixels while preserving the natural glint presence.

1 Introduction

Redeye is a common problem in consumer portrait photography. This artifact is caused by the flash light which enters through the pupil, and its reflection from the blood cells of the retina, making the pupil of the eye appear bright red, instead of the natural black. Redeye is most prevalent when the distance between the flash unit and optical axis of the camera is small. This is the case for many modern and easy to carry consumer cameras. Fig 1 depicts typical redeye artifacts. Many image enhancement techniques are available in software and hardware form to reduce this problem. One of the widespread solutions is to use multiple flashes which make the pupil to be contracted before the actual photography is taken. Although it prevents the redeye to some extent, this makes more power consumption which is not desirable. Redeye reduction softwares basically come in two forms; either as fully automatic or as semi-automatic where the user has to specify the regions to be corrected.

All the available methods contain generally two stages: detection and correction. Detection commonly involves finding the red candidates from the face and then confirms the detected part is eye. An automatic detection algorithm usually starts with skin and red pixel extraction and proceeded by confirmation and correction.

2 State of the Art

Most of the commercial image editing packages like Adobe Photoshop, Paintshop, Corel Draw, or Photopaint offers the facility to correct redeye artifacts and each of them has different degree of success. Almost all these require human intervention for the redeye detection. Even though many redeye removal algorithms are patented from industrial research groups; there is only few journal published works.

Figure 1: Typical redeye effect

Dobbs [1] published the first paper which detects the redeye itself using the roundness and the amount of red in a specific region selected by the user. Techniques proposed by Hardeberg [2, 3] to correct the redeye, are based on an image mask which is computed by calculating a colorimetric distance between a prototypical reference redeye color and each pixel of the image containing this artifact. Morphological operations are applied to the binary mask, followed by a blob analysis method to group the pixels of the mask into eight-connected components. The blob that has the highest probability of representing a redeye artifact, based on its size and shape is chosen. At the last stage the mask is smoothed, to achieve a softer correction that appears to be more natural to the human viewer. Gasparini and Schettini [4, 5] start their algorithm by detecting a face using color information (skin detection). They also use the edges on the intensity channel (detection of the nose or the eyes for instance) to ensure that the detected region is an eye. Only few solutions [6] have been developed specifically for the redeye confirmation stage. Most approaches simply de-saturate the pixels in the detected red regions. Despite the different methods present in state of the art, redeye detection and confirmation remains an open subject. Few conference publications based on the research results which make more efficient correction techniques, such as [7, 8].

3 Skin and Red Pixel Classification

Like many other human image processing algorithms, skin segmentation is the first stage of the proposed algorithm and a subsequent process for face detection depends on the skin/non skin classification. Due to the wide range of applications, skin detection has received considerable attention over the years. Research in skin detection has made use of many techniques like, neural networks [9], template matching, point distribution models (PDMs), Eigen faces, fisher faces and statistical approaches [10] like support vector machines (SVMs), higher order statistics [11], etc. It has been observed that color modeling [12] is an efficient tool to detect skin pixels. There has been various color space models are available and out of that HSV color space performance shows less error rate [13]. HSV color can most successfully segment the faces, while RGB or YCbCr color fails to do so depending on the image. Also, we can see that the skin color falls into a very small region in the entire HSV space. Skin region can be extracted by converting to HSV domain and with a suitable threshold operation.

In the cylindrical HSV color space, H is the hue component, which describes the shade of the color, S the saturation component, which describes how pure the hue (color) is, while V stands for the value component, which describes the brightness. The removal of V component takes care of varying lighting conditions. H varies from 0 to 1 on a circular scale i.e. the colors represented by H=0 and H=1 are the

same. S varies from 0 to 1, 1 representing 100 percent purity of the color. H and S scales are partitioned into 100 levels and a color histogram is formed using H and S.

A pixel is detected as 'Red' region if (Red/Green > Red-Green threshold) AND (Red > min(Red)). Likewise as 'Skin' region if V[0.4] AND S [0.2-0.6] AND (H[0-25] OR H[355-360]) is true. From experiments the Red-Green threshold is set to 1.8 (see Figure 2).

Color information alone will not give true output in all cases. We can improve the skin detection by using edge information and modifying the way we classify skin pixels. Since some of the false pixels will have a high gradient they can be found at the boundary of the face. We use the Sobel gradient to find the edges to classify pixel as skin, it should exceed the skin threshold for histogram as well as have a gradient less than a certain threshold called edge-threshold.

Using region growing technique, the skin region and red region are grown independently. After the primary skin non-skin classification, morphological operations dilation followed by erosion performed to clean up the skin area.

Figure 2: Skin detection stage

4 Redeye Confirmation

An eye confirmation module confirms each detected red region as being either part of an eye or not, and outputs an indication of those detected regions that are confirmed as being parts of eyes. The identified redeye regions can then be made available to other systems for correction. We have used multilevel redeye confirmation algorithm, which facilitates the error rate to be the least.

4.1 Confirm with Glint and Redness Variation

If a redeye is visible to a human observer, it is likely that the artifact is surrounded by pixels of a different degree of redness, creating an oscillation in the redness measure. One of the features that aids in the detection and confirmation of redeye is the glint (Fig. 3) caused by a camera flash. Since flash is the cause of redeye, reflection of the flash is often highly visible in the eyes of the subject. The Laplacian operator can easily locate such sharp changes in brightness. However, it will also emphasize pixels with excessive noise, which is undesirable. Hence, a low-pass-filtered version of the Laplacian, computed over the luminance plane is used to detect eye glint in the image.

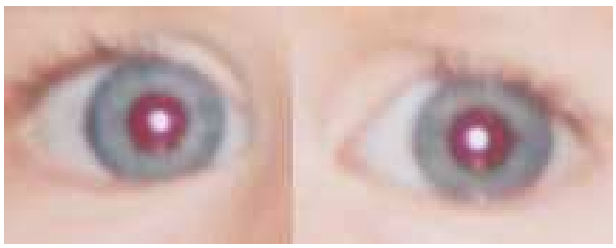


Figure 3: Glint formed during flash photography

4.2 Redeye Confirmation by Standard Deviation

This part of the algorithm performs a verification check to each candidate redeye area in order to ensure that it is “red enough” by itself, and more red than its surroundings. These verification tests are applied to candidate redeye areas. Candidate redeye pixel areas are filtered from the preliminary set based on redness contrast thresholds. Redness of the identified region is compared with four immediate blocks of given shape. If the standard deviation of the pixel content in each block is above a threshold then it is considered be as valid redeye region. Fig. 4 shows the selection of pixel blocks around the detected red area.

Figure 4: Standard deviation stage

4.3 Confirmation with Eye-Lips Geometrical Property

This approach is able to directly locate eyes, mouth, and face boundary based on measurements derived from the colorspace components of an image. Since the YCbCr colorspace has been identified as be best suited for classification of eye-lips, we use it for this part of the algorithm. An analysis of the chrominance components indicated that high Cb and low Cr values are found around the eyes. The Eye map in the chroma is constructed from Cb, the inverse of Cr, and the ratio Cb/Cr. The two resulting eye maps are combined by a multiplication operation. The resultant eye map brightens both the eyes and suppresses other facial areas. The mouth region contains excess red component compared to the blue component than other facial regions. Hence, the chrominance component Cr, is greater than Cb, near the mouth areas. We further notice that the mouth has a relatively lower response in the Cr/Cb feature but a high response in the Cr2. Therefore, the difference between Cr2 and Cr/Cb can emphasize the mouth regions. The eyes and mouth candidates are verified by checking (i) luma variations of eye and mouth blobs (ii) geometry and orientation constraints of eyes-mouth triangles and (iii) Based on the locations

of eye-mouth candidates, our algorithm first constructs a face boundary map from the luma. A triangle formed by the identified eye region and lips and, if the height to width ratio falls within the range of well-known golden ratio tolerance and if the percentage of skin is higher than a threshold called percentage threshold.

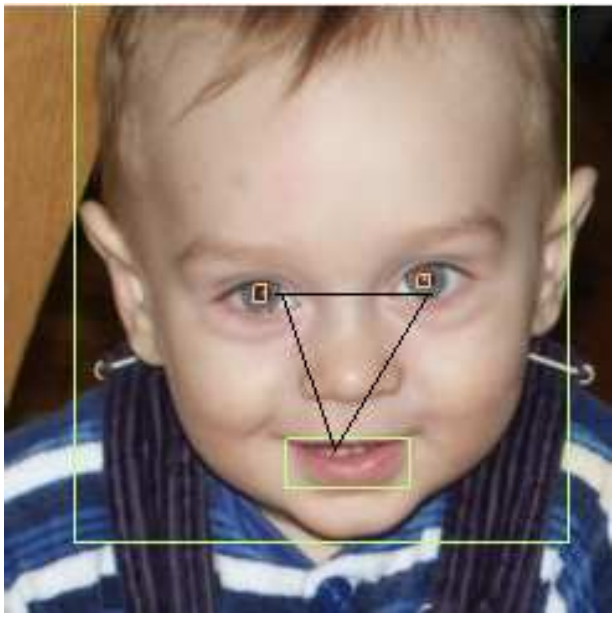


Figure 5: With eye lips detection

From the analysis of redeye images it has observed that, in a majority of cases, redeye artefact is generated when flash light falls in an acute angle and hence less probable for this to occur for images which has taken from side of the candidate. In order to avoid the error of detection with single eye images, we set less weight for the eye-lips triangle stage than other two.

5 Redeye Correction

The ultimate aim of any redeye correction algorithm is to eliminate red pixels in the pupils, and at the same time preserving important image details such as the glint (i.e. white curvature reflections in the cornea) and borders between pupil and iris. Only few solutions [14] have been developed specifically for the correction step. Most approaches simply de-saturate the pixels in the detected redeye regions.

In this algorithm, the detected redeye region, each pixel is corrected by first taking the average of 9 surrounding pixels and this reference is then processed according to Equation 1. From the confirmation step, the size of the red region is obtained and also the reddishness of the region. With reference to this data,a nd alpha value is determined and is used.

$$Rout[i,j] = alpha * [0.05 * Rin[i,j] + 0.6 * Gin[i,j]) + 0.3 * Bin[i,j] + (1 - alpha) * Rin[i,j]. \quad (1)$$

The correction is mainly performed by reducing the red content and by increasing the green and blue elements.

As shown in Fig 6, we checked the effect of the neighbour pixel averaging with 9 and 25 pixels and could not observe any differences, therefore we chose 9 pixels for averaging.

Figure 6: Corrected redeye region by using 9 pixels averaging and 25 pixels averaging

6 Results and Discussions

This paper presents an efficient technique to correct redeye images. More emphasis of this algorithm is on the detection and confirmation part. We have tested our algorithm over a large number of redeye images of various types and found it to give visually good results [Fig 7, 8]. As proposed in this paper multi stage confirmation avoids false detection and hence gives less chance for wrong correction. The proposed algorithm gives corrected image outputs for almost all the detected redeyes.

Reduction in computation and increase in speed could be achieved by reducing the number of confirmation stages. Once confirmed as redeye in the first or second stage, the third step could be skipped.

References

[1] M. Dobbs and R. M. Goodwin, Localized image recoloring using ellipsoid boundary function, U.S. Patent 5,130,789, (1992).

[2] Hardeberg, J.Y., 2001. Red eye removal using digital color image processing. In: Proc. Image Processing, Image Quality, Image Capture, Systems Conf., Montreal, Canada, April 22–25, 2001. pp. 283–287.

[3] Hardeberg, J.Y., 2002. Digital red eye removal. J. Imaging Sci. Technol. 46 (4), 375–381.

