POP image fusion - derivative domain image fusion without reintegration

Graham D. Finlayson
University of East Anglia
Norwich, UK.
G.Finlayson@uea.ac.uk

Alex E. Hayes
University of East Anglia
Norwich, UK.
alex.hayes@uea.ac.uk

Abstract

There are many applications where multiple images are fused to form a single summary greyscale or colour output, including computational photography (e.g. RGB-NIR), diffusion tensor imaging (medical), and remote sensing. Often, and intuitively, image fusion is carried out in the derivative domain. Here, a new composite fused derivative is found that best accounts for the detail across all images and then the resulting gradient field is reintegrated. However, the reintegration step generally hallucinates new detail (not appearing in any of the input image bands) including halo and bending artifacts. In this paper we avoid these hallucinated details by avoiding the reintegration step.

*Our work builds directly on the work of Socolinsky and Wolff who derive their equivalent gradient field from the per-pixel Di Zenzo structure tensor which is defined as the inner product of the image Jacobian. We show that the x- and y- derivatives of the projection of the original image onto the **P**rincipal characteristic vector of the **O**uter **P**roduct (POP) of the Jacobian generates the same equivalent gradient field. In so doing, we have derived a fused image that has the derivative structure we seek. Of course, this projection will be meaningful only where the Jacobian has non-zero derivatives, so we diffuse the projection directions using a bilateral filter before we calculate the fused image. The resulting POP fused image has maximal fused detail but avoids hallucinated artifacts. Experiments demonstrate our method delivers state of the art image fusion performance.*

1. Introduction

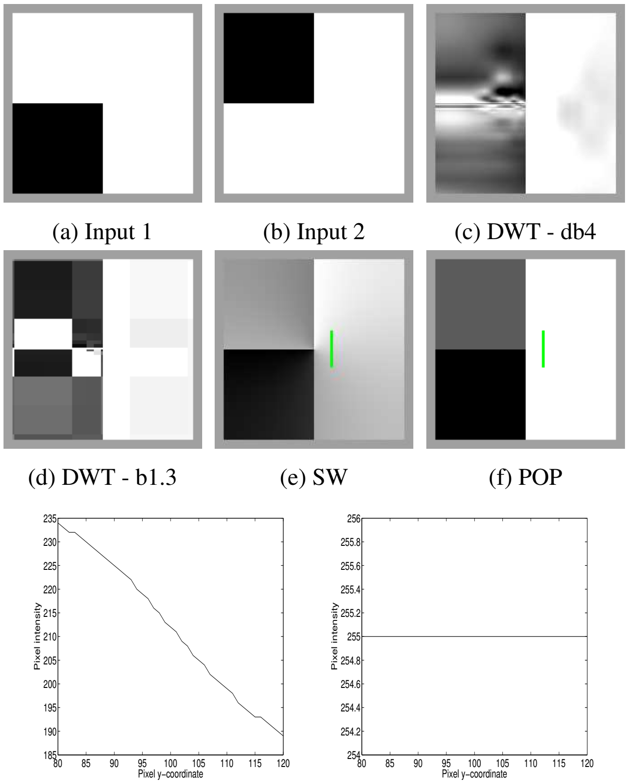
Image fusion has applications in many problem domains, including multispectral photography[7], medical imaging[34], remote sensing[23] and computational photography[18]. In image fusion we seek to combine image details present in N input images into one output image. Image gradients are a natural and versatile way of representing image detail information[8], and have been used as a basis for several image fusion techniques including [32] and [37]. Other image fusion methods include those based on wavelet decomposition[25], the Laplacian pyramid[31] and neural networks[17].

A powerful way of summarizing gradient information across N input image channels is the Di Zenzo structure tensor[9][13] (defined as the 2×2 inner product of the $N \times 2$ image Jacobian). Structure tensor based methods have many applications in computer vision[4], including in image segmentation[14] and, relevant to this paper, for image fusion[19].

The seminal image fusion method of Socolinsky and Wolff (SW) uses the structure tensor to find a 1-D set of *equivalent* gradients, which in terms of their orientation and magnitude, approximate the tensor derived from a multichannel image as closely as possible in a least-squares sense[30]. They show that the equivalent gradient is defined by the most significant eigenvalue and associated eigenvector of the structure tensor. Unfortunately, the derived gradient field of Socolinsky and Wolff is often non-integrable. Because the gradient field reintegration problem (of non-integrable fields) is inherently ill-posed, derivative domain techniques will always *hallucinate* detail in the fused image that wasn't present in the original image.

Modern techniques which apply additional constraints to the reintegration problem can sometimes mitigate but not remove these artifacts[2], [22], [10], [28] and [27]. In other work[29], the fused image is post processed so that connected components - defined as regions of the input multispectral image that have the same input vector values - must have the same pixel intensity. Unfortunately, this additional step can produces unnatural contouring and edge effects.

In this paper, we develop a derivative domain image fusion method which avoids the need for reintegration and, in so doing, we avoid reintegration artifacts. Our method begins by calculating the *outer product* of the Jacobian matrix of image derivatives (rather than the inner product that defines the structure tensor). We prove that the projection of the original multichannel image in the direction of the **P**rincipal characteristic vector of the **O**uter **P**roduct (POP) tensor results in the same equivalent gradient field defined



(a) Input 1 (b) Input 2 (c) DWT - db4

(d) DWT - b1.3 (e) SW (f) POP

(g) SW - plot (h) POP - plot

Figure 1: Image fusion example: (a) and (b) are fused by wavelet-based methods (c) and (d), resulting in severe image artifacts. The Socolinsky and Wolff gradient-based method (e) works better, but intensity gradients are hallucinated (g) where none appear in the input images. The POP method (f) captures all input detail with no artifacts or hallucinated detail.

in the Socolinsky and Wolff method. Of course this initial projected image is not well defined everywhere e.g. it can be non-zero only where the Jacobian has non-zero derivatives, and so we diffuse the POP projection directions that are available using a bilateral filter. The POP fused image is the per-pixel dot product of the projector image with the multichannel original. The resulting POP fused image has maximal fused detail but avoids entirely the hallucinated artifacts.

