EG UK Theory and Practice of Computer Graphics (2009)
Wen Tang, John Collomosse (Editors)

Calibrating a COTS Monitor to DICOM Standard

I. J. Grimstead[1] and N. J. Avis[1]

[1]Cardiff School of Computer Science, Cardiff University, Wales, UK

Abstract

We present a method for calibrating a commodity, off-the-shelf (COTS) monitor (costing in the region of £200) to produce a greyscale image approximately calibrated to the DICOM standard, rather than require a 10-bit radiology monitor (costing in the region of £10,000). We use the concept of PseudoGrey to extend the available shades of grey from 256 to 5,800, which is in excess of a 12-bit greyscale. The chromaticity of the resulting greyscale is analysed to verify that the colour introduced does not unduly detract from a pure greyscale image. The behaviour of low intensity levels in the COTS monitor is also analysed, showing that a naive approach to estimating luminance from individual passes through the red, green and blue components is insufficient to produce an accurate intensity range. The results show that we can achieve a basic DICOM calibration (with FIT and LUM tests), but we have yet to test for further variability (such as off-axis deterioration in brightness or inconsistent luminance across a display). As well as displaying medical images, this approach may be of use in other areas requiring a high dynamic range, such as thermal imagery or images taken through multiple alternative exposures.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Viewing algorithms I.4.3 [Computer Graphics]: Greyscale manipulation J.3 [Life and Medical Sciences]: Medical information systems

1. Introduction

For medical diagnosis, high image fidelity is required; otherwise, an incorrect diagnosis could result from an imaging artefact. When reviewing radiology images (namely: X-Ray or CT scans), radiologists use monitors calibrated to Digital Imaging and Communications in Medicine (DICOM) standard. The calibration is to the Grayscale Standard Display Function (GSDF), as defined in Part 14 of the DICOM standard [NEM08].

Monitors capable of being calibrated to the DICOM standard are specialised, available from manufacturers such as BARCO, NEC and National Display Systems. Such monitors are specialised and hence expensive (in the region of £10,000 compared to £200 for a commodity, off-the shelf monitor), restricting their use in hospitals to specialised workstations. The monitors are expensive as they are extremely bright (with high contrast) and can generate a 10-bit or 12-bit greyscale; namely, 1,024 to 4,096 shades of grey. With all but the latest graphics cards, the image is rendered internally to an 8-bit greyscale, and sent to the monitor. The monitor has been calibrated to select 256 shades of grey from its available range, the 256 shades carefully selected to produce a DICOM calibrated intensity curve.

Standard (COTS) monitors and graphics cards only produce an 8-bit greyscale, producing 256 shades of grey from a palette of 16 million colours. This is because to produce a grey shade, all three component colours (red, green, blue) are set to the same digital drive level. Hence red = green = blue; each channel has 8-bit resolution, so only an 8-bit greyscale can be produced.

We have recently published an approach [GAEB09] where the concept of PseudoGrey [TCL*92] has been applied to calibrate a COTS monitor to reproduce a near-perfect linear intensity curve. This is an improvement on the standard "Gamma Correction" which attempts to correct for the non-linear relationship between digital drive level (i.e. logical intensity) and produce a linear mapping. In this paper we use the calibrated PseudoGrey approach to produce a DICOM calibrated intensity curve and hence raise the possibility that a COTS monitor could be used for radiological imaging.

An overview of previous work is presented in the next section, followed by a description of our implementation of the calibration. Our results are then discussed, followed by our conclusions and future work.

DOI: 10.2312/LocalChapterEvents/TPCG/TPCG09/199-206

2. Previous Work

Previous work is divided up into work on PseudoGrey, and general work on monitor calibration.

2.1. PseudoGrey

The concept of "Pseudo Grey" [TCL*92] enables a standard monitor to display more than 256 shades of grey. This is achieved by using additional off-white shades of grey which are perceived as "grey" by the observer, but exhibit fractional levels of intensity between consecutive "pure" grey shades. Hence the display is no longer restricted to 256 levels of grey. For instance, rather than using red, green, blue values of (100,100,100), we may select (100,101,101) to produce a slightly brighter shade of grey. Consider the conversion from colour to greyscale using the YUV colour space presented in Equation 1.

$$Y = (0.299 \times R) + (0.587 \times G) + (0.114 \times B) \quad (1)$$

This equation defines the relative weightings of the colours in terms of luminance (brightness) (Y). From this, it can be seen that the blue component is weighted approximately at 11%, red 29% and green 58%. So, from a given pure grey colour (R,G,B) (where $R = G = B$), changing to $(R,G,B+1)$ would produce a luminance that was roughly 11% between (R,G,B) and $(R+1,G+1,B+1)$. In other words, incrementing just the blue channel would produce an luminance interpolation of 11% between two consecutive shades of pure grey. By using alternative combinations of red, green and blue, fractional increments of luminance can hence be generated between two consecutive shades of pure grey.

