This paper describes a novel approach to image fusion for colour display. Our goal is to generate an output image whose gradient matches that of the input as closely as possible. We achieve this using a constrained contrast mapping paradigm in the gradient domain, where the structure tensor of a high-dimensional gradient representation is mapped exactly to that of a low-dimensional gradient field which is then reintegrated to form an output. Constraints on output colours are provided by an initial RGB rendering. Initially we motivate our solution with a simple ‘ansatz’ (educated guess) for projecting higher-D contrast onto colour gradients, which we expand to a more rigorous theorem to incorporate colour constraints. The solution to these constrained optimisations is closed-form, allowing for simple and hence fast and efficient algorithms. The approach can map any \( N \)-D image data to any \( M \)-D output, and can be used in a variety of applications using the same basic algorithm. In this paper we focus on the problem of mapping \( N \)-D inputs to 3-D colour outputs. We present results in five applications: hyperspectral remote sensing, fusion of colour and near-infrared or clear-filter images, multi-lighting imaging, dark flash, and colour visualisation of MRI Diffusion-Tensor imaging. © 2015 Optical Society of America

OCIS codes: (100.0100) Image processing; (100.2000) Digital image processing; (100.2960) Image analysis; (100.2980) Image enhancement; (110.3080) Infrared imaging; (110.4234) Multispectral and hyperspectral imaging.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

As imaging technology has developed to solve a variety of problems, so the richness of imaging systems data has increased. Hyperspectral imaging systems used in remote sensing, for example, routinely capture > 200 channels of spectral data [1], while medical imaging systems capture multi-dimensional, and multi-modal image sets [2]. Ultimately these images are often interpreted by human observers for analysis or diagnosis, and it is therefore crucial that dimensionality is reduced such that the image can be displayed on an output device such as a colour monitor. This process is termed image fusion.

Thus, in the image fusion problem, there can be 10, or 20, or hundreds of values per pixel, and we are interested in reducing the number to 1 for a representative greyscale output or 3 for colour visualization. The simplest way to visualise the information is to simply average the values to produce a greyscale. This approach preserves basic scene structure but suffers from metamerism, where different multi-valued inputs are assigned the same output value.

Where the input values correspond to radiances at different wavelengths, a colour output can be generated by mapping the visible part of the spectrum to display RGB via projection onto a set of colour matching functions, which represent human sensitivity to wavelength [3]. At least such an approach produces a ‘naturalistic’ RGB image, where we define ‘natural’ as the colours that would be seen by a human observer, but it begs the question of how to take into account the influence of spectral values beyond the human visual system’s sensitivity. One idea is to simply stretch the colour matching functions over the full wavelength range of the data [3]; in this case the displayed output produces a false-colour RGB visualisation of the entire spectral range. In general false-colour visualisations can be hard to interpret when object colours are very different from their natural appearance. Furthermore, these spectral projection methods do not say how to fuse non-spectral multi-valued data, e.g. multi-modal medical data.

In order to incorporate non-spectral data a more general approach is required. Generic dimensionality reducing techniques such as principal components analysis (PCA) [4] or ISOMAP [5] can be applied to map multi-valued data to a 3-D space that is
then interpreted as colour values. These approaches maximise the separation of colours in the output image, i.e. minimise the incidence of metamerism, but again produce false colourings. Also, while the incidence of metamerism may be minimised relative to some global objective function, there often aren’t enough degrees of freedom to remove it completely.

To get closer to the preservation of all the multi-valued information in the output image, spatial information must be taken into account [6]. This can be done, for example, by transforming images into a multiscale representation, merging information at each spatial scale, and then inverting the multiscale transformation to produce an output image [7, 8]. Practically, while this has the potential to preserve more information, artefacts such as haloing and ghost images are common. Also, the outputs are rendered in greyscale, which is a disadvantage. One way around this is to retain the RGB colour information whilst swapping in the new greyscale to take the place of the original luminance (i.e., intensity) information [9, 10]. However, while such an approach does produce colour output in the fused image, the 3-D nature of colour is not fully harnessed.

An alternative approach to incorporating spatial information is to work in the gradient domain, where edge information is represented. Gradient domain processing has attracted significant interest due to the importance of edges in human perception [11] as well as its potential for noise reduction [12], and it has been applied in a range of fields such as high dynamic range (HDR) processing [13], image editing [14], and computational photography [15] among others. In the area of image fusion a key paper is the contrast preserving variational algorithm of Socolinsky and Wolff [16] who generate a greyscale image such that its gradient matches that of a multi-channel image as closely as possible. This approach preserves key image information in the output, but still generates a greyscale output.

In this paper we present a gradient domain approach to image fusion that: generates colour outputs, incorporates constraints to allow a more ‘natural’ colour labelling, and can be applied to both spectral and non-spectral data. The approach is motivated by the work of Socolinsky and Wolff and the colourisation work of Drew and Finlayson [17], who use a gradient domain decomposition to apply the gradient from a greyscale image to a colour image, which they use to regulate the output of a colorization algorithm. The key contribution in the present paper is a theorem similarly yielding a gradient decomposition, but one which can be applied to the more general N-D to M-D mapping. This result allows us generalise Socolinsky and Wolff’s work [16] to map N-D images to a colour, rather than just greyscale, output while also exactly matching contrast3.

