# Assessing the Effects of Gamut Compression in the Reproduction of Fine Art Paintings

*Lindsay MacDonald*
*Crosfield Electronics Ltd., England*
*and Ján Moroviç*
*London College of Printing, England*

## Abstract

This paper describes the results of experiments carried out at the National Gallery in London, comparing original paintings with printed colour reproductions made by processing digital images captured by a high resolution digital camera. A pairwise comparison method was used for three paintings under several different illumination conditions with a panel of observers. The results indicate that the optimum choice of colour gamut mapping procedure depends on the pictorial content of the painting.

## Introduction

The primary objective of the Esprit III project Methodology for Art Reproduction in Colour' (MARC) is the reproduction of colour paintings in print. The project aims to develop a system that can produce high quality facsimile reproductions from fine-art paintings, using direct digital image acquisition and improved methods of image processing, as per Figure 1. Each resulting print should visually match the original painting as closely as possible, without any visible defects in tone, colour or spatial characteristics. Specific aspects of the project have been the development of an ultra-high resolution digital camera, which can generate a digital image of up to 20,000 pixels square[8], and an optimal inkset for a seven-colour printing process to increase the colour gamut of the printed reproduction[9].

Ideally a reproduction should be the same size and have the same surface texture as the original, and at each point the tone and colour should be identical when viewed under any type of illumination. In practice the reproduction is generally less than a perfect facsimile, not only because of differences in surface texture but also because the reproduction colorants (printing inks) have different spectral absorption characteristics, leading to metameric differences under different illuminants, and because their tonal and colour gamut is generally less than the gamut of artists' pigments. Nevertheless, with careful processing of tone and colour it is possible to achieve a very good match between the original and reproduction under a specified illuminant[7].

Various techniques may be employed in the tone and colour correction and colour gamut mapping from original to reproduction as part of the colour separation processing, and these can result in subtle differences in the appearance of the reproductions. Of particular significance is the handling of out-of-gamut colours, depending on whether they are clipped to the gamut boundary or scaled in some manner. The effect of such changes on the final image cannot easily be predicted by computation or instrumental measurement. The best method remains visual assessment by a panel of observers, providing a suitable set of criteria for making judgements about colour fidelity.

In order to test the effectiveness of the MARC colour separation algorithms and of various methods of gamut mapping, we conducted a series of experiments at the National Gallery in London. Three paintings were captured as digital images, then reproduced as four-colour Cromalin proofs. Eight different reproductions of each painting were then compared side-by-side with the original under controlled viewing conditions. The pairwise preferences of observers were recorded and analysed to determine which gamut mapping methods produced the most pleasing overall reproduction.

## Assessment Procedures

Two different assessment procedures may be used for comparing the fidelity of a MARC printed reproduction with the original painting: instrumental measurement or visual assessment. The colorimetric match could formally be established by spot measurements with a reflection spectrophotometer. Comparison of the CIELAB values at corresponding locations of the original and the print would then allow colour difference metrics such as $DE_{LAB}$ or $DE_{CMC}$ to be calculated and averaged to give a quantitative measure of reproduction fidelity. Other image attributes could be quantified in a similar way, including density range, colour accuracy, dimensional accuracy, spatial resolution and signal-to-noise ratio.

Visual assessment may be performed by a panel of human observers, who look at both the original and the print and give their opinions about the goodness of the reproduction. The evaluation procedure consists of placing the original and the print side by side under the same conditions of illumination and comparing the two. For gallery viewing situations the print could be mounted in an identical frame and hung alongside the original at the same height, with the observers standing midway between the two. For more critical assessment, the original should be removed from its frame and positioned alongside the print in a standard graphic arts viewing cabinet with neutral grey walls and controlled lighting with a correlated colour temperature of 5000K and illuminance of 500±125 lux.[2]

For quantification of the visual match between the original and the print it is necessary to formalise the procedures for observation and recording of results. Quality criteria for visual assessment can be defined as subjective measures of spatial, tonal and colour attributes of the image, such as the following:

*Sharpness*
Edge quality and sharpness throughout the image.