2.2. Monitor Calibration

To calibrate a monitor to the DICOM GSDF standard, there are many levels of calibration that can be applied. At the individual pixel level, a monitor should produce consistent chromaticity and luminance. It has been shown [SFE99] that non-uniform luminance (i.e. noise) is disruptive to the diagnostic process; hence high-end medical monitors such as those from BARCO are calibrated at the pixel level at the factory, to ensure uniform intensity. Dallas et al. [DRF*09] have used a high-resolution camera to determine near-pixel-sized components of fixed pattern noise. A "noise map" is then generated, where each pixel has a luminance offset, so the digital drive levels of the monitor can be adjusted per pixel to produce a uniform intensity.

The non-uniform relationship between Digital Drive Level (DDL - the logical intensity output sent to a monitor) and physical luminance is also attended to with medical grade monitors. A photometer is attached to the screen to accurately read the luminance in response to requested DDLs. A map can then be made from available DDLs and physical luminance; given the map, any grey scale intensity sequence can be produced. For medical grade monitors, 10-bit or higher resolution greyscales are available, where 256 shades of grey are selected for use by the host workstation. Note that software is often limited by the host operating system and standard graphics APIs, so (for instance) a greyscale image is commonly handled as a 256 entry palettised image, hence selecting 256 greys. Latest developments enable a full 10-bit workflow [Xth08] without being restricted to 8-bit greyscales.

Rather than use hardware to produce the additional fraction levels of luminance, an alternative approach is to use PseudoGrey [GAEB09] to produce a 256 entry greyscale palette. This preliminary work was targeted at improving the perceived image quality in the operating theatre from existing software. A custom UltraVNC client received the screen information from a remote computer and modified its greyscale to use a calibrated PseudoGrey palette. The calibrated client produced a near perfect linear intensity curve, performing a perfect gamma correction. However, the display was not calibrated for DICOM, which is attended to in this paper.

3. Implementation

In this section we describe the implementation of our DICOM calibration for a COTS monitor using PseudoGrey.

3.1. Sampling Using Three Passes of Red, Green and Blue

In order to produce the fractional shades of grey, we need to know the luminance produced from each intensity of red, green and blue. In theory, these can then be selected and added together to determine the luminance this would produce. Hence we can scan 3×255 intensities, one each of red, green and blue, rather than scan each shade of PseudoGrey which is many thousands.

Taking an offset limit example of $0 \dots 2$ per channel from the "source" pure grey, this produces values such as (R,G,B), $(R,G,B+1)$, $(R,G,B+2)$, $(R,G+1,B)$, $(R,G+1,B+1)$, etc. We then restrict the selection of each offset such that the intensity of $(R+r,G+g,B+b)$ is less than $(R+1,G+1,B+1)$ (i.e. the next "pure grey" after (R,G,B)). This ensures that we produce incremental steps between two neighbouring shades of pure grey, and produces 12 additional fractional steps. Given that we can interpolate between 0..1, 1..2, ..., 253..254, but we cannot use an offset of $+2$ with intensity value of 254 ($254+2 = 256$, which is larger than maximum available intensity of 255). To interpolate between 254..255 we are limited to an offset range of 0..1, which gives 7 different shades of PseudoGrey. Hence we have a total of $((256-2) \times 12) + (1 \times 7) + 1 = 3{,}056$ possible shades of PseudoGrey for an offset limit of $0 \dots 2$.

Initially, we sampled each "pure" red, green and blue intensity. We then compared this to a "pure" white intensity.

In theory, the sum of intensities from $R+G+B=W$. To verify this, we sampled the luminance output of each pure red, green, blue and white intensity with DDLs in the range 0..255 using a Konica-Minolta Colour Spectrometer (CS-200). The CS-200 is accurate to $L_v \pm 2\%$, and $x, y \pm 0.003$. A DELL AS200-1905FP LCD monitor was positioned 2m away from the spectrometer, and both were positioned inside a blacked-out frame to remove the influence of ambient light. The CS-200 was set to use a measuring angle of 1°, and focussed on the monitor.