Our Spectral Edge (SpE) method is applicable to any domains where a) a transformation is required from an N-D space to an M-D space, b) the images in the individual channels are registered, and c) there is a putative, or guiding, M-D image available with plausible output values; this image may be captured by a separate device, or generated directly from the N-D image data (e.g. when $M = 3$ it could be a true-colour sRGB rendering of the image). The generality of the method makes it applicable to a wide range of problems, including: mapping multispectral / hyperspectral images to RGB; fusing RGB and near infrared (NIR) images; colour to greyscale; mapping 3D colour images to 2D to enhance images for colour-deficient observers; pan-sharpening; multi-exposure; dark flash; and visualisation of high-D medical image data such as magnetic resonance imaging (MRI) or time-activity curve data, to name a few. In this paper we report results for the applications of remote sensing, RGB / NIR fusion, dark flash, multi-lighting imaging, and medical diffusion tensor MRI (DTMRI) data, with the output a colour image ($M = 3$) and $N > M$. Clearly for visualising medical data there is no concept of a ‘natural’ colour image to form the putative 3D output; in these cases we can constrain the output colours using a putative false-colour labelling that is appropriate for the task.

The paper is organised as follows: in the next section we review related work in the application areas that we tackle in this paper; in §3 we describe the underlying mathematical formulation, and algorithmic details, of the method; in §4 we show the results of the method for three representative applications; and we conclude the paper in §5.

The overall schema in this paper is to begin by motivating a colour-edge change using an ansatz – an educated guess – in §D and we demonstrate that this approach does work to map higher-D contrast onto lower-D, colour image output. Then, we build a more elaborate and fuller version built on top of the ansatz method by bringing to bear further colour-space constraints, in §E. In an initial version of this paper [19], only the method with further constraints §E was introduced, without the opportunity to motivate the insight of the method §D along with its mathematical development; here, the motivating §D is included, along with new application paradigms and experiments as well as comparison with the bilateral filter and wavelets for two different approaches.

2. RELATED WORK

The image fusion literature encompasses a wide range of applications and techniques. Different channels are typically treated as independent greyscale images and mapped to a single greyscale output, e.g. by averaging them. A popular framework is to decompose each channel into a multi-scale representation, fuse the images at each scale – e.g. by choosing the maximum wavelet coefficient over all images for that pixel / region – and inverting the decomposition step to recover a greyscale output. This approach has been followed using Laplacian pyramids [7] and their variants [20], wavelets [8], complex wavelets [21], perceptual transforms using centre-surround filters [22], bilateral filtering [23], or multi-scale representations of the first fundamental form [24]. These methods are often complex and intensive to compute, as well as being prone to generating artefacts when conflicting information appears in different image channels, making them more suited to fusing pairs of images rather than multiple channels, although some recent approaches offer means for reducing artefacts [25]. Finally, the base layer of the pyramid, or wavelet decomposition, is often a low-pass average image, which can lead to poor colour separation at edges for low spatial scales.

Socolinsky and Wolff [16] cast image fusion as a variational problem, where the goal is to find a greyscale output with gradient information as similar as possible to the input image set. This approach solves the problem of greyscale separation at low spatial scales, but can also be prone to warping artefacts close to edges. These are exacerbated by the ambiguity of gradient ordering at each pixel [26]. Piella [27] uses a variational approach to generate an output that simultaneously preserves the underlying geometry of the multivalued image, similarly to Socolinsky and Wolff, and performs an edge enhancement to improve greyscale separation at object boundaries. The integration of gamut constraints means that potential for artefacts is

---

1U.S. patent granted March 2014 [18].
greatly reduced using this method, but necessitates that the objective function is minimised using an iterative gradient descent scheme, which restricts the speed of the method. As with the wavelet-based approaches, the outputs are in greyscale only.

Several strategies exist for mapping high-dimensional images to RGB, rather than just greyscale. Jacobson et al. [3] investigate different fixed projections; these have an advantage over adaptive methods that colours remain fixed across different visualisations, but the disadvantage that they preserve less information. Adaptive approaches using standard decompositions such as PCA and independent components analysis (ICA) have also proved popular. Tyo et al. [4] use PCA to extract a 3-D subspace from the spectral data, and then rotate the basis of this space so that the final 3D co-ordinates form a plausible RGB image. While this approach is information preserving, the false coloured output can deviate from the ‘natural’ representation, and the global nature of the transform means that localised metamerism may still be common.

In particular applications greyscale fusion schemes can also be applied to generate colour outputs. Schaul et al. [10] employ fusion of NIR and RGB images as part of a de-hazing scheme. They firstly decompose the RGB image into an opponent-based representation and then use an edge-aware multiscale representation to fuse the NIR and luminance channels into a single greyscale. This greyscale is then swapped into the original image as the luminance component. Our approach differs in that it maps the contrast of each of the R, G and B channels as well as the NIR image, rather than just luminance and NIR. Fay et al. [28] use dual-band RGB / long-wave infrared (LWIR) to improve night-vision in low-light settings. This work, which results in fused colour imagery, is specifically focused on a low-light-sensitive visible-light CCD imager.

The approach we outline here is generic, in that it can be applied to a range of input and output dimensions. In this respect our work is closely related to that of Lau et al. [29] who proposed an optimisation based approach to colour mapping. They firstly cluster input colours into groups, and then maximise separation of those groups in a target colour space. They also include constraints on how far resulting colours can deviate from the colours in the target colour space such that the output remains ‘naturalistic’. Although the goal and realm of application of our technique is similar, our approach is markedly different in that we work on gradients, thus focusing the colour separation on spatial boundaries between objects or segments. The speed and low complexity of our method also makes it more suitable for visualising hyperspectral images.