*Detail*
Rendition of fine detail in highlight, midtone and shadow regions of image.

*Smoothness*
Amount of noise 'texture' at all densities and enlargements.

*Spatial artefacts*
Any visible patterning or disturbances in the reproduction, such as streaks, asymmetry of solid lines, blocking, jitter, moiré and aliasing.

*Tonal rendition*
Tone reproduction over the full density range of original.

*Hue rendition*
Accuracy of rendition of hues over the full gamut of original.

*Colour saturation*
Degree of colour saturation, i.e. neither too much nor too little, over all hues and tones.

*Tone and colour artefacts*
Any visible disturbance or distortion in the colour reproduction, such as contouring, colour casts in certain tonal ranges or imbalance between differing hues.

Images are normally reproduced on different media from the originals and, because of spectral differences between the colorants, differing illumination levels and the influence of viewing conditions, an equivalent colour match is the best that can be expected. In this case human judgement is the only proper basis of assessment, at least until the advent of a fully comprehensive model of colour appearance. No two experts ever agree totally on colour fidelity, but by analysis of the results obtained from a panel of observers a trend can be established.

## Processing MARC Images for Reproduction

MARC images are captured and encoded colorimetrically in CIE L*a*b* colour coordinates, referenced to D65 white, which define precisely the colour stimulus at each point (pixel) of the image. The primary objective of the image processing software is to generate colour separation values (%dot) for the printing inks, which when printed will produce the identical colour stimulus at each point of the print. Considerable effort has been applied within the MARC project to achieving a high standard of colour calibration accuracy, so that any desired *L*a*b** value can be accurately reproduced in print over the full gamut of the printing inks.

For the MARC reproduction of fine art paintings it is theoretically possible to achieve an exact colorimetric match, provided that all colours of the original are within the gamut of the printing process. In general there will also be colours present in the original which, although they are captured by the MARC camera and are represented accurately in CIELAB coordinates in the digital image, lie outside the colour gamut attainable by the target printing process. Even the seven-colour process with six chromatic primary inks plus black cannot reproduce certain extreme pigment colours such as cadmium yellow or ultramarine, unless special inks are made up from the actual pigments in question. Thus in general there will exist out-of-gamut colours in the image which must be modified (or mapped) to produce the nearest colour within the printing gamut. This inevitably leads to differences in colour appearance between original and reproduction.

To compensate for differences between the colour gamuts of the original and the reproduction media, various correction techniques can be applied when processing the digital image. As changes in hue are more perceptible than changes of other colour attributes, all but one of the following techniques preserve hue and alter only lightness and colourfulness:

(a) *Paper cast removal* To account for the colour of the printing substrate (normally white paper), the *a** and *b** coordinates of the reproduction's substrate are subtracted from all colour coordinates in the original image. This is the only correction algorithm that alters hue, albeit to a minor extent.

(b) *Tonal mapping* To preserve the relative lightness values in the image, the larger range of densities in the original (typically up to density 3.0) is mapped onto the more limited density range of the reproduction (typically 2.0). This is in effect a one–dimensional transformation along the *L** axis, whereby the original's white point and black point are mapped onto the reproduction's white point and black point respectively.