A comparison of POP image fusion with antecedent methods is shown in fig. 1, where there are two uniform white images with respectively the top left and bottom left quarters removed. The discrete wavelet transform (DWT) images were produced using a wavelet-based method which merges the coefficients of the two images at different scales. We ran a standard DWT image fusion implementation using the CM (choose maximum) selection method, which is simple and one of the best performing in a comparison[25]. The input images are small so there is only a 7 level wavelet decomposition. In 1c and 1d we show the outputs using Daubechies 4 and Biorthogonal 1.3 wavelets, the best wavelet types as found in [25]. Clearly neither the basic wavelet method nor the Socolinsky and Wolff method (1e) work on this image fusion example. However the POP image fusion method (1f) - discussed in detail in section 3 - succeeds in fusing the images without artifact. The intensity profile of the green line in 1f, shown in 1h has the desired equiluminant white values, whereas the Socolinsky and Wolff intensity profile 1g shows substantial hallucinated intensity variation.

Section 2 discusses the background to our method. Our POP image fusion method is presented in section 3. In section 4, experiments - including comparisons with other methods - are presented. The paper concludes in section 5.

2. Background

Let us denote as $I(\mathbf{x})$ the multichannel image: $I(\mathbf{x}) : \mathcal{D} \subset \mathbb{R}^2 \rightarrow \mathcal{C} \subset \mathbb{R}^N$ ($\mathbf{x}$ is a 2-dimensional image coordinate and $I(\mathbf{x})$an N-vector of values). The Jacobian of the image I is defined as:

$$J = \begin{bmatrix} \frac{\partial I_1}{\partial x} & \frac{\partial I_1}{\partial y} \\ \frac{\partial I_2}{\partial x} & \frac{\partial I_2}{\partial y} \\ \cdots & \cdots \\ \frac{\partial I_N}{\partial x} & \frac{\partial I_N}{\partial y} \end{bmatrix} \tag{1}$$

The Di Zenzo structure tensor[9], which in differential geometry is known as the First Fundamental Form, is defined as the inner product of the Jacobian:

$$Z = J^T J \tag{2}$$

If $\mathbf{c} = [\alpha \;\; \beta]^T$ denotes a unit length vector then the squared magnitude of the multichannel gradient can be written as: $||J\mathbf{c}||_2^2 = \mathbf{c}^T Z\mathbf{c}$. That is, the structure tensor neatly summarizes the combined derivative structure of the multichannel image.

The singular value decomposition (SVD) of J uncovers structure that is useful both for the understanding of Socolinsky and Wolff's image fusion method and also our own POP image fusion algorithm presented in the next section:

$$J = USV^T \tag{3}$$

In Eq. 3, U, V and S are respectively $N \times N$ and 2×2 orthonormal matrices and a $N \times 2$ diagonal matrix. In the SVD decomposition - which is unique - the singular values are the components of the diagonal matrix S and are in order from largest to smallest. The ith singular value is denoted S_{ii} and the ith columns of U and V are respectively denoted U_i and V_i.

We can use the SVD to calculate the eigen-decomposition of the structure tensor Z:

$$Z = VS^2V^T \tag{4}$$

The most significant eigenvalue of Z is S_{11}^2 and the corresponding eigenvector is V_1. This eigenvector defines the direction of maximal gradient contrast in the image plane and S_{11} is the magnitude of this gradient.

In the Socolinsky and Wolff method[30], the 2-vector $S_{11}V_1$ is the basis of their *equivalent gradient* i.e. the derived gradient field that generates, per pixel, structure tensors that are closest to those defined from the multichannel image (Eq. 2). The per-pixel gradient field is written:

$$G(\mathbf{x}) = S_{11}^{\mathbf{x}}V_1^{\mathbf{x}} \tag{5}$$

In eq. 5 the superscript $^{\mathbf{x}}$ also denotes the x,y image location. We adopt this notation (rather than writing $S_{11}(\mathbf{x})V_1(\mathbf{x})$) to make the equations more compact., Respectively, $J^{\mathbf{x}}$, $Z^{\mathbf{x}}$, $U^{\mathbf{x}}$, $S^{\mathbf{x}}$ and $V^{\mathbf{x}}$ denote the per-pixel Jacobian, Di Zenzo tensor and the per-pixel SVD decomposition.

At this stage $G(\mathbf{x})$ in eq. 5 is ambiguous in its sign. Socolinsky and Wolff set the sign to match the brightness gradient (i.e. (R+G+B)/3) in the gradient orientation orientation V_1). The sign can also be optimized to maximize the integrability of the derived gradient field[10]. Once we fix the sign, we write

$$G(\mathbf{x}) = sign(\mathbf{x})S_{11}^{\mathbf{x}}V_1^{\mathbf{x}} \tag{6}$$

In general the derived gradient field $G(\mathbf{x})$ is not integrable (the curl of the field is not everywhere 0). So, Socolinsky and Wolff solve for the output image $O(\mathbf{x})$ in a least-squares sense by solving the Poisson equation:

$$G_{xx} + G_{yy} = \nabla^2 O(\mathbf{x}) \tag{7}$$

where $[G_{xx}\ G_{yy}]$ denotes the divergence of the gradient field. Unfortunately, because the gradient field is non integrable, O must have details (gradients) that are not present in the multichannel input I. For example, in Figure 1 we see, in 'SW', 'bending artifacts' which are not in either of the image planes that were fused. This kind of *hallucinated* artifact is common as are 'halos' at high contrast edges.

In [7] it was argued that so long as one expects the equivalent gradient to be *almost* integrable across different scales then the reintegrated image should be a global mapping of inputs. In effect, the reintegration step reduces to finding a global mapping - a look-up-table - of the original image that has derivatives close to the Socolinsky and Wolff's equivalent gradients[12]. Often the look-up-table reintegration theorem delivers surprisingly good image fusion (it looks like the Socolinsky and Wolff image but without the artifacts). Yet, sometimes the constraint that the output image is a simple global function of the output can result in a fused image that does not well represent the details in the individual bands of the multichannel image.

