Seeing Beyond Luminance: A Psychophysical Comparison of Techniques for Converting Colour Images to Greyscale

David Connah[1], Graham D. Finlayson[1] and Marina Bloj[2]; 1) Department of Computer Science, University of East Anglia, UK; 2) Bradford Optometry Colour and Lighting Lab, School of Life Sciences, University of Bradford, UK.

Abstract

Colours are usually described using three perceptual variables: an achromatic variable, such as lightness or luminance, and two "chromatic" varibles, such as chroma and hue. This means that the conversion of colour images to greyscale is often thought of as the removal of "chromatic information", which leaves a greyscale made by just the achromatic colour variable; i.e. luminance or lightness. One obvious problem with this approach is how to make a greyscale for equiluminant images, or images containing equiluminant object boundaries.

In this paper we review some of the more recent attempts to tackle the colour-to-greyscale problem. We then use an image preference experiment to test the performance of these methods on a selected set of images; some of which lose a significant amount of information when converted to greyscale using luminance, and others which do not. The results show that, in general, the newer techniques can provide a greyscale image which is preferred to that derived by luminance, especially for images that have prominent equiluminant boundaries. The results also show that this advantage is not guaranteed for every image, and that no particular algorithm provides consistently better performance than the others.

Introduction

The problem of converting colour to greyscale can be expressed in many different forms. In one sense the problem is a special case of the dimensionality reduction problem: how can we reduce a 3D signal (colour) to a 1D signal (greyscale) in a way that preserves all the important information? As such, the numerous techniques that have been employed in dimensionality reduction could also be applied to the problem at hand. Similarly, the colour to greyscale problem can also be thought of as a special case of gamut mapping, where the target gamut is the greyscale axis of a monochromatic output device. In this paper, we restrict ourselves to those dimensionality reduction and gamut mapping algorithms that have specifically been applied to the the colour-to-greyscale problem, although in future work we will explore the suitability of more general algorithms.

Luminance (LUM)

The most common technique for converting colour images to greyscale is to calculate the luminance value at each pixel. If the camera's colour space is *sRGB*, then the luminance value A_{lum} is given by:

$$A_{lum} = 0.2172 \times R + 0.7152 \times G + 0.0722 \times B, \tag{1}$$

where R, G and B represent the three colour channels of the camera. Mathematically, this operation projects the RGB vector in the direction of the luminance axis.

Alsam's method (ALS)

Instead of using this, image independent, luminance axis, Alsam [1] suggests projecting the image colours onto an image-dependent achromatic axis that captures the image's major colour variations, and hence colour contrasts. The method can be written mathematically as follows:

$$A_{als} = \omega_1 R + \omega_2 G + \omega_3 B, \tag{2}$$

where $\omega = [\omega_1, \omega_2, \omega_3]^T$ is the first eigenvector the image's raw cross product matrix (this is the same as the covariance matrix derived without subtracting the image mean), which is similar to the first principal component vector.

Alsam's method also includes a component for sharpening the resulting greyscale images by high-pass filtering the original three colour channels (R, G and B) and then adding this high-pass filtered information back onto the greyscale image.

Grundland's method (GRU)

Grundland *et al.* [2] also follow a projection based approach, although their technique differs significantly from both the LUM and ALS methods. They firstly transform the image into the opponent YIQ colour space. They treat the luminance (Y) component and chrominance (IQ) components separately. In the IQ plane they find a single vector that captures the important colour-differences between pixels; this part of the algorithm can be tuned to capture colour difference information over different spatial scales. When they have found this axis of the IQ plane, which they refer to as the predominant component, they then project the 2D IQ vectors at each pixel onto this vector, to give a single chrominance value, C. At this point the image has been reduced from 3 dimensions (YIQ) to 2 dimensions (YC). The final conversion from 2D to 1D (greyscale) is done by computing the weighted sum of the Y and C channels; that is:

$$A_{gru} = Y + \beta C, \tag{3}$$

where Y is the luminance image in the YIQ colour space, C is the chroma value derived from the algorithm and β is a weighting factor. The weighting factor β is determined to both ensure that the final image is within the displayable range of the device, and to emphasize colour details more or less strongly. Any image values that are outside the displayable range after this computation are then clipped to the boundary values.