(c) *Gamut mapping* The original's colour gamut is in most cases larger than that of the reproduction, and it is therefore necessary to devise some rules according to which out-of-gamut colours are going to be reproduced. The two gamut mapping algorithms used in this experiment were orthogonal mapping and chord mapping. *Orthogonal mapping* reproduces an out-of-gamut colour as the nearest colour on the reproduction's gamut boundary. In Figure 2 $c_2$ is the colour resulting from $c_1$ being

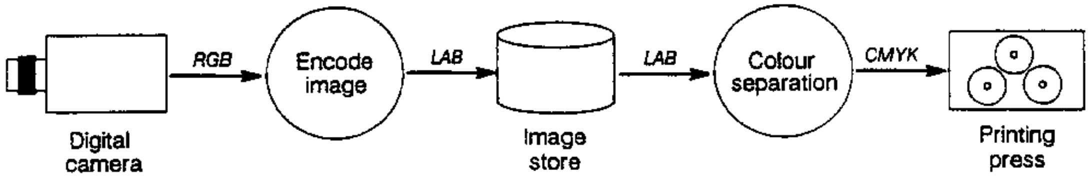


*Figure 1. Simplified MARC image processing path*

orthogonally mapped onto the gamut boundary (both colours are in a plane having constant hue). This gamut mapping technique provides the smallest $DE_{LAB}$ colour difference. *Chord mapping* reproduces the out-of-gamut colour as the colour which lies on the intersection of the gamut boundary with the line joining that colour's coordinates to the lightness coordinate of the colour with the highest saturation at the given hue angle. Using chord mapping $c_1$ is mapped in Figure 2 onto $c_3$. Chord mapping has a lesser effect on saturation, but results in a larger overall colour difference. These two techniques are in some sources referred to as gamut clipping, as they change only the coordinates of colours that are out of gamut.[10]

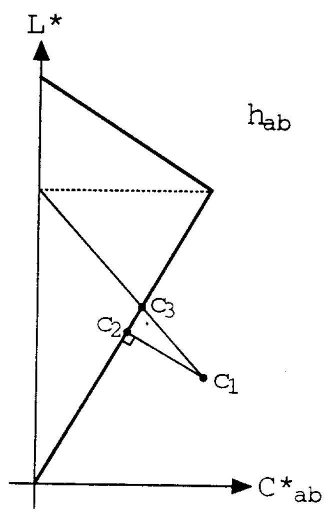


*Figure 2. Gamut mapping*

(d) *Gamut compression* It is generally regarded as more important to preserve the relative differences in colour saturation (or colourfulness) between all colours, rather than to preserve the absolute coordinates of those in gamut.[4] To do this, gamut compression can be applied to all colours in the original image by radial scaling of all coordinates in $L^*a^*b^*$ towards the 'gamut centre', assumed to be the point on the $L^*$ axis with value $L^*=50$. Another way to think of this process is applying an expansion to the gamut of the reproduction process, whereby all colours are moved outward from the central point on the $L^*$ axis (see Figure 3). The original's colours are then mapped onto the expanded gamut. For these experiments a fixed compression factor of 20% was used for the two cases of gamut compression tested.

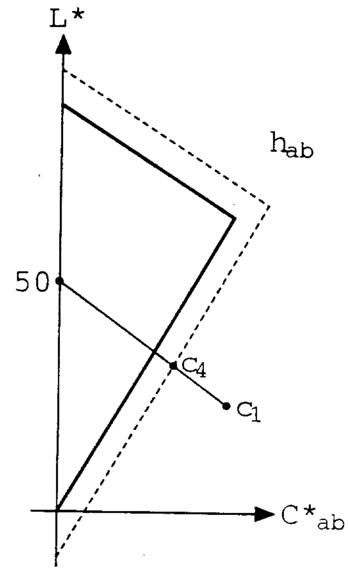


*Figure 3. Gamut compression*

To determine what influence these algorithms have on the extent to which the reproductions match the original, eight different combinations of the correction factors were tested. For each combination the original $L^*a^*b^*$ image was processed by Crosfield to produce four colour separations for cyan, magenta, yellow and black (CMYK) printing inks and a colour proof was made using the DuPont Cromalin process. To simulate the surface characteristics of the original oil paintings more closely, all proofs had a matte coating applied (DuPont Top Coat Matt). These reproductions will be referred to by their labels, as defined in Table 1.