The darkest luminance measured for each of red, green, blue and white was found, and assumed to be the ambient component. This was subsequently subtracted from each sample, and then entered into the graph as presented in Figure 1. In this graph, $R+G+B$ represents the sum of the luminance (measured in lumens) of the individual red, green and blue pure colours of a given intensity, and compared against the white of the same intensity, both measured against the left-hand Y axis. To aid in the visual analysis, white luminance was divided against $R+G+B$ luminance of the same DDL to provide a relative comparison; these values are plotted as crosses and are measured against the right-hand Y axis. It can be seen that the majority of the values are outside a relative value of 98%..102%, that is, outside the measurement error of ±2% (from the spectrometer). This graph shows that there is a slight error between the sum and pure white (which is not accounted for by the 2% error in luminance sampling), so our initial hypothesis that we can sample red, green and blue in isolation to perfectly model the resulting light output appears invalid. To further verify the potential error, the luminance measured for each red, green and blue colour is calculated as a ratio against the white of the same DDL. This should produce a constant factor (as in Equation 1, where the weights of individual red, green and blue determine the overall luminance). This ratio is presented in a graph in Figure 2, where the ratio of *red/white*, *green/white*, *blue/white* and the sum of the three ratios is displayed. In theory, the luminance of the sum of the three colours should equal the luminance of white—as white is simply a combination of the three colours. From Figure 2, it can be seen that the darker shades (below ~ 40 DDL) do not conform to our expectation; the sum of the ratios is at peak 3× what it should be. Also note, that the sum does not stabilise on the value 1; rather, it dips to a low of ~ 0.9 and slowly rises again back to 1.0 around a DDL of 200. This shows inherent instability in a COTS monitor at low DDLs. Hence we cannot simply analyse 3 × 255 colours to gather sufficient data to produce our PseudoGrey profile.

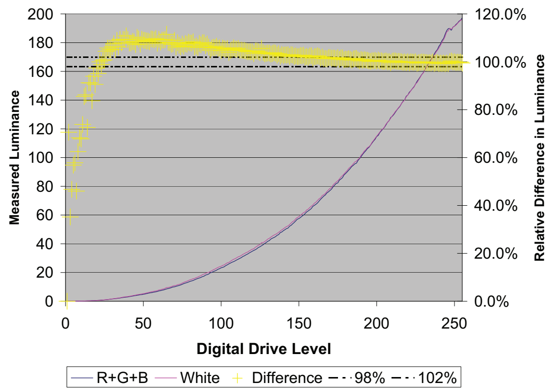

Figure 1: *Analysis of the sum of red, green and blue luminance against pure white luminance*

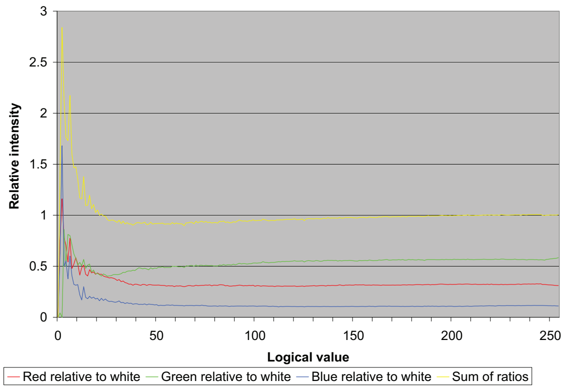

Figure 2: *Analysis of ratio of red, green and blue intensity against pure white intensity.*

3.2. Sampling Using Full PseudoGrey Range to Produce a DICOM Calibration

Given the issues regarding darker intensities of colour, we have to directly sample each PseudoGrey in turn to record a series of known luminescences. This will then capture any inherent discontinuity in produced luminance from a given monitor. This unfortunately results in far longer processing; a full range of 5,817 shades of PseudoGrey took 1 hour 6 minutes to sample.

Once we have recorded the luminance against each PseudoGrey entry, we then determine the brightest and darkest entries. Using the concept of "Just Noticeable Difference" from the DICOM standard [NEM08], we define the JND value from a given luminance (in lumens) in Equation 2. $j(L)$ returns the JND of a given luminance value L (in lumens), and A, B, C, D, E, F, G, H and I are constants defined in the standard (refer to Section 7.1 General Formulas on page 12 of the standard).

To produce a DICOM calibration, the displayed shades of grey must be at least 1 JND apart, for the shades of grey to be just noticeable by a human observer. Depending on the brightness of the monitor, it may be capable of displaying in excess of 256 JNDs. However, we restrict ourselves to 256 shades of grey, so distribute the intensities evenly throughout the JND space. To do this, we simply linearly interpolate between the lowest and highest JND value, and use the inverse of Equation 2 to produce a target lumen value from the given JND. We then select the PseudoGrey that has the nearest luminance to our desired value.

$$\begin{aligned} j(L) = {} & A + B \cdot Log_{10}(L) + C \cdot (Log_{10}(L))^2 \\ & + D \cdot (Log_{10}(L))^3 + E \cdot (Log_{10}(L))^4 \\ & + F \cdot (Log_{10}(L))^5 + G \cdot (Log_{10}(L))^6 \\ & + H \cdot (Log_{10}(L))^7 + I \cdot (Log_{10}(L))^8 \end{aligned} \quad (2)$$