2.1. SVD, PCA and Characteristic Vector Analysis

Finally we remark that $V_1^{\mathbf{x}}$ (the eigenvector associated with the largest eigenvalue) of $Z^{\mathbf{x}}$ is exactly the principal characteristic vector of the rowspace of $J^{\mathbf{x}}$. It is the vector direction along which the projection of the rows of $J^{\mathbf{x}}$ have maximum variance (Characteristic vector analysis is the same as principal component analysis where the mean is not subtracted from the data before the maximum variance direction is calculated[20]). All data matrices can be analyzed in terms of their row or column space. The vector $U_1^{\mathbf{x}}$ is the vector direction along which the projection of the columns of $J^{\mathbf{x}}$, the principal characteristic vector of the column space, have maximum variance. The vector $\acute{U}^x$ is simply the 1st column vector of U^x:

$$\acute{U}^{\mathbf{x}} = U_1^{\mathbf{x}} \tag{8}$$

3. POP image fusion

The derived gradient in the Socolinsky and Wolff method is a mathematically well founded fusion of all the available gradient information from a multichannel image. That said, Socolinsky and Wolff can produce poor looking results as a result of the ill-posedness of gradient field reintegration (of non-integrable fields).

The basic premise of our method is that we can carry out image fusion without the need to reintegrate. Rather, we seek only to find a per-pixel projection (linear combination) of the input channels such that if we differentiated the output projected image we would generate the equivalent gradients we seek. It turns out that not only can we take this projection approach, but that the projection direction is the **P**rincipal characteristic vector of the **O**uter **P**roduct of the Jacobian. The key intuition behind the method is to think of projection in image space (calculated using gradients), leading to an output scalar image, rather than the conventional projection in the gradient domain, which gives output gradients that are often impossible to reintegrate without artifacts.

POP Image Fusion Theorem: The scalar formed by the projection by the first characteristic vector of the outer product of the Jacobian at a single discrete location $\mathbf{x}$ (denoted $P(\mathbf{x}) = \acute{U}^{\mathbf{x}}.I(\mathbf{x}) = \sum_{k=1}^{N} \acute{U}_k^{\mathbf{x}} I_k(\mathbf{x})$) has, assuming the functions $I_k(\mathbf{x})$ are continuous, the property that re$[\frac{\partial}{\partial_x}(P(\mathbf{x})),\ \frac{\partial}{\partial_y}(P(\mathbf{x}))]^T = s^{\mathbf{x}}G(\mathbf{x})$ (where $s^{\mathbf{x}} = -1$ or 1)

Proof: Because differentiation and summation are linear operators, and because we are assuming the underlying functions are continuous,

$$\begin{array}{l}\frac{\partial}{\partial_x}(P(\mathbf{x})) = \sum_{k=1}^{N} \acute{U}_k^{\mathbf{x}} \frac{\partial}{\partial_x}(I_k(\mathbf{x})) \\ \frac{\partial}{\partial_y}(P(\mathbf{x})) = \sum_{k=1}^{N} \acute{U}_k^{\mathbf{x}} \frac{\partial}{\partial_y}(I_k(\mathbf{x}))\end{array} \tag{9}$$

Remembering that $U^{\mathbf{x}}$ is part of the singular value decomposition of the Jacobian - see Eq. 3 - and that, accordingly, $U^{\mathbf{x}}$ and $V^{\mathbf{x}}$ in this decomposition are orthonormal matrices and that $S^{\mathbf{x}}$ is a diagonal matrix, it follows directly that

$$[\frac{\delta}{\delta_x}(P(\mathbf{x}))\ \frac{\delta}{\delta_y}(P(\mathbf{x}))] = [S_{11}^{\mathbf{x}} V_1^{\mathbf{x}}] \tag{10}$$

Of course just as we have an unknown sign when we derive $G(\mathbf{x})$ from inner product tensor analysis the sign ambiguity remains here. We set $s^{\mathbf{x}}$ to 1 or -1 so that $[\frac{\delta}{\delta_x}(P(\mathbf{x}))\ \frac{\delta}{\delta_y}(P(\mathbf{x}))]^t = s^{\mathbf{x}} G(\mathbf{x})$.

■

While the sign in the proof is chosen to map the derived gradient of the Socolinsky and Wolff method we need not set it in this way. Indeed, because we are ultimately wanting to fuse an image that has positive image values we do not adopt the Socolinsky and Wolff[30] heuristic method. Rather we choose the sign so that the projected image is positive (a necessary property of any fused image):

$$s^{\mathbf{x}} = sign(\acute{U}^{\mathbf{x}}.I(\mathbf{x})) \tag{11}$$

Equation 11 always resolves the sign ambiguity in a well defined way (and as such is an important advance compared to Socolinsky and Wolff).

The POP image fusion theorem is for a single image point and assumes the underlying multichannel image is continuous. We wish to understand whether we can sensibly apply the POP image fusion theorem at all image locations and even when the underlying image is not continuous.