## The Experiment

The objective of the experiment was to compare the visual appearance resulting from different combinations of correction factors (i.e. to give information about their relationships) and not to ascertain the 'absolute' goodness of match of the reproductions.

A set of three different paintings from the National Gallery was chosen for the experiments. The first was a fragment of an altarpiece by the Master of Liesborn called *The Adoration of the Kings* believed to have been painted between 1470 and 1480. This painting has a wide range of hues, areas of high saturation, neutral areas and a wide tonal range. Analysis of the digital image showed that over 30% of all pixels had colours out of gamut. The second painting was the *Portrait of a Woman* from a painter of the Cologne School painted in 1495. It has large areas of dark, saturated red and considerable shadow and highlight detail. Its characteristics and the fact that 68% of the pixels in this image had colours out of gamut, make it particularly difficult to reproduce. The third painting was an 18th century 'Claudesque' landscape with colours predominantly in the desaturated shadow region. It is typical of many old oil paintings in that its original appearance is partially obscured by a layer of yellow-brown varnish, resulting in the desaturation and darkening of all the colours.

**Table 1. Eight Combinations of Correction Factors Used in the Experiments**

| *Label of Reproduction* | *Paper Cast Removal* | *Tonal Mapping* | *Gamut Mapping* | *Gamut Compression* |
|---|---|---|---|---|
| N | No | No | Orthogonal | No |
| NC | No | No | Chord | No |
| A | Yes | No | Orthogonal | No |
| AC | Yes | No | Chord | No |
| P | Yes | Yes | Orthogonal | No |
| PC | Yes | Yes | Chord | No |
| NG | No | No | N/A | Yes |
| PG | Yes | Yes | N/A | Yes |

Since the first painting (Master of Liesborn) could not be removed from its normal wall hanging location in the Sainsbury Wing at the National Gallery, the comparison was carried out under the illumination present in the Gallery. The room was illuminated by a limited amount of daylight admitted through overhead skylights, augmented by tungsten lights with filters to give them correlated colour temperatures of 4100K and 3000K. The two types of tungsten lights were mixed in a ratio of 1:1.

The second and third paintings were compared under standard viewing conditions[6] with three different illuminants: D50 (the graphic arts standard), D65 (the illuminant used as a reference white for the image capture and ink colorimetry) and A (the equivalent of a domestic tungsten light). These comparisons were carried out in a viewing booth at the National Gallery, in which the original painting and the reproduction proofs could be arranged side-by-side.

A number of observers took part in the comparison, all of whom were male with ages ranging from sixteen to fifty-two and had normal colour vision. The first painting was assessed by six different observers (resulting in six sets of data). The second painting was assessed by 10 observers, each making a different number of comparisons for the three different illuminants, resulting in a total of 24 sets of data. The third painting was assessed by only two observers for all three illuminants, giving a total of six sets of data.

The individual reproductions were presented to the observers in pairs alongside the original and they were asked to choose the reproduction which represented the closer match. So as not to force observers into a decision, they were also given the option to indicate that both reproductions matched the original to the same extent. This technique was chosen over other methods which require the observers to award points to the objects under comparison, or to rank them. The reason is that pairwise comparison asks the observer only to make a choice, and doesn't require him or her to specify or quantify the response, which means that less subjective judgement is involved. The individual choices can later be combined to provide the same type of result as techniques asking observers to rank the reproductions.[3]

Making a decision about which reproduction is the closer match involves the consideration of a range of variables, which are consciously or subconsciously given different weight when forming the overall choice. Observers were therefore asked specific questions for closely defined regions of each image, each having a dominant characteristic (e.g. hue, lightness, colourfulness). Even in these specific questions, however, the observer's choice may be influenced by variables other than the main one (according to observers, for example, questions about hue were often influenced by the amount of detail preserved in a particular region). The effect of these influences is probably smaller than in the case of asking one overall question for a reproduction.