[4] F. Gasparini, R. Schettini Color Balancing of Digital Photos Using Simple Image Statistics, Pattern Recognition, May 2003.

[5] R Schettini, F Gasparini. A modular procedure for automatic redeye correction in digital photos.

[6] Yi Wang, Fuhuei Lin, A novel automaic Red-Eye detection and removal method. International Symposium on Communications and Information Technologies (ISCIT 2007).

[7] Luca Marchesotti, Gabriela Csurka, Safe Red-Eye Correction Plug-in using Adaptive Methods., 14th International Conference of Image Analysis and Processing - Workshops (ICIAPW 2007)

[8] Li Chen, Jiliu Zhou, A skin detector based on neural network Communications, Circuits and Systems and West Sino Expositions, IEEE 2002 International Conference on Volume 1, 2002

[9] M. J. Jones and J. M. Rehg, Statistical color models with application to skin detection, Proc. CVPR'99, June 1999, pp. 274-280.

[10] N. Rajagopalan, K. Kumar. Finding faces in photographs. In Proceedings sixth IEEE Int'l Conference on Computer Vision, pages 640–645, 1998.

[11] T. Yoo and S. Oh. A fast algorithm for tracking human faces based on chromatic histograms. Pattern Recognition Letters, 20:967–978,1999.

[12] Osamu, Segmentation of faces in video footage using HSV color for face detection and image retrieval.

[13] R.Ulichney and M. Gaubatz, Perceptual-based correction of photo red-eye, IASTED Int. Conf. on Signal and Image Processing, 2005.

[14] F. Chazli, F. Gasparini and R. Schettini, A modular procedure for automatic redeye correction in digital photos, IS&T/SPIE Vol. 5293, 2004.

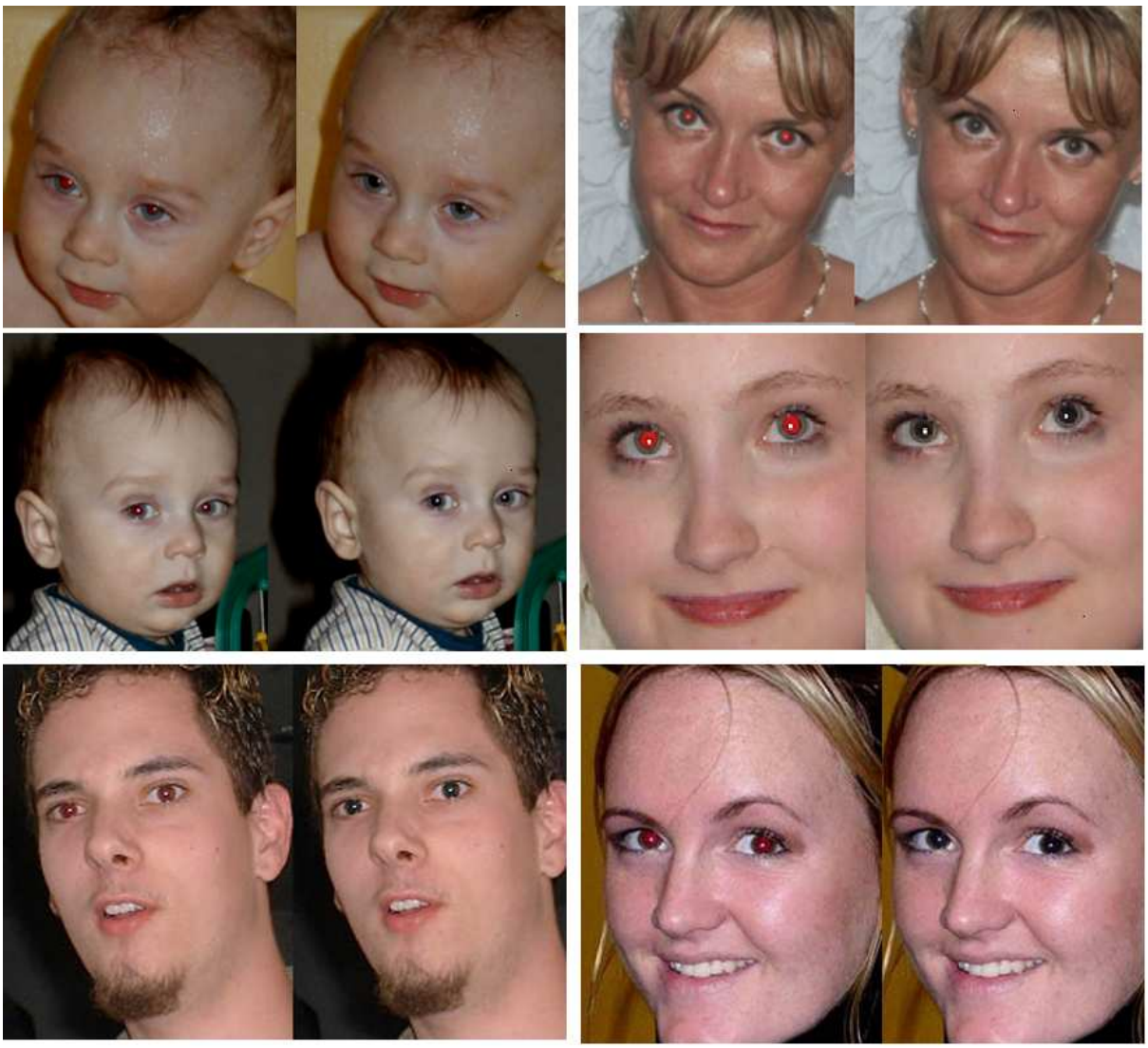


Figure 7: Redeye reduction results

Figure 8: Redeye reduction results

Finding Visually Interesting Regions using SURF Points

Patrick J.N. Harding
Heriot-Watt University
Edinburgh, United Kingdom
pjh3@hw.ac.uk

Neil M. Robertson
Heriot-Watt University
Edinburgh, United Kingdom
n.m.robertson@hw.ac.uk

Abstract

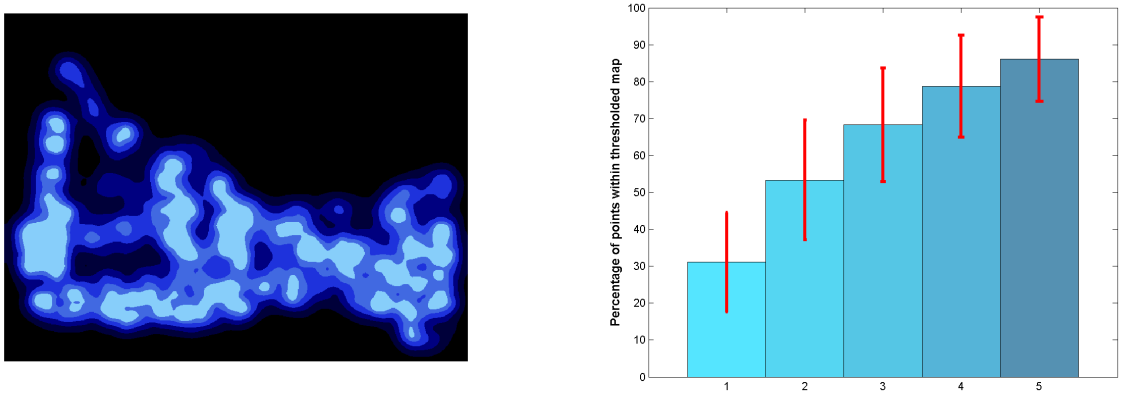
Recent work has shown that there is a varying coincidence of common interest-points towards the regions of an image that are visually salient. In this paper we compute a new saliency map which is derived from SURF interest points only. SURF points have been shown to be naturally distributed towards the visually salient regions in an image [4]. A probability map is computed which is then thresholded into the top 10 to 50% most salient pixels. We then compare the results with comprehensive eye-tracker data taken from human observers showing that up to 90% of the points attended by the observers can be recovered by our method. We then use this saliency map to perform more efficient image compression by extending the JPEG scheme to re-weight the image blocks by Q=50 or Q=5 depending on whether that region lies within a visually salient region of the image or not. We show the compression ratio is significantly higher *and* the more visually interesting regions are retained at higher resolution using our method.

(a) (b)

Figure 1: Transforming interest points into a surface is the first step in our method. Using robust interest points this surface can be expected to represent interesting regions even under different observer viewing angles and/or conditions. (a) Original image with SURF points superimposed, (b) interest surface computed from SURF points.

1 Introduction

There exist reliable models of visual saliency under passive viewing derived from bottom-up visual processes, such as described in [5]. Under observer tasking, there is a substantial shift in attention away from the passive case strongly driven by the nature of the task [4, 2]. There are models of attentive prediction under task (such as in [8, 6, 7], but they are not general models based on image content and involve prior learning of object categories and contexts. Given that certain interest-points have a high correspondence with the visually salient in both passive and task-based cases [4] we propose a method of construction for a general purpose "probability" map of what is visually interesting in an image based on the best-performing interest point scheme analysed in [4], Speeded



(a) (b)

Figure 2: Validating the proposed method by comparing with eye-tracker data: (a) shows the thresholded suface from Figure 1(b); chart (b) shows the overlap with the maps at different threshold levels for *all* eye-points, gathered over 8 experimental participants. The bar indices 1 to 5 correspond to the 10 to 50 surface percentage coverage of the masks. The bars indicate average overlap at each threshold (standard deviation in red).

Up Robust Features [1]. This has the advantage of highlighting a range of regions of interest, that *could* be attended under different viewing conditions and requires no prior learning. This is of value because we can use this interest-surface to apply image compresion based on image region importance such that if the viewing conditions by a subsequent analyst are changed, the key details of the image are preserved. In Section 2 we describe how the SURF points are used to compute an "interest" surface which can be thresholded to various degrees of visual saliency. This surface is then tested against a large set of eye-tracker data taken from human observers under strict experimental conditions. In Section 3 we introduce a new compression scheme based on JPEG where the salient regions are used to define the level of compression applied to each block depending on its coincidence with the saliency map.

2 Saliency Computation from SURF Points

An illustration of the SURF points on a test image is shown in Figure 1(a). Figure 1(b) shows the attentional surface derived from the distribution of the SURF points. The visual interest surface is built by calculating the Euclidean distance of each point in a blank template to each SURF point. The distances are then ranked in order and the probability map value for each pixel is assigned as the sum of the second to the tenth ranked distance elements. The map is then inverted and normalised to the interval [0.1 1]. This construction technique delivers an attention map that does not experience strong peaks at the points themselves and is also dependent upon the local density of SURF interest points. The lower bound further allows for the possibility of attention in non-SURF rich regions, which may be useful in any future combination with other attentional surfaces.

We tested our new visual attention probability map against eye-fixation data from observers under task. The eye-tracker data and image set from Torralba *et al.* has been used to validate the model [8]. The test image data set for this paper comprises 72 images and 108 search scenarios (3x36 tasks) performed by 8 observers performing a search-and-count task. (The tasks were count people in outdoor contexts and count paintings and cups in indoor contexts. Note that there are 36 images that have two different tasks applied to them). The visual-interest maps were constructed as described above for each image. The probability map was then thresholded to 10, 20, 30, 40 and 50% of the most salient pixels by image area, representing the supposedly more salient half of the image to finer

(a) (b) (c)

Figure 3: (a) The original image, (b) Standard Q50 JPEG compression, (c) Our new visually interesting region compression: Q50 in top 40% visually interesting regions, Q5 elsewhere. Note by inspection that the visually important regions in the image are generally preserved, while the contextual information outwith the core remains valuable, although highly degraded. The compression ratio is improved from 5.25 ± 0.5 to 6.0 ± 0.3 using our method without loss of detail at the salient locations (in this case the house).

degrees. Finally, the overlap % of the eye-tracker data was counted at each threshold level to assess the accuracy of the interest-point derived map at predicting human eye-fixations under task. The results are shown in Figure 2. The high coincidence of the eye-fixations with our attention based regions is a strong result. It remains consistent across different tasks, validating our assumption that the interest-points are a good way of assessing regions of visual interest under varying observation conditions.