3.3. Alternative Sets of PseudoGrey

PseudoGrey uses a range of red, green and blue steps in DDLs away from pure white. Given that (for instance) blue produces approximately 11% of the luminance for a pure white, we can use up to nine steps of blue to produce fractional 11% steps in luminance between two consecutive pure whites. Similarly, up to two steps in red and a single step in green can be used. This gives us several variations, such as just using blue, or a combination of all three colours. Three scenarios are presented in Figure 3. The "offset" in a single channel is restricted to the range 0..1, 0..2 and 0..9. It can

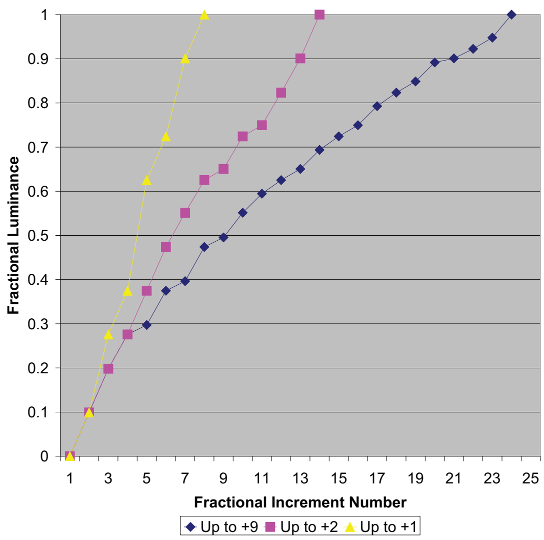

Figure 3: *Comparison of offset ranges to produce PseudoGrey steps between pure shades of grey*

be seen that the range 0..9 produces far more interpolation steps between neighbouring shades of pure grey, but at the expense of potentially introducing a visible colour tint to the ideal shade of grey. Table 1 shows the number of steps available.

PseudoGrey Range	0..9	0..2	0..1	Pure Grey
Number of Steps	23	13	7	1

Table 1: *Number of steps between neighbouring shades of pure grey*

4. Results

We now present and analyse our results of using PseudoGrey to produce a DICOM curve. The next section reviews the results of formal DICOM measurement of fit, followed by chromaticity analysis to determine how much colour we have introduced to a logically "pure" grey.

4.1. Direct 0..1735 PseudoGrey Approach and Variations

According to the DICOM standard [NEM08], the formal measurement of fit (to the DICOM GSDF) is defined in two stages:

1. FIT test. A graph of JNDs per luminance interval vs. luminance interval should have a horizontal line fitting the data. A third-order fitting curve should be used, which should be linear.
2. LUM test. Luminance uniformity; are the steps in the JNDs/luminance interval of uniform perceptual size? Measured by the RMSE (Root Mean Square Error) of the horizontal line fit.

We start with the FIT test, showing the number of JNDs between consecutive DDLs using different ranges of PseudoGrey. PseudoGrey range 0..9 is presented in Figure 4, range 0..2 in Figure 5, range 0..1 in Figure 6 and range 0..0 (i.e. pure grey) in Figure 7. The horizontal axis shows the DDL (i.e. logical intensity), and the vertical axis is the JND between consecutive DDLs (i.e. the number of JNDs between logical grey intensities). A third-order curve is also fitted through the plotted points, with its equation shown in the bottom-right hand corner of each graph.

To reach DICOM standard, the JNDs between DDLs must be at least 1 to produce a display that a human can differentiate between shades of grey. For a more refined display, a small spread of JNDs which are all as large as possible; this produces a consistent, high-contrast image. It can be seen that pure grey is unacceptable for DICOM calibration, given the number of 0 JND differences (hence located on the DDL axis) between consecutive DDLs. The range 0..1 has a few outliers that are very close (but larger than) a JND of 1.0. This is of concern, as such a point could fall below 1.0 and hence two shades of grey are indistinguishable by a human observer. However, the range 0..9 has a very tight fit, with all DDLs producing a JND in the range 1.7...2.5.