In common with the previous two methods, this results in a global colour mapping: a given colour will always be mapped to the same greyscale value regardless of its spatial location. It also has the feature that if the pixels do not have any chromatic content (they are grey in the original image) then these pixels remain unchanged.

Rasche's method (RAS)

Rasche and co-authors [3, 4, 5] pose the problem of converting colour to greyscale in a general mathematical form. They suggest that for a greyscale image to retain the same colour differences that exist in the original image the grey levels should be assigned by minimising the following objective function:

$$A_{ras} = \min_{\forall a_i, a_j} \left(\sum_{i \in S_1} \sum_{j \in S_2} \left[\left| a_i - a_j \right| - f(\vec{c}_i, \vec{c}_j) \right]^2 \right). \tag{4}$$

In this equation the variables a_i and a_j indicate the greylevel values assigned to colours $\vec{c}_i$ and $\vec{c}_j$ (the overhead arrow indicates a vector quantity). The set S_1 usually refers to all the colours in the image, while the set S_2 may be all the colours in the image, or may be a subset of colours in a local neighbourhood of the ith colour. The function $f(\cdot)$ refers to a colour difference calculation; Rasche *et al.* employ a proporitional difference calculation, that ensures both that the final image will be within the displayable greyscale range, and that large colour differences are not given excess weight in the optimisation.

Bala's Method (BAL)

In contrast to the global mappings of all the previous methods, Bala and Eschbach [6] use a local mapping: a given colour may be mapped to different greyscale values depending upon its spatial surround. Their idea is based upon the observation that the important loss of information in a luminance rendering occurs at the boundaries of coloured objects where the luminance component of the edge is small. Thus, in their technique they aim to enhance these boundaries. They do this by firstly transforming the image into an opponent colour space, consisting of a luminance channel and two chrominance channels. They then perform a high pass filtering on each of the three channels. The resulting high-pass filtered images indicate where edges occur in the original image, and the channel where the largest edge signal occurs indicates whether an edge is a luminance or chrominance edge. They then add the high-pass information from the chrominance images back onto the luminance image, but giving more weight to edges where the luminance component of the edge is small. In this way, the boundaries between equiluminant colours are enhanced to a greater degree than edges where there is also a strong luminance difference. In line with the earlier methods, we can summarise this operation mathematically as:

$$A_{bal} = A + \gamma(x,y)\, C_{hp}\,(x,y)\,, \tag{5}$$

where A is an achromatic channel (luminance or lightness), C_{hp} is a high pass filtered chrominance image, and $\gamma(x,y)$ is a spatially varying weighting function derived from the strength of the high-pass filtered achromatic image. Apart from the spatially varying weighting function, this method can be seen to be closely related to Grundland's method, in that it adds chromatic information back onto an ordinary achromatic image.

Socolinsky's Method (SOC)

Socolinsky and Wolff [7, 8] developed a technique of image fusion which also follows a local mapping regime (although their technique predates that of Bala and Eschbach). The goal of the technique is to find a greyscale which, when differentiated, returns gradients similar to the gradients of the colour image. This can be expressed mathematically as finding the greyscale A_{soc} that minimises the following equation:

$$\min_{A_{soc}} \left\| \nabla A_{soc}(x,y) - \nabla^c I(x,y) \right\|, \tag{6}$$

where ∇ and ∇^c are differentiation operators for the greyscale image and colour image, I, respectively. For the colour image, ∇^c is DiZenzo's structure tensor [9], which returns a single x and y derivative at each point that best captures separate derivatives in the three colour channels.