First, we remark that we can write $U^{\mathbf{x}}$ as

$$U^{\mathbf{x}} = J^{\mathbf{x}} V^{\mathbf{x}} [S^{\mathbf{x}}]^{-1} \tag{12}$$

that is, $U^{\mathbf{x}}$ is the product of the Jacobian and the the inverse of the square root of the structure tensor (the structure tensor decomposition in terms of the SVD is given in Equation 4 from which it follows that the inverse of the square root of the structure tensor is VS^{-1}). Because the structure tensor is positive-semidefinite the eigenvalues are always real and positive and, assuming the underlying multichannel image is continuous and that the eigenvalues are distinct then $\acute{U}^{\mathbf{x}}$ - the principal characteristic vector of the outer product matrix - will also vary continuously. However, in image regions with zero derivatives or where the structure tensor has coincident eigenvalues (e.g. corners) there may be a large change in the projection direction found at one image location compared to another (discontinuity). It follows then that we must interpolate, or diffuse, the projection vectors that are well defined across the image. We achieve this by applying a simple cross bilateral filter, which provides superior results to a standard Gaussian or median filter, as it uses the image structure contained in the input image channels to guide the diffusion of the projection vectors. While there are other ways of providing an 'in-filled' projection map including anisotropic diffusion[26], connected component labeling (enforcing the same projection for the same connected component in the input (in analogy to [29]) or enforcing spatial constraints more strongly than in bilateral filtering[21], we found the bilateral approach worked well for our purposes. After bilateral filtering - and a couple of additional post processing steps described in 3.1 - we have N values per pixel defining a projection direction along which we project the N-vector $I(x)$ to make a scalar output image. It could be said that this diffusion process is an analogue of gradient field reintegration, in that it combines gradient information across the image plane and produces a semi-global optimization - however, unlike previous methods it avoids halo and bending artifacts.

Let us denote $\mathcal{P}(\mathbf{x}) : \mathcal{D} \subset \mathbb{R}^2 \rightarrow \mathcal{C} \subset \mathbb{R}^N$ as the *projector image*. In POP image fusion the scalar output image $O(\mathbf{x})$ is calculated as a simple per pixel dot product.

$$O(\mathbf{x}) = \mathcal{P}(\mathbf{x}).I(\mathbf{x}) \tag{13}$$

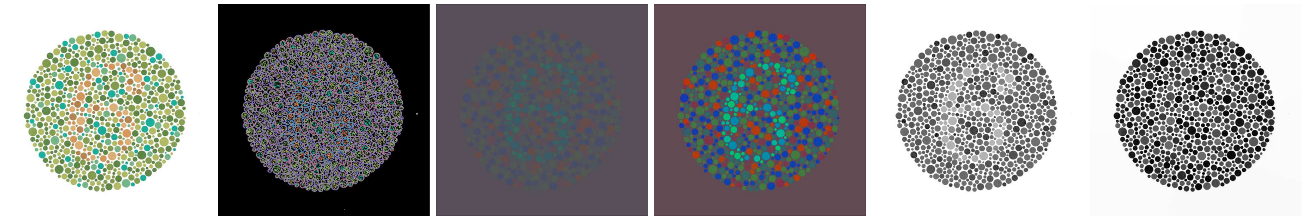
3.1. Algorithm for finding the Projector Image

initialize $\mathcal{P}(\mathbf{x}) = \mathbf{0}$ (initialize to the 0 projection at every pixel location).

1. For all image locations $\mathbf{x}$ calculate the Jacobian $J^{\mathbf{x}}$
2. If $min(S_{11}^{\mathbf{x}}, S_{22}^{\mathbf{x}}) > \theta_1$ and $|S_{11}^{\mathbf{x}} - S_{22}^{\mathbf{x}}| > \theta_2$ then $\mathcal{P}(\mathbf{x}) = \acute{U}^{\mathbf{x}}$ (at this stage $\mathcal{P}(\mathbf{x})$ is sparse).
3. $\mathcal{P}(\mathbf{x}) = BilatFilt(I(\mathbf{x}), \mathcal{P}(\mathbf{x}), \sigma_r, \sigma_d)$.
4. $\mathcal{P}(\mathbf{x}) = \mathcal{P}(\mathbf{x})/||\mathcal{P}(\mathbf{x})||$.
5. $\mathcal{P}(\mathbf{x}) = spread(P(\mathbf{x}))$.

Implementation Details

In step 2, θ_1 and θ_2 are set arbitrarily to .01 (assuming image values are in [0,1]). The function $BilatFilt()$ is a cross bilateral filter with the range term defined by the original image I. The filtering is carried out independently per channel with a Gaussian spatial blur with standard deviation σ_d and the standard deviation on the range parameterised by σ_r. With $\sigma_d = \sigma_r = 0$, no diffusion takes place. As $\sigma_d \rightarrow \infty$ and $\sigma_r \rightarrow \infty$, the diffusion becomes a global mean, and the projection tends to a global weighted sum of the input channels. If $\sigma_d \rightarrow \infty$ and $\sigma_r = 0$ each distinct vector of values in the image will be associated with the same projection vector and so the bilateral filtering step defines surjective mapping which could be implemented as a look-up table[12]. Excepting these boundary cases the



(a) RGB - input image to the POP method (b) Sparse $\mathcal{P}$ projectors (step 2) (c) $\mathcal{P}$ after normalization (step 4) (d) $\mathcal{P}$ after $spread$ operation (step 5) (e) POP image fusion result (f) SW image fusion result

Figure 2: In (a) we show an Ishihara plate. The initial projector image derived in POP image fusion is shown in (b) - note how edgy and sparse it is - and after bilateral filtering and normalization (steps 3 and 4) in (c). The $spread$ function is applied giving the final projector in (d). The per-pixel dot product of (a) with (d) is shown in (e). For comparison in (f) we show the output of the Socolinsky and Wolff Algorithm.

standard deviations in the bilateral filter should be chosen to provide the diffusion we seek, but we also need to make sure the spatial term is sufficiently large to avoid spatial artifacts. In our experiment σ_d and σ_r are set to $min(X, Y) * 4$ and $((max(I) - min(I))/4)$), these values were found empirically.

After the bilateral filtering, $\mathcal{P}$ is dense, but each projection direction is not a unit vector. This is remedied in step 4. Finally, we apply a spreading function $spread()$ to move each of the projection directions a fixed multiple of angular degrees away from the mean (the diffusion step pulls in the opposite direction and results in projection directions closer to the mean compared with those found at step 2 in the algorithm). By default, we simply compute the average angular deviation from the mean before and after the diffusion. We scale the post-diffusion vectors by a single factor k ($k \geq 1$) so that the average angular deviation is the same as prior to the diffusion step. If the spread function creates negative values we clip to 0. This scaling factor k can be varied according to the requirements of each application.