The third-order polynomials fitted to each PseudoGrey range shows a near-horizontal fit, with the x component less

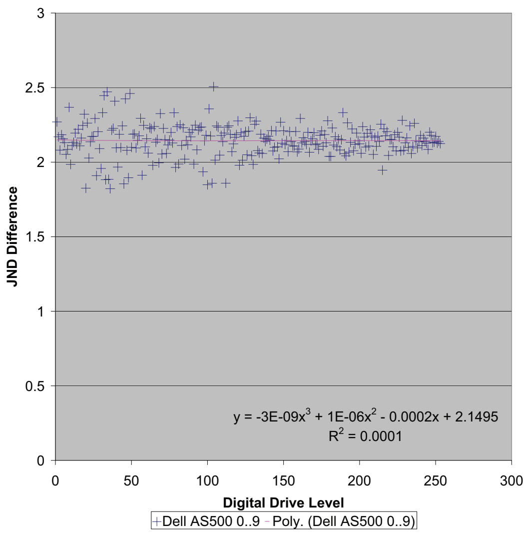

Figure 4: *JND differences between consecutive digital drive levels using PseudoGrey offset range 0..9*

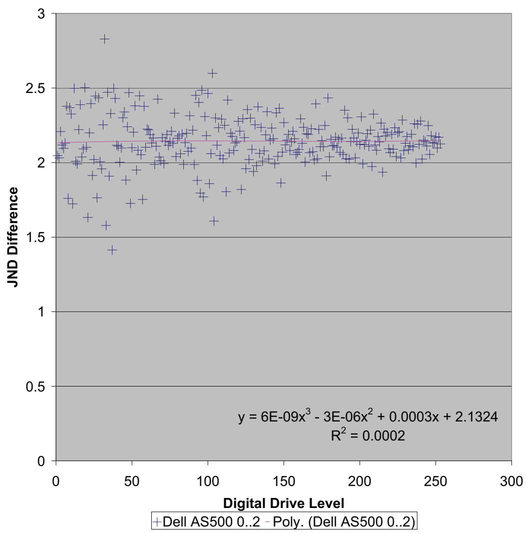

Figure 5: *JND differences between consecutive digital drive levels using PseudoGrey offset range 0..2*

than one-thousandth of a JND. Hence we have met the DICOM FIT test. Following the FIT test, we now carry out the LUM test. Table 2 presents statistical analysis of the different PseudoGrey ranges, including pure grey. *SD* is the Standard Deviation, and *RMSE* is the Root Mean Squared Error, comparing out measured luminance and comparing against desired luminance from the DICOM GSDF. The Pseudo-

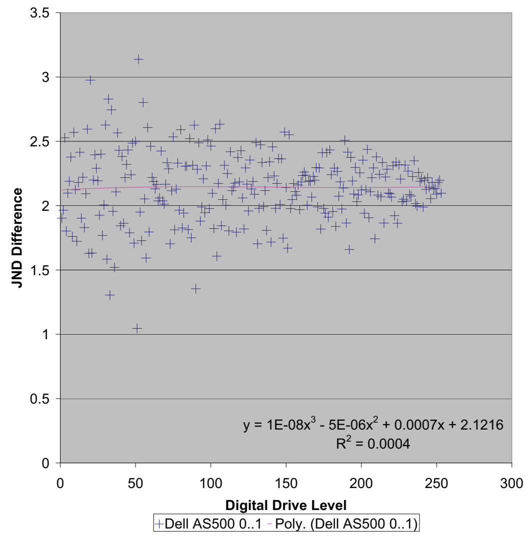

Figure 6: *JND differences between consecutive digital drive levels using PseudoGrey offset range 0..1*

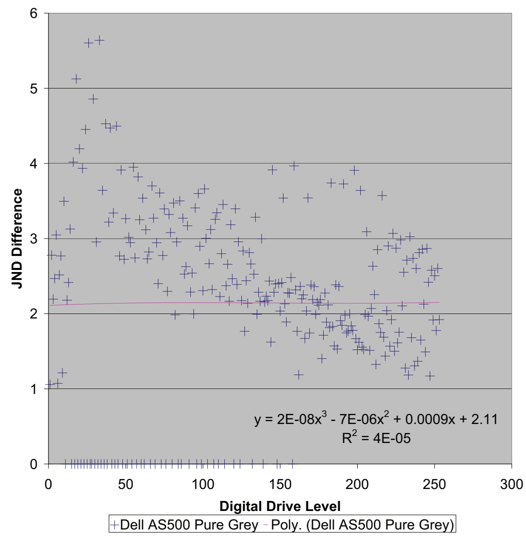

Figure 7: *JND differences between consecutive digital drive levels using pure grey*

Grey ranges 0..9 and 0..2 both show a reasonable fit and hence candidates for a DICOM display. The remaining question is whether we have introduced perceptible colour into a logical grey colour.