A related application area is ‘dark flash’ photography [30]. In this application two images are captured: a dark flash image, where the intensity of the flash is focused in NIR and UV wavelengths (hence the flash is not highly visible), and an image from ambient lighting which is captured with very low exposure time and hence has high noise levels. The two images are then fused to create a low-noise colour output. In the original work in this area by Krishnan and Fergus [30], the authors use sparsity constraints on output image gradients and edge locations to reduce noise, as well as constraints to ensure that output colours remain close to the ambient lighting image. In our approach we map the NIR / UV gradients directly to colour gradients, using the ambient lighting image as a guide. As a result we include both location and the magnitude of the gradients in the dark flash image to dictate image contrast. As a result our outputs contain more of the image information of the dark flash image than previous approaches. Our method is also significantly faster than the optimisation in [30].

In medical imaging, high-D information such as Time-Activity Curve multi-dimensional data is routinely reduced to RGB output, using various strategies such as false-colour renderings of the final sample or of the integral under the curve (cf. [2]). Our approach can be adapted to any application where a viable 3D colour output is available, whether one that is natural or colour obtained as a pseudocolour rendering. Our gradient domain approach focuses on separating colours at object boundaries, and can be used to improve colour visualisations derived from global mappings such as ICA or PCA. The application examined in this paper is visualisation of 6-D medical Diffusion Tensor Imaging data. We show that for both a simple PCA-based visualisation and for a more careful visualisation based on a multidimensional scaling approach [2] our method improves colour output by including more of the higher-D contrast information.

3. SPECTRAL EDGE IMAGE FUSION (SPE)

A. Definition of gradient and contrast

The goal of our method is to preserve the gradient of a high-dimensional image in a low-dimensional representation. The gradient of a multi-channel image \( C \) at a single pixel is given by the gradient matrix:

\[
\nabla C = \begin{bmatrix}
C^T_x & C^T_y \\
\vdots & \vdots \\
C^T_N_x & C^T_N_y
\end{bmatrix},
\]

where the function \( C^i \) is the \( i \)-th channel of an \( N \)-channel image \( C \) and subscripts \( x \) and \( y \) denote derivatives in the \( x \)- and \( y \)-directions. The gradient matrix \( \nabla C \) contains the partial derivatives of \( C \) in the \( x \) and \( y \) directions; the gradient in direction \( d = [\cos \theta, \sin \theta]^T \) is \( \nabla C d \). Assuming a Euclidean metric, the squared magnitude of the gradient in direction \( d \) is given by:

\[
m^2 = d^T (\nabla C)^T \nabla C d. \tag{2}
\]

The \( 2 \times 2 \) matrix \( Z_C = (\nabla C)^T \nabla C \) is known in differential geometry as the First Fundamental Form, and was introduced to the image processing literature by Di Zenzo [31] as the structure tensor. This can be written in full as:

\[
Z_C = \begin{bmatrix}
\sum_k C^i_k C^j_k & \sum_k C^i_k C^j_k \\
\sum_k C^i_k C^j_k & \sum_k C^i_k C^j_k
\end{bmatrix}
\]

The structure tensor representation is powerful because it encodes magnitude information for the \( N \)-dimensional matrix in 2 dimensions: given \( Z_C \) we can compute the gradient magnitude in any direction \( d \).

A fundamental idea behind our method, therefore, is: in order for a low-dimensional image (low-D) to have an identical contrast to a high-dimensional image (high-D), the structure tensor for both must be identical.

B. Exact contrast mapping

In Socolinsky and Wolff [16], the authors have a similar goal in mapping high-D contrast, defined by the the structure tensor, to a scalar image, approximately. In the first stage of their algorithm they define a scalar gradient field \( \nabla I \) by multiplying the first
eigenvector of $Z_C$ by the first eigenvalue of $Z_C$; the resulting gradient field has the closest possible possible structure tensor – $Z_I$ – that a scalar field can have to $Z_C$ in the least squares sense.

In the novel approach presented here, instead of creating a scalar gradient-field we create $M$ gradient fields, where $M$ is the number of channels in our output image; we refer to this set of gradient fields as an $M$-D gradient-field. By doing this we can now generate an $M$-D gradient field whose structure tensor matches the original structure tensor $Z_C$ exactly.

In order to ensure that the output is coloured naturally, we suppose that we have access to a putative low-D version $\tilde{R}$ of the high-D image data which has naturalistic colours: this image may either be captured by a specific device (e.g. an RGB camera), or generated from the high-D using some algorithm (e.g. a true colour rendering of remote sensing data). We then use the contrast information from the high-D image, and the colour information from the putative low-D image, to generate a new low-D gradient field, which we finally reintegrate to generate a colour output.

C. A gradient decomposition that motivates a simple ansatz for exact contrast mapping

To understand how the contrast can be mapped exactly between images, we start by representing the $N \times 2$ gradient matrix $\nabla C$ using its Singular Value Decomposition:

$$ \nabla C = U A V^T $$

In this equation the matrix $U$ represents an $N \times 2$ basis for a plane in the $N$-D range, or colour space; $V$ represents a $2 \times 2$, 2D basis for the spatial direction in the image domain; and the diagonal matrix $A$ contains the magnitude information (i.e., the square roots of the eigenvalues of $Z_C$) for each of the primary directions. Put simply, the decomposition from left to right describes the *colour*, *magnitude*, and *spatial* information in the gradient respectively.

From this decomposition we can see that Di Zenzo’s structure tensor at a given pixel is determined entirely by the magnitude and spatial components $A$ and $V$. That is:

$$ Z_C = \nabla C^T \nabla C = V A^T U^T U A V = V A^2 V^T $$

(4)

since $U$ is orthonormal and $A = A^T$. The insight motivating our ansatz in §D is that in this equation the colour component of the gradient $U$ can be replaced with any $M \times 2$ orthonormal matrix, where $M$ represents the number of output channels in a target output image (e.g., 3), without altering the structure tensor. A more elaborate and fuller version built on top of the ansatz method is given below in §E.

