



ELSEVIER

Computers in Biology and Medicine 34 (2004) 615–632

<http://www.intl.elsevierhealth.com/journals/cobm>

**Computers in Biology
and Medicine**

The effect of image colour distortion on evaluation of donor liver suitability for transplantation

Darryl A. de Cunha^{a,*}, Leila H. Eadie^b, John L. Barbur^a, David J. Hawkes^c,
Alexander M. Seifalian^b

^a*Applied Vision Research Centre, Department of Optometry and Visual Science, City University, Northampton Square, London EC1V 0HB, UK*

^b*University Department of Surgery, Royal Free and University College School of Medicine, Hampstead, London, UK*

^c*Department of Radiological Sciences, 5th Floor, Thomas Guy House, Guy's Hospital, London Bridge, London SE1 9RT, UK*

Received 19 April 2002; received in revised form 22 September 2003

Abstract

Some medical applications rely on the use of colour in the diagnostic process, even when poor colour reproduction can affect diagnosis. In this paper, we investigate the effect colour distortion can have on assessment of livers for possible transplantation. We compare the diagnostic effects for the likely colour shifts when illuminant and camera remain uncalibrated or when a monitor is incorrectly calibrated. We describe methods that can result in accurate reproduction of image colour on visual displays and determine whether accurate colour reproduction is necessary for effective liver assessment.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Colour distortion; Colour reproduction; Colour space; ROC analysis; Liver transplantation

1. Introduction

Liver transplantation is the only available treatment for end-stage liver failure and the number of patients waiting for transplant consistently exceeds the number of available livers [1,2]. One common problem in finding suitable livers for transplant is steatosis (accumulation of fat), which is well recognised as a risk factor for poor or non-function of the transplanted organ [3]. Severe steatosis can be identified by greasy liver texture and yellow discolouration. Such observations

* Corresponding author. Tel.: +44-20-7040-0262; fax: +44-20-7040-8355.

E-mail address: d.decunha@city.ac.uk (D.A. de Cunha).

have been used to establish a relationship between subjective colour appearance and fatty change graded by histology [4]. The explicit use of colour for medical diagnosis by video imaging, has been used in applications such as assessment of wound healing [5,6], tongue diagnosis [7], skin erythema [8] and dermatology [9]. Telemedicine applications make the use of colour cues more difficult, largely because of poor reproduction of colour on display devices. It is of interest to determine the diagnostic consequences of inaccurate colour reproduction in the video image. The aim is to specify the tolerances needed to achieve an effective system. We have assessed the effects of colour distortion through the different electronic imaging devices as part of a project on remote assessment of livers for transplantation using video imaging. The magnitude of typical colour changes resulting from the use of uncalibrated illuminant and cameras have been measured and compared to the colour difference required for correct identification of normal and fatty livers. Our aim was to establish whether the use of uncalibrated systems were likely to affect diagnosis based on colour.

Studies reported in the literature [5–9] evaluate image colour using red, green, blue (r, g, b) colour representation of pixels. Various methods have been used to convert r, g, b values to (Commission Internationale de l’Eclairage (CIE) [10]) colour spaces that are accepted internationally. One slightly misleading but commonly used term in video colour representation is to describe a 24 bit per pixel r, g, b colour display as ‘true colour’ [11]. Although 24 bit ‘true colour’ r, g, b does allow the full gamut of monitor colours to be displayed simultaneously (as opposed to a colour palette which displays a subset of possible colours), the use of such a system with an uncalibrated camera and illuminant does not result in accurate colour reproduction. For the CIE ‘standard eye’ ‘accurate colour’ reproduction can be expected to generate the same cone receptor signals irrespective of whether the eye views the object directly or its corresponding image. The diagnostic consequences of using the same ‘true colour’ image representation on different monitors but without correcting for the different monitor characteristics is also examined in this study.

A device-independent conversion of r, g, b to CIE colour space was made by Vander Haeghen et al. [12] using the sRGB colour space [13], for use in dermatology. This standard default monitor colour space is becoming widely available as a configuration option for good-quality CRT monitors. An independent study [14] of sRGB compatible monitors from different manufacturers, has shown good monitor adherence to the sRGB standard for colour reproduction across all manufacturers.

2. Methods

To successfully complete this study, we needed to assess the likely colour distortions that occurred when an object (a liver) was imaged by a camera and illuminant of unknown characteristics and compare these distortions to the colour changes in assessing the livers’ fatty content. If the magnitude of colour distortion was small compared to the colour changes used in diagnosis, then knowledge and calibration of the camera and illuminant would not be important in the overall process. However, if the colour distortion was equal or larger in magnitude to the colour changes noted between normal and fatty livers, uncalibrated systems could lead to errors in liver assessment. When livers are physically assessed for fatty content, the assessment is based on the colour of the liver, the greasy texture of the surface and the ‘roundness’ of the liver edges. However, when assessed

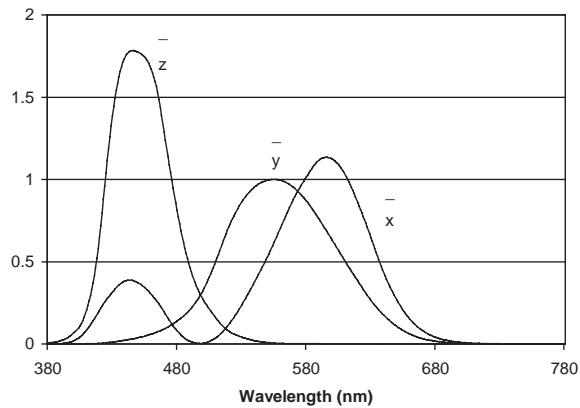


Fig. 1. CIE colour-matching functions.

remotely using minimally invasive imaging techniques, the fat level estimate is based solely on colour.

The methods of determining image colour distortions and the amount of colour change used in identifying normal or fatty liver are described below.

2.1. Colour reproduction and colour spaces

Before describing the measurement of colour distortions and the methods to relate colour to diagnostic accuracy, it is worth reviewing some background on colour reproduction and colour spaces.