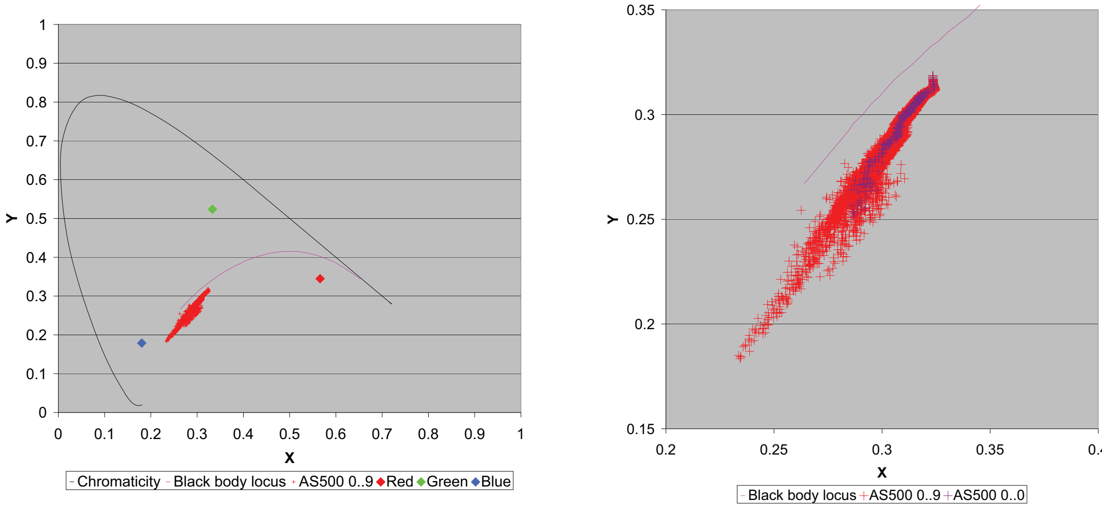

Figure 8: *Analysis of DELL AS500 monitor using PseudoGrey offset range 0..9, showing full view and close-up with pure grey superimposed*

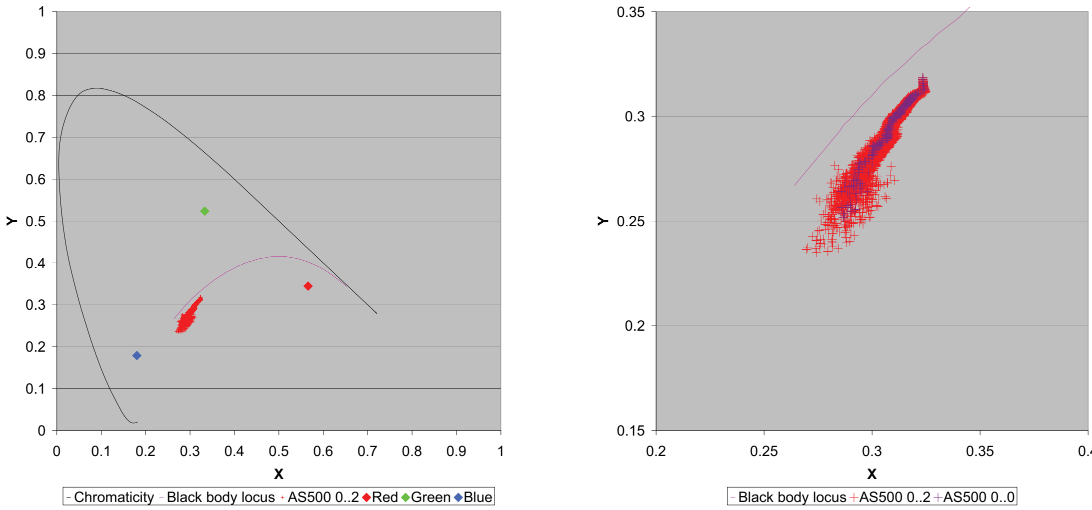

Figure 9: *Analysis of DELL AS500 monitor using PseudoGrey offset range 0..2, showing full view and close-up with pure grey superimposed*

4.2. Chromaticity Analysis of Direct 0..1735 PseudoGrey Approach and Variations

In this section, we plot all of the available PseudoGrey shades for a given range (such as 0..9) into a 1931 CIE Chromaticity chart. Pure red, green and blue produced with maximum DDLs are presented (to show the limit of the display's colour gamut), along with the black body locus (where our shades of grey should lie). Range 0..9 is presented in Figure 8, range 0..2 in Figure 9 and range 0..1 in Figure 10. Each figure contains a pair of graphs; the left-hand graph shows the entire 1931 CIE chromaticity chart, whereas the right-hand graph shows a close-up of the area of interest (same scale axis for each figure). Note that pure grey is also plotted for comparison on the close-up charts, to reveal how much PseudoGrey varies from pure grey (where $R = G = B$). PseudoGrey range 0..1 does not exhibit a large deviation in chromaticity from pure grey; range 0..2 appears to widen

