New Techniques in Derivative Domain Image Fusion and Their Applications

by

Alex E Hayes

A thesis submitted to the
School of Computing Sciences
in conformity with the requirements for
the degree of Doctor of Philosophy

University of East Anglia
Norwich, United Kingdom
May 2018

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 (based on image gradients). In this thesis, we propose new derivative domain image fusion methods and metrics, and carry out experiments on a range of image fusion applications.

After reviewing previous relevant methods in derivative domain image fusion, we make several new contributions. We present new applications for the Spectral Edge image fusion method, in thermal image fusion (using a FLIR smartphone accessory) and near-infrared image fusion (using an integrated visible and near-infrared sensor). We propose extensions of standard objective image fusion quality metrics for M to N channel image fusion - measuring image fusion performance is an unsolved problem.

Finally, and most importantly, we propose new methods in image fusion, which give improved results compared to previous methods (based on metric and subjective comparisons): we propose an iterative extension to the Spectral Edge image fusion method, producing improved detail transfer and colour vividness, and we propose a new derivative domain image fusion method, based on finding a local linear combination of input images to produce an output image with optimum gradient detail, without artefacts - this mapping

can be calculated by finding the principal characteristic vector of the outer product of the Jacobian matrix of image derivatives, or by solving a least-squares regression (with regularization) to the target gradients calculated by the Spectral Edge theorem. We then use our new image fusion method on a range of image fusion applications, producing state of the art image fusion results with the potential for real-time performance.

Acknowledgments

I would like to thank my supervisor, Prof. Graham D. Finlayson, for his many ideas and valuable support. Sometimes my mind drifts off into the clouds, and he brings me back to reality and mathematical rigour.

Spectral Edge Ltd. has been a great company to work for, and has provided some exciting opportunities to use image fusion in the 'real world'. I'd like to thank the company for its images and algorithms used in this thesis.

I would also like to thank Davide Eynard, for the code for the Laplacian colourmaps image fusion method and assistance in using it.

Publications

The following publications are related to this thesis:

- Hayes, A. E., Finlayson, G. D., & Montagna, R. (2015). RGB-NIR color image fusion: metric and psychophysical experiments. In SPIE/IS&T Electronic Imaging (pp. 93960U-93960U-9). International Society for Optics and Photonics.

- Finlayson, G. D., & Hayes, A. E. (2015). Iterative Spectral Edge image fusion. In Color and Imaging Conference (pp. 41-45). Society for Imaging Science and Technology.

- Finlayson, G. D., & Hayes, A. E. (2015). POP image fusion – derivative domain image fusion without reintegration. In Proceedings of the IEEE International Conference on Computer Vision (pp. 334-342).

- Hayes, A. E., Montagna, R., & Finlayson, G. D. (2016). New applications of Spectral Edge image fusion. In SPIE Defense+Security (pp. 984009-984009-11). International Society for Optics and Photonics.

Glossary

- **POP** - **P**rincipal characteristic vector of the **O**uter **P**roduct
- **SW** - **S**ocolinsky and **W**olff
- **LLC** - **L**ocal **L**inear **C**ombination
- **SE** - **S**pectral **E**dge
- **DWT** - **D**iscrete **W**avelet **T**ransform
- **ROLP** - **R**atio of **L**ow-**p**ass **P**yramid
- **DOLP** - **D**ifference of **L**ow-**p**ass **P**yramid
- **NIR** - **N**ear **I**nfra**r**ed
- **LUT** - **L**ook**u**p **T**able
- **MEF** - **M**ulti-**e**xposure **F**usion

Mathematical Notation

- **J** - Jacobian matrix of image derivatives.
- **Z** - inner product of the Jacobian matrix of image derivatives ($J^T J$).
- **H** - input high-dimensional image.
- ∇**D** - derived gradients of fused image.
- **O** - output fused image.
- **R** - input RGB guide image.
- $\mathcal{P}$ - local linear combination (LLC) weights.
- P_n - polynomial expansion of order n.
- **A, B, F** - for image fusion metrics: A, B - input images, F - fused image.
- **w** - sliding image window.
- **W** - the set of windows across the image plane.
- **Q** - image quality.
- J^x - compact superscript notation used to indicate a particular x, y image location (in this case the Jacobian matrix at that location).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Image fusion can have many definitions, but in this thesis it is the task of combining the most salient details of two (or more) input images into one composite image - saliency can be measured in many ways, such as image derivatives, wavelet coefficients, image Laplacian coefficients, and many more. Compressing high-dimensional image data into a smaller number of dimensions will always involve a loss of information, but minimizing this loss is the task of image fusion algorithms - as well as, in many cases, optimizing the output image in terms of subjective quality and lack of artefacts.

The topic of this thesis is exploring new ways of doing this, focusing on image fusion techniques based on first order image derivatives, and then assessing the performance of these methods.

There are various uses of image fusion, but the classic examples involve two greyscale images being fused into a single summary greyscale image. As shown in figure 1.1 the 'UN Camp' image pair [18] consists of thermal IR and visible images, in which a person is clearly visible in the thermal image, but not in the visible image, while the terrain is more detailed in the visible image. In a surveillance setting, the fusion of these images is clearly advantageous, to see both the figure and the background in maximum detail. Another

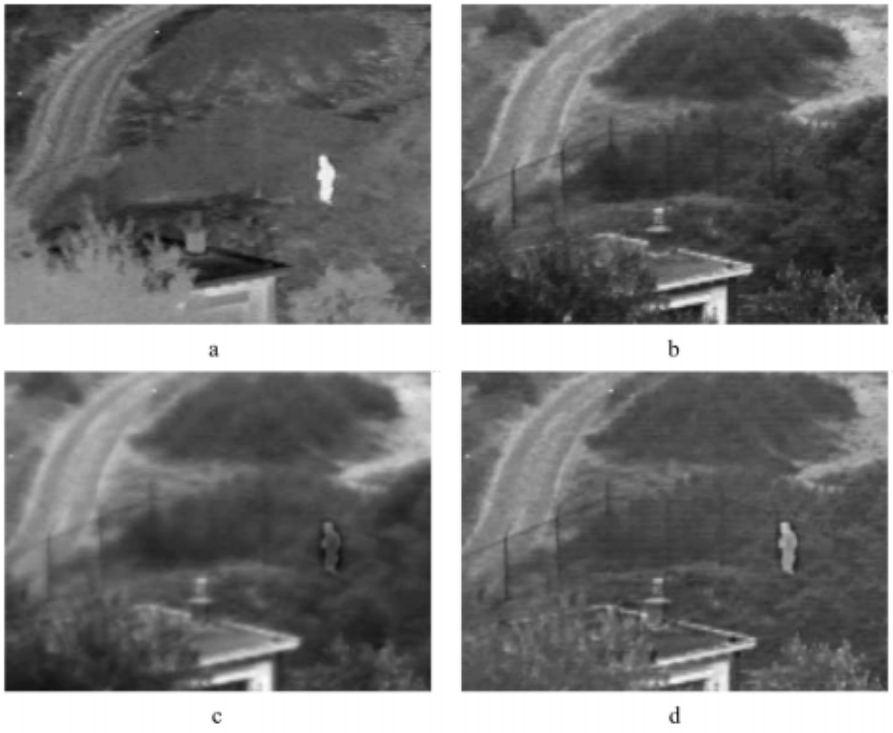


Figure 1.1: Introduction: visible-thermal image fusion (“UN Camp”) - (a) thermal IR (b) visible (c) and (d) two fusion results (images and results from [20])

popular image fusion example involves two greyscale images of clocks, shown in figure 1.2, in each of which approximately half of the image is in focus. Multifocus image fusion can be used to create an image in focus at every point from images with different depths of focus [52]. New results for this task are shown in section 6.3 of this thesis, in which our proposed method performs better in a majority of cases, on standard image fusion metrics.