3.2. Fast Implementation

To speed up the technique, the input images may be downsampled and $\mathcal{P}$ calculated only for the thumbnail image. We then use joint bilateral upsampling[16] to find the full resolution projector image. We remark that this thumbnail computation also has the advantage that the projector image can be computed in tiles i.e. we never need to calculate the full resolution projector image.

An example RGB-NIR image pair at 682 × 1024 resolution (the image pair shown in fig. 5), fused as separate R, G and B channels for a total of 3 fusion steps, takes 54.93 seconds at full resolution, and 2.82 seconds when calculated on 68 × 102 downsampled thumbnail images, using a MATLAB implementation of our method. This increase in speed does not significantly affect the resulting image - the mean SSIM[33] between the full resolution and downsampled results over the corresponding image channels is 0.9991. In general we found that we could downsample aggressively to 10K pixel thumbnails (or, even slightly less as in this example) with good results. Though, almost always if we downsized to approximately VGA[3] resolution then the results we computed on the thumbnails would be close to identical as those computed on the full resolution image.

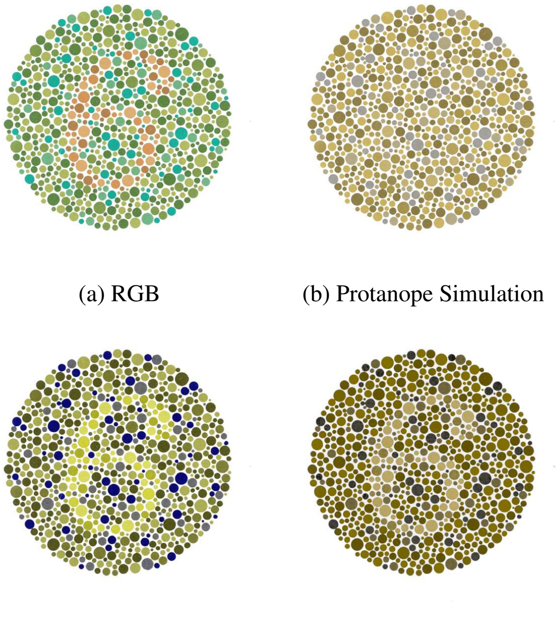
4. Experiments

We compare our method against two state-of-the-art algorithms, the image fusion method of Eynard *et al.*, based on using the graph Laplacian to find an M to N channel color mapping, and the Spectral Edge (SE) method of Connah *et al.* [7], which is based on the structure tensor together with look-up-table based gradient reintegration[12].

In our supplementary material, we include the full-size images from this paper, several more RGB-NIR image fusion comparisons, and color to greyscale conversion examples of the entire Ĉadík data set[6].

In Fig. 2 we show the colour to greyscale image fusion example of an Ishihara plate used to test colour blindness (if the reader cannot see a number look at 2e for a greyscale representation of what most people see!). In 2f we show the output of the Socolinsky and Wolff image fusion algorithm. Socolinsky and Wolff fails here because the image is composed of circles of colour on a white background. Because all edges are isolated in this way, the equivalent gradient field exactly characterizes the colour gradients and is integrable and the output in 2f does not have integration artifacts. Yet, the fused images does not capture the actual look and feel of the input. In contrast POP image fusion which produces an output 2e (see Fig 2 caption for description of the intermediate steps) the initial projection directions are diffused with the bilateral filtering step enforcing the projection directions calculated at a pixel to be considered in concert with other image regions.

We can use this greyscale output for image optimization for color-deficient viewers, as shown in fig. 3. The POP re-


(a) RGB (b) Protanope Simulation

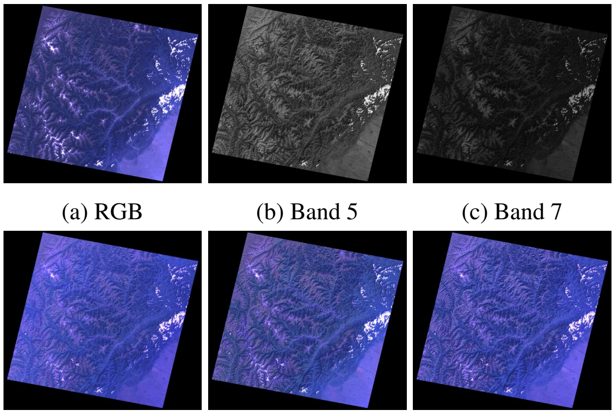
(c) Eynard *et al.* (d) POP

Figure 3: Image optimization for color-deficient viewers: Protanope Comparison (Ishihara plate 3). The orange digit 6 disappears in the Protanope simulation image, but is clearly visible in both the results of Eynard *et al.* and the POP method.

sult 2e is used as a luminance channel replacement in LUV color space for the Protanope simulation image 3b, mapping color variation in the original RGB image 3a that is invisible to color-deficient observers, to luminance channel detail which they can perceive. For this application we use a downsampling ratio of 0.5 and a k stretching parameter of 2. The result of Eynard *et al.* is also presented as a comparison - both methods achieve the desired result, although the result of Eynard *et al.* produces a higher level of discrimination, as their fusion changes the output color values, whereas our method only affects luminance.

4.1. Remote Sensing

Images captured for remote sensing applications normally span the visible and infrared wavelength spectrum. We use data from Landsat 5's Thematic Mapper (TM)[1]. The Landsat 5 TM captures 7 image channels - 3 in the visible spectrum and 4 infrared images. The three visible images are captured from 0.45-0.51μm (blue), 0.52-0.60μm (green), and 0.63-0.69 μm (red), and we use these as the B, G and R channels respectively of the input RGB image. In fig. 4a, we show an input RGB image from the Landsat image set, and in 4b and 4c the infrared bands 5 and 7 which include extra detail not present in the RGB bands. All 4 infrared channels are used in the fusion, but only 2 are shown here for reasons of space. The 4 infared channels are fused with the R, G and B channels in turn using the Socolinsky and Wolff method in 4d and the POP method in 4f, and then the output RGB channels have high and low quantiles matched to the input RGB channels. In fig. 4e we show the result of the Spectral Edge method[7], which directly fuses the RGB image and all 7 multiband images. For this application we use a downsampling ratio of 0.5 and a k stretching parameter of 2.