Colour reproduction in image displays is based on the fact that there are only three different types of colour photoreceptors (cones) in the human retina. Equal ratios of cone stimulation will result in equivalent colour perception, even when the underlying spectral distributions causing the stimulation are different, an effect known as metamerism. This phenomena was used to establish an internationally accepted set of tricolour functions to match any colour, using specified amounts of three monochromatic stimuli (tristimulus R, G, B) at 700 (red), 546.1 (green) and 435.8 nm (blue). The results across the visible range produced three colour matching functions defined for the CIE 1931 standard calorimetric observer [15]. One problem with the wavelength of the chosen stimuli was that the green stimulus at 546.1 nm also stimulated a large ‘red’ response from the red sensitive cones. Therefore, green colours with no red component could only be matched by adding red to the colour sample. For the purpose of producing the colour matching functions, the process of adding red to a colour sample can be interpreted as negative values of the red colour matching function. Because negative values in the colour matching functions were considered by the CIE to be detrimental to their system’s wide adoption, the R, G, B tristimulus ratios were transformed to a set of non-negative X , Y , Z tristimulus values at each wavelength [16].

The X , Y , Z values across all visible wavelengths define the CIE colour matching functions (denoted $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$) and are shown in Fig. 1. A given object of wavelength radiance distribution,

$L(\lambda)$, has X, Y, Z tristimulus values that can be computed by multiplying and summing each colour matching function with $L(\lambda)$, over the visual wavelengths. The $\bar{y}(\lambda)$ function was carefully chosen to be proportional to perceived luminance, therefore calculated Y tristimulus values for any given spectral distribution are also proportional to luminance.

Normalisation of the X, Y, Z tristimulus values produces a set of (x, y, z) chromaticity co-ordinates that represent the *relative* X, Y, Z tristimulus values:

$$x = \frac{X}{X + Y + Z}, \quad y = \frac{Y}{X + Y + Z}, \quad z = \frac{Z}{X + Y + Z}. \quad (1)$$

The normalisation process ensures that the sum of the three chromaticity co-ordinates equals unity, i.e., $x + y + z = 1$, and has the advantage of specifying a chromaticity using only two co-ordinates. A plot of x (abscissa) and y (ordinate) produces the simplest, standard device-independent colour space called the CIE (x, y) —chromaticity diagram. The CIE (x, y) —chromaticity diagram has the undesired characteristic that perceptually equal shifts in colour are not represented by equal distances on the diagram. Therefore, a further transformation from x, y space to a more perceptually uniform u', v' colour space is usually employed, defined by (2)

$$\begin{aligned} u' &= 4x/(-2x + 12y + 3), \\ v' &= 9y/(-2x + 12y + 3). \end{aligned} \quad (2)$$

2.2. Measurement of colour distortion

The method of measuring image colour distortion is shown in Fig. 2. The method utilised a Macbeth Colour checker chart [17] that contained 18 coloured patches with known spectral reflectance u', v' chromaticity co-ordinates. There is no generally accepted standard illuminant under which livers should be viewed when determining their level of steatosis. The first step was then to record an image of the chart with the camera and illuminant used for subsequent images of liver samples in this study. The u', v' co-ordinate of each patch in the final recorded image depended not only on the patches' spectral reflectance and the spectral distribution of the illuminant, but also the spectral response of the camera and any subsequent post-processing of the image during digitising (quantisation effects, for example). For this study and in general surgical applications, the camera and illuminant spectral characteristics remain unknown. Recording the Macbeth chart was therefore effectively a system colour calibration and further liver images taken for this study had to either keep the camera and illuminant constant or recalibrate.

To determine the displayed chromaticities of each patch, the pixel r,g,b values of each patch in the image were used to calculate the displayed monitor phosphor luminance (L_R, L_G and L_B), using a previously taken monitor calibration of r,g,b to luminance. This calibration measured the visible light output (using a photometer) from each phosphor colour in cd/m^2 as a function of the monitor r,g,b look-up table (lut) value from 0 to 255. This calibration function has the general form

$$L_i = (A_i \text{ lut}_i)^{\gamma} + b_i, \quad (3)$$

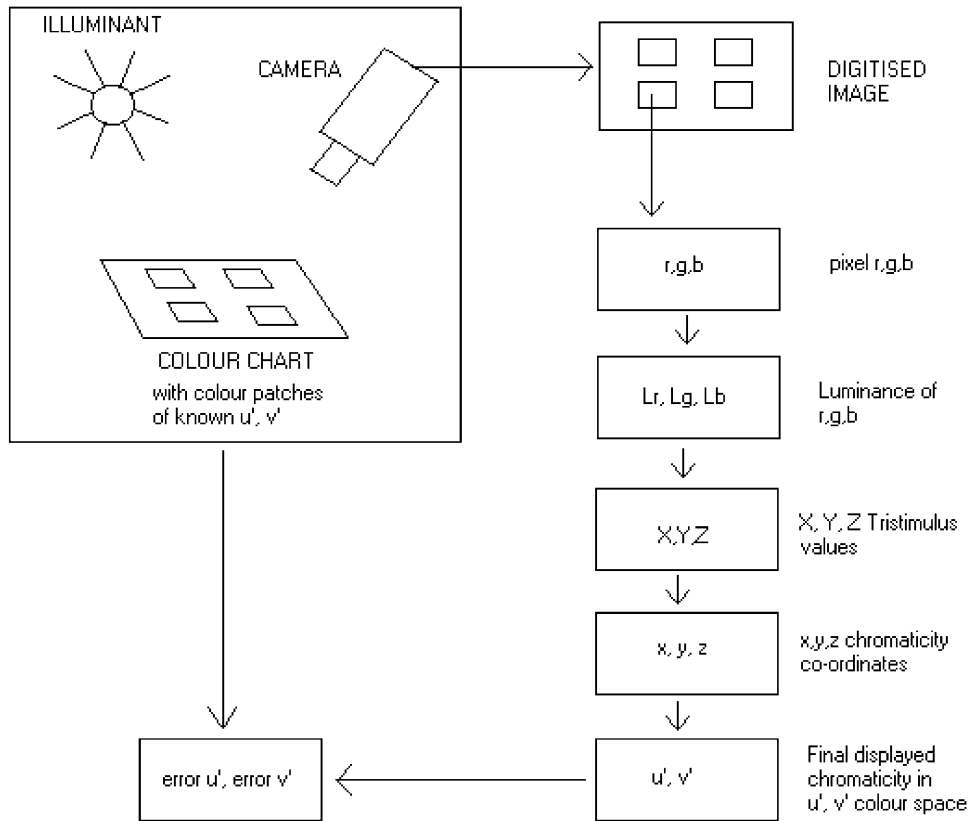


Fig. 2. Block diagram showing the measurement of colour distortions caused by camera and illuminant. The r , g , b , values of the digitised image of the Macbeth chart were read from computer memory. The luminance (L_r, L_g, L_b) of r , g , b was computed from Eq. (3), whose parameters A_i , b_i and γ had been determined from earlier calibration measurements. X , Y , Z tristimulus values were computed from Eq. (4). The x , y , z chromaticity co-ordinates were computed from Eq. (1) and the final chromaticity in (u', v') colour space was computed from Eq. (2).

where $i = \text{red, green or blue phosphor}$, $lut_i = r, g \text{ or } b$ pixel value (between 0 and 255 for our system), b_i is a constant specified by the luminance level with no applied voltage (dark level), A_i and γ (γ generally = 2.2 for CRT type displays) are constants.