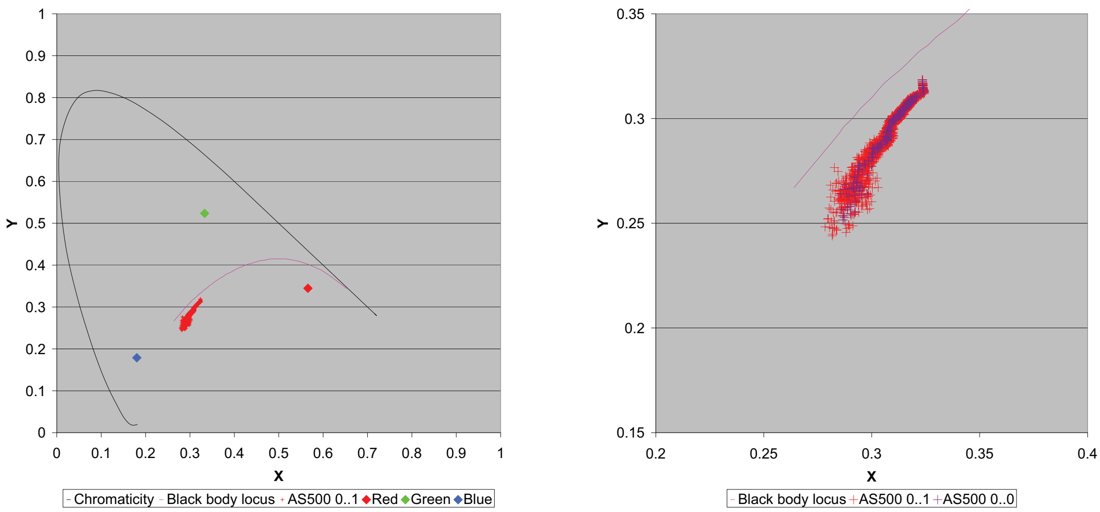

Figure 10: *Analysis of DELL AS500 monitor using PseudoGrey offset range 0..1, showing full view and close-up with pure grey superimposed*

PseudoGrey Variant	Range 0..9	Range 0..2	Range 0..1	Pure Grey
Mean	2.153	2.153	2.153	2.153
SD	0.187	0.247	0.332	1.254
Minimum	1.822	1.413	1.045	0
Maximum	4.572	4.934	5.078	5.639
RMSE	0.109	0.176	0.276	1.237

Table 2: *Comparison of JND between consecutive intensity levels*

its scatter of potential greys, but range 0..9 has a noticeable "tail" that creeps towards blue. This is as expected, as the range 0..9 is mainly enabling an additional 9 steps of blue away from pure grey. However, the angle of the "tail" appears to be a continuation of the black body locus, so may be perceived as a shade of grey, rather than a shade of blue. Rendering a linear grey scale (parallel lines using grey intensity 0..255) using each of the approaches is presented in Figure 11. The range 0..9 does have observable blue stripes (although this may not be noticeable when printed); such stripes are located when the offset of $R, G, B+9$ was used—which is infrequent, but noticeable.

To highlight the issue of colour introduction, the PseudoGrey palette used for each range is mapped to HSV colour space. The saturation component (S) of HSV is in the range 0..1, where 0 indicates a pure grey tone, and 1 indicates a colour at maximum off-grey that can be created. Figure 12 shows the saturation values, and it can be seen that ranges 0..0, 0..1 and 0..2 remain within 0.05 saturation for the ma-

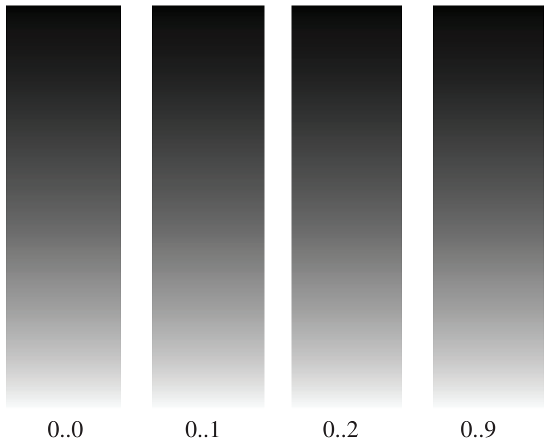

Figure 11: *Analysis of DELL AS500 monitor using PseudoGrey offset ranges 0..0, 0..1, 0..2 and 0..9*

jority of logical device driver levels (namely: between 20 and 255 DDL). All ranges appear to introduce a significant saturation with low DDL (less than 10), but these are probably not observed due to the dark shade of grey. However, the range 0..9 can be seen to regularly produce saturation up to 4× higher than the other ranges. This explains the blue fringes observed. When observed on-screen in colour, the extreme case of the range 0..9 produces easily observed stripes of colour (blue stripes). This shows that chromaticity as well as luminance must be considered when creating a PseudoGrey palette. Such striping is not easily visible in the