Since the three chosen paintings had very different characteristics, primarily in terms of gamut and tonal range, a different questionnaire was designed for each painting. All the questionnaires consisted of the following general components: identity of the observer, explanation of the observation's aim, a halftone reproduction of the painting, and a list of the regions to be compared, e.g. the red, green, blue, yellow, saturated, highlight, neutral and shadow areas. Each choice was represented by two boxes, representing the pair of reproductions. One box was to be ticked to indicate the preferred choice, or both boxes were to be ticked if no difference was perceived. To carry out the full pairwise comparison of all eight reproductions required 28 pairs of paintings to be compared, with the pairs presented to the observers in a random order.

The Master of Liesborn painting was assessed on the wall of the Gallery, so that it was not possible to have one reproduction either side. The pair of reproductions thus had to be hung on the wall to the right of the painting. The assessment of this painting was carried out simultaneously by six observers standing at a distance of approximately two metres from the painting. The Cologne School and Claudesque paintings were placed in a standard viewing booth with the original in the centre and one reproduction on either side. The comparison of all 28 pairs was then conducted under each of the three illuminants.

## Evaluation of Results

Since the observers were given the choice to declare both reproductions in a pair as being equal, the following scoring system was used. The chosen reproduction was given two points and the other one zero; when both were selected, they were awarded one point each. For evaluation the scores were transcribed into a spreadsheet in the form of an 8x8 matrix for all pairings of the reproductions for each question, where the points given to each reproduction were placed in its column and in the row of the reproduction to which it was compared (e.g. if A is chosen over C then +2 is written into column A row C and -2 in column C row A).

The following procedure was used to obtain a ranking of the reproductions. For each questionnaire a matrix was set up containing the summary of answers to all questions. Each score in this 8 x 8 matrix was obtained by summing the scores for that particular combination and dividing by the maximum possible score (i.e. the number of questions times two). All questions were given an equal weight, since the introduction of a weighting system for the questions would only add another subjective influence and thus distort the results.

An average score was then calculated from all observers' questionnaires under the same viewing conditions. To determine the statistical significance levels, standard deviations were calculated for each reproduction's total score, namely the sum of all scores in its column. For the differences between two reproductions to be statistically significant at a 95% confidence level, the mean total score of one reproduction had to be outside the other reproduction's 95% confidence interval, which is delimited by the mean of the total score plus and minus 1.96 standard errors. To obtain the final scores for a given painting the results from all illuminants were averaged, which also increased the precision of the results since they were obtained from a higher number of observations.

The results were used to establish a ranking of the reproductions, and also converted into an interval scale, where the differences between individual reproductions can be compared. This conversion is suggested by Thurstone's law of comparative judgement which is 'based on the notion that the proportion of times stimulus A will be judged greater than stimulus B is determined by the degree to which sensation A and sensation B differ'.[3]

## Analysis of Results

### Master of Liesborn

The reproductions of this painting can be divided into three significantly different groups, as shown by the scores in Figure 4. Those obtaining the highest scores were P and PC, which have both had tonal mapping and paper cast removal applied. Coming last was PG, being the reproduction where gamut compression was applied in addition to tonal mapping and paper cast removal. The remaining reproductions were not significantly different from each other and were in between the other two groups. It is interesting to note the disagreement between observers regarding the reproduction NG. This was possibly caused by some observers being more influenced by the tonal characteristics of the reproduction, whereas others based their choice on the colour of the area under comparison, which was what they were asked to do.

The discrepancies between observers were probably not due their age, professional background or other characteristic, since such a correlation is not apparent from their responses. A more likely reason is the difficulty experienced by an observer to base his judgement solely on the criteria he was asked to judge. In this case part of the task given to the observers was to compare certain areas of a characteristic colour, where it was difficult not to be influenced by the amount of tonal detail maintained in those areas. This influence might have been reduced, but not eliminated, by giving this potential problem more emphasis when the task was explained to the observers.