3 JPEG Encoding Reweighted towards Salient Regions

We next use this result for a practical application. We choose to demonstrate a scheme for compression based on the JPEG algorithm which is designed for good visual quality in photo-real images. JPEG relies on quantisation of the Discrete Cosine Transform applied to 8 by 8 pixel blocks of an image. This reduces the relatively unimportant high frequency components in each block, allowing for efficient huffman or arithmetic coding. The quantisation is performed using a quantisation matrix derived from psychovisual tests and this matrix can be weighted to provide the required degree of compression in the block. The reverse process decodes the image [3, 9]. The heavier this quantisation, the larger the compression ratio achieved, however this is tempered by the fact that over-quantisation will produce blocking atifacts that significantly reduce image quantity and can damage real information within the image. In regular JPEG, the quantisation is fixed across the whole image. In our case, however, we have a reliable method of selecting regions of visual interest. In this example presented here, we choose to preserve the top 40% of the image by "visual interest" from SURF-point distribution, which we can expect in a probabilistic sense to attract 80% of eye fixations under task. We will compress the other 60% to a much higher degree. This information will not be lost altogether and will be available for contextual information.

We use a greyscale copy of the image and choose two quality factors to impose a high or low quality on the image region. The quality factor (Q) of 50 uses an unweighted matrix which is the original matrix derived from psychovisual experiments to give acceptable compression. The quantisation matrix we use is that specified in Annex K of the JPEG standard for the luminance component of images [3], appropriate for grayscale. We choose a low value of Q=5 for the outlying regions and weight the quantisation matrix according to the following relationship: $(50/Q) * Qmatrix$.

We set a threshold such that if the pixel in the image was in the most "visually interesting" 40% of the image it would have the Q50 weighting applied, Q5 otherwise. The quantised set of blocks were vectorised according to the jpeg zig-zag pattern [9] and the DC components were encoded dif-

ferentially according to the previous element. We appended each DCT block with one more element to give a block length of 65 - 0 if the block was for high compression and 1 if the block was for normal compression. Since the DCT-quantisation process generally results in large numbers of zeros at the highest of frequencies, the huffman encoding scheme that we used is generally efficient. We do add one more piece of information per 64 pixels, but we are able to discard more information with confidence than otherwise and there is a net gain from the efficiency in the huffman coding process.

We applied this visually-interesting region compression over all of our 72 images. As a comparison, we also performed a normal JPEG process at Q50. An illustration of the output is shown in figure 3. Over all of the 72 images the average compression ratio achieved by our Q50 JPEG compression was 5.25 ± 0.5 and for our visually interesting region based compression the achieved compression ratio was 6.0 ± 0.3. From the image set statistics above and from the illustration in figure 3, it is clear that there is an advantage in the method in terms of performance over regular JPEG as well as being capable of producing usable images where the core details of the scene survive the compression. e.g. in figure 3(b) the sky and bland textural details have suffered strong degradation, but the interesting content of the scene is largely preserved.

4 Conclusion and Future Work

In conclusion, we have used a robust interest-point detector (known to coincide with the visually salient parts of an image under different observer conditions) to construct a map of visually interesting regions of an image. We have validated this technique against observers acting under different tasks and the method is a good predictor of eye scan points under object count tasking. We have further demonstrated a compression scheme using the visual interest map as a guide that offers advantage in terms of filesize while preserving the core details of images. A strong advantage of the method is that it is simple and can be performed "live", being based on existing image content and not requiring any prior learning stages.

Avenues for future work include introducing a spectrum of Q values for the encoding of the JPEG compression rather than the binary approach taken here. It is likely that compression is not the only image processing technique which would benefit from intelligent application based on salient regions. We propose next to investigate segmenting an image using saliency then discriminately performing image enhancement on regions of interest.

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Computer Vision and Image Understanding (CVIU)*, 110(4):346–359, 2006.

[2] M. S. Castelhano, M. L. Mack, and J. M. Henderson. Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):1–15, 3 2009.

[3] JPEG Committee. Iso/iec 10918-1. ISO Standard, 1994.

[4] P. Harding and N. M. Robertson. A comparison of feature detectors with passive and task-based visual saliency. *LNCS*, 5575:716–725, 2009.

[5] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *Advances in Neural Information Processing Systems 19*, pages 545–552, 2007.

[6] V. Navalpakkam and L. Itti. Modeling the influence of task on attention. *Vision Res*, 45(2):205–231, January 2005.

[7] R.J. Peters and L. Itti. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2007.

[8] A. Torralba, A. Oliva, M.S. Castelhano, and J.M. Henderson. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 113(4):766–786, October 2006.

[9] G.K. Wallace. The jpeg still picture compression standard. *Commun. ACM*, 34(4):30–44, 1991.

Study of Spatio-Temporal Influence on Color Appearance

Olivier Tulet, Mohamed-Chaker Larabi and Christine Fernandez-Maloigne
XLIM-SIC, Poitiers University
Poitiers, France
name@sic.univ-poitiers.fr

Abstract

This paper deals with the study of the influence of spatiotemporal effect on the color appearance of a stimulus. Color appearance models (e.g. CIECAM02) have achieved to very interesting results for flat stimulus. However when a stimulus is modulated spatially, temporally or spatiotemporally these models fail in correcting its appearance. The aim of this work is to study the spatiotemporal influence on the color appearance of a stimulus and to model its behavior in order to extend the CIECAM02 color appearance model.

1 Introduction

Nowadays the color quality is a very important challenge in the industry framework. According to the media, color does not seem always simlar as for example the printed image could be different from the same displayed on a screen. This problem concerns the WYSIWYG (what you see is what you get) and for it a number of Color Appearance Models (CAM) have been developed. The main objective of a CAM is to ensure a reliable reproduction of colors through various devices by introducing some characteristics of the Human Visual System (HVS).This means that the corrected stimulus will be perceptually similar but physically different.
These models allow correcting many phenomena that modify the appearance of a color stimulus[2, 3]. However these models do not take into account the spatial aspect that can be contained in a stimulus or spatio-temporal aspect that can be contained in a video[5, 8].

Thus, figure 1 shows an example of the influence of the spatial modulation on human color perception. Indeed, the colors on the right and the left are physically identical but they appear different. This effect is due to the spatial modulation that affects our color perception. However, this effect is not taken into account by any CAM and many authors have mentioned this lack [5, 8].
In a first work[7], we have conducted a psychophysical experiment that aimed at measuring the difference between the perceived-similar stimulus and the physically-similar one. For this study a model taking into account the spatial modulation have been developed. This model, named S-CIECAM02, was build by exploiting the perceptual components of the CIECAM02: lightness (J), chroma (C) and hue (h). The obtained results are very interesting and an extension to images has been developed to serve as a

Figure 1: Influence of spatial modulation on color perception.

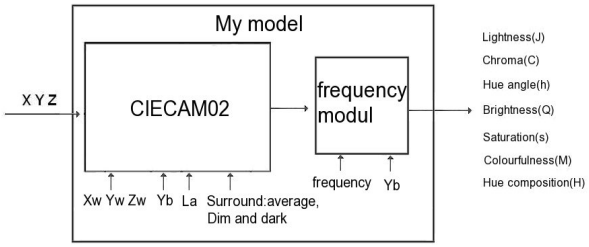
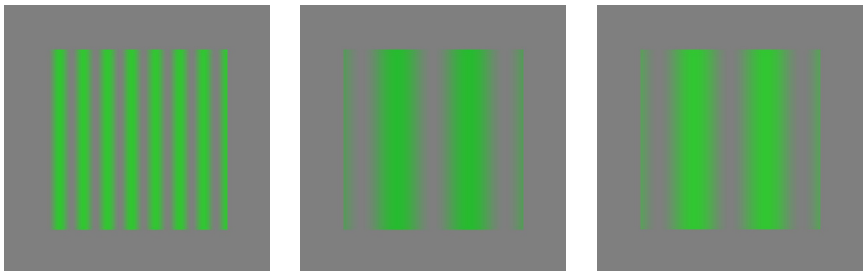


Figure 2: Representation of S-CIECAM

color rendering model.
This paper is an extension of the previous study dedicated to the prediction of the spatiotemporal effect on color appearance. So another psychophysical study has been prepared and conducted in order to measure the gap between perceived and physical spatial-temporally modulated stimulus.
The remainder of this paper is organized as follows: Section 2 briefly describes the S-CIECAM and its extension to images. Section 3 is dedicated to explain the spatio-temporal experiment and its results. Once those results obtained, they were modeled and integrated into CIECAM02 to create a new model: the ST-CIECAM. Some results of this new model are presented. We finish this contribution by some conclusions and we introduce some future works.

2 S-CIECAM

In a first work, an extension of the CIECAM2 color appearance model was developed with the help of a psychophysical experiment. This experiment was based on tuning the hue, the Chroma or the Lightness of modulated stimuli in order to match with flat stimuli in terms of appearance. The obtained results were modeled and integrated into the CIECAM02. The obtained model, called S-CIECAM, allow CIECAM02 to predict the appearance of a color stimulus whatever is its modulation frequency. Figure 2 shows the description scheme of the integration of the spatial model into CIECAM02.



-a- -b- -c-

Figure 3: Example of corrected pattern. (a): input stimulus at frequency f. (b): stimulus at frequency f' and same color as (a). (c): Spatially corrected stimulus from f to f'.

Figure 3 shows a result appearance prediction obtained by the S-CIECAM for a green stimulus on a gray background. The input stimulus is at a given frequency f. When this frequency is decreased to f', we obtain the stimulus of figure 3-b. It is easy to notice that the two colors seem different. By using the s-CIECAM, for the spatial correction, the corrected stimuli (figure 3-c) are closer to the input one.

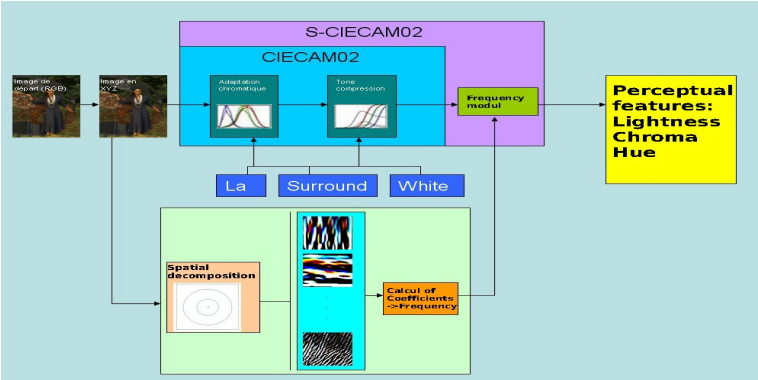


Figure 4: Rendering model: extension of S-CIECAM to images

2.1 Extension to Images

Even if the obtained results are quite encouraging, the s-CIECAM is designed to correct only one stimulus and is not adapted to the correction of images. This is why we have envisaged to extend it in order to deal with images. Based on a decomposition of the Fourier spectrum, this model is looking at the pixel as a stimulus and its neighborhood as the background and to estimate the average frequency to be given to the model. Figure 4 shows the flowchart of the image extension model and an example of obtained results is presented in figure 5.

In order to estimate the quality of the obtained results, a quality assessment session has been managed in order to compare original images to rendered ones.75% of observers prefered rendered images and the reason is tat it looked more natural than the original one.