The phosphor luminance values L_R , L_G and L_B were then used to calculate the device-independent CIE X , Y , Z tristimulus values that characterise the colour. It should be noted that each phosphor will emit light over a range of the visible spectrum and not just at its specified colour of red, green or blue. Therefore, the monitor phosphor spectral distributions were measured and their chromaticity co-ordinates (relative tristimulus values) determined, to indicate the relative contribution of each phosphor to the X , Y and Z tristimulus value of the displayed colour. With this information of the fundamental characteristics of the phosphors and the known phosphor luminance levels for pixels in the Macbeth chart image, the displayed tristimulus value X , Y , Z for each image pixel (and therefore each Macbeth patch) could be immediately written down in terms of the L_R , L_G , L_B

phosphor luminance as shown in Eq. (4):

$$\begin{aligned} X &= \frac{x_r}{y_r} L_R + \frac{x_g}{y_g} L_G + \frac{x_b}{y_b} L_B, \\ Y &= \frac{y_r}{y_r} L_R + \frac{y_g}{y_g} L_G + \frac{y_b}{y_b} L_B, \\ Z &= \frac{z_r}{y_r} L_R + \frac{z_g}{y_g} L_G + \frac{z_b}{y_b} L_B, \end{aligned} \quad (4)$$

where x_i , y_i , z_i are the chromaticity co-ordinates of each phosphor ($i = \text{red, green or blue}$). The chromaticity co-ordinates for a given phosphor have been divided by its y value so that the matrix coefficients are independent of luminance.

The X , Y , Z for each displayed Macbeth colour patch were converted (normalised) to x, y, z chromaticity co-ordinates and then to uniform colour space u', v' values. These final values for each colour patch in the image were compared to the manufacturer-specified u', v' values for the reflectance function of each patch, to give an estimate of the change of chromaticity caused by the specific illuminant and camera system used to obtain the image (see Table 4 and Fig. 6).

2.3. Measurement of normal and fatty liver colour

To measure the mean colour of normal and fatty livers, a series of photographs were taken of livers that had been donated for transplantation. The livers were assessed at the point of harvesting to classify them as normal or to assign a level of fattiness. The camera and illuminant were calibrated by the methods described in Section 2.2 and the colours in the images were measured on a pixel-by-pixel basis to record their distribution in u', v' colour space. The colours were corrected for distortion using the distortion measured for the Macbeth colour chart. For any given pixel u', v' point in an image of a liver, its likely error in magnitude and direction in colour space was found by simple linear interpolation between the four nearest calibration points. This interpolation method depended on the fact that the calibration points in the region of the liver colours showed approximately similar magnitudes and directions of colour errors, as shown by the six calibration points surrounding the mean liver colours in Fig. 6.

2.4. Method to determine the amount of colour change needed to distinguish between normal and fatty liver

To determine the threshold colour change between normal and fatty livers, we presented surgeons with liver images of different colours along the normal to fatty colour vector. We then measured their certainty in correctly identifying a normal and fatty liver using Receiver Operating Characteristic (ROC) analysis [18]. Fig. 3 shows the u', v' chromaticity diagram with a vector along the direction from normal (N) to fatty (F) and the threshold point T to be determined, where livers are identified as changing from normal to fatty. The distance in colour space between N and T then gave the diagnostic threshold shift in colour to compare to the colour distortions found in the imaging systems.

The methods of determining the mean colour points N and F and calculating the threshold colour by ROC analysis are described below.

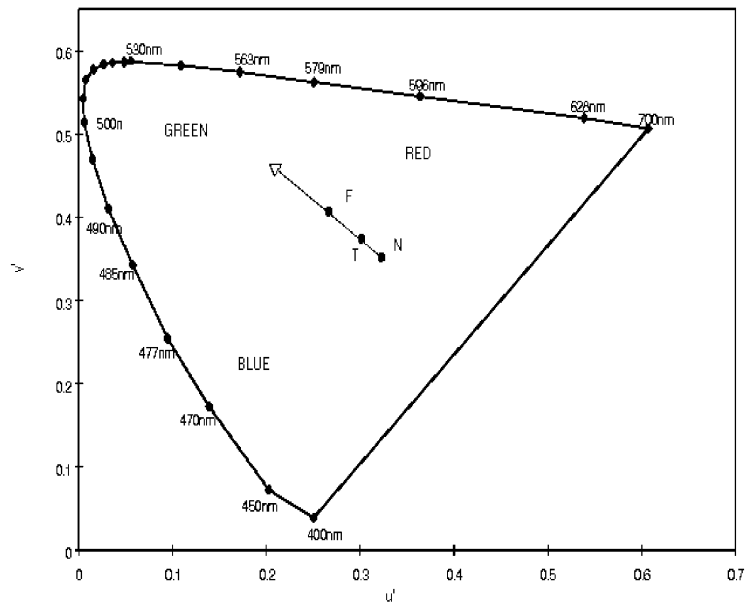


Fig. 3. CIE chromaticity diagram (for uniform colour space u', v') showing the vector colour shift between normal and fatty livers. Point N shows the mean colour of a normal liver, F shows the mean colour of a fatty liver and T is the threshold where livers are identified as changing from normal to fatty (see Section 2.4). The distance N to T was determined in the study. The diagram also shows the boundary of visible colours and the general location of red, green and blue in the colour space.