(a) RGB (b) Band 5 (c) Band 7

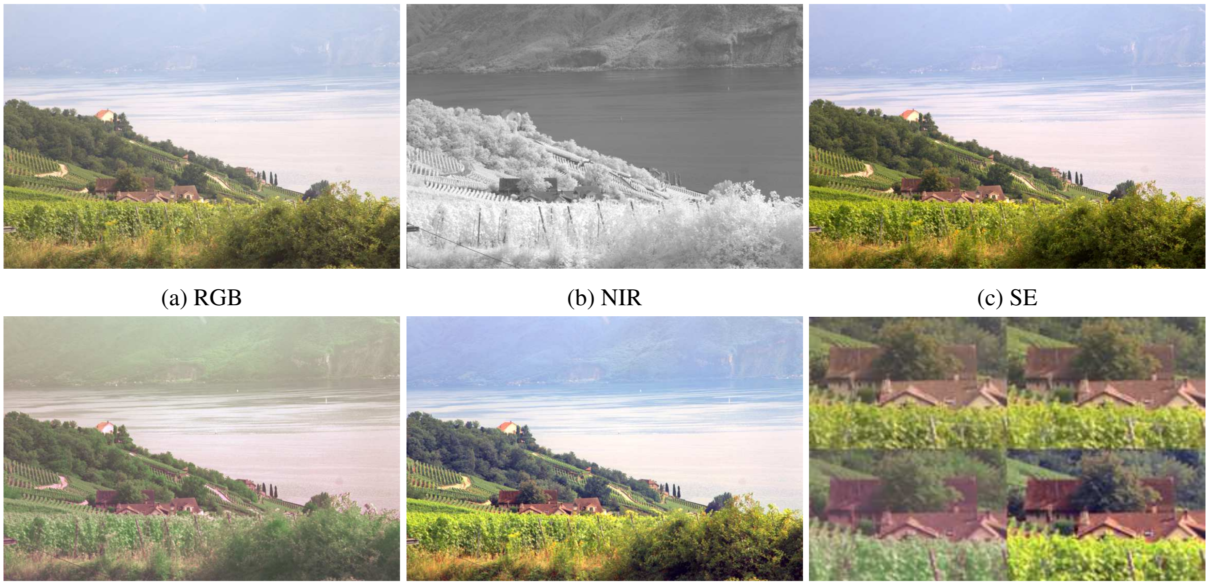
(d) SW (e) SE (f) POP

Figure 4: Remote sensing image fusion - Landsat 5[1]: original RGB image, bands 1-3 (a), infrared bands 5 (b) and 7 (c) capture extra detail, which is fused with the RGB by the SW (d), SE (e) and POP (f) methods.

Both the SE and POP method produce significantly more detailed results than the SW method. In our opinion, the POP method's result is slightly preferred to that of SE, as its details are sharper and more crisp.

4.2. RGB-NIR Image Fusion

In fig. 5 we wish to fuse the conventional RGB image (5a) with an near-infrared (NIR) image (5b). We apply POP image fusion 3 times - we fuse the R-channel with the NIR, the G-channel with the NIR and the B-channel with the NIR. We perform post-processing in which we stretch the images so that their 0.05 and 0.95 quantiles are the same as the original RGB image. The POP Image fusion result is shown in fig. 5e. For comparison we show the Spectral Edge output, Fig. 5c and the Eynard *et al.* output 5d. In the same image order we show a magnified detail inset in 5f. The output image of the POP method captures more NIR detail than the SE result, while producing more natural colors than the result of Eynard *et al.*, which has a green color cast and a lack of color contrast. The POP result shows good color contrast, naturalness and detail. For this application we use a downsampling ratio of 0.1 and a k stretching



(a) RGB (b) NIR (c) SE

(d) Eynard *et al*. (e) POP (f) Detail Comparison

Figure 5: RGB-NIR Image Fusion: 'Water47'[5] Comparison - original RGB and near-infrared input images, Spectral Edge, Eynard *et al*. and POP results (detail, top-left: RGB, top-right: SE, bottom-left: Eynard *et al*., bottom-right: POP). The POP result has superior contrast and detail compared to the other methods. The SE result is natural and adds extra detail, while the result of Eynard *et al*. transfers NIR detail effectively, but suffers from a green color cast.

parameter of 1.

4.3. Multifocus image fusion

Multifocus image fusion is another potential application, which has typically been studied using greyscale images with different focal settings[17][18]. Standard multifocus image fusion involves fusing two greyscale input images with different focal settings. In each input image approximately half the image is in focus, so by combining them an image in focus at every point can be produced.

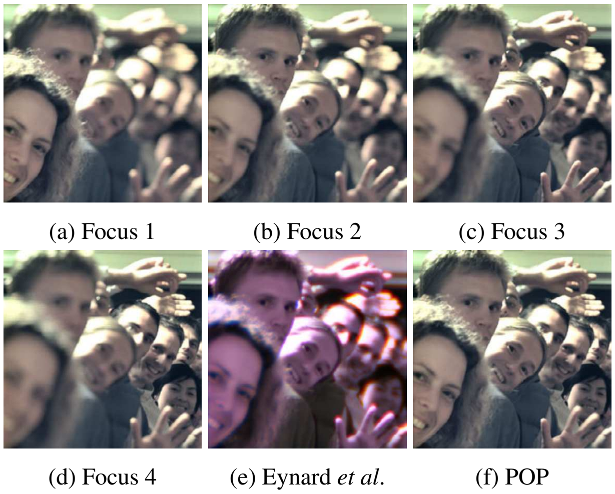
Table 1 shows a comparison of the performance of the POP image fusion method on this task, on several standard multifocus image pairs, using standard image fusion quality metrics. The $Q_{XY/F}$ metric is based on gradient similarity[35], the $Q(X, Y, F)$ metric is based on the structural similarity image measure (SSIM)[33][36], and the M_F^{XY} metric is based on mutual information[15]. The results are compared to the method of Zhou and Wang, based on multi-scale weighted gradient-based (MWGF) fusion[38], as well as a standard DWT fusion, using a Daubechies wavelet and CM (choose max) coefficient selection - the POP result comes out ahead in the majority of cases.