The solution to Equation 6 is given by the solution to Poisson's equation:

$$\nabla^2 A_{soc} = \mathrm{d}iv\,(\nabla^c)\,, \tag{7}$$

where ∇^2 denotes the Laplacian operator and $\mathrm{d}iv\,(\cdot)$ is the divergence.

Algorithm implementation and image details

The LUM, and SOC methods were both implemented using our own MATLAB code, following published descriptions of the methods [7, 10]. Code for the GRU method was provided from the author's website [1], as was the code for RAS method [2]. Code for the ALS method was provided by the author. In our implementation we omitted the last sharpening stage so as to avoid biasing observers preferences, as there is evidence [11] that observers tend to prefer sharpened images and we wanted to investigate preferences for greyscale conversions only. In the case of the BAL method we used the author's own implementation. Whenever possible we used the methods' default parameters, only adjusting them when a clearly unsatisficatory image was generated.

All the methods assumed that the images were encoded in a linear sRGB space. Thus, where the algorithm did not explicitly invert the sRGB gamma, we inverted the gamma ourselves before processing the image, and then added it on to the result for display. In addition, our implementation of the SOC algorithm did not contain an explicit gamut-mapping step; i.e. it was possible for values in the resulting greyscale to fall outside the range $[0,1]$. To overcome this we used a simple linear tone-mapping operator that matched, as closely as possible, the histogram of the SOC output to the histogram of the LUM output [12].

The six images chosen for testing included artwork (a painting by Patrick Heron [3] and Claude Monet's "Impression Sunrise" [4]), cases with iso-luminant edges (including the two art images and a photograph of a field of poppies), and a human face, as well as natural (an image of a parrot) and man-made (an image of some hats) scenes . The six images are shown in Figure 1.

[1] http://www.cl.cam.ac.uk/ mg290/Portfolio/TurnColorsGray.html

[2] http://www.fx.clemson.edu/ rkarl/c2g.html

[3] Rumbold Vertical Three: Orange Disc in Scarlet with Green, http://www.waddington-galleries.com/artitsts/heron/

[4] http://ibiblio.org/wm/paint/auth/monet/first/impression/sunrise.jpg

In Figure 2 we show the rendering of the Heron painting by each of the six methods. It is clear from the luminance (LUM) version of this image (far left) that a large feature is lost from the bottom-right of the image. This corresponds to a green patch that is equiluminant to the surrounding area. Similarly, when using luminance the sun in the "monet" image becomes indistinguishable from the sky, while in the image of the field of poppies the red poppies are indistinguishable from the green grass.

The greyscale images produced by the six methods differ qualitatively in several ways. Figure 2 demonstrates differences in terms of average brightness, contrast, and even greyscale ordering (the RAS method makes the feature in the bottom-right lighter than its background, while the other methods make it darker). When applied to the "parrot" image several of the methods produced visible artefacts: the GRU method produced many saturated pixel values on the parrot's head, while the SOC method produced significant edge smearing and the RAS method produced an image that was relatively dark.

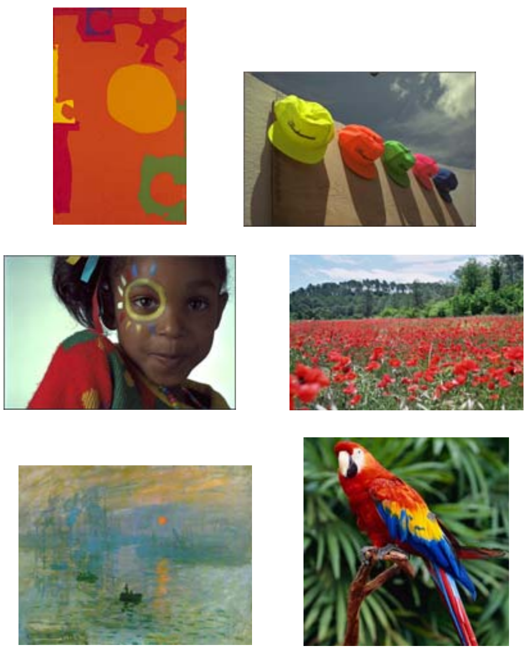

Figure 1. *6 different images used in the experiment. From the top-left, we label the images "heron", "hats", "girl", "poppies", "monet" and "parrot".*