To calculate and transform the colour of liver images to any specified point in colour space, we used Eq. (4) in its inverted form shown in Eq. (5) to give the required (relative) luminance of the screen phosphors in terms of the relative X, Y, Z tristimulus value (given by the x, y, z chromaticity co-ordinates of the point)

$$L_R = \frac{\begin{vmatrix} X & \frac{x_g}{y_g} & \frac{x_b}{y_b} \\ Y & 1.0 & 1.0 \\ Z & \frac{z_g}{y_g} & \frac{z_b}{y_b} \end{vmatrix}}{D}, \quad L_G = \frac{\begin{vmatrix} \frac{x_r}{y_r} & X & \frac{x_b}{y_b} \\ 1.0 & Y & 1.0 \\ \frac{z_r}{y_r} & Z & \frac{z_b}{y_b} \end{vmatrix}}{D}, \quad L_B = \frac{\begin{vmatrix} \frac{x_r}{y_r} & \frac{x_g}{y_g} & X \\ Y & Y & Y \\ \frac{z_r}{y_r} & \frac{z_g}{y_g} & Z \end{vmatrix}}{D}, \quad (5)$$

where D is the determinant of the matrix coefficients in Eq. (4).

The relative luminance values were then scaled to give the required luminance of the displayed colour and inserted into the inverse of Eq. (3), to give the required screen r, g, b values to display the particular colour.

2.5. Measurement of diagnostic accuracy using receiver operating characteristic (ROC) analysis

To analyse the diagnostic effect of a shift in colour (incorrect identification of a normal liver as being fatty), 100 small image sections were taken at random positions from images of normal and

fatty livers (50 from each liver type). The original whole liver images from which the small image sections were taken, came from four human livers donated for transplantation (two normal and two fatty). Each small image section represented approximately an area of 3.0 cm² on the liver surface. Any features that could identify the livers had previously been carefully removed and liver edges were also excluded because ‘roundness’ was a confounding factor. The colour of the fatty liver images was first changed to be the same as normal liver. The fatty liver images were then changed again with a defined colour shift back along the direction of increasing ‘fatty’ colour. Therefore, normal liver images were presented unchanged but fatty liver images were all adjusted in colour. Each set of 50 fatty images were given the same percentage colour shift of either 0%, 10%, 20%, 30%, 40% or 50% of the distance between normal and fatty colours. The colour shift was applied on a pixel-by-pixel basis with the resulting image maintaining the original achromatic luminance level for each pixel. The 100 images were then displayed in random order (50 normal liver and 50 fatty liver images at one of the specified percentage colour shifts), surrounded by uniform grey at 30 cd/m² on a Sony Trinitron monitor to two experienced liver transplant surgeons. The process was repeated for each percentage colour shift, so that a total of 600 images were shown to each surgeon. The subjects were required to rate each image using a slider control, giving a percentage value according to their certainty that the liver was normal. Information was not given to the subjects about the origin or amount of colour shift in the images, but the subjects were told to judge the images on their normal or fatty appearance. This represented the criteria that would be used in any remote assessment of livers by electronic imaging. An ROC curve was then generated by using a threshold from 0% to 100% in unit steps and at each step the fraction of fatty colour images and normal colour images below threshold was calculated. To be consistent with the normal ROC classification, the correct identification of fatty colour images was taken as a positive test result, therefore the ‘fatty’ fraction above threshold gave the true positive fraction (TPF) and the fraction of ‘normal’ below threshold gave the false positive fraction (FPF). The ROC curve for each colour shift was plotted with TPF as ordinate and FPF as abscissa. The area under the curve was measured and the ratio of this area to maximum possible area gave the probability of a correct rating of the images [19]. The measurements were repeated with different colour shifts along the ‘normal’ to ‘fatty’ colour vector and the probability of correct identification of normal and fatty livers was calculated for each shift. The colour shift corresponding to 0.95 correct rating of the images as ‘normal’ or ‘fatty’ was taken as the minimum (threshold) colour shift required for accurate diagnosis. This threshold also represents the maximum amount of allowable colour distortion for a normal liver to be still correctly identified as being normal.

3. Results

3.1. Monitor tristimulus values and luminance

The phosphor spectral luminance distributions for the monitor (Sony Trinitron G500) used as the main image display for this project is shown in Fig. 4. These spectral distributions were measured using a Minolta CS-1000 Telespectroradiometer. The pixel phosphor r,g,b and b/w (black and white) to luminance relation, measured using a Megatron DL3/L Digital Lightmeter (photometer), is shown in Fig. 5.

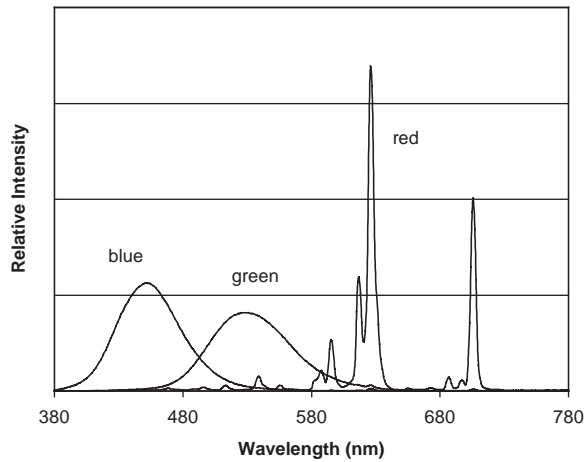


Fig. 4. Sony G500 phosphor spectral distributions.

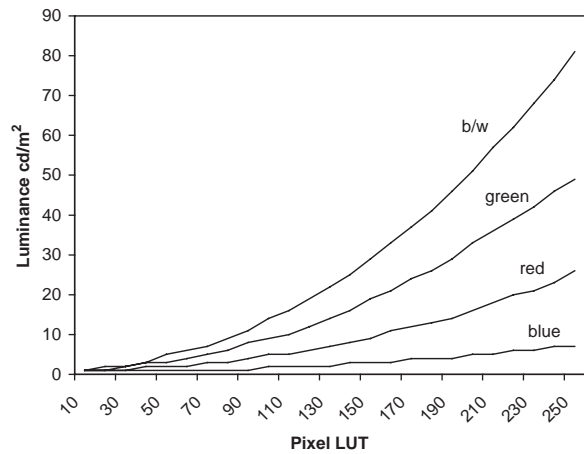


Fig. 5. Sony Trinitron G500 monitor pixel phosphor red, green, blue and black and white level (0–255) to luminance functions. These calibration values were used to derive the constants A_i , b_i and γ in Eq. (3).