D. Contrast mapping ansatz

For our 3 gradient fields, the Di Zenzo matrices are defined as:

$$ Z_H = (\nabla H)^T (\nabla H) $$

$$ Z_R = (\nabla R)^T (\nabla R) $$

(6)

$$ Z_R = (\nabla \tilde{R})^T (\nabla \tilde{R}) $$

where the tilde refers to the input putative RGB image and without a tilde refers to the desired output RGB image.

Now, the goal of our method is to create a new gradient $\nabla R$, whose structure tensor matches that of the high-D image $H$ exactly, i.e. $Z_R = Z_H$. To do this we use the SVD of the putative image and the high-D image:

$$ \nabla R = \tilde{U} \tilde{A} \tilde{H} \tilde{V}^T $$

(7)

Then to generate our ansatz we simply exchange the $3 \times N$ matrix $U_H$ with the $3 \times 2$ matrix $\tilde{U}_R$ to give:

$$ \nabla R = \tilde{U} \tilde{A} \tilde{H} V^T $$

From equation (4) we get:

$$ Z_R = (\nabla R)^T \nabla R = V H A^2 V^T = Z_H $$

which shows that the structure tensor matches that of the high-dimensional image exactly and the output gradient $\nabla R$ is the gradient matrix for a 3 channel RGB output. The final step of the algorithm is to reintegrate the 3-D gradient field to produce an output image $R$.

E. Spectral Edge image fusion

Using the above simple scheme, the only constraint on the output colour gradients is that they must fall within the column space of $U_R$ and of course have a Di Zenzo matrix matching that of the input. It turns out that this is a relatively weak constraint on the image output colours, and can result in desaturated outputs (see Section 4). This leads us to look for a more constrained principled method that better maps the high-D contrast to the putative image colours. This idea motivates the following theorem:

**SPECTRAL EDGE (SpE) PROJECTION THEOREM:**

Given a multidimensional image $C$ and a putative RGB “guiding” image $\tilde{R}$, we can generate a new RGB gradient matrix $\nabla \tilde{R}$ that is as close as possible to the gradient of the RGB image, and
whose contrast matches that of \( \mathbf{C} \) exactly.

In this theorem we aim to satisfy two conditions: (1) For a generated \( \nabla \mathbf{R} \), i.e. the result of the theorem, we wish \( \mathbf{Z}_R \) to equal \( \mathbf{Z}_H \), the structure tensor for the higher-D image, so that contrast is mapped exactly from high-D to low-D; and (2) the output gradient \( \nabla \mathbf{R} \) should approximate as closely as possible the putative gradient \( \nabla \mathbf{R} \), so that no large colour shifts are obtained. That is, we desire an altered colour gradient \( \nabla \mathbf{R} \approx \tilde{\nabla} \mathbf{R} \), subject to (1) and (2).

A solution obeying (1) can be found easily if we keep only within the span of colour gradient \( \nabla \mathbf{R} \), and seek a \( 2 \times 2 \) linear matrix transform \( \mathbf{A} \) such that

\[
\nabla \mathbf{R} = \tilde{\nabla} \mathbf{R} \mathbf{A}
\]

so that the colour gradient will not differ greatly from the approximation\(^2\).

In that case the desired relation between Di Zenzo matrices is as follows:

\[
\mathbf{Z}_R \equiv \mathbf{Z}_H \\
\Rightarrow \mathbf{Z}_R = \nabla \mathbf{R}^T \nabla \mathbf{R} = \mathbf{A}^T \nabla \mathbf{R}^T \nabla \mathbf{R} \equiv \mathbf{Z}_H \quad (9)
\]

\[
\Rightarrow \mathbf{A}^T \mathbf{Z}_R \mathbf{A} \equiv \mathbf{Z}_H
\]

Given this relation, we satisfy (1) above provided matrix \( \mathbf{A} \) is any solution of (9). For example, one solution is given by:

\[
\mathbf{A} = \left( \sqrt{\mathbf{Z}_R} \right)^+ \sqrt{\mathbf{Z}_H}
\]

where in this equation the symbol \( \sqrt{\cdot} \) denotes the unique symmetric root \([32]\) of the real positive semi-definite symmetric matrices \( \mathbf{Z}_R \) and \( \mathbf{Z}_H \), and \( ^+ \) indicates the Moore-Penrose pseudoinverse operator. Even though \( \sqrt{\mathbf{Z}_R} \) is square, nonetheless we guard against instability by using the pseudoinverse rather than the inverse.

To show that \( \mathbf{A} \) is indeed a valid solution we can see that:

\[
\mathbf{A}^T \mathbf{Z}_R \mathbf{A} = (\sqrt{\mathbf{Z}_H} \sqrt{\mathbf{Z}_R})^+ \tilde{\mathbf{Z}}_R (\sqrt{\mathbf{Z}_R} \sqrt{\mathbf{Z}_H}) = \mathbf{Z}_H
\]

since \( \sqrt{\mathbf{Z}_H} \) and \( \sqrt{\mathbf{Z}_R} \) are symmetric.