More recently, colour image fusion has become more of a topic of interest. Remote sensing produces visible colour images with lower resolution and multispectral greyscale images with higher levels of detail - image fusion can be used to produce a colour image with maximum detail. Visible and near-infrared (RGB-NIR) colour image fusion involves fusing a colour image taken in the visible spectrum with a greyscale near-infrared image,

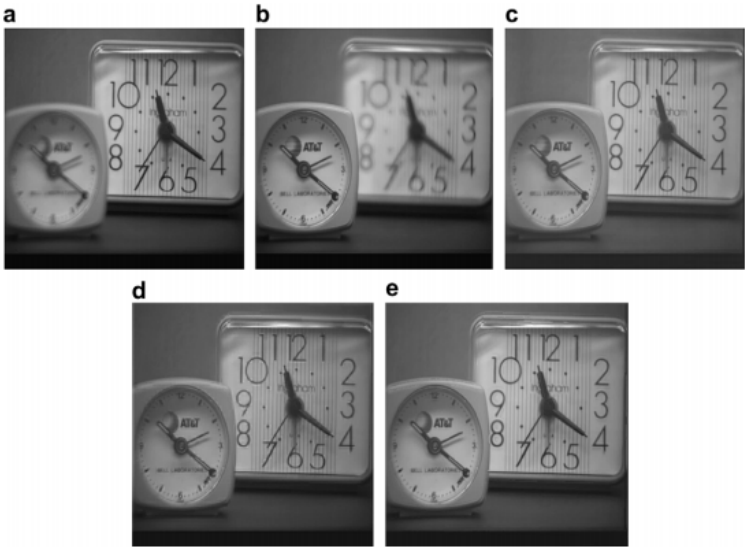


Figure 1.2: Introduction: multifocus image fusion (“clocks”) - (a) right-side focus (b) left-side focus (c), (d) and (e) various fusion results (images and results from [52])

to produce a colour output image with combined details. Near-infrared light can penetrate haze, allowing the RGB-NIR image fusion to be used for dehazing [79], and skin, leading to the application of skin smoothing [33]. New results for RGB-NIR fusion, as well as metric and psychophysical assessments, are shown in sections 4.3, 4.5 and 6.1 of this thesis. Multispectral image fusion has also been used to analyze medieval manuscripts [48]. Multi-exposure image fusion is also an image fusion application taking as input colour images with different exposure levels, and producing an output colour image which is well-exposed and detailed in all parts of the image [62] - new results for this task are shown in section 6.4 of this thesis.

Standard image fusion techniques include the ratio of low-pass pyramid [87] (ROLP) technique, which represents detail as the ratio between levels of a scale-space pyramid, and discrete wavelet transform (DWT) fusion, which encodes image detail as wavelet coefficients at different scales [66]. A common way of fusing using these methods is to take

the maximum detail coefficients from each input image channel, before undoing the multi-scale decomposition to produce an output image. These common methods, as well as other image fusion methods not based on image derivatives, are surveyed in Chapter 2.

First order image derivatives - the change in pixel intensity between one pixel and the next, measured in the x and y directions - are a natural and intuitive image detail representation, and have been strongly linked to image saliency in human perception [88]. However, it is not immediately obvious how to combine input image gradients into a single output gradient, and how to go from these output gradients to an output image. To solve the first part of this problem, Di Zenzo proposed the structure tensor, a way of combining gradients from any number of input image channels [22]. The structure tensor is defined as the inner product of the Jacobian matrix of derivatives across all image channels in the x and y directions. The eigendecomposition of this matrix produces an eigenvector pointing along the line of maximal gradient variation, and the square root of the corresponding eigenvalue represents the magnitude of that gradient - this produces a gradient which best captures the detail of all the input gradients. Unfortunately, this square root is ambiguous in its sign - the gradient could point either direction along the line defined by the eigenvector. Socolinsky and Wolff chose to assign a sign to the gradient by looking at the direction of the gradient in the mean input channel, or luminance channel [83], but no definitive method has been found.

To go from the gradients produced by the structure tensor to an output fused image, the standard method is gradient field reintegration. This involves solving a Poisson equation, which can be accomplished through iteration or Fourier deconvolution. Unfortunately, some gradient fields are inherently non-integrable, meaning there is no possible image with their exact gradients - in these cases, the reintegration finds a least-squares optimal output

image, which may produce halo or bending artefacts (see section 3.3 for examples of these artefacts). There has been much research in trying to solve this problem [5] [63], but a definitive solution has not been found. Chapter 3 of this thesis presents previous derivative domain image fusion methods (and their problems) in more detail.

Spectral Edge image fusion is a derivative-based image fusion method which avoids reintegration artefacts [16]. It does this using lookup-table-based gradient field reconstruction, a method which find a global lookup-table between the N input image channels and a 3-dimensional output image, which produces an artefact-free output image which best matches the target gradients of the structure tensor in a least-squares optimal way [31]. The other key idea of the Spectral Edge method is that it maintains the colour of a given putative RGB image, by producing target RGB colour gradients, which have a structure tensor at every pixel which is simultaneously exactly equal to the input N-dimensional input image channels, and as close the putative RGB image as possible, combining input detail and colour.

Chapter 4 of this thesis explains the Spectral Edge image fusion method in detail, and presents an iterative extension to the method. It compares the Spectral Edge image fusion method with the methods of Fredembach and Süsstrunk [34] and Schaul *et al.* [79], for the application of RGB-NIR image fusion. The goal of this image fusion method is photographic enhancement, so the best way to compare the methods is a forced choice pairwise psychophysical preference experiment - in this experiment, the Spectral Edge method is the most preferred, followed by the method of Schaul et al. Attempts are also made to produce an objective image fusion metric which will produce similar results to the experiment. Previous metrics have been for 2 to 1 channel image fusion, so they are extended to M to N channel fusion (in this case 4 to 3 channels), and their quality assessments are

compared to the psychophysical experiment. Colourfulness and contrast-based metrics are also considered. In the end, a combination of metrics based on gradient and colourfulness gives the result closest to the psychophysical experiment. New applications of Spectral Edge image fusion are also presented, using a combined visible and near-infrared sensor to create inputs for RGB-NIR image fusion, and a thermal smartphone camera to create inputs for both natural and false colour RGB-thermal image fusion.

In chapter 5, a new image fusion model is proposed, based on a local linear combination (LLC) of the input images, which avoids problems and artifacts inherent in previous gradient reintegration techniques, by transforming the task of gradient reintegration into a simple local linear combination of the input image channels. Two methods of calculating local linear combination coefficients are explained: the first is based on the **P**rincipal characteristic vector of the **O**uter **P**roduct (**POP**) of the Jacobian matrix of derivatives, which produces a mapping that creates an output image with derivatives equal to those calculated by Socolinsky and Wolff, and the second is based on finding a mapping by fitting a least-squares regression (with regularization) from the input image derivatives to the Spectral Edge calculated colour derivatives.

Various applications of the new image fusion model are explored in chapter 6, and the results presented and compared with existing image fusion methods. These applications are RGB-NIR image fusion, RGB-thermal image fusion, colour to greyscale conversion, flash and no flash image enhancement, multifocus image fusion, multi-exposure image fusion, and image fusion for astronomical visualization. The sheer variety of applicable image fusion tasks for our method is unusual - most image fusion methods are designed to perform well at one or two specific tasks.

There are several potential areas of future work. Firstly, it would be useful to create an

RGB-NIR image set, with high-quality, well-registered images of a variety of scenes and types - these could be obtained from dual RGB and NIR sensors (requiring registration), or a single RGB-NIR sensor. The EPFL RGB-NIR data set [8] has been very useful, but has problems with registration, and not many of its images are photographically pleasing. Secondly, more work could be done to find a definitive objective metric to measure the performance of image fusion tasks. Finally, the local linear combination image fusion methods presented here could be further refined, and other coefficient diffusion techniques used. The mathematics behind the method could also be explored in more depth.