Image Pair	Metric	DWT	MWGF	POP
Book	$Q_{XY/F}$	0.8208	0.8327	$\mathbf{0.8332}$
	$Q(X, Y, F)$	$\mathbf{0.8053}$	0.8027	0.8008
	M_F^{XY}	0.9738	$\mathbf{1.227}$	1.057
Clock	$Q_{XY/F}$	0.7860	0.7920	$\mathbf{0.7956}$
	$Q(X, Y, F)$	$\mathbf{0.8008}$	0.7955	0.7910
	M_F^{XY}	0.7475	1.142	$\mathbf{1.248}$
Desk	$Q_{XY/F}$	0.7907	$\mathbf{0.8287}$	0.8242
	$Q(X, Y, F)$	0.7933	0.7978	$\mathbf{0.7979}$
	M_F^{XY}	0.7261	1.072	$\mathbf{1.248}$
Pepsi	$Q_{XY/F}$	0.8648	0.8800	$\mathbf{0.8820}$
	$Q(X, Y, F)$	$\mathbf{0.7950}$	0.7725	0.7792
	M_F^{XY}	0.8751	1.196	$\mathbf{1.210}$

Table 1: Multifocus Fusion: table of metric results.

Plenoptic photography provides various refocusing options of color images, allowing images with different depths of focus to be created from a single exposure[24]. The POP method can be used to fuse these differently focused images into a single image wholly in focus. Our method can be fine tuned for this application, due to the knowledge that only one of the images is in focus at each pixel. Here we apply a large k scaling term in the $spread$ function, and we use a downsampling ratio of 0.5. This allows a crystal clear output image, in focus at every pixel, to be created.

Fig. 6 shows an image (from Ng *et al*. [24]), in which four different refocused images are created from a single exposure. The POP method is used to fuse these differently focused images into a single image in focus at every point - in comparison the result of the method of Eynard *et al*.


(a) Focus 1 (b) Focus 2 (c) Focus 3

(d) Focus 4 (e) Eynard *et al.* (f) POP

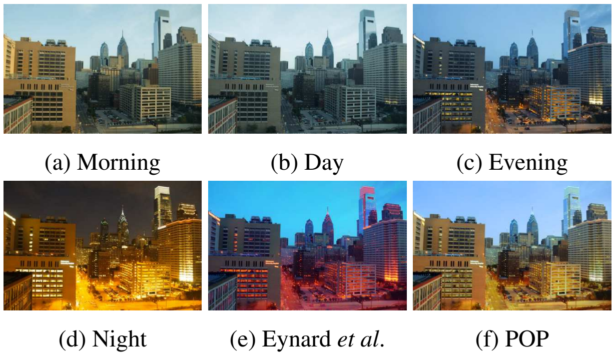
Figure 6: Multifocus Fusion: four color input images with different points of focus captured with one exposure using a plenoptic camera, and the fusion results of Eynard *et al.* and the POP method. The POP result brings details across the image into sharper focus with natural color.

does not show perfect detail in all parts of the image, and has unnatural color information.

4.4. Merging time-lapse photography

Time-lapse photography involves capturing images of the same scene at different times[11]. These can be fused using the standard POP method in the case of greyscale images, and for RGB images the stacks of R, G and B channels are fused separately. This fusion result creates an output image which combines the most salient details of all the time-lapse images. For this application we use a downsampling ratio of 0.5 and a k stretching parameter of 2.

Fig. 7 shows a series of time-lapse images (from Eynard *et al.* [11]) from different parts of the day and night, and results of POP fusion and the method of Eynard *et al.* The details only visible with artificial lighting at night are combined with details only visible in the daytime in both results, but the POP result produces far more natural colors. It must be noted that the result presented here for Eynard *et al.* is different in color to that presented in their paper - we ran the code they provided on the input images ourselves and it produced this result. We believe the POP result is a more natural fusion in either case.


(a) Morning (b) Day (c) Evening

(d) Night (e) Eynard *et al.* (f) POP

Figure 7: Time-lapse Photography - Fusion of Multiple Illuminations: four color input images captured at different times of day and night, and the fusion results of Eynard *et al*. and the POP method. The POP result has far more natural colors and detail.

5. Conclusion

In this paper, we have proposed a new image fusion method based on image derivatives. It avoids integrability problems with gradient reintegration methods, by calculating a projection of the input image channels per-pixel based on the principal characteristic vector of the outer product of the Jacobian matrix of image derivatives. We have proved that, in the case of continuous multichannel images with derivatives at every point, this produces an output projected image with derivatives equal to the equivalent gradients found by Socolinsky and Wolff from the Di Zenzo structure tensor. In real images, derivative information is sparse, so we diffuse the projection coefficients among similar image regions using a joint bilateral filter, before projecting the input image channels to produce an output image. We call this the **P**rincipal characteristic vector of the **O**uter **P**roduct image fusion method.

We have explained how the POP method can be optimized to improve its performance, and how it can be applied to RGB-NIR image fusion, color to greyscale conversion and multifocus image fusion. We have compared our method to state of the art methods for RGB-NIR image fusion, image optimization for color-deficient viewers, and remote sensing, and provided illustrative results for multifocus image fusion based on plenoptic imaging and the fusion of time-lapse images.

The POP method produces results visually superior to the other methods we have tested - its output images have high levels of detail with minimal artifacts.