The complete set of solutions solving (9) then consists of all matrices \( \mathbf{A} \) that are any \( 2 \times 2 \) orthogonal transform \( \mathbf{O} \) away from (10):

\[
\mathbf{A} = \left( \sqrt{\mathbf{Z}_R} \right)^+ \mathbf{O} \sqrt{\mathbf{Z}_H}, \quad \mathbf{O}^T \mathbf{O} = \mathbf{I}_2
\]

(12)

since any such solution satisfies (9):

\[
\mathbf{A}^T \mathbf{Z}_R \mathbf{A} = (\sqrt{\mathbf{Z}_H} \mathbf{O}^T \sqrt{\mathbf{Z}_R})^+ \tilde{\mathbf{Z}}_R (\sqrt{\mathbf{Z}_R} \mathbf{O} \sqrt{\mathbf{Z}_H}) = \mathbf{Z}_H
\]

(13)

To produce realistic colours we also wish to fulfil constraint (2), that the adjusted gradient \( \nabla \mathbf{R} \) approximates as closely as possible the putative colour gradient \( \nabla \mathbf{R} \). From (8), this implies a constraint on rotation \( \mathbf{O} \) as follows:

\[
\nabla \mathbf{R} \approx \tilde{\nabla} \mathbf{R} \Rightarrow \nabla \mathbf{R} \mathbf{A} \approx \tilde{\nabla} \mathbf{R}
\]

\[
\Rightarrow \nabla \mathbf{R} \mathbf{A} \approx \mathbf{A} \approx \mathbf{I}_2
\]

\[
\Rightarrow \sqrt{\mathbf{Z}_R} \mathbf{O} \sqrt{\mathbf{Z}_H} \approx \mathbf{I}_2
\]

\[
\Rightarrow \mathbf{O} \sqrt{\mathbf{Z}_H} \approx \sqrt{\mathbf{Z}_R}
\]

with \( \mathbf{I}_2 \) the \( 2 \times 2 \) identity matrix. The last line of (14) says that \( \mathbf{O} \) should be chosen to rotate \( \sqrt{\mathbf{Z}_H} \) such that it is as close as possible to \( \sqrt{\mathbf{Z}_R} \). This problem is known as the Orthogonal Procrustes Problem \([32]\); the solution in the least-squares sense is to firstly use a singular value decomposition to express the product of square roots of \( \mathbf{Z}_R \) and \( \mathbf{Z}_H \):

\[
\sqrt{\mathbf{Z}_R} (\sqrt{\mathbf{Z}_H})^T = \mathbf{D} \Gamma \mathbf{E}^T
\]

where \( \mathbf{D} \), \( \Gamma \), and \( \mathbf{E} \) are the three matrices of the SVD. We note here that the transpose on \( \sqrt{\mathbf{Z}_H} \) above is actually unnecessary since \( \sqrt{\mathbf{Z}_H} \) is symmetric, but we include it to agree with the formulation in [32]. Then the solution \( \mathbf{O} \) that minimises the last line of (14) in terms of Least Squares is given by:

\[
\mathbf{O} = \mathbf{D} \mathbf{E}^T
\]

(16)

We can now obtain \( \mathbf{A} \) by substituting this solution for \( \mathbf{O} \) into equation (12), and then directly derive a modified colour gradient \( \nabla \mathbf{R} \) using (8).

F. Relationship between SpE projection theorem and the ansatz

The gradients from the original ansatz and the subsequent SpE projection theorem both have structure tensors that match a high-D image, and therefore match each other, but produce different outputs. To understand the relationship mathematically firstly we recall that the ansatz is given by:

\[
\nabla \mathbf{R}_{\text{ansatz}} \equiv \tilde{\mathbf{U}}_R \Lambda_H \mathbf{V}_H^T
\]

The meaning of the projection theorem in this framework is in fact the insertion of an additional rotation matrix that maps the output more closely to the input putative image gradients. We now show that this can be written as:

\[
\nabla \mathbf{R}_{\text{ansatz}} = [\tilde{\mathbf{U}}_R] [\Lambda_H \mathbf{V}_H^T]
\]

(17)

\[
\Rightarrow \nabla \mathbf{R}_{\text{SpE}} = [\tilde{\mathbf{U}}_R] (\mathbf{V}_H^T \mathbf{O} \mathbf{V}_H) [\Lambda_H \mathbf{V}_H^T]
\]

which gives an improved RGB gradient.

\footnote{It can be proven that the optimal solution must fall within the span of \( \nabla \mathbf{R} \), although the proof is too long to include here.}
G. Proof of relationship between ansatz and SpE projection theorem

From eq. (12) the $2 \times 2$ matrix $A$ is given in terms of the two symmetric square root matrices for contrast from hi-D, $Z_H$, and from the putative colour, $Z_R$.

First, we can make use of the SVD decomposition of $\nabla R$ to derive the decomposition of $\sqrt{Z_R}$ as well as the decomposition of its inverse, or pseudo-inverse, as follows:

$$\sqrt{Z_R} = \tilde{V}_R \tilde{A}_R \tilde{V}_R^T \Rightarrow \left( \sqrt{Z_R} \right)^+ = \tilde{V}_R \tilde{A}_R^+ \tilde{V}_R^T$$

Then merging the representation in eq. (12) for matrix $A$ with the decomposition of colour gradients (7), the colour gradient solution is as follows

$$\nabla R = \tilde{\nabla R} A$$
$$\left[ \begin{array}{c} \tilde{u}_R \\ \tilde{\lambda}_R \tilde{V}_R^T \end{array} \right] \left( \begin{array}{c} \sqrt{Z_R} \\ O \sqrt{Z_H} \end{array} \right)$$
$$= \left[ \begin{array}{c} \tilde{u}_R \\ \tilde{\lambda}_R \tilde{V}_R^T \end{array} \right] \left[ \begin{array}{c} \tilde{V}_R \tilde{A}_R^+ \tilde{V}_R^T \\ 0 \end{array} \right]$$
$$= \tilde{u}_R \left[ \tilde{V}_R^T O V_H \right] \left[ \tilde{\lambda}_R V_H^T \right]$$

and indeed we arrive at the last line in (17).