Psychophysical experiment

The purpose of the experiment was to find out which of the methods, if any, produced the most pleasing greyscale reproductions. To do this a balanced pairwise comparison was used, in which the greyscale image produced by each of the different methods is compared to all the other greyscale versions of the same colour original. This results in $N(N-1)/2$ comparisons, where N is the number of different methods to be tested (in our case $N = 6$, which gives 15 comparisons in total).

For a single comparison the original colour image was displayed in the centre of the screen with two, different, greyscale reproductions shown flanking the original. Subjects were asked to consider all three images, and to report which of the greyscale reproductions they preferred. The explicit instructions were: "Consider the colour image and the two grey versions; which grey image would you prefer as a copy of the colour image?". No further instructions were given to the subjects, who were left free to choose their own criteria on which to form their judgements. When they had made their decision, they used a response box to indicate whether it was the left or right image that they preferred. The experiment was performed by 6 paid subjects, naive to the purpose of the study. All were colour normal and had normal or corrected-to-normal acuity.

For each of the six original test images, six greyscale versions were produced (one for each algorithm tested). Each subject judged a given pair of algorithms 4 times for each test image. In total, each subject made $15 \times 4 = 60$ comparisons for each image, making 360 comparisons over all six images. For a given test image, each pair of algorithms were compared 24 times (4×6 observers). To prevent subjects becoming tired or bored, the trials were split into two sessions, each lasting 20-25 minutes.

The experiment took place in a darkened room, so as to avoid the effects of ambient lighting. Subjects were seated approximately 1.14 metres from the monitor, and their head positions were stabilised using a chin-rest. The average size of the images was 6.5 degrees (visual angle) by 5.5 degrees and they were displayed against a checkerboard background to reduce contrast effects at the boundaries of the image. At the start of each session the subjects adapted for 60 seconds to a neutral grey background. In a single trial, each image triad was displayed for exactly 8 seconds. After 8 seconds the images disappeared, leaving only the complex background. Subjects were allowed to respond as soon as they had made a decision, whether 8 seconds had elapsed or not.

The computer control software and monitor were part of a ViSaGe system, supplied by Cambridge Research Systems. The monitor was calibrated and characterised beforehand and the white-point set to CIE D65. The images used were assumed to be stored in sRGB format, and thus had to be colour-corrected to be displayed on the monitor in question. To avoid out-of-gamut colours, the images were scaled by the same scaling factor (after removing the sRGB gamma) until they were guaranteed to be within the monitor gamut.

Statistical methods

The result of the paired comparison experiment is a matrix of scores for each image. An example is shown in Table 1, which has been compiled for the "poppies" image. Each row and column of the matrix indicates one of the six different algorithms. After each comparison judgement an element of this matrix is incremented: the row of the element corresponds to the preferred method, while the column corresponds to the method that was not preferred. This process is repeated until all the comparisons, for all subjects, have been completed. The column to the far right of this matrix shows the sum of each of the rows. This indicates how many times that particular method was preferred.