Chapter 2

An Overview of Image Fusion

This chapter presents a short survey of the field of image fusion, excluding derivative domain image fusion methods, which are described in the next chapter. It then goes on to cover image fusion quality assessment, through psychophysical experiments and objective metrics.

2.1 Definition

Standard definitions of image fusion include: "the objective of image fusion is to combine information from multiple images of the same scene. The result of image fusion is a new image which is more suitable for either human perception or machine perception and further image-processing tasks such as segmentation, feature extraction and object recognition," [66], and "the goal in pixel level image fusion is to combine and preserve in a single output image all the important visual information that is present in a number of input images." [93]. The input images may be, for example, the bands of a multispectral or hyperspectral image[64], medical images captured with different scanning modalities [16], multifocus images with different points of focus [52], or images with different illumination

[25]. Blum and Zheng define image fusion as fusing images from different sensors (multi-sensor image fusion) [7]. If we define image fusion more widely, as any task involving image dimensionality reduction, we can include colour to greyscale conversion (although it is not commonly considered an image fusion problem in the academic literature) [37]. In all these cases, one image, traditionally a greyscale image, but now also commonly a colour image, must be produced as an output. This output image should in some way capture the detail and structure of the input images in a way which summarizes them most fully - as defined by some measure of image detail or saliency.

2.2 Advantages

In most cases, different input images (from different sensors or times) will have information which is shared and repeated across the images. This redundant information is unnecessary, and image fusion can produce a single image with less redundancy. In other cases, the input images contain different salient information, This complementary information can (hopefully) be fused to produce an image with all (or as many as possible) of the salient details from each input image [7]. However, image fusion will always result in some loss of complementary information - dimensionality reduction is a problem with no perfect solutions - the task is to minimize this loss.

Image fusion can improve the efficiency of human operators - if human operators are monitoring image streams, their workload increases with the number of images that must be monitored. Image fusion can reduce this workload, by combining images and thereby reducing the number of images necessary [7]. The RGB-thermal image fusion results shown in sections 4.5 and 6.2 of this thesis could be used by security camera operators, for example, to reduce the number of image streams it is necessary to monitor.

Subjective image quality can be improved by image fusion. Near-infrared (NIR) images can be used to enhance visible spectrum colour images to produce a superior fused colour image, as shown in sections 4.3-5 and 6.1 of this thesis. Multi-exposure image fusion produces an output image with a greater subjective image quality than any of the single input images, as shown in section 6.4 of this thesis. Flash and no-flash image enhancement, shown in section 6.7, produces an output image of higher subjective quality than either the flash or no-flash images alone. Image fusion can also be used on astronomical images, to produce a more pleasing image for astronomical visualization.

2.3 Techniques

In this section we briefly cover the background of widely-used image fusion techniques, such as pyramidal approaches, methods based on the discrete wavelet transform, optimization-based methods, and sparse representations - several of these methods are compared to our proposed methods in later chapters. Neural networks are also beginning to be used for image fusion, but are not covered here.

2.3.1 Pyramidal Approaches

The Gaussian scale space is the most basic and fundamental way of representing information in an image at different scales[55]. To see information at different scales within an image, the image is convolved with a Gaussian filter kernel with different standard deviations[68]:

$$I(x, y, t) = I_0(x, y) * G(x, y, t). \tag{2.1}$$

Where $I_0(x, y)$ is the original image at a certain x and y pixel location, and $G(x, y, t)$

represents the Gaussian kernel with scale space parameter t, corresponding to its variance.

The ratio of low-pass pyramid (ROLP)[87], and the difference of low-pass pyramid (DOLP)[10] create a pyramid of input image channels at different scales:

$$I(x, y, k) = I_0(x, y) * G(x, y, t_k). \tag{2.2}$$

Where t_k is the appropriate scale parameter for level k. The standard deviation is doubled at each level, and as high frequency information has been removed, the image can be downsampled by half.

The ratio components of the ROLP pyramid are defined as a ratio of two levels of the pyramid:

$$R(x, y, k) = I(x, y, k)/I(x, y, k + 1). \tag{2.3}$$

This ratio is equal to the Weber contrast plus one (contrast is centred around zero because where the background and foreground are equal there is zero contrast, whereas ratios are centred around one):

$$C(x, y, k) = I(x, y, k)/I(x, y, k + 1) - 1. \tag{2.4}$$

The difference components of the DOLP pyramid are defined as a difference of two levels of the pyramid:

$$D(x, y, k) = I(x, y, k) - I(x, y, k + 1). \tag{2.5}$$

The original image can be reconstructed from the top level of the pyramid by multiplying (for ROLP) by or adding (for DOLP) it to the ratio or difference coefficients per pixel

at each level.

The fusion step comes from choosing ratio or difference coefficients across several input channels using some selection rule - typically picking the coefficients with the largest magnitude (max selection)[11]. One is taken away from the ratio coefficients before the selection takes place, as the largest contrast will be selected. For ROLP:

$$R_{\text{OUT}}(x, y, k) = R_{\lambda}(x, y, k) \tag{2.6}$$

where

$$\lambda = \underset{\{\forall n \in \mathbb{N}: 1 \leq n \leq N\}}{\arg\max} \left(|R_n(x, y, k) - 1| \right), \tag{2.7}$$

And for DOLP:

$$D_{\text{OUT}}(x, y, k) = D_{\gamma}(x, y, k) \tag{2.8}$$

Where

$$\gamma = \underset{\{\forall n \in \mathbb{N}: 1 \leq n \leq N\}}{\arg\max} \left(|D_n(x, y, k)| \right), \tag{2.9}$$

Where R_n and D_n are the ratio and difference coefficients for the nth input image channel, out of N total input image channels. The top level images are typically averaged:

$$I_{\text{OUT}}(x, y, K) = \frac{\sum_{n=1}^{N} I_n(x, y, K)}{N}, \tag{2.10}$$

Where K is the number of levels in the pyramid, and then the output image is produced using the new ratio or difference coefficients at each level:

$$I_{ROLP}(x,y) = I_{\text{OUT}}(x,y,K)\prod_{k=1}^{K} R_{\text{OUT}}(x,y,k) \tag{2.11}$$

$$I_{DOLP}(x,y) = I_{\text{OUT}}(x,y,K)\sum_{k=1}^{K} D_{\text{OUT}}(x,y,k) \tag{2.12}$$

This image should have the maximum detail across all the input images at each scale.

Another variation of image fusion using a pyramid decomposition is based on mean filtering instead of Gaussian filtering. The detail coefficient combination method is also more complex [50]: first a Laplacian filter is convolved with the input images to produce a saliency map, then weight maps are constructed by setting the weight for the input image with maximum saliency at a particular pixel to 1 and the others to 0. These weight maps (coefficient images) are then diffused and smoothed by using a guided filter, with the input corresponding source image used as the guide image.

2.3.2 Discrete Wavelet Transform

The discrete wavelet transform (DWT) is another multiscale representation of an image [59]. Wavelets are waves (or functions) with compact support, i.e. waves only defined for a certain domain of input values. Wavelets are comprised of scaling and wavelet functions - the scaling function changes the scale of the input signal (often similar to a low-pass filter), then the wavelet function describes the input signal at that scale (somewhat analogous to a high-pass filter). The scaling and wavelet functions are also known as approximation and detail functions. These wavelets are used to calculate coefficients at different scales, of which the approximation coefficients represent low-frequency information, and the detail coefficients represent high-frequency information.