Thanks to the obtained results, the next section will deal with the extension to temporal modulation in order to be able to predict spatio-temporal appearance of moving stimuli.

Figure 5: Example of results obtained by our model. (A): Original image. (B): Rendered image.

3 SpatioTemporal Appearance Prediction

3.1 Experimental Protocol

This new experiment looks like its predecessor, but this time the observer has to evaluate the color of a scrolling patch. On the left observer can see a reference patch which is uniform and not moving. On the left, a patch is spatially-modulated and its content is horizontally moving, more or less fast. For each test

Figure 6: Psychophysical test room

only one of the perceptual features of the modulated patch is increased or decreased until the observer judges it as perceptually similar to the reference stimulus. With this method, a gap of perception between an uniform patch and a spatio-temporal modulated one is obtained. Different parameters are presented to the observer:

Color stimulus To measure the spatio-temporal effect on color appearance, the three primaries have been chosen. With these colors we can obtain all the others.

Perceptual features Because of the use of the CIECAM02, we have chosen three of its perceptual feature : J for the Lightness, C for the Chroma and h for the hue. The observer in the test should judge of the appearance of a stimulus three times: one for each perceptual feature.

Spatial frequencies To limit the number of tests, we have chosen the following spatial frequencies: 3, 5, 10 and 20 cpd.

Temporal frequencies Because this parameter is the goal of the study, ten temporal frequencies were chosen: 0.4, 0.6, 0.8, 1.3, 1.5, 2.6, 3, 5.3, 10.6 and 21.6Hz. These values have been selected based on the study of a TCSF [6] and because they cover a large range of temporal values. After 20 Hz the display cannot reproduce correctly stimulus and some artifacts appear.

Background color One more time, in order to limit the number of test, only one background was chosen: the neutral gray.

3.2 Test Description

A panel of 15 observers has been selected. These observers have a correct color vision validated by a Snellen test for acuity and Ishihara test for color blindness. The experiments were conducted in a dedicated room respecting the standard conditions [1, 4] in terms of lighting, display calibration...

4 Results

The results obtained for the lightness shows:

- For low frequencies it seems that the maximum of difference between a flat stimulus and spatiotemporally modulated one is obtained near 3Hz.
- For high frequencies we have an increase of the perceived difference with the temporal frequencies.

The chroma and hue show the same behavior as the lightness. Standard deviations of the test are low for the lightness and the hue but the chroma seems more difficult to evaluate and its standard deviation is a slightly higher than the two others perceptual features. Figures 7-a, 7-c and 7-e gives the average experimental results for respectively chroma, lightness and hue obtained for the spatiotemporal experiments for the red primary.

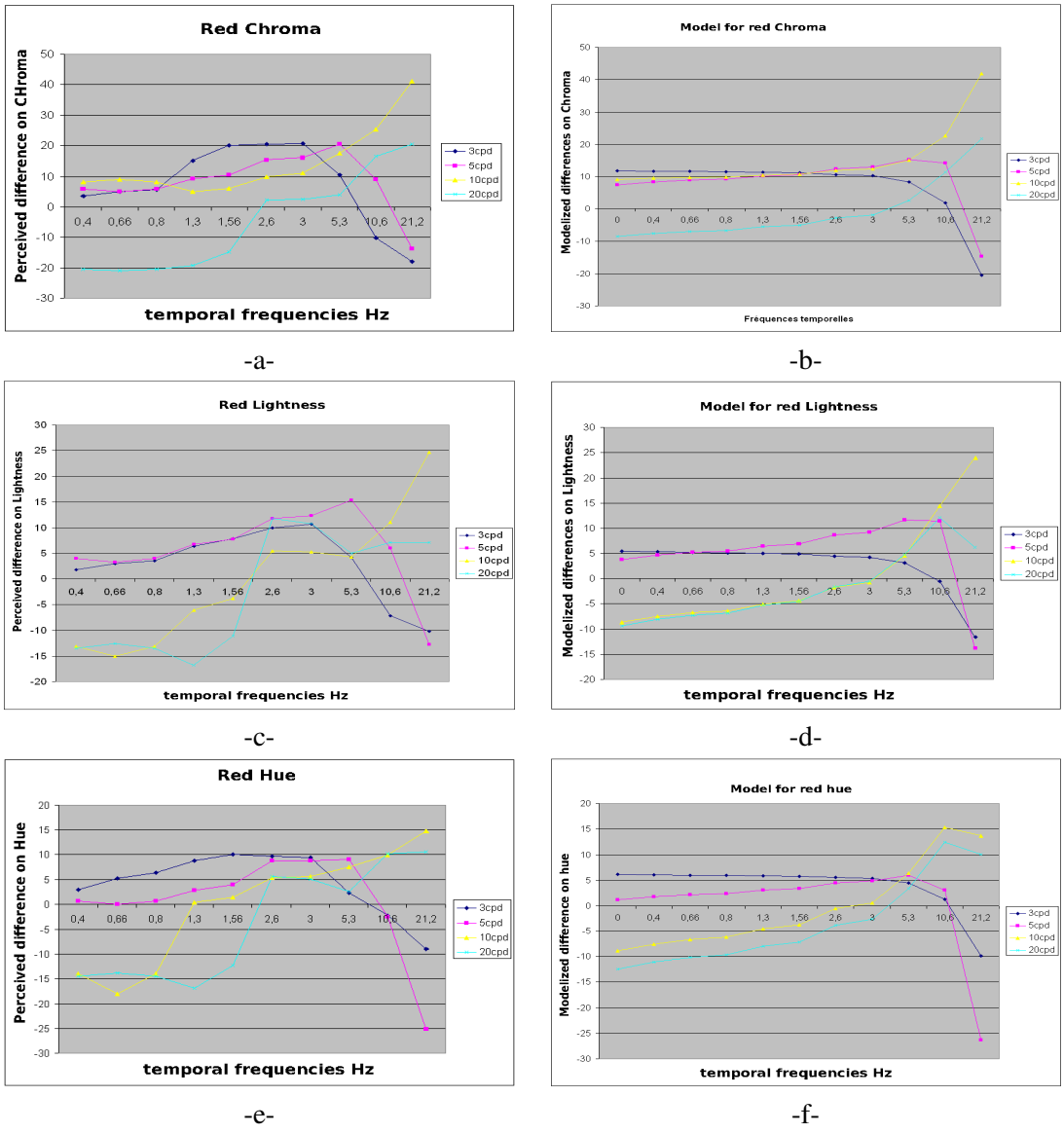


Figure 7: Obtained results and their associated models for the red primary.

4.1 Integration into CIECAM02

For modeling the results, curves of degree two were chosen and have been fitted using the least square method. Figures 7-b, 7-d and 7-f show the models respectively for the red chroma, lightness and hue. We can see that the models curves obtained represent a coherent fit of the obtained results. Once this step realized they were integrated in the CIECAM02.

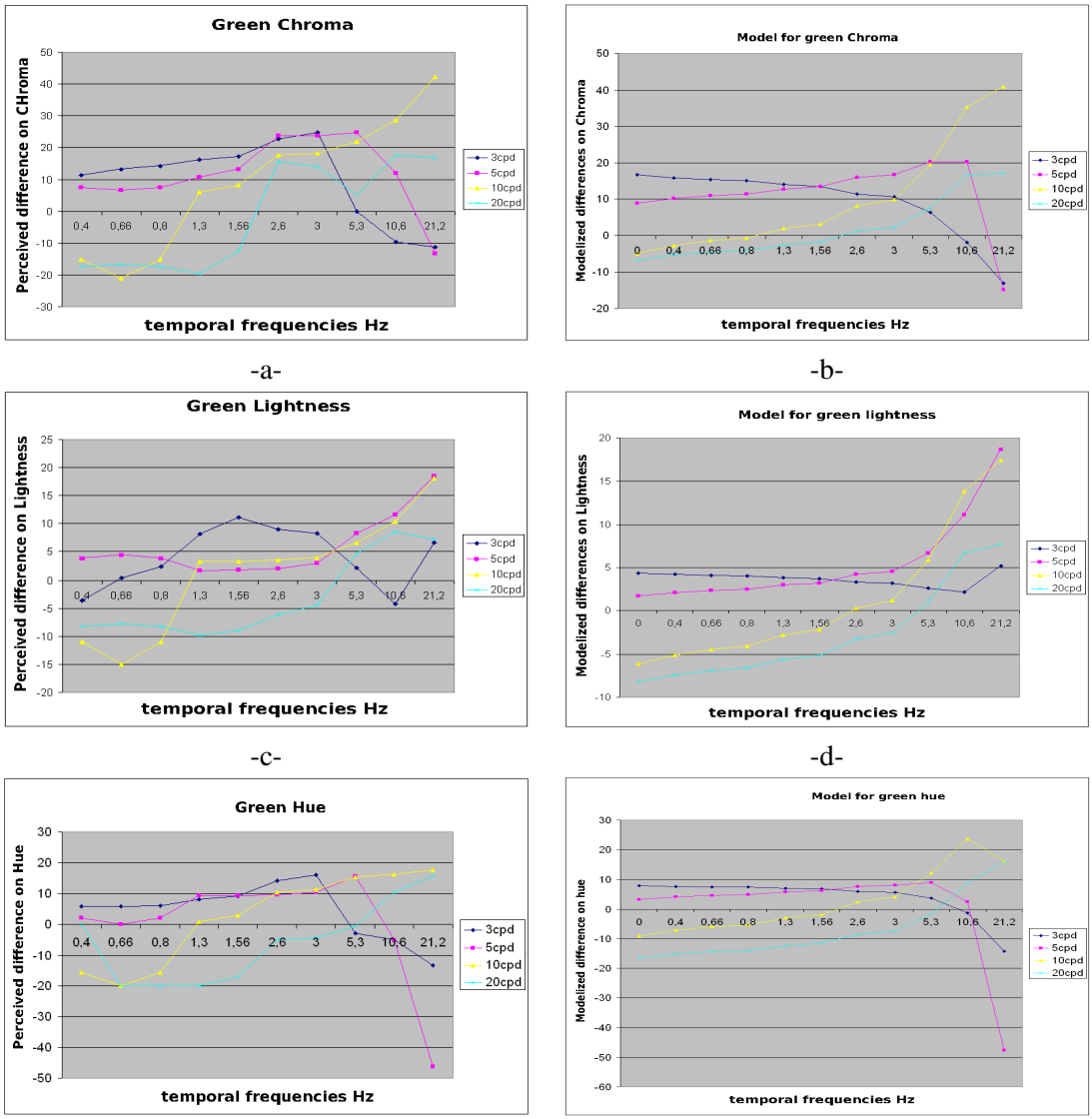


Figure 8: Obtained results and their associated models for the green primary.

The obtained model is named ST-CIECAM02 and is able to predict all effect predicted by the CIECAM02 as chromatic adaptation, simultaneous contrast, crispening, spreading...
and is described by figure 10.

4.2 ST-CIECAM Results

This new model is able to correct only the appearance of moving stimuli. Figure 11 shows an example of correction.