The "Total" column in Table 1 suggests that, for the "poppies' ' image, the SOC method was the most preferred method, and the LUM method was least preferred. The raw scores, however, do not tell us whether the difference between the scores is significant or not. To get an estimate of this we employ Thur-

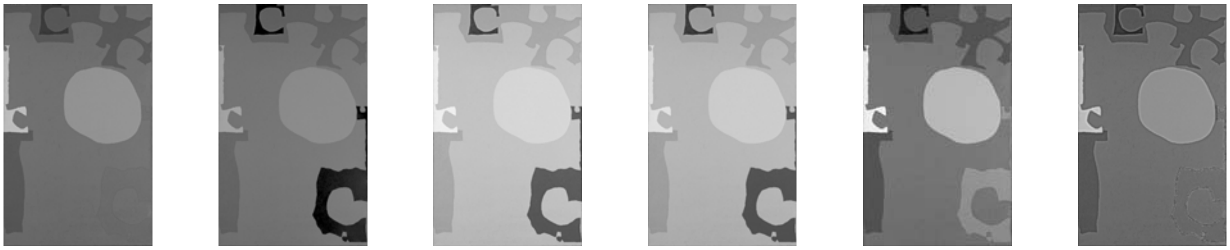

Figure 2. *6 different renderings of a colour image (original not shown). From left to right: LUM, SOC, ALS, GRU, RAS, BAL.*

Method	LUM	SOC	ALS	GRU	RAS	BAL	Total
LUM	0	3	4	4	13	12	36
SOC	21	0	17	18	19	21	96
ALS	20	7	0	17	16	16	76
GRU	20	6	7	0	19	19	71
RAS	11	5	8	5	0	13	42
BAL	12	3	8	5	11	0	39

Matrix of scores for the “poppies” image.

stone’s law of comparitive judgement, case V (see [13] for more details) [5]. The result of applying Thurstone’s law is a normalised-score for each of the algorithms: positive values suggest that the method is generally preferred, while negative values suggest the opposite. This analysis also produces 95% confidence limits for each score, which assist in judging the significance of the results.

To find out if one method outperforms another, some measure of the agreement between observers is needed. Here we follow the method of Ledda *et al.* [14] and calculate Kendall’s coefficient of agreement [15]. To describe this statistic we firstly label the ith row and jth column of the matrix shown in Table 1 as p_{ij}. The next step is to calculate the total number of agreements between pairs, Σ, which is given by:

$$\Sigma = \frac{1}{2}\sum_{i\neq j} p_{ij}\left(p_{ij}-1\right) \tag{8}$$

Kendall’s coefficient of agreement u can now be defined as:

$$u = \frac{8\Sigma}{N(N-1)T(T-1)} - 1 \tag{9}$$

where N is the number of methods tested and T is usually the number of observers taking part in the experiment. In our calculations, however, we define T as the number of observations, which is the number of observers multiplied by the number of repetitions ($4\times 6 = 24$). The closer the value of u to 1, the greater the agreement between observations. The minimum possible value for this statistic is $-1/(T-1)$ when T is even, and $-1/T$ when T is odd. It is also possible to examine whether this agreement is significant or not, i.e. to test the null hypothesis that observers do not agree with one another. We again employ the methodology of Ledda *et al.* [14] to do this, and use the χ^2 statistic applied to Kendall’s coefficent of agreement.

[5] We used the implementation in Green’s toolbox, which can be found at http://www.digitalcolour.org/toolbox.htm

As well as finding out if observers agree with one another, it is important to examine their *constistency*. The idea of consistency can be captured by a simple example: if we have three algorithms, say A, B and C, and we compare each method with each of the others, then the judgements are consistent if, for example $A > B$, $B > C$ and $A > C$ (where the $>$ symbol means “preferred to”). This consistency can be violated if, for example, $C > A$. Kendall and Babington-Smith [15] provide a statistic, the coefficient of consistency, to measure this property. This statistic, Ω, is given by:

$$\Omega = 1 - \frac{24c}{N^3-4N}, \tag{10}$$

where c is the total number violations of consistency, per observer, per image. The value of c is given by:

$$c = \frac{N}{24}\left(N^2-1\right) - \frac{1}{2}z, \tag{11}$$

where $z = \sum\left(score_i - (N-1)/2\right)^2$ and $score_i$ is the total score for the ith method (i.e. the ith entry in the “Total” column of Table 1). The value of Ω is 1 when consistency is perfect, in which case the methods can be ranked reliably. In our experiments we measure the average value of Ω over all observations.