The Haar wavelet is defined as a step function:

$$\Psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2} \\ -1 & \text{if} \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

This, in practical terms, translates to a $\left(1/\sqrt{(2)}\right)[1-1]$ image filter, which must be convolved with the input image signal at the appropriate scale to obtain high-frequency information. The corresponding low-frequency wavelet filter is $\left(1/\sqrt{(2)}\right)[1\ 1]$. These filters must be applied in both x and y directions - first x, then y, in all sequential combinations of the high- and low-frequency wavelet filters. After each application of the filter, the image is downsampled by a factor of 2 in the corresponding direction. This process begins at the largest (raw input) image scale, then the image obtained from convolving low-frequency filters in both directions and downsampling is used as the input for the next scale. Figure 2.1 shows a diagram illustrating one level of this process, where H represents high-frequency wavelet filtering, and L low-frequency filtering.

The inverse process is used to reconstruct an image from wavelet coefficients. The coefficient images are first upsampled by a factor of 2 in the y direction, then inverse L or H filters are applied (depending on which filters were used to create the current coefficient image). The same is done is the x direction, and then the coefficients are summed to produce the output image. This must be done at every scale before the final image is reconstructed. Figure 2.2 shows an illustration of one level of this process.

Image fusion is performed by using a selection rule to combine the coefficients from multiple input image channels[66]. A simple selection rule might be to take the mean of the approximation coefficients at the largest scale, and the detail coefficients at every scale

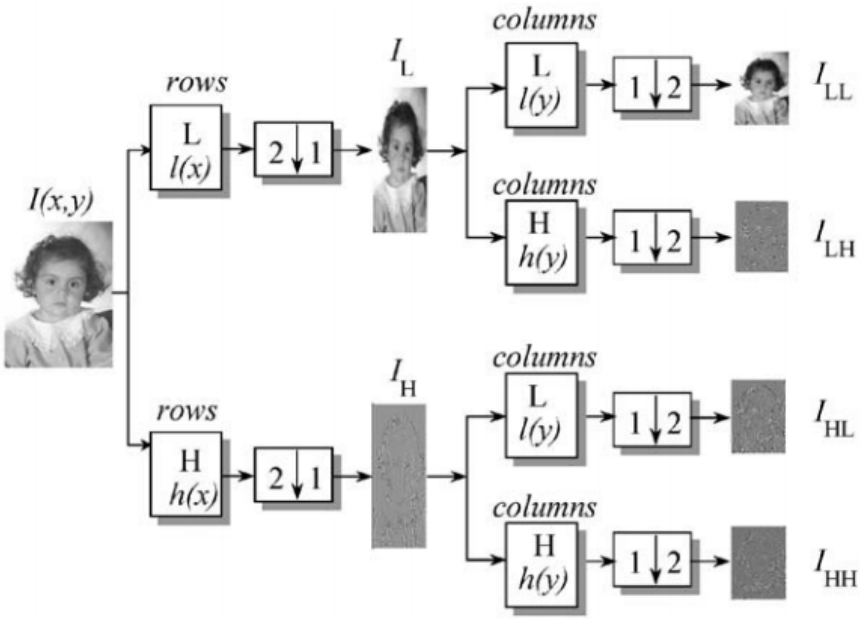


Figure 2.1: Overview of image fusion: DWT Image Fusion - process of obtaining wavelet coefficients (from [66]).

with the greatest magnitude, this is known as the Choose Max (CM) selection rule. More complex selection rules are also used, using activity measures (such as variance) either in a local window in the image, or in the set of coefficients.

Further steps can be added, such as consistency verification, in which nearby image regions are assumed to contain the most salient image details from the same input image. This can be accomplished by, for example, smoothing the decision map of which input image to take coefficients from using a majority filter. Consistency can also be extended between scales, enforcing coefficients from a similar image area at different scales to use input information from the same image.

Once the output coefficients are combined from the selection of input coefficients, the output image is then formed by inverting the wavelet decomposition from the fused coefficients.

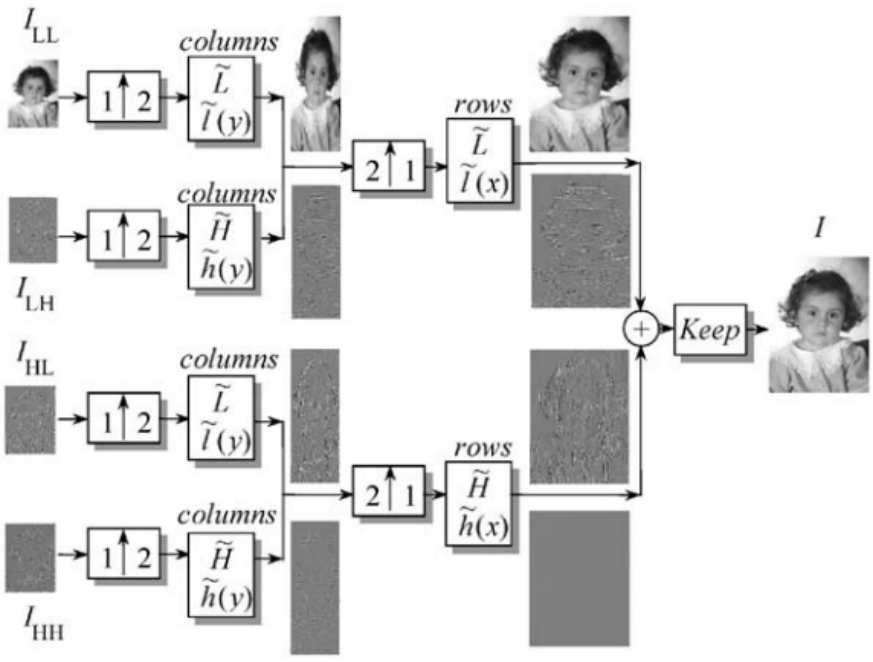


Figure 2.2: Overview of image fusion: DWT Image Fusion - process of reconstructing an image from wavelet coefficients (from [66]).

Image fusion has also been implemented with complex wavelets [39], and curvelets [64], which reduce the artefacts inherent in wavelet image fusion.

2.3.3 Optimization-based Methods

Many recent methods in image fusion are based on the optimization of one or more key variables to fit an objective function.

One method based on optimization is that of Laplacian colourmaps [25]. In this method, they define an image Laplacian, creating a sparse matrix representation of the image structure. The commutativity of two images' Laplacians is proposed as a measure of image similarity. This is then used as the primary target of minimisation, to find a global M-to-N channel mapping - the Laplacian of the output of the mapping should be as similar to the

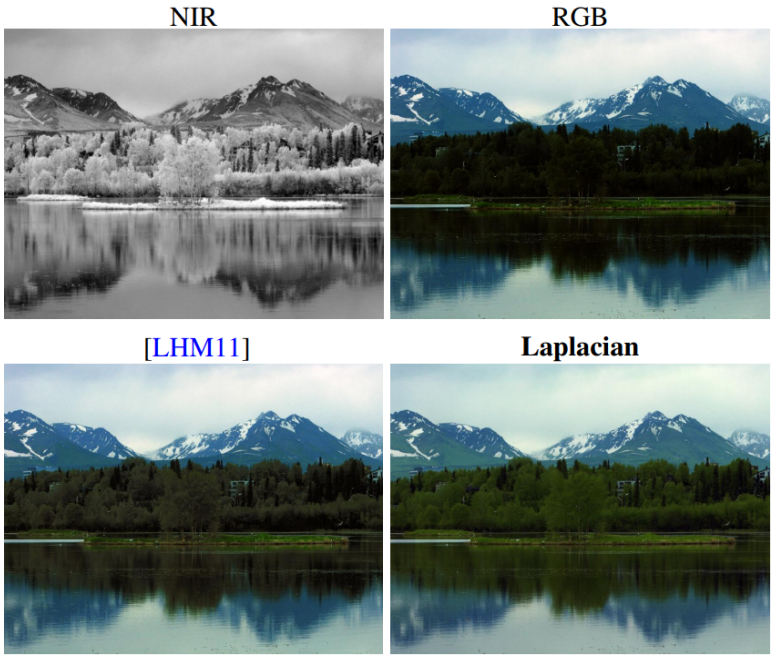


Figure 2.3: Overview of image fusion: Laplacian colourmaps image fusion - RGB-NIR (from [25] - [LHM11] is the method of [47]).