Tables 1, 2 and 3 show the measured chromaticity values for the CRT monitor used in this study, a typical LCD monitor (Sony L150) and an LCD projector (Sony Proxima) display device.

The CRT monitor was connected to an Elonex MTX-6400/I computer with NT operating system and ATI Technologies 3D RAGE PRO graphics card.

3.2. Camera, illuminant and monitor colour shift

Results of the measured u', v' values of the Macbeth Colour Checker chart patches, compared to the theoretical values supplied by the manufacturer are given in Table 4 (imaged using Storz 30° Forward-Oblique Laparoscope with Storz single chip Endo-Camera and Xenon 180 W light source).

Table 1

Phosphor spectrum measurements for Sony Multiscan G500 CRT monitor used in this study, configured with factory default colour temperature of 9300 K and zero luminance offset (luminance = 0 when r, g and b = 0)

	Red	Green	Blue
x	0.6265	0.2748	0.1470
y	0.3418	0.6116	0.0647
z	0.0317	0.1136	0.7883
Maximum luminance (cd/m ²)	26	49	7
Monitor white point	$x = 0.3228$	$y = 0.3082$	

Table 2

Measured red, green and blue display, spectral distribution tristimulus values for Sony L150 LCD monitor

	Red	Green	Blue
x	0.5678	0.2805	0.1499
y	0.3307	0.5280	0.1091
z	0.0015	0.1915	0.741
Maximum luminance (cd/m ²)	26	37	8
Monitor white point	$x = 0.3402$	$y = 0.3467$	

Table 3

Measured red, green and blue display, spectral distribution tristimulus values for LCD projector type (Proxima) display

	Red	Green	Blue
x	0.61	0.207	0.141
y	0.377	0.759	0.07
z	0.031	0.034	0.789
Maximum luminance (cd/m ²)	21	68	6
Monitor white point	$x = 0.2954$	$y = 0.3470$	

The colour shifts in relation to the monitor gamuts are shown in Fig. 6. The circles represent the correct position of Macbeth chart colour patches in u', v' space. Each line indicates the amount of colour distortion from the correct colour co-ordinate, produced by the camera. The Trinitron triangle shows the position of the monitor gamut (r, g, b boundary) for the Sony Trinitron G500. In addition to the Trinitron gamut, the gamut triangle of a Sony L150 LCD monitor and a Proxima LCD projector are shown. Also shown in Fig. 6 by the filled circles are the mean colours of normal 'N' (0.284, 0.490) and fatty 'F' (0.281, 0.521) human liver images, taken using a Nikon Coolpix 995 camera under operating theatre illuminant. The Nikon Coolpix camera was also calibrated using the Macbeth chart as described in Section 2.2. The separation between each gamut is a measure of the distortion expected when image r, g, b values were correctly calculated for one display, but then displayed unchanged on the different display type. The typical magnitude of the colour shift between

Table 4
Theoretical and measured CIE u', v' mean and s.d. values of Macbeth Colour Checker Chart patches

No.	Patch name	Theor. u'	Meas. u'	Theor. v'	Meas. v'
1	Dark skin	0.25	0.273 ± 0.0058	0.492	0.484 ± 0.0031
2	Light skin	0.236	0.229 ± 0.0050	0.486	0.440 ± 0.0046
3	Blue sky	0.179	0.185 ± 0.0036	0.409	0.389 ± 0.0023
4	Foliage	0.182	0.213 ± 0.0025	0.513	0.485 ± 0.0036
5	Blue flower	0.198	0.207 ± 0.0021	0.403	0.413 ± 0.0024
6	Bluish green	0.158	0.187 ± 0.0023	0.468	0.427 ± 0.0025
7	Orange	0.294	0.292 ± 0.0036	0.533	0.525 ± 0.0042
8	Purplish blue	0.180	0.173 ± 0.0042	0.336	0.350 ± 0.0098
9	Moderate red	0.314	0.324 ± 0.0072	0.477	0.483 ± 0.0036
10	Purple	0.234	0.248 ± 0.0032	0.374	0.421 ± 0.0026
11	Yellow green	0.187	0.243 ± 0.0048	0.542	0.521 ± 0.0054
12	Orange yellow	0.258	0.278 ± 0.0033	0.539	0.523 ± 0.0028
13	Blue	0.179	0.174 ± 0.0048	0.278	0.373 ± 0.0025
14	Green	0.150	0.163 ± 0.0061	0.529	0.488 ± 0.0034
15	Red	0.379	0.371 ± 0.0045	0.496	0.507 ± 0.0002
16	Yellow	0.231	0.271 ± 0.0138	0.546	0.530 ± 0.0013
17	Magenta	0.287	0.274 ± 0.0032	0.413	0.415 ± 0.0022
18	Cyan	0.139	0.162 ± 0.0043	0.402	0.394 ± 0.0026

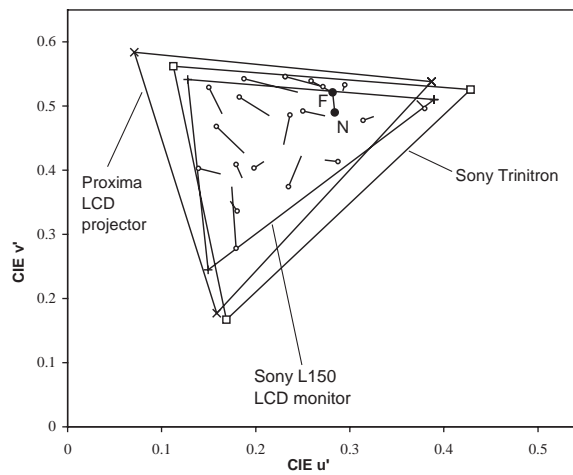


Fig. 6. Camera colour shift calibration and monitor gamuts. The mean colours for normal and fatty livers are shown by the filled circles (F fatty and N normal). The open circles are the u', v' co-ordinates of the Macbeth colours, as supplied by the manufacturer (the assumed 'correct' co-ordinates). The lines indicate the amount of colour distortion from the 'correct' colour co-ordinate.

each gamut can be seen in Fig. 6 from the relative shift of the triangle vertices for each gamut. The average shift in u', v' colour space between the monitor gamuts was found to be 0.0425, compared to the shift in liver colours of magnitude 0.0298 in an uncalibrated system. Therefore, variations