Results

In Figure 3, the data from Table 1 are plotted in terms of the normalised scores derived from applying Thurstone’s law. Very broadly, we can say that where the error bars of two methods do not overlap, there is a significant difference between them; i.e. we can reject the null hypothesis that the two scores are not different from one another. Thus, for this image, we may conclude that the SOC, ALS and GRU methods tend to be preferred reproductions, while the RAS, BAL and LUM reproductions are not preferred. A similar trend was also found for the “monet” and “heron” images (although for these images the RAS method performed significantly better). Interestingly, the “heron”, “monet” and “poppies” images are those that lose salient details when luminance is used to convert them to greyscale.

Figure 4 shows the results pooled for all images. Here, the methods SOC, ALS, GRU and RAS perform equally well, while the methods BAL and LUM tend not to be preferred. These results provide some experimental evidence to back up the claim that the luminance transform can be improved upon when rendering colour images in greyscale. They do not tell the whole story however. Figure 5 shows the results for the “hats” image. In this

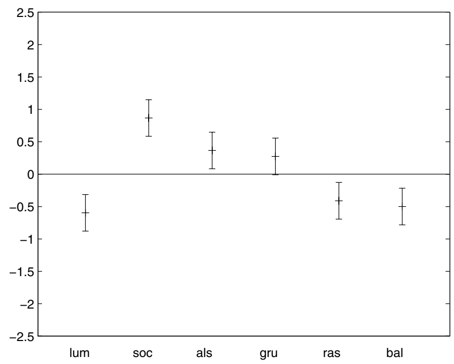

Figure 3. *Outcome of applying Thurstone's law to preference judgements on the "poppies" image. The x-axis shows the six different colour-to-grey algorithms, while the y-axis represents the normalised preference-score for each method (positive values indicate the method was preferred, while negative values suggest it was not preferred). The error-bars represent 95% confidence limits for the normalised scores.*

case the LUM method performs as well as most of the other algorithms. Nonetheless, there appears to be a slight advantage of the BAL method. Similar results were also found for the "girl" image, where this time the RAS method was the only one to outperform LUM.

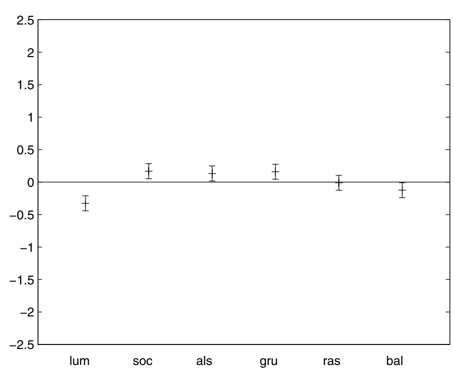

Figure 4. *Results for all the images pooled together (see Figure 3 for details).*

In Figure 6, the results for the "parrot" image are shown. The trend for this image is different to those for the other images, with the LUM, ALS and BAL methods clearly outperforming the GRU, SOC and RAS methods. We recall, however, that the GRU, SOC and RAS methods all produced visible artefacts for this image, while the LUM, ALS and BAL images were artefact-free. This is sure to have contributed to the preference results.