input image's Laplacian as possible, as judged by their commutativity. Figure 2.3 shows the Laplacian colourmap result for RGB-NIR colour image fusion of the NIR and RGB images,

Lau et al. use a clustering approach to define information loss between the original input images and the output fused image [47]. They create a graph that whose nodes represent clusters of similar pixels, and whose edges represent the difference between spatially similar clusters. They then formulate a minimization which seeks to create a mapping which creates an output image with as similar a graph as possible to that of the input images, which also being as similar as possible to an initial mapping.

$$\mathbf{x} = \arg\min_{x} E_{\Upsilon} + wE_{M} \tag{2.14}$$

Where x are the optimal output cluster colours, E_Υ is the term that preserves contrast, and E_M is a constraint to keep the mapping close to the original image mapping, and w is a parameter weight.

The target term is

$$E_\Upsilon = \sum_{(i,j)\in\varepsilon} \Upsilon_{ij}((\mathbf{x}_j - \mathbf{x}_i) - \mathbf{t}_{ij})^2, \tag{2.15}$$

Where $t_{i,j}$ are the target vectors for all edges (i, j) in the graph edge structure ε, and Υ_{ij} is a weight on edge (i, j).

An application-specific optimization method is that of Feng et al. [30] uses a generalized haze model, defined by Fattal [27] as:

$$\mathbf{I}(x) = \mathbf{t}(x)\mathbf{J}(x) + (1 - \mathbf{t}(x))\mathbf{A}, \tag{2.16}$$

where I is the image captured by the camera, J is the dehazed scene image, t is the transmission map, and A is the airlight colour. It then utilises the NIR image to help determine the amount of haze at each location in the RGB image. They then formulate dehazing as an optimization problem, simultaneously finding the original colour image and the transmission map which best fit the RGB and NIR images. They solve this using iteratively reweighted least squares (IRLS) - the optimization problem is defined as

$$(\hat{\mathbf{J}}, \hat{\mathbf{t}}) = \underset{(\mathbf{J},\mathbf{t})}{\arg\min} ||\mathbf{t}\mathbf{J}+(1-\mathbf{t})\mathbf{A}-\mathbf{I}^{RGB}||^2+\lambda_1 w|\nabla\mathbf{J}-\nabla\mathbf{I}^{NIR}|^\alpha+\lambda_2|\nabla\mathbf{J}|^\beta+\lambda_3||\nabla\mathbf{t}||^2 \tag{2.17}$$

The first part of the minimization is based on the single image dehazing model. The second term uses the NIR detail as a constraint for the calculated original dehazing image,

as the NIR image should be more detailed in hazy areas. The last two terms are smoothness constraints for the dehazed image and the transmission map. The NIR detail is weighted by w to be more significant at lower values of t - i.e. where the objects are further away.

Optimization-based methods can often produce excellent output results with minimal artefacts, but in our experiments they often take a large amount of computational time. This is due to the nature of optimization, which demands computational complexity, in contrast to a closed-form solution.

2.3.4 Sparse Representations

A sparse representation of an image describes the image in terms of sparse coefficients of an overcomplete dictionary of prototype signal atoms. Overcompleteness means that the number of basis atoms in the dictionary exceeds the number of pixels in the image (number of image dimensions). Sparseness means that number of descriptors required to describe the image is less than the number of pixels[94].

If we describe a signal as $\gamma \subset R^n$, sparse representation theory proposes a dictionary $D \in R^{n \times T}$, where T is the number of prototype signals, referred to as atoms. For a given signal $x \in \gamma$, there is a linear combination of atoms from D that approximates it closely. Formally, $\forall x \in \gamma, \exists s \in R^T$ such that $x \approx Ds$. Usually $T > n$, meaning the dictionary is overcomplete. Finding s, which is usually not unique, involves solving this optimization problem:

$$\min_s ||s||_0 \text{ subject to } ||Ds - x|| < \varepsilon \tag{2.18}$$

Where $||s||_0$ is the number of nonzero components in s. This optimization is NP-hard, so an approximate solution is found. The typical algorithms used to select the dictionary

are matching pursuit (MP) or orthogonal MP.

The input images are first decomposed into patches, from which dictionaries of atoms are calculated. Next, the patches are decomposed by prototype signal atoms into their corresponding coefficients. A larger coefficient means more salient features are present. The coefficients are combined between the input images by the choose-max (CM) selection rule. Finally, the output fused image is created by applying the fused coefficients to the dictionary.

Image fusion using sparse representations has been used for applications including multifocus image fusion [94], remote sensing [53], and hyperspectral image fusion [91].

2.4 Applications

The image fusion applications in this section are all widely explored in the academic literature, and typically image fusion algorithms are designed to solve a specific one of these problems. In Chapter 6 of this thesis, we show how the proposed LLC method can be used for a wide variety of these applications.

2.4.1 Multifocus Image Fusion

In digital photography, when a lens is focused on a certain object in a scene (using a certain focal length), at a certain distance from the camera, parts of the scene that are at different distances can become out of focus.

Multifocus image fusion involves fusing images taken with different focal lengths. Figure 1.2 shows a classic example, in which two greyscale images are each in focus in around half the scene. The fused result is in focus at every point [52]. Many of the standard image fusion methods have been used for this task, such as the discrete wavelet transform [49]

and complex wavelets [39], as well as other methods such as neural networks [51].

Plenoptic photography involves using a microlens array to measure not only the total amount of light arriving at each pixel, but also the amount from each light-ray direction. This provides various refocusing options of color images, allowing images with different depths of focus to be created from a single exposure [65]. These can be fused to produce a color image in focus at every point, for example using Laplacian colourmaps [25], as in fig. 6.7.

2.4.2 RGB-NIR Colour Image Fusion

The goal of RGB-NIR colour image fusion is to combine the detail of the input colour RGB and greyscale NIR images, while maintaining or even improving the colour of the original image. This can be done for purposes such as dehazing or photographic enhancement.

Fredembach et al. introduced the idea of decomposing the input RGB image into its luminance and chrominance, in a colour space such as HSV or CIELUV, and then replacing the luminance channel with the input NIR image, before converting the image back to RGB colour space [34]. This method provides high levels of detail transfer, but gives very unnatural and strange looking results. A logical next step is to take an average of the RGB luminance channel and the NIR input image, and use this as a luminance channel replacement. If we define V as the RGB luminance channel, and N as the NIR image, the new luminance image is $(V + N)/2$.

Schaul et al. developed this idea of combining the RGB luminance channel and NIR image, by fusing them using an undecimated pyramidal image fusion technique. Pyramidal image fusion techniques, of which Toet's ratio of low-pass pyramid (ROLP) is the standard example [87], obtain a pyramid of approximation images by blurring at different scales (see

section 2.1.1). In the case of Schaul et al. this filter is an edge preserving filter, as proposed by Farbman et al. [26].