On this example (figure 11), one can see a reference stimulus on fig 11-a. The spatial modulation is increased and the appearance is corrected with help of ST-CIECAM. It works exactly like the S-CIECAM which is able to correct the spatial effect. Due to the medium, it is impossible to show temporal modulation, This is why it is difficult to show coherent temporal appearance correction. On figure 11-c, we can see a result of spatiotemporal correction of the first stimulus. Even if it looks very different from

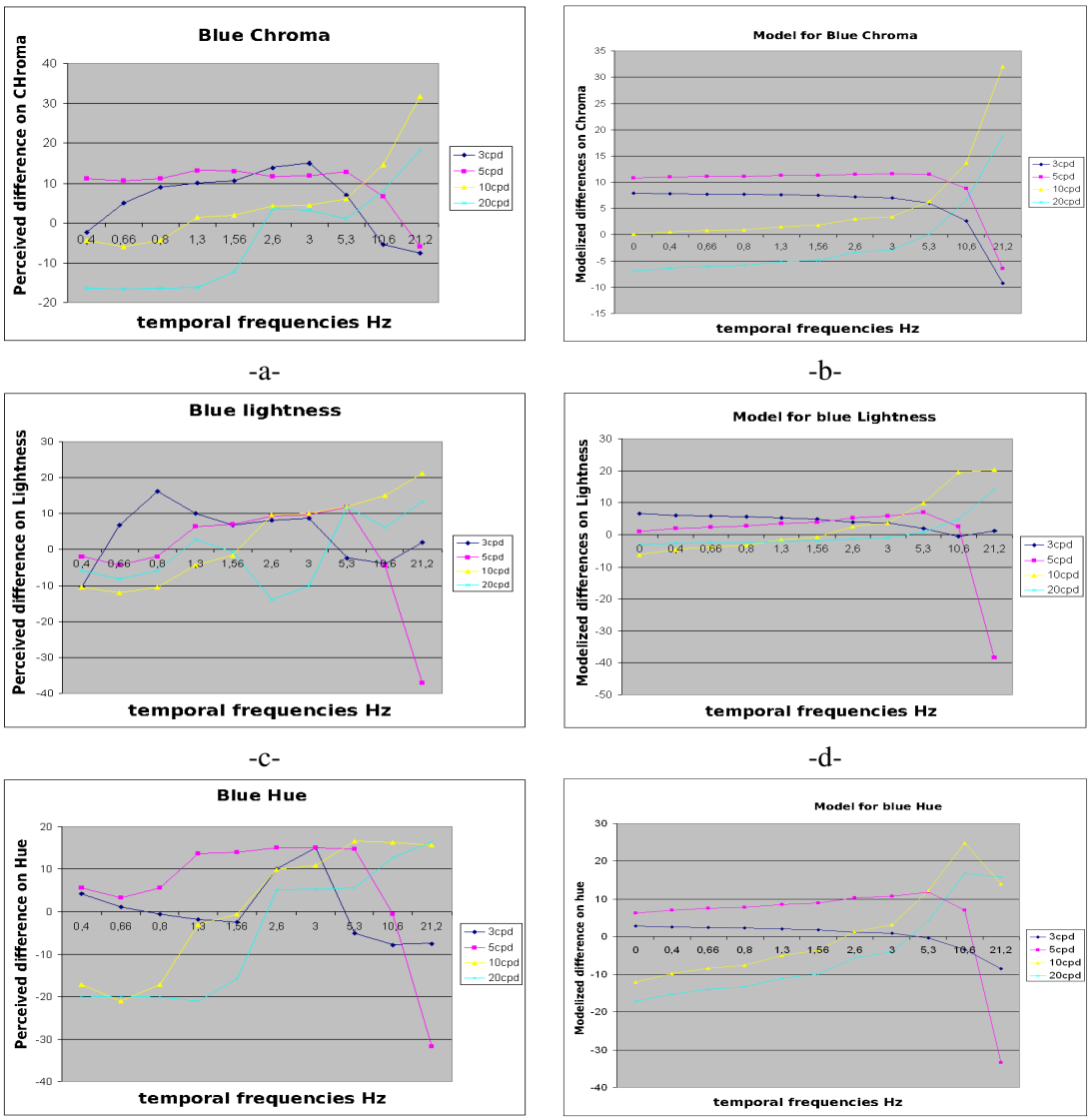


Figure 9: Obtained results and their associated models for the blue primary.

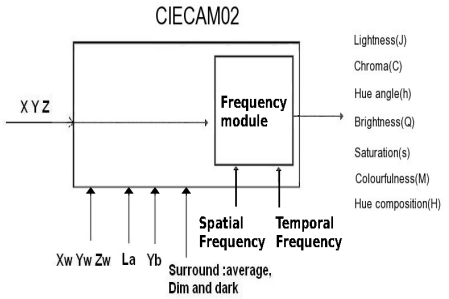
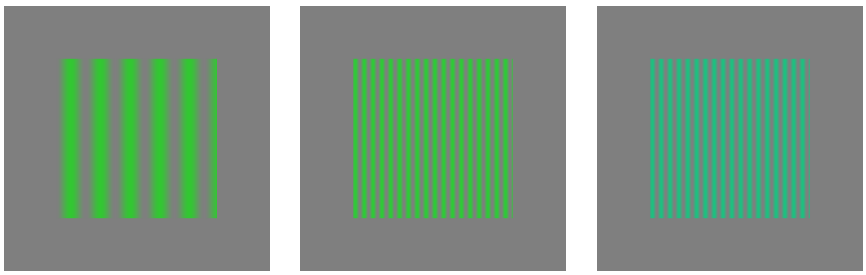


Figure 10: ST-CIECAM



-a- -b- -c-

Figure 11: Use of ST-CIECAM: (a) reference stimulus. (b) stimulus with an other spatial frequency but zero temporal frequency . The correction of ST-CIECAM was applied on this stimulus and it seems like (a). (c): stimulus with an other spatial frequency than (a) and with a high temporal frequency. The correction of ST-CIECAM was applied on this stimulus and it seems different from (A) but if moving it will look similar

the original when it is displayed as a still stimulus, the moving stimulus is closer to the original than the one of figue 11-b.

5 Conclusion

In this paper, we tried to measure the influence of spatiotemporal modulation on color appearance. The results given by this new model are encouraging. Indeed, the ST-CIECAM02 allows to correct the appearance of a dynamic stimulus which is difficult to show on this paper. But not all parameters which can have an influence of spatio-temporal frequencies have been tested like the environment or the background luminance. However this model is only able to correct stimuli. An extension of this model to video sequences is envisaged.

References

[1] Methodology for the subjective assessment of the quality of television pictures. Technical Report ITU-R Recommendation BT.500-10, Switzerland, 2000.

[2] M. D. Fairchild. *Color Appearance Models*. Addison-Wesley, Massachusetts, 1997.

[3] R.W.G. Hunt and M.R. Luo. The structure of the cie 1997 colour appearance model (ciecam97). *manuscript handed out at the AIC Conference*, 1997.

[4] ISO. Viewing conditions for graphic technology and photography. Technical Report ISO3664, ISO, 1999.

[5] G.M. Johnson. The quality of appearance. In *AIC*, Granada Spain, 2005.

[6] H. De Lange. Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light, i. attenuation chracteristics with white and colored light. *Journal of the Optical Society of America*, 48:777–784, 1958.

[7] Tulet O., Larabi M.-C., and Fernandez-Maloigne C. Use of spatial adaptation for image rendering based on an extension of the ciecam02. In *3rd International Conference on Computer Vision Theory and Applications*, January 2008.

[8] B.A. Wandell. *Foundations of vision*. Sinauer Associates, Massachussetts, 1995.

Perceptual Uniformity in Digital Imaging

Charles Poynton
Simon Fraser University
Vancouver, BC, Canada
charles@poynton.com

Abstract

Digital image coding is *perceptually uniform* if a small perturbation to a component value is approximately equally perceptible across the range of that value. Most digital image coding systems – including sRGB used in desktop graphics, and BT.709 used in HDTV – are perceptually uniform, but this fact is often shrouded in confusion. This document surveys perceptual uniformity in digital image coding at the turn of the millennium, and attempts to clarify some aspects of image coding that are widely misunderstood.

1 Luminance

Absolute luminance, defined by the CIE, is proportional to optical power across the visible wavelengths, weighted according to a standardized spectral weighting that approximates the spectral sensitivity of normal human vision. Luminance has units[1] of $\mathrm{cd} \cdot \mathrm{m}^{-2}$ ("nit," or nt); its symbol is L_v. The spectral weighting is denoted $V(\lambda)$ or $\bar{y}(\lambda)$.

The term *luminance* and its symbol Y are well established in colour science; however, the term and the symbol are widely misused in the fields of video, computer graphics, and digital image processing. Workers in those fields commonly use the term "luminance" – or worse, "luminosity" – to refer to a weighted sum of nonlinear (gamma corrected) red, green, and blue tristimulus signals instead of the linear-light quantities defined by the CIE [3]. The nonlinear quantity is properly termed *luma* and given the symbol Y' [8].

A set of three signals proportional to intensity, and having specific spectral weighting, are called tristimulus values. They are pure numbers with no units [2]. *RGB*, *LMS*, and *XYZ* are all examples of tristimuli. A suitably-weighted sum of tristimuli yields luminance [4].

Tristimulus values and luminance are what I call *linear-light* measures, directly proportional to light power. Cameras typically depart from the spectral sensitivities prescribed by CIE standards, so tristimulus values and luminance in video are usually estimated, not exact. Instead of using my informal term linear-light, some practitioners use the term *photometrically linear.* The adjective *photometric* properly refers to use of the CIE standard luminance spectral weighting. Practical cameras don't closely approximate the CIE spectral weighting, so the term photometrically linear shouldn't be used to describe them.

In image capture – including photography, cinema, video, HDTV, digital cinema, and graphics arts – we are rarely, if ever, concerned with the absolute luminance of the original scene. Instead, we characterize scene luminance relative to an "adopted" scene white luminance associated with the state of visual adaptation of an actual or hypothetical person viewing the scene. Subsequent processing and display involves *relative luminance,* whose symbol is *Y*, and whose value according to CIE conventions is a pure number ranging 0 through 100. (Some practitioners, including me, prefer a range from 0 to 1.) Image scientists and engineers ordinarily call this quantity *luminance,* even though properly speaking it is relative luminance.

[1] The foot-lambert unit [fL] once used for luminance is now deprecated. I use the SI unit, $\mathrm{cd} \cdot \mathrm{m}^{-2}$ ("nit," or nt). In my view, using foot-based units such as foot-Lambert [fL] and foot-candle [fc] impedes the understanding of radiometry and photometry.

2 Introduction to Perceptual Uniformity

I introduce perceptual uniformity in Chapter 1 of my book [9]. Put briefly:

> Vision cannot distinguish two luminance levels if the ratio between them is less than about 1.01 – in other words, the visual threshold for luminance difference is about 1 percent.

The 1% value that I mention is the Weber contrast. Image coding whereby a constant ratio is maintained from code to code across the tone range from some minimum representable luminance up to white is effected by a logarithmic transform. Log transforms are rare in practical image coding.

For a true logarithmic law having a 1.01-ratio between adjacent codes, the relative luminance difference between codes is 1% across the whole range. There are 463 codes between relative luminance of 0.01 and 1 – that is, 463 codes cover a contrast ratio of 100:1. A photographer or cinematographer is interested in how many codes cover each "stop" (factor of two) of luminance. For pure logarithmic coding with a Weber contrast of 1%, there are 232 codes per decade, equivalent to 69 codes per stop – six bits of data per stop.

An estimate of vision's lightness response, denoted L^*, was standardized by the CIE in 1976 [3]: Given relative luminance, CIE L^* returns a value between 0 and 100; a "delta" (difference) of 1 lies approximately at the threshold of vision. The L^* function is basically a power function with what I call an "advertised" exponent of $^1/_3$ – that is, a cube root. The technical literature is rife with statements that L^* is a cube root. However, a linear segment is inserted near black, below relative luminance of about 1%. The power function segment is scaled and offset to maintain function and tangent continuity at the breakpoint. The scaling and offset cause the function to approximate an "effective" 0.42-power over its entire range. See Figure 1.