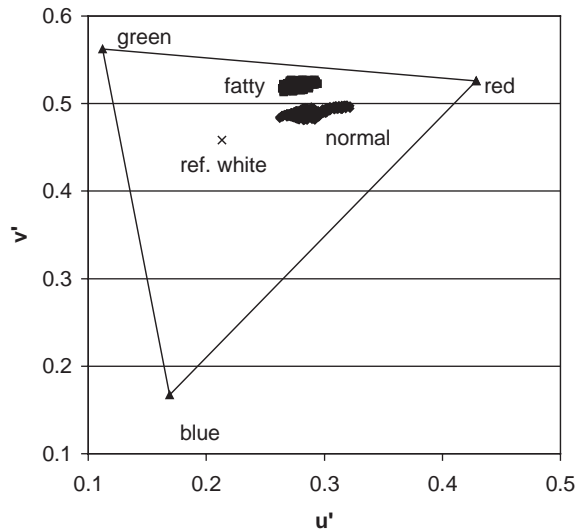


Fig. 7. Colour distribution of typical normal and fatty livers measured from 10,000 pixel points, relative to the monitor gamut and reference white.

between these three monitor gamut triangles show that replacing a display by a different type, without recalibration, can produce colour distortions larger than those produced by an uncalibrated camera and illuminant alone.

The colour distributions from the photographs of normal and fatty liver images, used in our ROC analysis are shown in Fig. 7. The reference white was taken as the monitor display white, as this colour was always visible surrounding the displayed images and was spatially invariant across the screen (measured luminance variation < 4%). A mean colour value for a normal liver and a liver with mild to moderate fat content were calculated and used to adjust the colours of images presented to surgeons for assessment by ROC analysis, as described in Section 2.5.

A calibration using the Macbeth chart was made for each illuminant and camera configuration used in this study.

3.3. ROC analysis results of diagnostic accuracy with different colour separations between normal and fatty livers

The ROC results for both surgeons (WY and RK) are shown in Figs. 8 and 9 for images of fatty liver with colour the same as normal liver colour, or shifted in 10% steps towards fatty liver colour. The percentage difference in colour between the normal and fatty liver images is marked next to each curve, with 50% difference meaning fatty liver images were presented with a colour half way between normal and the true colour of fatty livers.

The areas of the ROC curves (as a fraction of the maximum possible area) are given in Table 5 and express the probability of a correct diagnosis.

The average probability for a correct diagnosis with different colour separations between normal and fatty livers is plotted in Fig. 10. The colour separation in u', v' space between N and F was

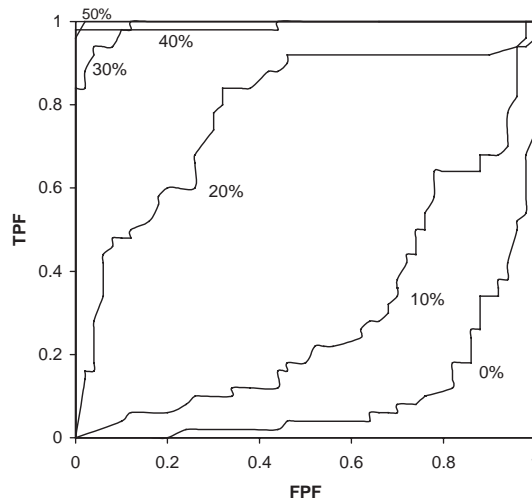


Fig. 8. ROC results for surgeon 1 (WY).

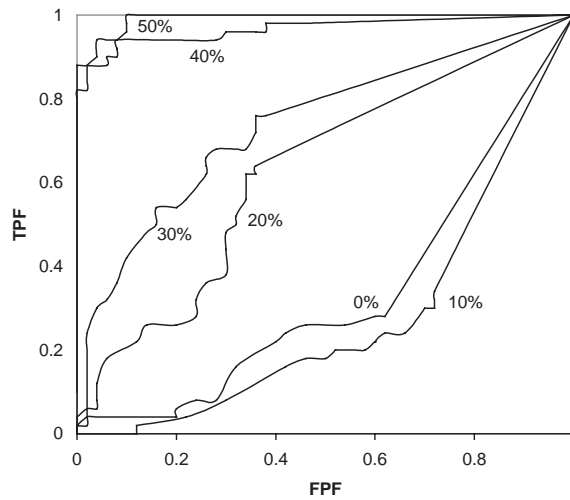


Fig. 9. ROC results for surgeon 2 (RK).

0.031 and defines the 100% level in Fig. 10. The worst direction of colour distortion from camera, illuminant and monitor for correct diagnosis, would be a shift of the colour of livers along the line of increasing or decreasing fatty colour. For normal livers, if the distance to the threshold is reduced, the probability of a correct diagnosis decreases as the colour position approaches threshold. The probability of correct diagnosis as a function of colour distortion in u', v' space along the normal to fatty colour line (given by 1.0 minus the function shown in Fig. 10) is given in Fig. 11.

Table 5

Probability of correct diagnosis for colour differences between normal and fatty livers, as a percentage of the true colour difference between normal and ‘mild to moderate’ fatty livers. A plot of this data is shown in Fig. 10

Difference in colour of fatty image from normal (as percentage of normal to mild/moderate fatty colour change)	Distance (magnitude) in u', v' space	Probability of correct diagnosis surgeon 1 (WY)	Probability of correct diagnosis surgeon 2 (RK)
0	0	0.09	0.33
10	0.0031	0.29	0.27
20	0.0062	0.78	0.62
30	0.0093	0.99	0.74
40	0.0124	0.99	0.96
50	0.0155	0.99	0.99

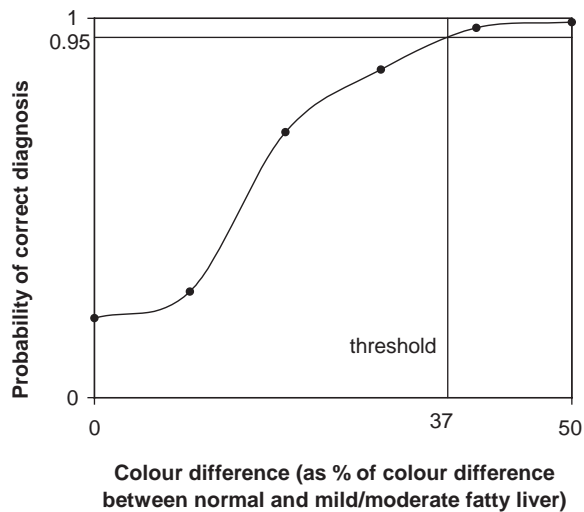


Fig. 10. Average probability of a correct diagnosis with a displayed colour difference between normal and fatty livers, from the data shown in Table 5.