Their fusion works by taking the maximum detail coefficient at each level and at each pixel, across the two input images (the RGB luminance and NIR), and then undoing the pyramid transform, to produce the output luminance image:

$$F_0 = V_n^a \prod_{k=1}^{n} (\max(V_k^d, N_k^d) + 1) \tag{2.19}$$

where V_k^d is the detail image of the RGB luminance channel, and N_k^d is the detail image of the NIR image, both at scale k. Only the approximation coefficients from the RGB luminance channel are used, in order to maintain natural low-frequency image details. This luminance channel is then used as a replacement luminance channel for the RGB image to produce an output image.

RGB-NIR image fusion methods which do not use a replacement luminance channel include the Spectral Edge image fusion method (see chapter 4 of this thesis), and optimization-based methods such as Laplacian colourmaps [25] and cluster-based colour space optimizations [47].

Skin smoothing is another use of RGB-NIR image fusion. Near-infrared light penetrates further into skin than visible light, so the skin appears smoother, with fewer blemishes. Fredembach *et al.* use an image fusion technique based on using a bilateral filter to separate images into their base and detail layers, and combining the visible base layer with the NIR detail layer [33]. Figure 2.4 shows an example of the method's input and output, along with two other possible methods, using luminance transfer (NIR image used as replacement luminance channel) and DWT fusion of the two luminance channels.

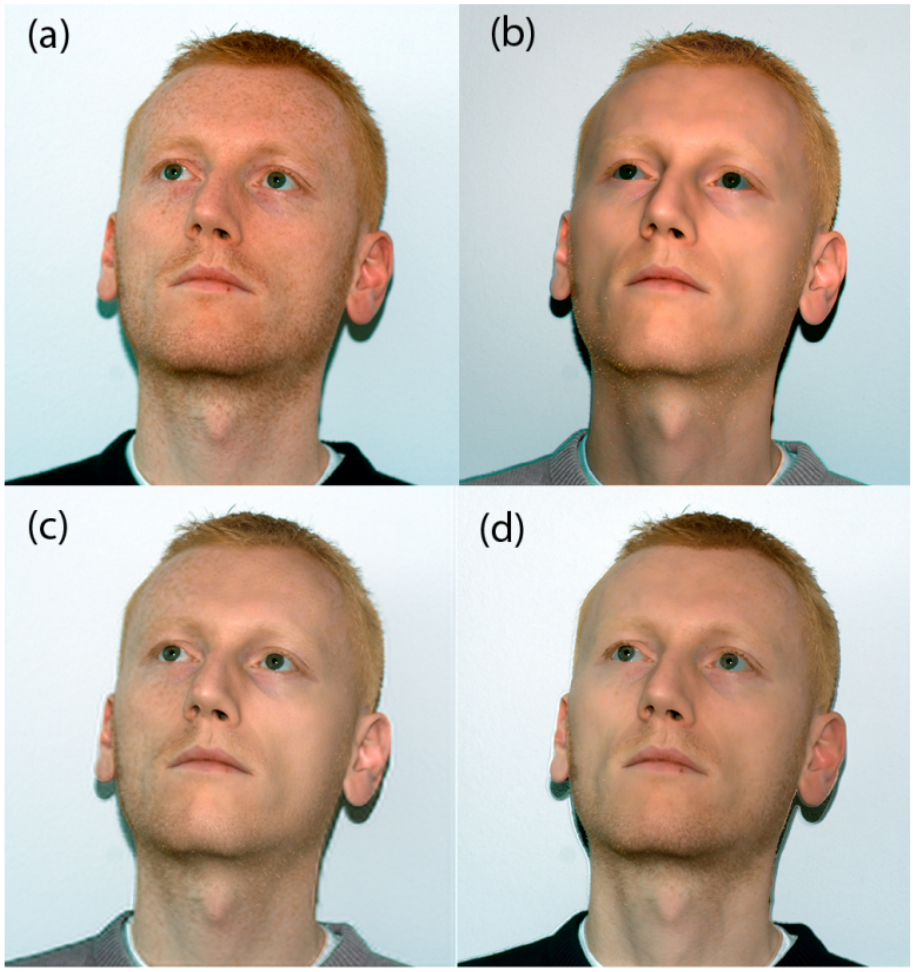


Figure 2.4: Overview of image fusion: RGB-NIR Image Fusion for skin smoothing: a) visible RGB, b) luminance transfer, c) DWT fusion, d) method of [33] (from [33]).

2.4.3 Multi-exposure Image Fusion

Multi-exposure fusion (MEF) fusion is a simple and practical alternative to high dynamic range (HDR) imaging. HDR in typical consumer cameras (e.g. smartphone cameras) usually involves taking multiple exposure images, converting these to an HDR radiance image, then converting this HDR image to a low dynamic range (LDR) image. MEF fusion avoids

the step of creating an HDR image by going directly from a set of input images with different exposures to an output fused image. This method assumes all input images are perfectly registered, and is widely used in consumer photography [77].

A comparison of MEF algorithms [58] poses MEF fusion as a weighted average problem:

$$O(\mathbf{x}) = \sum_{n=1}^{N} W_n(\mathbf{x}) I_n(\mathbf{x}) \tag{2.20}$$

Where O is the fused image, N is the number of multi-exposure input images, $I_n(\mathbf{x})$ is the luminance (or other coefficient value) and $W_n(\mathbf{x})$ the weight at the $\mathbf{x}$-th pixel in the n-th exposure image. The weight factor $W_n(\mathbf{x})$ may be spatially varying or global.

In a subjective comparison, based on assessing 8 MEF algorithms by their mean opinion score (MOS), rated from 1 to 10 [58], the best performing algorithm was that of Mertens *et al.*[62] (as explained in section 2.5 of this thesis, paired comparisons are a superior method of subjective quality comparison, so this result is questionable). This is based on a multiscale Laplacian pyramid decomposition of the input images, with the coefficients from each image weighted by a combination of contrast, saturation and well-exposedness, and then reintegrated to produce a fused image. The next best performing was the method of Li *et al.*, which takes the results of Mertens' method and applies extra detail enhancement. Closely behind the top two methods is an image fusion method based on guided filtering [50], in which approximation and detail coefficients are calculated using an averaging filter, then a weight map to combine these coefficients is calculated from taking the difference between saliency measures at each pixel between the different input images, then using a guided filter (with the input images using as the guide images) to smooth the resulting

weight maps. Another method for MEF fusion is that of Shen *et al.*, who perform multi-exposure image fusion using generalized random walks [80].

2.4.4 Colour to Greyscale Conversion

Colour to greyscale conversion is the process of converting a color RGB image to a summary greyscale, which should represent all of the intensity and colour details of the original as closely as possible. This is a dimensionality reduction problem, from three dimensions to one, and therefore it is impossible to preserve all the input information. Therefore, the most important information must be selected in some way, and transferred without artefacts - a difficult task.

There have been various methods proposed to accomplish this goal. The "color2gray" method of Gooch *et al.* converts the colour input image into a luminance-chrominance colour space, then sets up an optimization problem, to find an output greyscale image which captures the most significant luminance and chrominance differences. This is then solved using conjugate gradient iterations [36].

Another colour to greyscale conversion method is that of Rasche *et al.* [74]. This method sets up an optimization problem, in which colour contrasts and luminance consistency are the two components in the objective function. It then uses a constrained, multivariate optimization to solve this function and produce a greyscale output image. Figure 2.5 shows some examples of their method's results, compared to the original luminance channels. Colour contrasts are maintained much more in their output images.