Acknowledgements

Much gratitude goes to the EPSRC, whose funding supports the studentship of Alex E. Hayes. This research is also, in part, funded by EPSRC grant M001768. Many thanks to Davide Eynard, for providing code for the Laplacian colormaps method, as well as example fusion results. Also, thank you to Spectral Edge Ltd., for providing the image fusion results of their method used in this paper.

References

[1] NASA: Landsat 5 imagery. http://landsat.usgs.gov/. Accessed: 2015-04-22. 6

[2] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field? *Computer Vision, European Conference on*, pages 578–591, 2006. 1

[3] W. Berry. Vga controller card, Sept. 22 1992. US Patent 5,150,109. 5

[4] J. Bigun. *Vision with direction*. Springer, 2006. 1

[5] M. Brown and S. Susstrunk. Multi-spectral sift for scene category recognition. *Computer Vision and Pattern Recognition, IEEE Conference on*, pages 177–184, 2011. 7

[6] M. Ĉadík. Perceptual evaluation of color-to-grayscale image conversions. *Computer Graphics Forum*, 27(7):1745–1754, 2008. 5

[7] D. Connah, M. S. Drew, and G. D. Finlayson. Spectral Edge image fusion: Theory and applications. *Computer Vision, European Conference on*, pages 65–80, 2014. 1, 3, 5, 6

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition, IEEE Conference on*, 1:886–893, 2005. 1

[9] S. Di Zenzo. A note on the gradient of a multi-image. *Computer vision, Graphics, and Image Processing*, 33(1):116–125, 1986. 1, 2

[10] M. S. Drew, D. Connah, G. D. Finlayson, and M. Bloj. Improved colour to greyscale via integrability correction. *IS&T/SPIE Electronic Imaging*, pages 72401B–72401B, 2009. 1, 3

[11] D. Eynard, A. Kovnatsky, and M. M Bronstein. Laplacian colormaps: a framework for structure-preserving color transformations. *Computer Graphics Forum*, 33(2):215–224, 2014. 8

[12] G. D. Finlayson, D. Connah, and M. S. Drew. Lookup-table-based gradient field reconstruction. *Image Processing, IEEE Transactions on*, 20(10):2827–2836, 2011. 3, 4, 5

[13] W. Förstner. A feature based correspondence algorithm for image matching. *International Archives of Photogrammetry and Remote Sensing*, 26(3):150–166, 1986. 1

[14] S. Han, W. Tao, D. Wang, X.-C. Tai, and X. Wu. Image segmentation based on grabcut framework integrating multiscale nonlinear structure tensor. *Image Processing, IEEE Transactions on*, 18(10):2289–2302, 2009. 1

[15] M. Hossny, S. Nahavandi, and D. Creighton. Comments on information measure for performance of image fusion. *Electronics letters*, 44(18):1066–1067, 2008. 7

[16] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Transactions on Graphics*, 26(3):96, 2007. 5

[17] S. Li, J. T. Kwok, and Y. Wang. Multifocus image fusion using artificial neural networks. *Pattern Recognition Letters*, 23(8):985–997, 2002. 1, 7

[18] S. Li and B. Yang. Multifocus image fusion using region segmentation and spatial frequency. *Image and Vision Computing*, 26(7):971–979, 2008. 1, 7

[19] B. Lu, H. Wang, and C. Miao. Medical image fusion with adaptive local geometrical structure and wavelet transform. *Procedia Environmental Sciences*, 8:262–269, 2011. 1

[20] L. T. Maloney. *Computational approaches to color constancy*. PhD thesis, Stanford University, 1984. 3

[21] L. Meylan and S. Süsstrunk. High dynamic range image rendering with a retinex-based adaptive filter. *Image Processing, IEEE Transactions on*, 15(9):2820–2830, 2006. 4

[22] R. Montagna and G. D. Finlayson. Reducing integrability error of color tensor gradients for image fusion. *Image Processing, IEEE Transactions on*, 22(10):4072–4085, 2013. 1

[23] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone. Remote sensing image fusion using the curvelet transform. *Information Fusion*, 8(2):143–156, 2007. 1

[24] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report*, 2(11), 2005. 7

[25] G. Pajares and J. M. De La Cruz. A wavelet-based image fusion tutorial. *Pattern recognition*, 37(9):1855–1872, 2004. 1, 2

[26] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, 1990. 4

[27] G. Piella. Image fusion for enhanced visualization: a variational approach. *International Journal of Computer Vision*, 83(1):1–11, 2009. 1

[28] D. Reddy, A. Agrawal, and R. Chellappa. Enforcing integrability by error correction using 1-minimization. *Computer Vision and Pattern Recognition, IEEE Conference on*, pages 2350–2357, 2009. 1

[29] D. A. Socolinsky. *A Variational Approach to Image Fusion*. PhD thesis, John Hopkins University, 2000. 1, 4

[30] D. A. Socolinsky and L. B. Wolff. Multispectral image visualization through first-order fusion. *Image Processing, IEEE Transactions on*, 11(8):923–931, 2002. 1, 3, 4

[31] A. Toet. Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters*, 9(4):245–253, 1989. 1

[32] C. Wang, Q. Yang, X. Tang, and Z. Ye. Salience preserving image fusion with dynamic range compression. *Image Processing, IEEE International Conference on*, pages 989–992, 2006. 1

[33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004. 5, 7

[34] Z. Wang and Y. Ma. Medical image fusion using m-pcnn. *Information Fusion*, 9(2):176–185, 2008. 1

[35] C. Xydeas and V. Petrović. Objective image fusion performance measure. *Electronics Letters*, 36(4):308–309, 2000. 7

[36] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu. A novel similarity based quality metric for image fusion. *Information Fusion*, 9(2):156–160, 2008. 7

[37] W. Zhang and W.-K. Cham. Gradient-directed multiexposure composition. *Image Processing, IEEE Transactions on*, 21(4):2318–2323, 2012. 1

[38] Z. Zhou, S. Li, and B. Wang. Multi-scale weighted gradient-based fusion for multi-focus image. *Information Fusion*, 2014. 7