4. Discussion

The objectives of this study were to determine the magnitude of colour distortion likely when using an uncalibrated illuminant, camera and monitor, determine the required colour difference for correct identification of normal and fatty livers and conclude whether colour distortions from uncalibrated systems are likely to affect diagnosis or not. If colour distortions from illuminants, cameras and monitors are small compared with the colour differences required for correct diagnosis, then it is likely that systems can remain uncalibrated without adversely affecting diagnostic outcome. The results presented in Section 3.2 show that colour distortion from a typical CCD camera and illuminant in medical applications does not vary in a systematic way over the whole colour space, but may exhibit

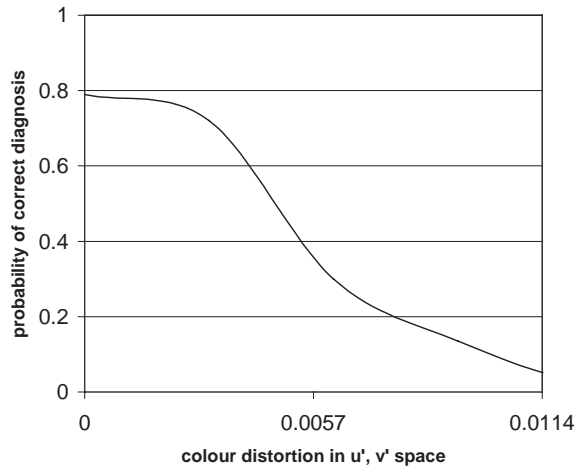


Fig. 11. Probability of correct identification of a normal liver as the colour is distorted towards the colour of fatty livers.

some localised areas with similar colour distortions. As a consequence, the applied interpolation scheme to correct for errors in colour away from the calibration points must depend on the pattern of errors. For results similar to ours it may be necessary to apply different interpolation schemes in different areas of the colour gamut. Any interpolation scheme that attempts to use all calibration points regardless of the pattern of errors may have little value and could produce errors as large as the effects being studied. The average colour distortion for the camera and illuminant used in this study was $\Delta u' = 0.0175$ and $\Delta v' = 0.0242$, calculated from the values given in Table 4. Therefore the average magnitude of colour error = 0.0298 in u', v' space (average error = 0.0301 for the eight closest calibration points to the mean liver colour), but was effectively in a random direction in the colour space.

The mean colours of livers were: for normal livers $u' = 0.284$, s.d. ± 0.026 , $v' = 0.490$, s.d. ± 0.005 and for fatty livers $u' = 0.281 \pm 0.015$, $v' = 0.521 \pm 0.010$, giving a colour difference of 0.031 in magnitude. At the co-ordinate position of normal livers, the magnitude of colour shift¹ for an LCD monitor relative to a CRT Trinitron monitor = 0.015 and an LCD projector relative to a CRT Trinitron monitor = 0.041. It is also interesting to note from Fig. 6, in the case of the LCD monitor, the mean value colour of fatty livers and several Macbeth patch colours are either on or over the colour gamut boundary. This suggests that 'clipping' of the colour distribution of fatty livers or other colours near the gamut boundary is likely because of the wide spread of colours as shown in Fig. 7, which is typical of colour distributions in natural images [20].

From the ROC analysis, the threshold colour difference in u', v' space required to enable correct identification of liver types = 0.011. This is approximately 15 times larger than the colour difference that is just notable to the human observer, at the liver colour [16].

¹ Relative colour shift was the change in colour found when displaying a pixel r,g,b value on a system calibrated for a given monitor and then replacing that monitor with another display device, without recalibration. The actual colour shift values were calculated by applying each monitor calibration with the mean normal liver colour in Eq. (4) and comparing the results.

The colour distortion from camera or monitor may be in any direction in colour space, with the direction producing most diagnostic distortion along the line of normal to fatty liver colour. Along this colour direction, the camera distortion (0.0298) was approximately three times the threshold colour difference (0.011) for correct diagnosis. For CRT to LCD display colour shift, the distortion is between one and four times the threshold distance, depending on the LCD display type. This indicates that the colour distortions from an uncalibrated camera, illuminant and monitor are larger than the colour differences used in identifying liver types. This also implies that colour calibration is likely to be critical in cases when colour distortion is along the line of colour change from normal to fatty.

For the best calibration of a working system, the colour calibration chart should contain as many colours as possible spread over the range of reproducible colours in the system. For example, the GretagMacbeth [17] 'Colorchecker DC' chart contains 237 patches with 177 different colours, to produce a suitable spread of colours for an effective calibration. A more specialised colour scale to help visual assessment of liver fatty content can easily be generated on a calibrated monitor. Ideally, this scale would be presented next to the liver images being assessed and would show colours that progressively change from normal to fatty, to help diagnosis using direct colour comparison.

5. Summary

The principle objective of this study was to investigate whether diagnosis from medical images based on colour information was adversely affected when cameras, illuminants and monitors were incorrectly calibrated or remain uncalibrated. The study aimed to establish the requirements for diagnostic accuracy when assessing images of livers for transplantation. This is a typical example of a telemedicine application where colour is considered important for correct diagnosis. The results show that typical colour distortions that result from the use of an uncalibrated CCD camera and illuminant can be significantly larger than the colour difference thresholds required for accurate diagnosis. In this study, images presenting a correct diagnostic rate of 95% were considered to be reliable for 'accurate' diagnosis. Further, the directions of colour distortion showed no systematic variation, suggesting that colours within a single image may be given random and opposite colour distortions.

In the case of uncalibrated or incorrectly calibrated monitors, the colour distortions were systematic with respect to distortion direction, but the distortion magnitude varied between 1.3 and 3.7 times the threshold distance of correct diagnosis. Therefore, we conclude that accurate colour calibration and reproduction on monitors is critical if diagnostic accuracy is to be maintained. Colour distortions in uncalibrated systems can be several times larger than the colour differences used in assessing normal and steatotic livers, a task where colour is a critical component in the assessment process. By chance, uncalibrated colour acquisition systems may not affect the diagnostic process, depending on the system distortion and image colour, but only accurate calibration and reproduction can guarantee diagnostic accuracy in colour critical applications.