Grundland and Dodgson proposed the "decolorize" method for colour to greyscale conversion [37]. They claim this to be superior to previous methods, due to fulfilling their design objectives, which include global consistency, greyscale preservation, and luminance

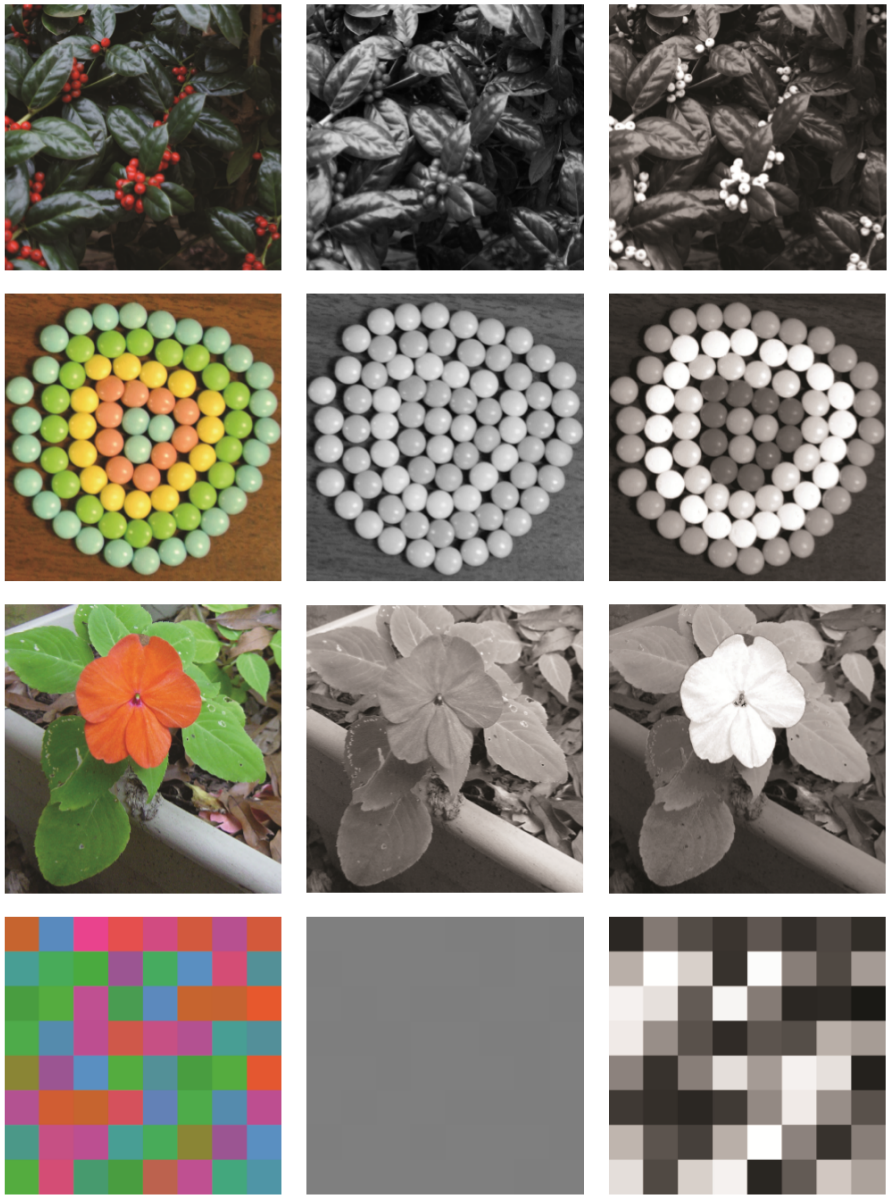


Figure 2.5: Overview of image fusion: colour to greyscale conversion: a) original colour image, b) luminance channel, c) result of [74].

ordering. Their method is based on a unique dimensionality reduction technique, which aims to preserve a maximal amount of chrominance contrast in the output greyscale image.

These methods and others were presented and compared in the work of Eynard *et al.* [25], and their proposed method was found to be the best performing, based on the root weighted mean square RWMS metric [45] (described in section 2.5 of this thesis) and a psychophysical experiment.

2.4.5 Other Applications

Medical image fusion involves registering and fusing several images from a single or multiple medical imaging modalities, to produce a single image with the combined salient details. A review of the field shows that all the common image fusion methods are used for medical image fusion [41]. Magnetic resonance imaging (MRI) can be processed using image fusion, particularly for 3D conformal radiation therapy and prostate studies. Computerized tomography (CT), positron emission tomography (PET), single photon emission computed tomography (SPECT) and ultrasound are other common medical image modalities which can be fused. The brain is the most common organ which benefits from medical image fusion, for such purposes as segmentation of brain tissues, image guided neurosurgery, and decoding brain visual states.

Concealed object or weapon detection is an application of image fusion, in which a visible image is fused with an infrared (IR) image. In the visible image, realistic details are present, while in the infrared image the concealed object or weapon is apparent. Xue and Blum detect concealed weapons by fusing the visible RGB luminance channel with the greyscale IR image, as well as its negative (in case the concealed weapon is more visible in the negative image) [92]. In the fused results, concealed weapons are visible in

(a) Visual image (b) IR image (c) Fused image

Figure 2.6: Overview of image fusion: concealed weapon detection by image fusion - from [92]

a naturalistic colour image, as shown in figure 2.6 - the advantage of this is that one video feed can simultaneously provide both identification and concealed object detection.

An artistic application of image fusion is presented by Raskar and Yu [75], in which gradient domain image fusion is used to stitch together different scenes. The input gradients are linearly combined based on image saliency into desired output gradients, and then reintegrated using an alternative to Poisson gradient reintegration. Fig. 3.8 shows an example result, in which the foreground figure is added to a different background scene.

2.5 Image Fusion Quality Assessment

2.5.1 Psychophysical Experiments

Psychophysical experiments are widely used in image fusion and image processing, as a means of identifying the best method from a choice of several. Single stimulus categorical rating involves asking an observer to rate a single image on a scale of 1 to 5, from excellent to bad. Double stimulus categorical rating is similar, but two images are shown and then rated at a time, a reference image and a test image. Pairwise similarity judgements involve rating the relative quality of two images. The observer can decide that the images are

equal in quality, or that one is of higher quality by any amount on a continuous scale. The problem with these methods is that different observers may have different internal scales by which they measure quality.

Forced-choice pairwise comparison involves a pair of images being displayed to the observer, who picks the image with the highest perceived quality. The forced-choice element is that they must pick one of the images, even if they cannot discern any difference in quality. The theory behind this is Thurstone's law of comparative judgment, which assumes that the quality of each image is a random variable with Gaussian distribution (different observers may rate the quality of a single image differently) [85]. If we denote two arbitrary images as having quality distributions Q_A and Q_B, then we seek to find the difference between their means, which is an approximation to the 'true' quality difference between them. Given many pairwise comparisons, each of which is a comparative observation of which image is of greater quality, we can estimate the quality difference $Q_A - Q_B$.

The general case of the model requires the standard deviation of these Gaussian quality distributions, and the correlation between them, to be estimated. Case 5 of Thurstone's Law is often used to simplify the model, in which the discriminal dispersions are specified to be uniform and uncorrelated.

To calculate the quality difference $Q_A - Q_B$ from a set of comparisons of multiple images, we use the equation

$$Q_A - Q_B = H^{-1}(P(Q_A > Q_B)). \tag{2.21}$$

where H is the normal cumulative distribution function. To find $P(Q_A > Q_B)$, the probability that image A is of higher quality than B, we first create a frequency matrix. In this matrix, each element ij is a tally of the number of times image i is preferred over

image j. We then normalise this by dividing by the total number of observers to produce (in the case of comparing the perceived quality of three image categories)

$$P = \begin{bmatrix} P(Q_1 > Q_1) & P(Q_1 > Q_2) & P(Q_1 > Q_3) \\ P(Q_2 > Q_1) & P(Q_2 > Q_2) & P(Q_2 > Q_3) \\ P(Q_3 > Q_3) & P(Q_3 > Q_2) & P(Q_3 > Q_3) \end{bmatrix} \tag{2.22}$$

Our quality matrix is then created by applying H^{-1} to P, and the final estimated quality differences are calculated by summing the rows of the quality matrix.