Table 2 shows the results of the observer-agreement analysis. The column headed "u" reports the values of Kendall's coefficient of agreement. This value is relatively high for four of the images, but is low for both the "hats" and "girl" image. In the cases where agreement was high, three of the four images have prominent isoluminant features, while the fourth showed visible artefacts for

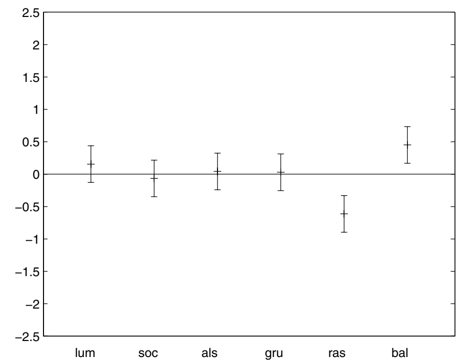

Figure 5. *Results for the "hats" image (see Figure 3 for details).*

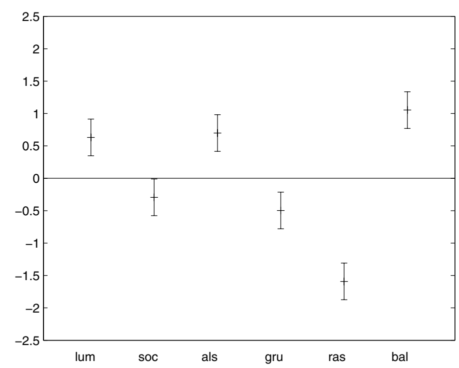

Figure 6. *Results for the "parrot" image (see Figure 3 for details).*

several methods. It seems reasonable that when when a particular feature is prominent (e.g. the sun in the "monet" image) this will tend to be used by observers to guide preference, and thus different observers will agree on the preferred images. For the "hats" and "girl" images, the low agreement scores support the earlier result that most algorithms perform similarly for these two images, and suggests that different observers may be using different criteria to judge these images.

The next two columns of the table test the significance of the u scores, i.e. test the null hypothesis that observers do not agree against the alternative hypothesis that they do agree. The first column shows the raw χ^2 score, while the second shows the probability with which we can accept the null hypothesis. In general the u scores are significant, and we can say that observers agree with one another. Even for the "hats" and "girl" images the agreement is significant at the 0.01 and 0.05 confidence levels respectively. This suggests that, while observers may not agree on some judgements, there is good agreement on other judgements, e.g.for the "hats" image most observers agree that the BAL method is preferred to the other methods, and that the RAS method is not preferred (see Figure 5).

The final column shows the average consistency score. These values are relatively high, and show good observer consistency. Again, however, the values are lower for the "hats" and "girl" image. This suggests that not only do observers not agree with one another, but they themselves are relatively inconsistent, and may be using changing their criteria for different individual judgements.

It is also interesting to note that the average response time of observers was 3.25 seconds, which suggests that they were using salient features to guide their decisions, rather than taking a long time to inspect the detail in the images.

Image	u	χ^2	Confidence level	Consistency
Heron	0.52126	194.8333	$0.001 > p$	0.88542
Hats	0.06087	36	$0.01 > p$	0.60417
Girl	0.040097	28.8333	$0.05 > p$	0.71354
Poppies	0.2256	92.8333	$0.001 > p$	0.75521
Monet	0.43527	165.1667	$0.001 > p$	0.80729
Parrot	0.38551	148	$0.001 > p$	0.81771

Summary statistics for the six different images used in our experiment.

Discussion and conclusions

In this paper we have both reviewed the existing state of the art in colour-to-greyscale conversion, and used an image preference experiment to differentiate between the performance of the methods. In general the LUM reproduction was not preferred when signficiant equiluminant features were found in the original scene (i.e. for the "heron", "monet" and "poppies" images). For these scenes there was also better agreement between observers, suggesting that the accurate reproduction of equiluminant details may be important for the observers, and hence drive their decision making. For images where strong equiluminant boundaries were not present the LUM rendering performed at least as well as most of the other methods. The results for the "parrot" image suggest a note of caution in the use of more advanced colour to greyscale techniques, with 3 of the 6 methods producing unwanted artefacts that counteracted the positive effect of maintaining colour contrast. When rendering individual images, however, these artefacts may be mitigated by a more careful choice of algorithmic parameters.