References

- [1] C.E. Price, D. Lowe, A.T. Cohen, F.D. Reid, G.M. Forbes, J. McEwen, R. Williams, Prospective study of the quality of life in patients assessed for liver transplantation: outcomes in transplanted and not transplanted groups, *J. R. Soc. Med.* 88 (3) (1995) 130–135.

- [2] The report of the working party to review organ transplantation, The Royal College of Surgeons of England, 1999, <http://www.rcseng.ac.uk>.
- [3] T. Fukumori, N. Ohkohchi, S. Tsukamoto, S. Satomi, Why is a liver with steatosis susceptible to cold ischemic injury, *Transplant. Proc.* 31 (1999) 548–549.
- [4] A.M. Seifalian, V. Chidambaram, K. Rolles, B.R. Davidson, In vivo demonstration of an impaired microcirculation in steatotic human liver grafts, *Liver Transplant. Surg.* 4 (1) (1998) 1–8.
- [5] M. Herbin, F.X. Bon, A. Venot, F. Jeanlouis, M.L. Dudertret, L. Dubertret, G. Strauch, Assessment of healing kinetics through true color imaging processing, *IEEE Trans. Med. Imaging* 12 (1) (1993) 39–43.
- [6] G.L. Hansen, E.M. Sparrow, J.Y. Kokate, K.J. Leland, P.A. Iaizzo, Wound status evaluation using color image processing, *IEEE Trans. Med. Imaging* 16 (1) (1997) 78–86.
- [7] C.H. Li, P.C. Yuen, Regularized color clustering in medical image database, *IEEE Trans. Med. Imaging* 19 (11) (2000) 1150–1155.
- [8] M. Nischik, C. Forster, Analysis of skin erythema using true color images, *IEEE Trans. Med. Imaging* 16 (6) (1997) 711–716.
- [9] P. Schmid, Segmentation of digitized dermatoscopic images by two-dimensional color clustering, *IEEE Trans. Med. Imaging* 18 (2) (1999) 164–171.
- [10] <http://www.cie.co.at>.
- [11] J. Miano, *Compressed Image File Formats*, Addison-Wesley, Reading, MA, 1999.
- [12] Y.V. Haeghen, J.M.A.D. Naeyaert, I. Lemahieu, W. Philips, An imaging system with calibrated color image acquisition for use in dermatology, *IEEE Trans. Med. Imaging* 19 (7) (2000) 722–730.
- [13] International Electrotechnical Commission (IEC) 61966-2-1, *Multimedia systems and equipment—colour measurement and management—Part 2-1: colour management—default RGB colour space—sRGB*, 1999, <http://www.iec.ch/>
- [14] R. Reháč, P. Bodrogi, J. Schanda, On the use of the sRGB colour space, *Displays* 20 (1999) 165–170.
- [15] R.W.G. Hunt, *Measuring Colour*, Ellis Horwood Ltd., Chichester, England, 1987.
- [16] G. Wyszecki, W.S. Stiles, *Colour Science: Concepts and Methods, Quantitative Data and Formulas*, Wiley, New York, 1964.
- [17] <http://www.munsell.com/munchcheck.htm>.
- [18] A.R. van Erkel, P.M.T. Pattynama, Receiver operating characteristics (ROC) analysis: basic principles and applications in radiology, *Eur. J. Radiol.* 27 (1998) 88–94.
- [19] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Diag. Rad.* 143 (1) (1982) 29–36.
- [20] M.A. Webster, J.D. Mollon, Adaption and the color statistics of natural images, *Vision. Res.* 37 (23) (1997) 3283–3298.

Darryl A. de Cunha is a Research Fellow in the Applied Vision Research Centre at City University, London. His research interests include colour vision, stereopsis, computer imaging and telemedicine. He has an M.Sc. in Astrophysics from London University and a Ph.D. in Visual Science from City University. He is also a member of the Applied Vision Association, the Colour Group (of Great Britain) and the IEEE.

Leila Eadie received her Master of Sciences degree in Clinical Neuroscience from the Roehampton Institute in 1996, and is currently studying for a Ph.D. at the Royal Free and University College Medical School's Department of Surgery. Her research interests include the assessment of livers for transplantation using colour science and ultrasound, and the use of ultrasound for surgical guidance when splitting donor livers for two recipients. To this end she is creating a system to link planning ultrasound with live intraoperative images.

John Barbur is Professor of Optics and Visual Science and Director of the Applied Vision Research Centre at City University. His interests cover both fundamental studies of visual mechanisms as well as applied and clinical research. During the past 25 years he has pursued the development of research instrumentation and novel measurement techniques and this has resulted in new methods and instrumentation for analysis of pupil response components, spatial vision and chromatic sensitivity and the measurement of scattered light in the eye. He is a Fulbright Scholar and spent time as

Visiting Professor at the Center for Visual Science at the University of Rochester, NY where he worked on colour vision and visual performance in the mesopic range. The studies carried out over several years by John and colleagues at City University formed the basis for the formation of the Applied Vision Research Centre in 1986. The AVRC has since then developed as an important centre for both fundamental and applied vision research.

David J. Hawkes is Professor of Computational Imaging Science and Chairman of the Division of Imaging Sciences at Guy's Hospital, Kings College London. He graduated in Natural Sciences (Physics) from Oxford in 1974 and obtained his Ph.D. in X-ray computed tomography in 1981. His current research interests include image matching, data fusion, visualisation, shape representation, surface geometry and modelling tissue deformation with applications in image guided interventions, augmented reality in surgery, 3D ultrasound and interventional MRI. He is a Fellow of the Royal Academy of Engineering, Institute of Physics and Institute of Physics and Engineering in Medicine.

Alexander M. Seifalian is Reader in Biophysics and Tissue Engineering in the Department of Surgery, Royal Free Hospital, University College London. He was trained at London University (KCL) and obtained a Ph.D. from Royal Free Hospital School of Medicine. He is interested in the development of tissue engineering organs and his current projects include teleplanning surgery, application of optics in assessment of tissue oxygenation and microcirculation, application of ultrasound in assessment of cardiovascular risk vascular and, cardiovascular and hepatic haemodynamic. He currently supervises three postdoctoral and eight Ph.D. students.