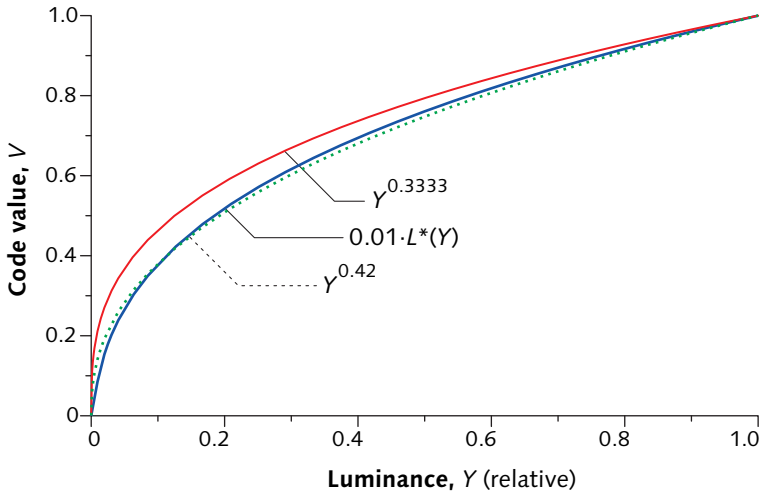


Figure 1: **CIE Lightness,** denoted L^*, estimates the perceptual response to light intensity (technically, relative luminance). Here L^*, scaled to the range 0 … 1, is overlaid by power function having an exponent 0.42, the exponent that best fits L^*. The L^* function involves a cube root – that is, a $^1/_3$-power function – but L^*'s power function is scaled and offset. I also overlay a cube root onto the plot: A pure cube root is a poor approximation to L^*.

In capturing, processing, storing, and transmitting image data, a limited number of bits are most effectively used by perception if coding of luminance values (or tristimulus values) is nonlinearly mapped,

like L^*, to mimic the lightness response of human vision. Mappings based upon power functions are most common, though mappings based upon logarithms are sometimes used.

In nearly all commercial imaging systems, an optoelectronic conversion function (OECF) – or loosely, "gamma correction" – is imposed at encoding. Gamma correction takes R, G, and B (linear) tristimulus estimates, and forms (nonlinear) R', G', and B'. The primes signify the nonlinear relationship to light power. To achieve perceptual uniformity, the OECF roughly approximates vision's lightness sensitivity (e.g., L^*). Decoding and display of digital image data involves an electro-optical conversion function (EOCF) that approximates the inverse of lightness sensitivity.

In a CRT display, the electrostatic characteristics of the electron gun cause the CRT to impose an EOCF that is approximately a 2.4-power function from voltage input to light output. The symbol γ (gamma) represents the exponent at the display: A studio reference display is said to have gamma of about 2.4. In non-CRT display devices, signal processing provides an equivalent nonlinear function. A 2.4 power is a near-perfect match to the inverse of the L^* function; see Figure 2.

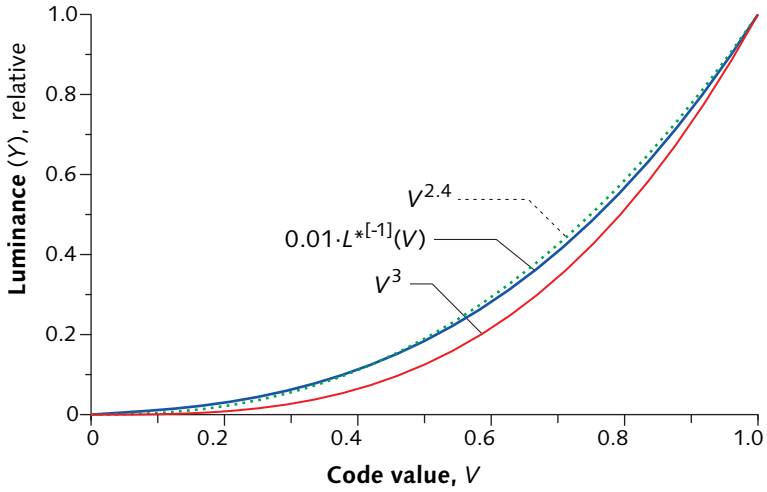


Figure 2: **EOTF of a typical CRT** is approximated by a 2.4-power function from video signal in to light power out. The gamma of a display system – for example, a CRT, or the reference sRGB EOCF – is the numerical value of the exponent of the power function. I overlay the inverse of the CIE L^* function: It is evident that a 2.4-power function is a very close match to the inverse of L^*. I also overlay a 3.0-power function; clearly, a cube function is a rather poor match to the inverse of L^*.

It is frequently claimed that 8-bit imaging has a "dynamic range" of 255:1 or 256:1. Such claims arise from the assumption that image data codes are linearly related to light. However, nearly all 8-bit image data is coded perceptually, like sRGB, assuming a 2.2- or 2.4-power function at the display: The dynamic range associated with code 1 is close to a million to one, not just $^1/_{255}$. A related claim [6] is that 8-bit imaging has an optical density range of about 2.4, where 2.4 is the base-10 log of $^1/_{255}$. This claim similarly rests upon the assumption of linear-light coding – an assumption which, for 8-bit coding, is nearly always false.

Figure 3 plots L^* as a function of code value for linear-light coding, a 1.8-power coding typical of graphics arts, and pure power functions having exponents of 2.2 (sRGB), 2.4 (studio video), and 2.6 (digital cinema, to be discussed). EOCF power function exponents of 2.2, 2.4, and 2.6 are all quite perceptually uniform.

As I mentioned earlier, ΔL^* of unity is widely agreed to approximate the threshold of vision. The ratio of luminance between L^* values of 99 and 100 is about 1.025 – that is, the relative luminance difference

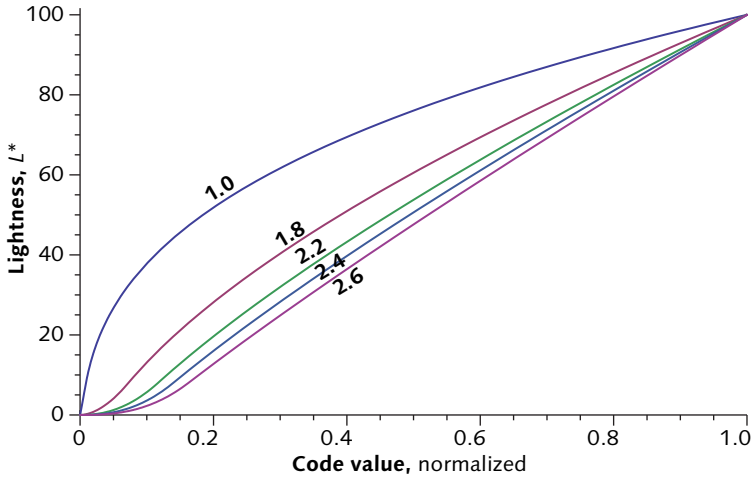


Figure 3: **Various pure power function EOCFs**

at threshold is 2.5% (the Weber contrast). The difference increases as relative luminance decreases; see Figure 4. At relative luminance of 0.01, L^* is about 8, and the relative luminance difference at threshold has reached 12.5%. The L^* scale assigns 92 levels – or 93, including the endpoints – across a 100:1 range of luminance. Seven bits suffice.

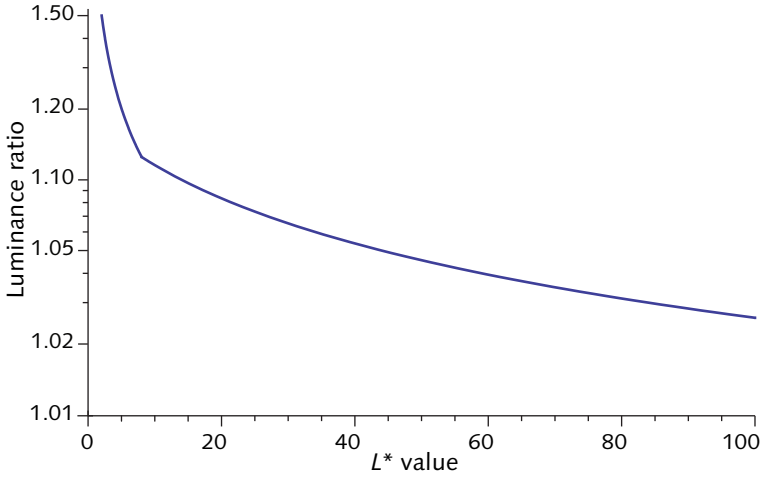


Figure 4: **Ratio of relative luminance values** for unit ΔL^*, across the L^* range from 1 to 100. Starting at the right, between L^* values 99 and 100, there is a 2.5% difference between relative luminance values at the assumed threshold of unity ΔL^*. As L^* decreases, the "delta" increases. At L^* of 8 – corresponding to relative luminance of about 1%, or contrast ratio of 100:1 – the difference has increased to 12.5%.

Coding L^* values produces considerably larger luminance ratios than logarithmic coding with a Weber contrast of 1%. Digital studio video has 219 steps over a comparable contrast ratio; sRGB has 255. These numbers are intermediate between the 463 codes of pure log coding (at a Weber fraction of 1.01) and the 92 codes of L^* coding. In Photoshop LAB coding, and in the LAB PCS of the ICC standard [5], L^*'s range of 0 to 100 is coded digitally into the range 0 through 255: The coding has about 2.5 digital code

values per L^* unit – that is, a Weber contrast of about 1% at white.

I have been discussing the number of codes across 100:1 contrast ratio, or two decades of luminance. A particular imaging application may require a range less than or greater than 100:1. Also, typical photographic images have a certain amount of noise; visibility of contouring will be reduced by noise, and quantization will be less demanding.

If an imaging application is required to maintain relative luminance values from an encoder to a decoder, then the OECF (at encoding) should be chosen as the mathematical inverse of the EOCF that will be imposed at decoding and display. For the near-ideal 2.4 power used in studio video display, you would expect the encoder to have as its exponent the reciprocal of 2.4 – that is, 0.42. (For an example of perceptually uniform decoding in a different domain, medical imaging, see the DICOM standard [1].)

3 Picture Rendering

All imaging applications involve non-ideal displays, and almost all applications involve image viewing in conditions different from those in effect at the time of image capture. In most applications the goal is to not to match relative luminance values between the scene and the display, but to match the appearance of the scene. Engineers and scientists unfamiliar with colour science are usually surprised to learn that the intended appearance is not achieved by matching relative luminances between scene and display: Preserving appearance almost always requires manipulating the tristimulus values between the scene and display.

In many commercial imaging systems, including video and digital still photography, the intended appearance is often obtained by using an OECF that approximates a 0.5-power function, rather than the 0.42 that would perfectly invert a 2.4-power at decoding and display.

Perceptual encoding for distribution is performed in virtually all commercial image systems. In applications where image data is manipulated for creative purposes between capture and display – for example, in graphics arts, or in video post-production – perceptual uniformity is imposed at capture to the extent required for the image manipulation. Suppose that processing requires a linear-light gain of 4 to overcome a two-stop underexposure owing to poor lighting or incorrect aperture setting. Capture must then have quantization four times finer than the quantization required at the display. Where 8-bit $R'G'B'$ components might suffice for distribution of consumer video or commodity JPEG imagery, to enable manipulation in post-production, 10 bit $R'G'B'$ components might be required at capture. Video is typically processed in the camera to produce perceptually uniform signals; the recorded image data is quite close to the required final product, and not much processing headroom is needed. However, digital cinema capture typically involves downstream processing for creative purposes; more severe constraints are thereby placed on perceptual uniformity. Put simply, more bits per component are required.