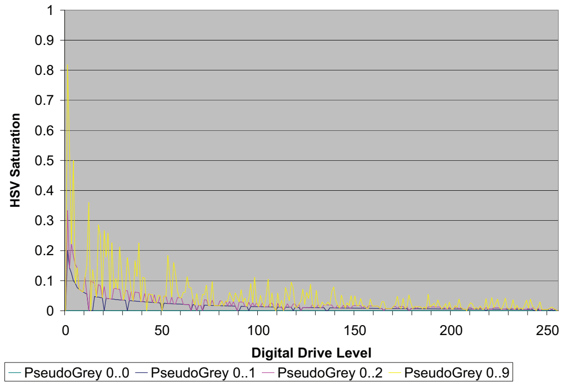

Figure 12: *HSV saturation analysis of DELL AS500 monitor using PseudoGrey offset ranges 0..0, 0..1, 0..2 and 0..9*

range 0..2, but the range 0..9 produces the best luminance fit with the DICOM curve. Hence some form of compromise has to be reached between luminance perfection and noticeable chromaticity.

5. Conclusions

PseudoGrey can be used to produce an approximately DICOM compliant display, and shows potential for a COTS monitor to be used reliably for viewing medical images. We suggest "approximately DICOM" as the required tolerances for FIT and LUM tests are not explicitly defined in the DICOM standard, instead left defined as "Clinical practice is expected to determine the tolerances for the FIT and LUM values". However, we feel that our approach certainly shows promise given a near-linear result from the FIT test and a minimum of 1 JND between each intensity of the LUM test.

The introduction of colour can adversely affect the colour gamut when compared to pure grey, but only in extreme cases. Further investigation is required to optimise the balance between chromaticity and luminance when selecting shades of PseudoGrey to produce a DICOM calibrated curve.

Finally, this approach may be of use in displaying high dynamic range images on a standard monitor (such as thermal images, or images generated through multiple alternative exposures). We aim to carry out further visual tests using high dynamic range imagery to reveal potential improvements obtainable with this technique.

6. Future Work

We wish to compare our COTS output with that of a DICOM calibrated medical-grade monitor, to verify if indeed we are reaching a comparable standard (such as the FIT and LUM test results).

Medical grade monitors are specially constructed to provide a near-uniform intensity when viewed off-axis (i.e. the viewer is not perpendicular to the display), with each pixel being treated to given even luminance across the display. Given such issues (which are not corrected with COTS monitors, human observer tests will be required to determine how closely a COTS monitor can meet the same standard as a medical monitor.

Given the length of time to analyse a monitor (approximately 1 hour for the full 0..9 range), a more optimal approach is required. A directed search would probably produce as good a fit as a full search, but we have yet to investigate this.

7. Acknowledgements

This work was made possible initially through the support of VizNET and related funding from JISC and laterly through support from the Welsh Institute for Visual Computing (funded by Welsh Assembly Government).

The authors would finally like to thank the reviewers for their insightful and supportive comments.

References

[DRF*09] DALLAS W. J., ROEHRIG H., FAN J., KRUPINSKI E. A., JOHNSON P.: Spatial Noise Suppression for LCD Displays. In *Proceedings of SPIE Medical Imaging 2009* (2009).

[GAEB09] GRIMSTEAD I. J., AVIS N. J., EVANS P. L., BOCCA A.: High-Quality Remote Interactive Imaging in the Operating Theatre. In *Proceedings of SPIE Medical Imaging 2009* (2009).

[NEM08] NEMA: *Digital Imaging and Communications in Medicine (DICOM) Part 14: Grayscale Standard Display Function*. Tech. rep., National Electrical Manufacturers Association, 2008.

[SFE99] SAMEI E., FLYNN M. J., EYLER W. R.: Detection of Subtle Lung Nodules: Relative Influence of Quantum and Anatomic Noise on Chest Radiographs. *Radiology 213* (1999), 727–734.

[TCL*92] TYLER C. W., CHAN H., LIU L., MCBRIDE B., KONTSEVICH L. L.: Bit-stealing: How to get 1786 or more grey levels from an 8-bit color monitor. In *Proceedings of SPIE 1666* (1992), pp. 351–364.

[Xth08] XTHONA A.: *Medical Displays Give Radiologists an Edge*. Tech. rep., BARCO Medical Imaging Division, 2008.