H. Summary of SpE approach

In summary, starting from a lower-D image containing a naturalistic rendering of the scene $R$, at each pixel we find a transform $A$ of the $M \times 2$ gradient matrix of the lower-D image such that (i) the altered gradient has an identical contrast as that for the higher-D image — i.e. we transfer the higher-D contrast to the lower-D image; and (ii) the altered lower-D gradient $\nabla R$ remains in the span of the unaltered gradient, at each pixel; i.e. the new $M \times 2$ gradient is a $2 \times 2$ linear transform away from the putative gradient.

I. Reintegration

The contrast mapping process results in an $M$-D gradient matrix $\nabla R$ at each pixel location. We would like to treat $\nabla R$ as a set of $M$ gradient fields, one for each output channel, defined by the rows of $\nabla R$. The final phase of the algorithm is to reintegrate each gradient field in turn to generate $M$ new output channels. However, in general each of the approximate gradient fields will be non-integrable, i.e. will not in fact be the gradient for a scalar image. An output image must therefore be reconstructed by computing an image whose gradient matches that of the target field as closely as possible, by minimising some error function. Interestingly, however, we have more information available here than in the traditional reintegration problem of forming a greyscale image $I$ from a gradient-approximation — we have the actual, $N$-D image dataset itself.

If we denote the approximate gradient field from the $i$-th channel of $\nabla R$ as $P^i = \left( R_{i,x}^i R_{i,y}^i \right)$, then we seek a scalar image $I$ such that:

$$R^i = \arg \min \left\| P^i - \nabla I \right\|_n$$

where $n$ defines the norm used in the error function. For $n = 2$ the solution could be given by the solution to Poisson’s equation, and a number of approaches have been applied to do this, e.g. [33, 34]. However since here we also have the $N$-D data $H$, we can use the look-up-table approach of Finlayson et al. in [26, 35], which minimises the error function in (19) for $n = 2$ using a LUT mapping from the high-D image $H$ to each $R^i$. This constraint means that the final image is guaranteed to be free of artefacts, and facilitates the operation of the algorithm in real time. Importantly, in [26] it was shown that if a multi scale gradient is approximately integrable across multiple scales then a LUT mapping is the correct reintegrating function.

J. Implementation details

To compute the gradient matrices $\tilde{\nabla R}$ and $\nabla H$ we use local finite differencing, i.e. for an image $C$ at pixel $(x, y)$ and channel $i$, $C_i^p(x, y) = C'(x - 1, y) - C'(x, y)$ and $C_i^q(x, y) = C'(x, y - 1) - C'(x, y)$, although other gradient operators, e.g. Sobel operators, would serve the purpose just as well. Furthermore, given the global nature of the reintegration approach in [26], the gradient operator could also be applied at different spatial scales, and reintegrated simultaneously. For other reintegration techniques the finest spatial scale is advised to reduce blurring in the output.

There is a potentially large discrepancy in image dimensionalities between input and output, i.e. $N \gg M$ for input dimensionality $N$ and output $M$, and as a result the total high-D contrast may not be displayable within the low-D gamut. Here, we mitigate this with a simple contrast scaling approach whereby 99% of pixel values are mapped within the image gamut, although more complex gamut mapping strategies could also be employed [36] as post-processing after applying the algorithm. We note here that gamut mapping is not necessary for all applications, and we explicitly state where it was required in the experimental section.

As stated earlier, the method assumes pre-registered images. However, in practice we have found the method to be robust to mis-registration. Furthermore, using a difference-of-gaussians (DoG) or multi-scale derivative operator further reduces the effect of mis-registrations.

The complexity of the contrast projection algorithm is $O(P)$, where $P$ is the number of pixels. The complexity of the reintegration is also $O(P)$ [26], although using other approaches, such as iterative Poisson solvers, can increase the complexity. Memory requirements are low, since most of the calculations are performed on $2 \times 2$ structure tensor matrices. In our case the chosen reintegration [26] increases memory requirements since the high-dimensional image needs to be stored and used in the reintegration. But the chief advantage of this method is its ability to remove artefacts.

The method is general in the choice of output colour space. We represent images in sRGB for the applications here, but the putative low-D image could be represented in a different space, e.g. a perceptually uniform space such as CIELAB, and then mapped to sRGB for display. This would be a good approach in applications where Euclidean distances in sensor-space should correlate with the magnitude of perceived differences in the output.

4. EXPERIMENTS

A. Experiment paradigms

In this paper we show results of our method in five application areas: i) hyperspectral / multispectral remote sensing; ii) fusion of RGB+NIR / LWIR (thermal imaging), or RGB+clear-filter imaging; iii) dark-flash 6-D image sets; iv) multi-lighting image
capture, and v) medical MRI diffusion-tensor imaging. Each of the different applications falls naturally within the same computational framework; we explain in the next section how to adapt this framework for each application.

We note here that the results we present are limited to qualitative comparisons, as there is no "ground truth" fused image with which to compare our outputs quantitatively. The purpose of this section is also to show the wide applicability, and effectiveness, of the general method and as such we do not compare with state of the art for all applications. Rather, engineering an optimal solution in different applications is a subject for future work.