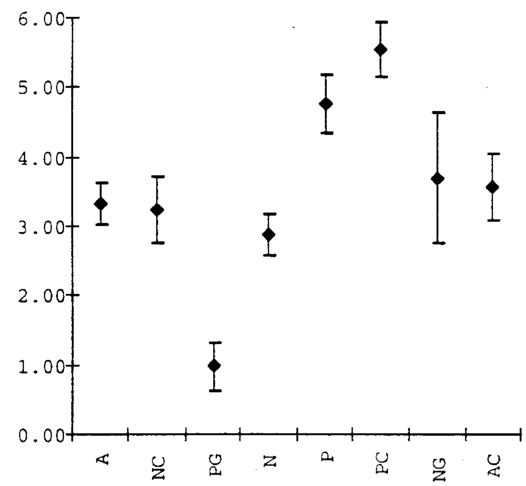


*Figure 4. Scores for Master of Liesborn*

Since this painting has a wide tonal range, the corrections modifying tonal reproduction were preferred most often. The failure of gamut compression techniques seems to be caused by the fact that the painting's gamut did not exceed the reproduction's gamut significantly and there was therefore no real need for this correction.

### Cologne School

This painting differs from the first, since 68% of its pixels have colours out of the CMYK gamut. The need for gamut compression was confirmed by the results, shown in Figure 5, indicating that the two reproductions with gamut compression (NG and PG) had the highest scores, alongside reproduction P. These three reproductions are followed by reproduction PC, which differs from P only in the gamut mapping algorithm. The wide tonal range of this painting explains why the two reproductions with tonal mapping are among the most preferred in this case. The least preferred reproductions were those without correction or only with paper cast removal.

Since results of the comparisons under the different illuminants do not differ significantly, only their average is shown here. The differences in results for the different illuminants can be explained in terms of metamerism between the pigments of the painting and the dyes of the Cromalin proofs, and also changes in appearance due to chromatic adaptation. Some differences between the results from different illuminants can also be accredited to the small number of observations made under the individual illuminants.

**Claudesque**

Due to the limited contrast and small tonal range of this painting, and the fact that the colours were predomi nantly in gamut, the reproductions with the highest scores were N, NC and AC, which either make no correction or only remove the paper cast. The second group of reproductions is made up of PC, NG and A closely followed by P, as shown in Figure 6. The least preferred reproduction for this painting was PG, which affects both the gamut and the tonal characteristics of the image.