4 Modern Misconceptions

Astonishingly, since about 1960 to the present, the significance of perceptual uniformity has been largely forgotten! Engineers, always desirous of linearity, apparently came to believe that gamma correction was necessary to overcome a supposed deficiency – that is, nonlinearity of the CRT. They realized that the sensible place to perform the ”correction” was close to the transmitter, so as to avoid millions of nonlinear circuits in receivers; however, the link to perceptual uniformity was forgotten. Widespread misunderstanding among television engineers of the fundamental reason for ”gamma correction” remains rampant even today. As I stated on page 258 of my book [9]:

> If gamma correction were not already necessary for physical reasons at the CRT, we would have to invent it for perceptual reasons.

You can test your colleagues: Ask, "If television displays in 1953 had exhibited a linear relationship between applied voltage and light output, would television standards have included gamma correction?" Anyone who answers "Of course not!" does not, in my view, appreciate the importance of perceptual uniformity.

Electrical engineers, video engineers, and digital image processing practitioners often claim that their systems are "linear." However, if gamma correction has been imposed at image capture or encoding, and an approximate inverse is imposed at decoding or display, then linearity in the R', G', and B' signal domain does not extend to luminance or tristimulus values! In other words, you can treat calculations in the tristimulus domain as linear, and you can treat calculations in the $R'G'B'$ (video signal, code, or voltage) domain as linear, but values in one domain are clearly not proportional to values in the other.

In my paper "The rehabilitation of gamma" [7], I reviewed several widely-held misconceptions concerning gamma, including these:

- The nonlinearity of a CRT display is a defect that needs to be corrected.
- The main purpose of gamma correction is to compensate the nonlinearity of the CRT.
- Ideally, linear-intensity representations should be used to represent image data.

My paper then presents what I consider to be the facts of the situation:

- The nonlinearity of a CRT is very nearly the inverse of the lightness sensitivity of human vision. The nonlinearity causes a CRT's response to be roughly perceptually uniform. Far from being a defect, this feature is highly desirable.
- The main purpose of gamma correction in video, desktop graphics, prepress, JPEG, and MPEG is to code luminance or tristimulus estimates (proportional to intensity) into a perceptually-uniform domain, so as optimize perceptual performance of a limited number of bits in each of the RGB components.
- If a quantity proportional to intensity represents image data, then 12 bits or more would be necessary in each component to achieve high-quality image reproduction. With nonlinear (gamma-corrected) coding, just 8 bits are sufficient.

In my 1998 paper, I referred to 8 bits per component being sufficient for video distribution purposes. In order to provide some measure of protection against roundoff error liable to be introduced by video processing, today's studio video standards – and most studio equipment – have 10 bits per component. CCD and CMOS sensors used in cameras are intrinsically linear-light devices; it is necessary to capture at least 12 bits per component to maintain 10-bit accuracy once the signals are gamma-corrected [10]. Several digital cinema cameras offer 14 bit linear-light components, and thereby offer about 12 bits of quantization performance when coded perceptually (for example, by the $XYZ^{1/2.6}$ function specified in SMPTE/DCI standards for digital cinema). Roughly speaking, representing colour components in a perceptually uniform manner saves 2, 3, or 4 bits per component compared to representation in linear-light form.

5 Modern Practice

Today's studio reference displays have gamma very close to 2.4, reference white luminance of between 80 and 120 $\mathrm{cd} \cdot \mathrm{m}^{-2}$, and a contrast ratio of about 250:1. They are viewed with a dim surround, illuminated such that the surround luminance is about 5% of the reference white luminance.

Creative approval of program material in the studio environment causes not only the studio EOCF but also the studio viewing conditions to be implicit in the definition of the $R'G'B'$ exchange standard: It is implicit that the intended picture appearance at the consumers' premises is obtained from a comparable EOCF in a comparable environment. Should the consumer's display characteristics or viewing conditions differ substantially from the studio – for example, if the consumer display is brighter, or has inferior contrast ratio, or is located in a lighter or darker surround than the studio – then image data should be altered at the consumer's premises to yield a closer match to the intended appearance.

CRTs are now essentially obsolete, and several display technologies such as LCD, PDP, DLP, and LCoS are vying to replace them. None of these technologies involves a physical 2.4-power law like that of a CRT. Some people argue that emergent display technology gives us a chance to adopt linear-light encoding; however, perceptual uniformity remains important for these reasons:

- Perceptually uniform coding maximizes the perceptual utility of a limited number of bits – usually 8, or 10, or 12 – per component;
- Nearly all commercially important digital image storage and exchange standards call for perceptual uniformity; and
- Billions of stored images incorporate perceptual uniformity.

Emergent, non-CRT display devices incorporate signal processing circuits that apply a transfer function to impose the difference between the device's native, physical response and the behaviour required to mimic the electro-optical conversion function (EOCF) implicit or explicit in exchange standards. DLP displays and PDP displays both have physical linear-light response; display systems incorporating these displays incorporate a power function, or a function approximating one, to convert $R'G'B'$ signals (presented at the interface) to linear-light RGB that modulates the display itself.

6 Perceptual Uniformity in D-Cinema

Standards for digital cinema distribution such as SMPTE 431-1 [10] and SMPTE 431-2 [11] call for $R'G'B'$ or $X'Y'Z'$ components (at the reference projector interface, or the digital cinema distribution interface, respectively) to be raised to the power 2.6 for display. The 2.6-power is imposed to invert perceptually uniform encoding. Compared to the 2.4-power OECF of studio video, the 2.6-power offers improved visual performance in the low luminance and dark surround situation of the cinema.

There are no SMPTE/DCI standards for digital cinema acquisition; many techniques are in use. The basic principles that I have outlined outlined apply when the cinematographer decides, based upon the scene being captured, upon a diffuse white reference near the top end of the digital coding scale. If specular highlights beyond diffuse white are to be accommodated, then the cinematographer may impose what an engineer might call a distortion of the code scale above diffuse white. The cinematographer may have reason to acquire a scene while deferring any decision about reference white – that is, the decision may be deferred until post-production. In that case there is an argument to have an acquisition standard that uses a pure logarithmic code, or a pseudolog code with an appropriate number of digital code values per stop of scene-space luminance ("exposure"); for example, see SMPTE RDD 2 [12].)

7 Conclusion

Perceptual uniformity is a tremendously important aspect of digital image coding, particularly video, HDTV, digital cinema, and digital still photography. Without it, we would need 11, 12, or 13 bits per component, instead of 8 or 10. Perceptual uniformity was appreciated half a century ago, yet is either poorly understood or not recognized at all by a surprisingly large number of image scientists and engineers working today.

References

[1] ACR/NEMA PS 3.14 (2006), *Digital Imaging and Communications in Medicine (DICOM), Part 14: Grayscale Standard Display Function*.

[2] BRILL, MICHAEL H. (1996), "Do tristimulus values have units?", in *Color Research and Application* **21** (4): 310 – 313 (Aug.).

[3] CIE 15 (2004), *Colorimetry* (Vienna, Austria: Commission Internationale de L'Éclairage).

[4] HUNT, ROBERT W. G. (1997), "The Heights of the CIE Colour-Matching Functions," in *Color Research and Application* **22** (5): 337 – 313 (Oct.).

[5] ISO 15076 (2005), *Image technology colour management – Architecture, profile format and data structure – Part 1: Based on ICC.1:2004-10.*

[6] KIM, MIN H. and MACDONALD, LINDSAY W. (2006), "Rendering High Dynamic Range Images," in *Proc. 17th Annual EVA Conference* (EVA 2006): 22.1 – 22.11 (July).

[7] POYNTON, CHARLES (1998), "The rehabilitation of gamma," in ROGOWITZ, B. E. and PAPPAS, T. N., eds., *Human Vision and Electronic Imaging III, Proc. SPIE/IS&T Conf. 3299*: 232 – 249 (San Jose, Calif., Jan. 26 – 30).

[8] POYNTON, CHARLES (1999), "YUV and luminance considered harmful," available at www.poynton.com.

[9] POYNTON, CHARLES (2003), *Digital video and HDTV algorithms and interfaces* (San Francisco: Morgan Kaufmann).

[10] SMPTE 431-1 (2006), *D-Cinema Quality – Screen Luminance Level, Chromaticity and Uniformity.*

[11] SMPTE 431-2 (2007), *Reference Projector and Environment for Display of DCDM in Review Rooms and Theaters.*

[12] SMPTE RDD-2 (2007), *Use of Logarithmic Non-Linear Transfer Characteristic for Transmission of Linear Signals through Limited-Bit-Depth Image Representations and Interfaces.*

Colour and Lightness Perception in Low and High Dynamic Range Scenes

Alessandro Rizzi
University of Milano
Milano, Italy
alessandro.rizzi@unimi.it

Aditya Sole
Gjøvik University College
Gjøvik, Norway
adityas@hig.no

Peter Nussbaum
Gjøvik University College
Gjøvik, Norway
peter.nussbaum@hig.no

Abstract

This paper presents a simple experiment aiming at checking how our vision system deals with big changes in illumination intensity and geometry. Two identical scenes created from painted wooden blocks have been set in a big dark room. Observers have been asked to compare the appearance of the faces of the painted wooden blocks. Data confirms that HVS does not extract or estimate any reflectance, nor discount the illuminant.

1 Introduction

Studying, estimating and assessing the appearance of color in context is a relatively recent and very interesting research field. The attempt is to overcome the limits of classic colorimetry that is expressly designed to work in "aperture mode", that is color isolated from any possible context. However, in everyday situations it is well-known that context plays a big role in changing color appearance [3].

Several models attempted to deal with appearance so far, trying to simplify the complexity of the context description [2]. Context complexity is not only the geometrical arrangement of the scene or the foreground-background segregation problem. Lighting geometry plays a very important role. This paper presents a simple experiment aiming at checking how the illumination geometry can affect color and lightness appearance. It is the remake of a very similar previous experiment [4] made during the CREATE FP6 European Project [1].

2 Experiment Setup

The experiment was carried out using two identical scenes created from painted wooden blocks. Both the scenes were setup in a big dark room. Pair comparison technique was used in which the observers were asked to compare the appearance of a pair of faces of the painted wooden blocks at a time.
Figure 3 shows a pencil sketch of the scene with the face numbers. There were a total of 35 colored faces in the scene painted with 11 different acrylic paints. A color wheel was created using the 11 colors: red, green, blue, cyan, magenta, yellow white, black, light grey, mid grey and dark grey. Only these 11 colors were used to paint the faces of the wooden blocks.

Scene 1 was illuminated with a tungsten bulb of 150 watts from the right side and a monochromatic green covered tungsten bulb from the left side. At the back of scene 1 a black chamber, in which a color wheel is posed, it was built. The aim of this black chamber was to create a high dynamic range lighting situation (see Fig. 1), but this has not been used in the presented experiment. The uneven distribution of illumination is the main tested characteristic.
To present a strong difference from the illumination point of view, scene 2 was illuminated by a set of uniform diffuse photographic tubes approximately at 45◦ from the centre of the scene from both sides (see Fig. 2) and no back dark chamber, obtaining in this way a much more uniform and diffused illumination.

(ed.): easychair 1.0, 2008, volume 1, issue: 1, pp. 1-7

The experiment was performed in a wide dark ambient room. A total of 14 observers participated in this experiment, mostly young students. Out of the 14 observers, 10 were male and 4 female. No colorblind test was conducted on the observers, however, no one declared any color deficiency.