A.1. Remote Sensing Applications

Images captured for remote sensing applications, e.g. from satellite or airborne imaging systems, typically span the visible, near infra-red and far-infra red wavelength spectrum. Here we use data from two publicly available datasets: a) Landsat 7 [37], and b) AVIRIS [36]. The Landsat 7 satellite captures 8 separate images; 3 in the visible range, 4 IR images (including one thermal image) and a panchromatic detail image; these images are captured using a scanning radiometer. The three visible images are captured from 450-515nm (blue), 525-605 (green), and 630-690 nm (red), and we use these as the B, G and R channels respectively of $\hat{R}$; $H$ then consists of the three RGB channels, and three IR images captured at: 750-900nm (NIR); 1550-1750nm (SWIR); and 2090-2350nm (SWIR). We omit the thermal and panchromatic channels as they have different spatial resolutions than the other images. The AVIRIS data is captured from an airborne imaging system, and uses a "sweep-broom" hyperspectral camera with 224 adjacent spectral channels, which span a spectral range 380-2500 nm and are sampled at approximately 10nm intervals. To generate $\hat{R}$ in this case we project the visible wavelengths, 380-730nm, onto the sRGB colour matching functions [39], to generate a true-colour sRGB rendering; $H$ is composed of all 224 channels.

In these applications, as $N \gg M$, we employ simple gamut mapping procedure (see Section J) to keep outputs within the displayable range. We apply this to all outputs (Spectral Edge and other methods) to ensure a fair comparison.

A.2. Visualising RGB+NIR / LWIR and RGB+Clear images

Pairs of RGB and NIR (or thermal) images can be captured using different methods, e.g. using a beamsplitter and two CCD arrays to capture registered NIR and RGB, or taking successive photographs with an IR filter (“hot mirror”) present and absent.

To apply our technique to this problem we construct a 4-D image $H$ by appending the NIR channel to the colour image. This 4D image is used to calculate the high-D gradient $\nabla H$ while the original RGB image is used to calculate the putative gradient $\nabla R$.

We compare our technique with: (a) “alpha blending”, where the RGB outputs, $R_{\text{out}}$, $G_{\text{out}}$, and $B_{\text{out}}$ are constructed as convex combinations of the RGB and NIR input images, e.g. $R_{\text{out}} = \alpha R + (1 - \alpha)NIR$ for $0 \leq \alpha \leq 1$; (b) the max-wavelet approach of [8], where the RGB image is firstly mapped to YIQ space, and the luminance component, $Y$, is then replaced by the output of the max-wavelet algorithm; (c) the colour-cluster optimisation method of Lau et al. [29].

RGB+Clear imaging, commonly used in astronomy [40] is another 4-D to colour application. An RGB-Clear camera has the usual Bayer colour-filter array replaced by an array with one extra channel with no filtering applied in the visible range. The clear channel has better performance in low light.

A.3. Multi-lighting imaging

Multi-lighting imaging is becoming a popular technique for capturing multiple wavelength images of a scene without the use of movable filters or complex imaging setups. This can be considered a branch of multispectral imaging, where the goal is to both improve colour imaging workflows to reduce the effects of metamerism, and sometimes to introduce wavelengths from outside the visible range. In this paper we present results from two different systems.

In the first a monochrome camera is used to capture multiple images under LED illumination with varying peak wavelengths; the image captured under each LED becomes a different image channel in $H$. The putative RGB $\hat{R}$ is then generated by projecting the visible wavelength images onto the sRGB colour matching functions.

In a second application an image is captured with a given RGB camera under two separate illuminants (e.g. a fluorescent illuminant and a daylight simulator). Concatenating the two RGB images leads to 6 independent channels, which comprise $H$. The putative RGB $\hat{R}$ is taken as one of the initial RGB images, preferably that captured under a standard white illuminant such as D65.

A.4. Dark Flash

In dark flash photography we have two images: an RGB image captured under NIR / UV flash (with wavelength sensitivity of approximately 350-400nm and 700-850nm), and a low-exposure image captured in ambient lighting (400-700nm). We would like our output image to be smooth and so contrast magnitude should be guided by the low-noise NIR / UV flash image. In general the relative weight of the NIR/UV image in the output can be altered by applying a simple gain factor to the NIR / UV gradients. A weighting could also be applied locally, for example in order to attenuate shadow regions in the dark flash image that are not present in the original image, as in [30]. Here we adhere to a simple implementation where the image $H$ is actually the NIR / UV image, and the putative RGB $\hat{R}$ is the ambient light high-noise image. This is an example of a case where $H$ can actually have equal or lower dimension than the output $R$. Images are taken from [30].

A.5. Medical applications

In some fusion applications there is no “true-colour” rendering of the input image available, but labelling the input data using colour still has value for interpreting the data. In medical imaging, for example, multimodal and multidimensional imaging devices such as positron emission tomography PET, MRI and diffusion tensor imaging (DTI) systems are used to gather physiological data that is displayed as an image, and used by clinicians to aid diagnosis.

Here we apply our algorithm to the problem of visualising MRI Diffusion-Tensor data. In this application the data consists of $3 \times 3$ symmetric positive semi-definite matrices at each spatial location, and is hence 6-D. To preserve its character, 6-D vectors are formed respecting a Log-Euclidean metric [41]. The most common method for visualising such data is to display loadings on the first 3 principal component vectors [4]. A more perceptually meaningful approach than PCA is to carry out multi-dimensional scaling (MDS) on the 6-vectors, descending to 3-D [2]; then the result is conceived as approximately perceptually uniform CIELAB colour and then mapped to standard gamma-corrected sRGB display space. We apply our algorithm to both PCA and MDS approaches and generate different puta-
ative RGB outputs $\nabla R$, using the 6-D tensor data to calculate $\nabla H$.

B. Results

Results from the multispectral Landsat data are shown in Figs. 1 and 2, and a result from the hyperspectral AVIRIS data is shown in Fig. 3. Each example includes the RGB rendering, an example IR image, and the output of our approach. For comparison in Figure 2 we show outputs for a bilateral filtering approach [23], and for Figure 3 we show the result of using a stretched colour-matching function approach (cf. [3]). The bilateral filtering approach implicitly produces a greyscale output. In [23], the authors suggest partitioning the wavelength domain into three disjoint subsets and fusing each separately to create three greyscale outputs that are interpreted as R, G and B (see Figure 2 (b)). Here we also apply a more standard method of producing a single greyscale output and substituting this for the luminance component of the putative RGB image; the result of this approach (denoted ‘luminance replacement’) is shown in 2 (e).