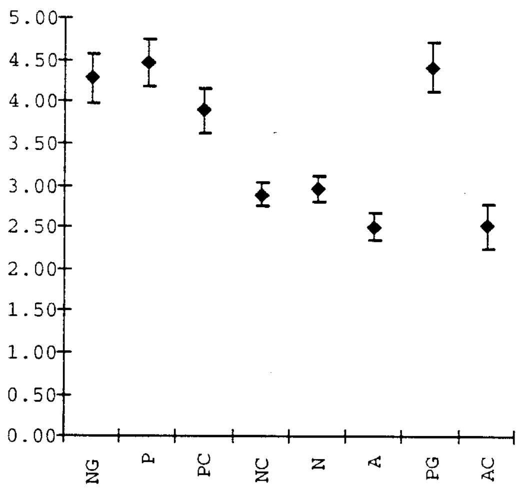


*Figure 5. Scores for Cologne School*

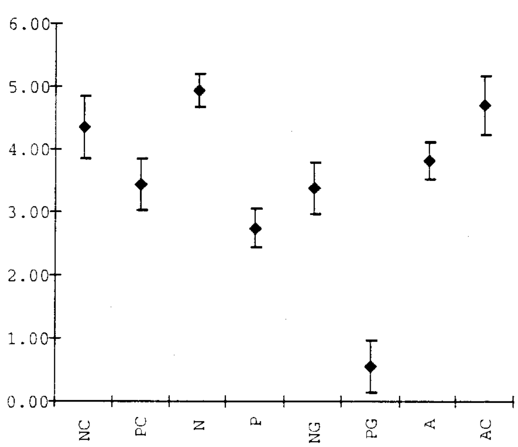


*Figure 6. Scores for Claudesque*

## Conclusions

Since the comparisons of each painting suggest a different ranking for the preferred types of corrections under examination, it is not possible to suggest one correction method that is optimal for all images. Rather, the results of this experiment suggest that the success of a particular correction algorithm depends primarily on the characteristics of the image to which it is applied.

As can be seen, gamut compression achieves good results if the original painting's gamut significantly exceeds the gamut of the reproduction process, but it degrades the quality of images having gamuts comparable to the gamut of the reproduction process. Further experiments could be conducted, comparing a series of paintings with a gradually increasing number of out-of-gamut colours, to determine a threshold level above which gamut mapping should be applied.

The tonal mapping techniques improved the match of paintings having a wide tonal range, and consequently paintings with large gamut and a wide tonal range achieved the closest match when a combination of gamut compression and tonal mapping was applied to them. In most cases the type of gamut mapping used (i.e. either orthogonal or chord) did not have a significant influence on the success or failure of a particular algorithm.

From the results of this experiment it is clear that a particular correction should be only used if there is need for its effect. The results also confirm that if the correct algorithm is used it can produce a positive influence on the reproduction's visual match with the original painting. These results therefore support the approach of providing the user of the MARC software with a choice of correction methods, to be selected according to the pictorial content of the painting to be processed. There is no substitute for human judgement in making a choice of this kind.

## Acknowledgements

We are indebted to Crosfield Electronics and DuPont for providing the facilities for image processing and preparation of Cromalin proofs. We would like to thank Dr. David Saunders and Dr. John Cupitt (National Gallery) for assistance and valuable suggestions in setting up the experiments and making the paintings available. Also thanks are due to Dr. Bob Thompson (London College of Printing) and Dax Rughani (Crosfield Electronics Ltd) for their contributions. The project was funded under the European Commission's *Esprit* collaborative research programme.

## References


1. J. Bristow and P.A. Johansson, "Subjective Evaluation by Pair Comparison: Pitfalls to Avoid and Suggestions for the Presentation of Results", *Advances in Printing Science and Technology* **17**, Pentech Press, pp. 278–291 (1983).
2. G. G. Field, "Color and its Reproduction", *GATF Press*, Pittsburgh PA (1988).
3. G. A. Gescheider, *Psychophysics, Method and Theory*, Lawrence Erlbaum Associates, pp. 84-102 (1976).
4. R. W. G. Hunt, *The Reproduction of Colour in Photography, Printing and Television*, 4th Ed, Fountain Press, Tolworth UK, p. 43 (1987).
5. A. J. Johnson, Device Independent Colour ... Is It Real?", *Proc. TAGA Conference*, Vancouver, April 1992 (1992).
6. A. J. Johnson and M. Scott–Taggart, *Guidelines for Choosing the Correct Viewing Conditions for Colour Publishing* PIRA Press, UK (1993).
7. L. W. MacDonald, "Colour Fidelity Issues in Image Reproduction for Print", *Proc. EOS/SPIE Europto Symp.*, SPIE Vol. **1987**, pp 77-87 (1993).
8. L. W. MacDonald, J. D. Deane, and D.N. Rughani, "Extending the Colour Gamut of Printed Images", *J. Phot. Science* **42**, No 3, pp 97-99 (1994).
9. L. W. MacDonald and R. Lenz, "An Ultra-High Resolution Digital Camera", *J. Phot. Science* **42**, No 2, pp 49-51 (1994).
10. P. K. Robertson, "Perceptual Color Spaces", *IEEE Computer Graphics & Applications* **8**, No 5, pp 50-64 (1988).