The experiment consisted of 3 tasks to be completed by the observers. Tasks 1a, 1b were carried out on scene 1 while for Task 2 scene 2. The observers were provided with an instruction manual. It consisted of a pencil diagram of the experiment scene setup with the face numbers on it as in Figure 3. This diagram was used to identify the faces in the scene and record the observations. This pencil diagram was used for all the tasks in this experiment.

Figure 1: Scene 1

Figure 2: Scene 2

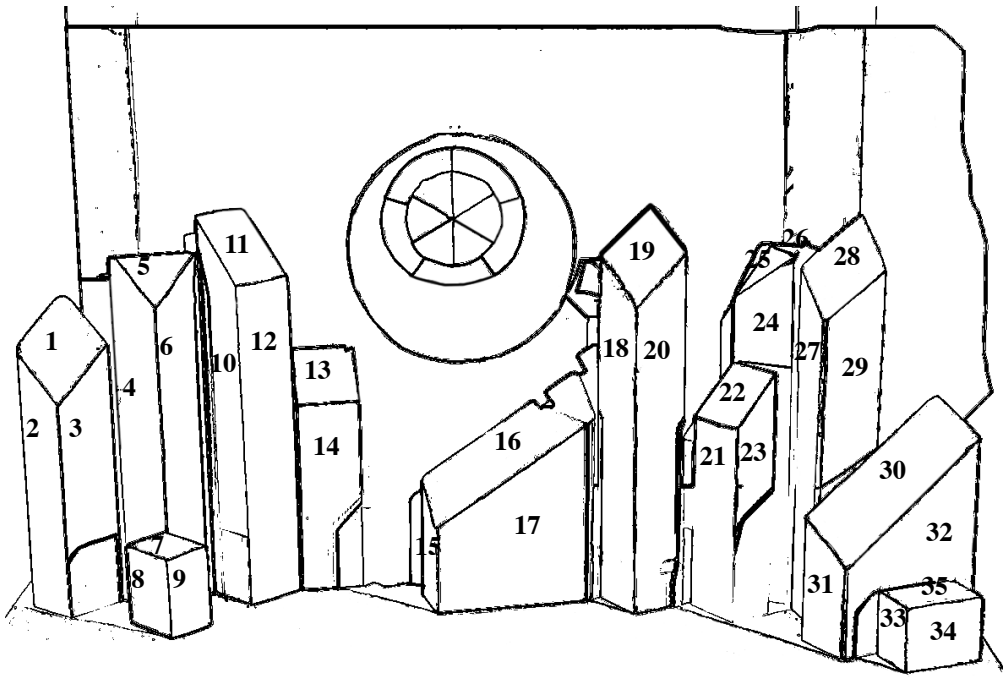


Figure 3: Figure A

3 Task 1a

3.1 Experiment

In task 1a, the observers were asked to arrange the given achromatic faces in an ascending order from lightest face to darkest face. A total of 9 faces were to be ranked. Illuminance on these faces was measured in lux using a telespectro-radiometer Minolta CS1000 on scene 2 (refer Fig. 3) to obtain the actual illuminance values under the reference white lighting conditions. (second line of Table 1). The observers were asked to rank the faces using scene 1 (refer Fig. 1). Scene 1 was illuminated with a tungsten bulb of 150 watts from the right side and a monochromatic green covered tungsten bulb from the left side.

3.2 Observation

Table 1a, below shows the face numbers in the first row (in ascending order of its illuminance), its illuminance in lux in the second row and the readings given by the observers in the remaining rows.

3.3 Result

Figure 4 shows a graph plot of the readings obtained from the observers in task 1a. X-axis plots the faces in an ascending order in terms of their illuminance values obtained from the TSR. First face in this graph plot is Face 4, which shows an illuminance of 4.54 lux while the last face is Face 9, which shows an illuminance of 340 lux (refer Table 1). Y-axis plots the observer readings given by the observers from 1 to 9

Table 1: Observer readings for Task 1a.

Observer	4	34	20	11	7	14	17	3	9
TSR(lux)	4.54	64.4	69	87.5	139	178	253.3	298.5	340
C	9	8	7	5	6	4	3	2	1
J	9	7	6	5	8	4	3	2	1
Ch	9	7	6	5	8	4	3	1	2
G	3	8	7	6	9	5	4	2	1
H	9	8	6	5	7	4	3	2	1
F	9	8	7	5	6	4	3	1	2
K	9	8	7	5	6	4	3	2	1
W	9	8	7	4	6	5	3	2	1
R	9	8	7	6	5	4	3	2	1
T	9	8	7	6	4	5	3	1	2
Me	9	4	8	6	5	7	3	2	1
A	9	8	7	6	4	5	3	2	1
Th	9	7	6	5	8	4	3	1	2
Ma	9	8	7	5	6	4	1	2	3

to arrange the faces in descending order from dark to light with darkest face as 9 and the lightest face as 1.

It is observed from the graph that only one observer (R) readings match exactly to those measured with the TSR, but the overall trend is not too irregular. 5 observers out of 14 have observed Face 7 (139 lux) darker than Face 11 (87.5 lux) while 1 observer has observed Face 7 (139 lux) darker than Face 34 (64.4 lux).

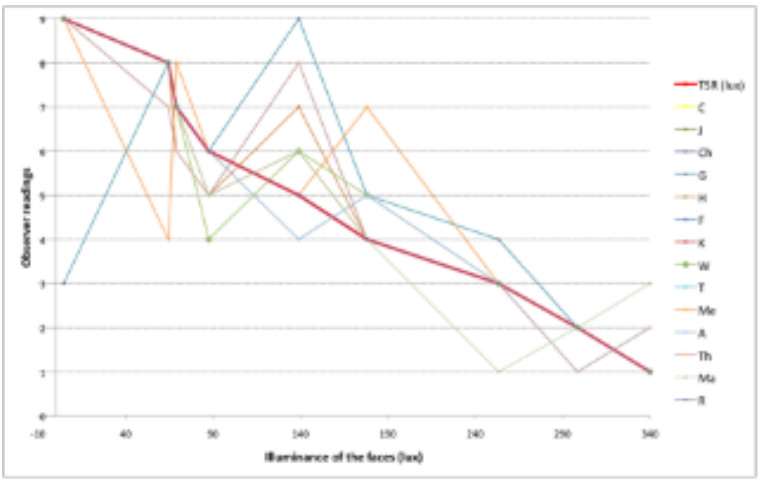


Figure 4: Task 1a Observer readings

4 Task 1b

4.1 Experiment

In Task 1b, the observers were given pairs of faces and were asked to compare the faces given in the pairs with each other and assign the number according to the following categories:

1. No match at all
2. Very poor match

3. Poor match
4. Close match
5. Very close match
6. Exact match

[For example in a pair say $1 - 2$, if the color of these faces appear exactly the same you will write the number '6' i.e. Exact match between the pair].

4.2 Observation

Table 2 shows the pairs of faces and the observer readings on the scale of 1 to 6 given by the observers.

Table 2: Observer readings for Task 1b.

Observer	06-10	10-12	01-02	01-07	16-23	22-24	22-29	08-15	13-28	32-33	09-17	17-11	01-20	04-34
C	2	2	3	4	3	5	5	3	3	2	5	3	3	2
J	2	2	2	4	3	5	6	4	4	3	4	3	4	1
Ch	2	2	1	2	3	4	4	3	4	3	4	3	2	1
G	2	2	1	4	3	6	6	2	2	3	5	1	4	2
H	2	2	1	3	4	5	5	1	4	3	5	1	2	1
F	3	3	2	4	3	6	6	5	2	4	5	2	2	1
K	3	3	1	5	4	5	5	3	1	2	5	1	1	1
W	2	2	1	4	2	6	5	1	3	2	5	2	1	1
R	4	5	2	3	4	5	5	3	4	3	4	4	2	1
T	3	3	2	4	3	5	5	3	1	3	5	2	2	1
Me	2	2	1	2	1	5	5	2	1	2	5	1	1	1
A	2	2	3	4	2	6	5	4	2	2	4	1	2	3
Th	4	4	3	5	4	6	6	4	4	5	5	5	4	3
Ma	4	4	2	5	2	5	5	1	3	3	6	2	2	1

5 Task 2

5.1 Experiment

Task 2 was exactly the same as task 1b but was performed on scene 2 (the scene illuminated with diffuse reference lighting conditions). Same pairs and categories (as used in task 1b) were used in this task.

5.2 Observation

Table 3 below shows the pairs of faces and the observer readings on the scale of 1 to 6 given by the observers.

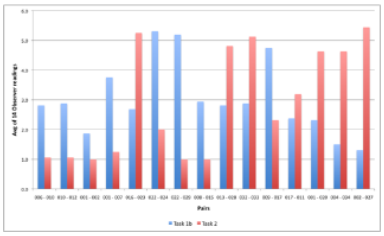


Figure 5: Average observer readings for task 1b and task 2

Table 3: Observer readings for Task 2.

Observer	06-10	10-12	01-02	01-07	16-23	22-24	22-29	08-15	13-28	32-33	09-17	17-11	01-20	04-34
C	1	1	1	3	5	2	1	1	5	6	4	4	5	4
J	2	2	1	1	5	2	1	1	5	5	3	4	5	5
Ch	1	1	1	1	4	1	1	1	3	4	2	2	1	3
G	1	1	1	1	5	1	1	1	5	6	1	1	5	5
H	1	1	1	1	5	2	1	1	5	5	3	3	4	5
F	1	1	1	1	6	3	1	1	4	6	3	2	5	6
K	1	1	1	1	6	1	1	1	5	1	1	1	4	3
W	1	1	1	1	6	1	1	1	5	1	1	1	4	3
R	1	1	1	2	5	3	1	1	4	6	5	5	5	5
T	1	1	1	2	5	3	1	1	4	6	5	5	5	5
Me	1	1	1	1	6	2	1	1	5	6	1	3	5	4
A	1	1	1	1	4	2	1	1	6	6	2	4	6	6
Th	1	1	1	2	5	4	1	1	5	1	3	4	5	1
Ma	1	1	1	1	6	1	1	1	5	6	1	4	5	5

Figure 5 shows the graphical representation (bar graphs) to compare the average observer readings of task 1b and task 2.

6 Discussion

Task 1b and task 2 were very identical tasks given to the observers with the same number and pairs of faces to compare by the same method (pair comparison). The only difference between the two scenes was the illumination. In task 1b, scene 1 was used ((refer Fig. 1)). Scene 1 was illuminated in a very uneven way with uneven intensity, shadowed and with inter-reflections. Task 2 was performed using scene 2. Scene 2 was illuminated with a very uniform and diffused light setup (refer Fig. 2).

These tests aim at checking to what extent our vision system adjusts to the scene lighting. Some works about color and lightness constancy of humans assume that our vision system extract object reflectance or alternatively discount the illuminant component.

Data from this test confirm the one obtained in [4], proving that HVS does not extract or estimate any reflectance, nor discount the illuminant. As you can see from figure 5 pair comparison score completely changed from scene1 to scene 2.

Appearance of each surface comes from the interaction among surfaces and illumination and in this process the geometry and the non-uniformity of the light can strongly modify the final appearance. Also the reciprocal disposition in the scene can affect the final appearance.

A CAM or a model of human visual perception should try to deal as much as possible with the complexity of illumination geometry and of the scene spatial arrangement.

References

[1] Create (colour research for european advanced technology employment. MSCF-CT-2006-045963, FP6 European Community Marie Curie Conferences and Training Courses Project.

[2] *Color Appearance Models*. John Wiley & Sons, 2005.

[3] J. Albers. *Interaction of Color*. Yale University Press, 1971.

[4] C. Parraman, A. Rizzi, and J.J. McCann. Measuring hdr rendering for colour appearance: an experiement. In *Electronic Imaging*, San Jose, California, USA, January 2009.