For both Figure 2 and 3 the content of the output SpE image shares the same colour scheme as the putative true-colour RGB output, and as well integrates the information from the additional channels. In particular, because of the inclusion of IR data, the presence of bodies of water becomes more pronounced than in the original.

Figures 4 and 5 show results for the problem of merging RGB and NIR images. In Figs. 4(c,d) alpha-blending and luminance replacement outputs significantly alter the natural colouring. The method of Lau et al. Fig. 4(e) attempts to incorporate detail from the NIR image and does produce naturalistic colours (i.e. colours that are close to the original RGB). Our approach focuses on preserving the information content in all four input image planes; as a result the presence of the NIR image is much more noticeable in regions of low contrast in the original RGB, e.g. around the trees. In Fig. 5 we succeed in keeping colour information intact while displaying NIR information more visibly. As in the non-gradient approach [42], age-spots are removed, along with freckles; but as well, more of the NIR content is displayed using the SpE method.

In Figure 6 we show results for another RGB + NIR application from Cultural Heritage Imaging. In this figure we compare the output of our ansatz (c) with the SpE projection theorem (d), which shows a clearly more colourful output, with a greater amount of NIR detail preserved.

Fig. 5. RGB + NIR fusion application.  

Fig. 6. Master of the Retable of the Reyes Católicos, Spanish, active 15th century. The Annunciation, late 15th century. Oil on wood panel. 60 3/8 x 37 in. The Fine Arts Museums of San Francisco, gift of the Samuel H. Kress Foundation, 61.44.2. Figs.6(a,b) courtesy of Cultural Heritage Imaging and Fine Arts Museums of San Francisco.

We also demonstrate that our method can be used to fuse RGB with longer-wave, even thermal, IR (wavelengths > 10 $\mu$m). Figure 7 shows fusion results for an image from the OTCBVS dataset [43], which contains registered RGB and thermal images. The fusion is successful, with hidden structures made visible.

Fig. 7. Example of thermal + RGB fusion; images taken from OTCBVS dataset [43].

Figure 8 shows images as captured from (a) a 3-sensor RGB

---

3Data taken from http://ivrl.epfl.ch/research/infrared/skinsmoother and used with permission.
camera and (b) by a 4-sensor RGB-Clear camera, with the 4th channel appended for the higher-D input data for Spectral Edge. Fig. (c) shows the improvement over RGB when using Spectral Edge, especially in the top-right where we can now see the objects inside the window.  

Figure 9 shows images captured from using an LED based multi-lighting imaging system, where the illumination is altered frame-by-frame, but the monochrome camera remains fixed. Images are captured at approximately 30 nm intervals between 365nm and 760nm, with additional NIR channels at 870 and 940nm; these images comprise the high-D image $H$. A comparison of SpE projection with the ansatz shows that SpE produces colours that are closer in hue and saturation to the original image; this is particularly evident in the colour chart to the side of the image.

Figure 10 shows images captured from an RGB camera under two illuminations, from [29]. In our approach the contrast from both images is weighted equally, while the image under the white illuminant $l_1$ (far left) is used to guide the output colours. While in the original RGB image Fig. 10(a) it is difficult to disambiguate the colours of the yellow lemon and red grapefruit, this is more salient in (c), and better again in our method (d).

In Figure 11 we show outputs for our method for the ‘dark flash’ application. The RGB images captured in ambient light (left) are noisy, whereas the dark flash image (centre) is less noisy but has very little colour information. The SpE output (right)
Fig. 3. Example of hyperspectral image fusion; images taken from AVIRIS dataset [38]. In (b), the largely blue output is due to most of the energy measured in each pixel spectrum residing in the visible band, which is on the small-wavelength end in the full measured spectrum extending from 370.5nm to 2507.6nm.

Fig. 4. Comparison of SpE with other methods for an RGB + NIR fusion application.

preserves the gradient magnitude, and hence the smoothness of the dark flash image, but retains a good approximation to the original image colours.

In Fig. 12 we show results for the medical, DTI, application. This data consists of 55 axial images of brain slices, each representing a different depth plane. In Fig. 12(a), we use PCA
Fig. 8. Synthetic RGBC image.

Fig. 9. Visualization of a multi-lighting image of the Archimedes Palimpsest

Fig. 10. Application to mapping multispectral images from multi-illuminant capture setups: a) and b) show images captured under illuminants $I_1$ and $I_2$ respectively, where the image under $I_1$ is used to constrain output colours; original images taken from Lau et al.\[29\].

Fig. 11. Application of our method to “Dark Flash” Photography (input images taken from [30]).

Fig. 12. Visualization of 6-D DTMRI data: (a,b) PCA approach, (c,d) MDS method.
5. CONCLUSION
In this work we have presented a novel, gradient-domain, approach for mapping images of any dimension to images of any other dimension. The method is based on mapping contrast, defined by the structure tensor matrix, exactly onto a low-dimensional gradient field, and incorporates constraints on the naturalness of output colours borrowed from a putative RGB rendering. The approach is formulated as a constrained optimisation with a closed-form solution, making the method both fast and efficient. We have demonstrated applications in mapping high-dimensional images to RGB outputs for display, and will expand the applicability to new areas in future work, as well as engineering optimised solutions for different applications (e.g. reducing the effect of conflicting shadows in the dark-flash application).

REFERENCES