The results show that the performance of colour-to-greyscale algorithms is strongly image dependent: one algorithm may outperform another for a particular image, but this does is not hold over all images. One possible strategy for coping with this might be to implement the algorithms in two stages: firstly to classify an image based upon its content, and then to apply the algorithm that is likely to give the best reproduction.

It is also important to point out that our results relate only to image preference judgements. Furthermore, we found that subjects registered their responses fairly quickly, which suggests that they were using broad details to guide their preferences. Both these factors are likely to influence the results. In future work we intend to test the algorithms in more functional settings, e.g. interpreting information in pie charts or bar charts rendered in greyscale; we expect that if people are given such a task, then the relative performance of the different algorithms will change.

Acknowledgments

We thank Karl Rasche for his assistance in implementing his algorithm, Ali Alsam for providing the code for his method, and to Raja Bala who kindly ran our test images through his algorithm. We also thank Diego Socolinsky for providing us with a copy of his thesis, and all the observers who gave us their time to take part in the experiment. This work is supported by joint EPSRC grants number EP/E012248/1 and EP/E12159/1.

References

[1] Ali Alsam and Øyvind Kolås. Grey color sharpening. In *IS&T/SID 14th Color Imaging Conference*, pp. 263–267, Scottsdale, Arizona (2006). IS&T/SID.

[2] Mark Grundland and Neil A. Dodgson. The decolorize algorithm for contrast enhancing, color to grayscale conversion. Technical Report UCAM-CL-TR-649, Computer Laboratory, University of Cambridge (2005).

[3] Karl Rasche, Robert Geist, and James Westall. Detail preserving reproduction of color images for monochromats and dichromats. *IEEE Computer Graphics and Applications*, **25**(3), 22–30 (2005).

[4] Karl Rasche, Robert Geist, and James Westall. Re-coloring images for gamuts of lower dimension. *Computer Graphics Forum*, **24**(3), 423 – 432 (2005).

[5] Karl Rasche. *Re-coloring images for gamuts of lower dimension.* PhD thesis, Clemson University (2005).

[6] Raja Bala and Reiner Eschbach. Spatial color-to-grayscale transform preserving chrominance edge information. In *IS&T/SID 12th Color Imaging Conference*, pp. 82–86, Scottsdale, Arizona (2004). IS&T/SID.

[7] Diego A. Socolinsky and Lawrence B. Wolff. Multispectral image visualization through first-order fusion. *IEEE Transactions on Image Processing*, **11**(8), 923–931 (2002).

[8] Diego A. Socolinsky. *A variational approach to image fusion.* PhD thesis, Johns Hopkins University (2000).

[9] Silvano Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, **33**(1), 116–125 (1986).

[10] Michael Stokes, Matthew Anderson, Srinivasan Chandrasekar, and Ricardo Motta. A standard default colour space for the internet: srgb, http://www.w3.org/graphics/color/srgb.html.

[11] Buyue Zhang, Jan P. Allebach, and Zygmunt Pizlo. An investigation of perceived sharpness and sharpness metrics. In *Image Quality and System Performance II. Edited by Rasmussen, Rene; Miyake, Yoichi. Proceedings of the SPIE, Volume 5668, pp. 98-110 (2004).* (2004).

[12] David Connah, Graham Finlayson, and Marina Bloj. Coding contrast as brightness to convert colour images to grey-scale. *Perception*, **36**, 303 (2007).

[13] Peter G. Engeldrum. *Psychometric Scaling: A Toolkit for Imaging Systems Development*. Imcotek Press (2000).

[14] Patrick Ledda, Alan Chalmers, Tom Troscianko, and Helge Seetzen. Evaluation of tone mapping operators using a high dynamic range display. In *Proceedings ACM SIGGRAPH 2005, Annual Conference on Computer Graphics*, pp. 640–648, Los Angeles, CA, USA (2005).

[15] M. G. Kendall and B. Babington-Smith. On the method of paired comparisons. *Biometrika*, **31**, 324–345 (1940).