The study of Mantiuk et al. [61] found that the forced-choice pairwise comparison produced the most accurate and efficient results, compared to three other types of psychophysical experiment (single stimulus, double stimulus and similarity judgments) - therefore we use this method in the experiments in this thesis. Connah et al. provide a practical example of a forced-choice pairwise comparison psychophysical experiment, in which they compare several methods of colour to greyscale conversion [17].

2.5.2 Objective Metrics

Image fusion quality metrics can be divided into two main types - metrics with a reference image and non-reference metrics. Both types are based on measures of image similarity, but they are used in different ways depending on whether a reference image is available.

In some cases, a reference image is available - for example in multifocus image fusion, an image with everything in focus may have been taken of the same scene as the input images.

With a reference image available, the metric is simply a measure of similarity between the output fused image and the reference image. This measure could be root mean squared

error (RMS) [89], signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR), mutual information (MI), the structural similarity image measure (SSIM) [90], or another measure.

We examine non-reference metrics here. These are common metrics used for image fusion tasks, as no reference image is typically available for tasks such as RGB-NIR image fusion. They typically consist of some measure of the image similarity between each input image and the fused image, often weighted by a measure of image salience. These metrics make the assumption that an ideal fused image will transfer maximum detail from all input images - as shown in later chapters of this thesis, this is not the case for all image fusion applications.

Xydeas and Petrovic defined a metric based on gradient similarity [93]. This metric measures the amount of gradient detail transferred from each of the input images to the output image. Its first part is a gradient similarity measure

$$Q^{AF}(x,y) = Q_g^{AF}(x,y)Q_\alpha^{AF}(x,y) \tag{2.23}$$

which measures the gradient similarity between images A and F, at pixel location with a particular x and y coordinate. The first part, $Q_g^{AF}(n,m)$, measures the similarity in gradient magnitude, and the second part, Q_α^{AF}, measures the similarity in gradient angle.

These gradient similarity measures are added up across the image plane (with dimensions X and Y), and calculated between both input images and the fused image. The results are weighted by the gradient magnitude at that pixel, $|G(x,y)|$.

$$Q_G = \frac{\sum_{x=1}^{X}\sum_{y=1}^{Y} Q^{AF}(x,y)|G^A(x,y)| + Q^{BF}(x,y)|G^B(x,y)|}{\sum_{x=1}^{X}\sum_{y=1}^{Y}(|G^A(x,y)| + |G^B(x,y)|)} \tag{2.24}$$

Where A and B are the two input images, and F is the fused image.

Several measures have been proposed based on the structural similarity image measure (SSIM), defined by Wang et al.[90] as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2))}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2.25}$$

Where μ_x is the mean of image x, σ_x^2 is the variance of image x, σ_{xy} is the correlation coefficient between images x and y, and C_1 and C_2 are constants to provide stability for the division. The SSIM combines measures of luminance, contrast and structure similarity.

Piella defined several similar metrics based on SSIM: Q, Q_W, and Q_E [70]. The Q metric is an average of the SSIM values between the input images and the fused image, in all possible windows of a certain size across the image plane, weighted by the salience of the input images. The Q_W metric adds a weighting per window location based on its local saliency, and the Q_E metric adds SSIM values between the edge images of the input images and the edge image of the fused image.

$$Q(a,b,f) = \frac{1}{|W|}\sum_{w\in W}(\lambda_A(w)Q_0(A,F|w) + \lambda_B(w)Q_0(B,F|w)) \tag{2.26}$$

$$Q_W(a,b,f) = \sum_{w\in W} c(w)\,(\lambda_a(w)Q_0(a,F|w) + \lambda_b(w)Q_0(b,F|w)) \tag{2.27}$$

$$Q_E(A,B,F) = Q_W(A,B,F)^{1-\alpha}\dot{Q}_W(A'.B'.F')^{\alpha} \tag{2.28}$$

Where $Q_0(A,F|w)$ is the SSIM between a certain window of A and F, $s(A|w)$ is a measure of image saliency of the window w in image A, $C(w) = \max(s(A|w), s(B|w))$, and $c(w) = C(w)/(\sum_{w\in W} C(w'))$. The images A', B', F' are defined as edge images, the norm of the gradient of the images. The weight $\lambda_A(w)$ is defined as

$$\lambda_A(w) = \frac{s(A|w)}{s(A|w) + s(B|w)} \tag{2.29}$$

The parameter α in Q_E can be varied to change the influence of the edge images.

Cvejic et al. added a correlation measure to the metric, defined as

$$Q_C = \sum_{w \in W} sim(A, B, F|w)Q(A, F|w) + (1 - \text{sim}(A, B, F|w))Q(B, F|w), \tag{2.30}$$

where $\text{sim}(A, B, F|w)$ is

$$\text{sim}(A, B, F|w) = \begin{cases} 0, & \text{if } \frac{\sigma_{AF}}{\sigma_{AF}+\sigma_{BF}} < 0, \\ \frac{\sigma_{AF}}{\sigma_{AF}+\sigma_{BF}}, & \text{if } 0 \leq \frac{\sigma_{AF}}{\sigma_{AF}+\sigma_{BF}} \leq 1, \\ 1, & \text{if } \frac{\sigma_{AF}}{\sigma_{AF}+\sigma_{BF}} > 1. \end{cases} \tag{2.31}$$

and

$$\sigma_{uv} = \frac{1}{N-1} \sum_{i=1}^{N} (u_i - \bar{u})(v_i - \bar{v}) \tag{2.32}$$

The metric is defined as

$$Q_C = \sum_{w \in W} sim(X, Y, F|w)Q_0(X, F|w) + (1 - sim(X, F, F|w))Q_0(Y, F|w) \tag{2.33}$$

Where X and Y are the input images, F is the fused image, w is the current window, and W is the set of windows across the image plane [19].

Yang et al. proposed a variant with a threshold based on the similarity between the

two input images. Where they are similar (SSIM ≥ 0.75) the metric is the same as Piella's Q metric, and where they are different (SSIM < 0.75) the maximum similarity between the input images and the fused image is taken as the result [95]. The intuition behind this method is that where the input images are very different, output image details should be taken mostly from only one input image - the value of 0.75 is chosen arbitrarily.

$$Q_Y = \begin{cases} \lambda(w)\text{SSIM}(A, F|w) + (q - \lambda(w)\text{SSIM}(B, F|w), \\ \qquad\qquad SSIM(A, B|w) \geq 0.75 \\ \max\{\text{SSIM}(A, F|w), \text{SSIM}(B, F|w)\}, \\ \qquad\qquad SSIM(A, B|w) < 0.75 \end{cases} \tag{2.34}$$

Mutual information (MI) has been used in objective metrics, first by Qu et al.[73], and then an updated metric proposed by Hossny et al. [40], who add normalisation by image entropy.

$$M_F^{AB} = 2\left[\frac{I(F, A)}{E(F) + H(A)} + \frac{I(F, B)}{E(F) + E(B)}\right] A \tag{2.35}$$

Where $I(A, B)$ is the mutual information between images A and B, and $E(A)$ is the entropy of image A.

Liu et al. performed a large-scale comparison of objective metrics for use in measuring the performance of image fusion for context enhancement in night vision [56]. They compared 12 image fusion metrics, as used to measure the performance of 6 image fusion algorithms. They concluded that no one metric was universally useful and reliable, but gave several recommendations, which included the gradient metric Q_G, and two SSIM-based metrics.

Kuhn et al. proposed a metric to measure the performance of colour to greyscale mappings [45]. This metric, which they term the root weighted mean square (RWMS) metric, calculates the error at each pixel between the distance between it and its neighbours in the input image in CIE L*a*b* colour space, and the corresponding distances in the output greyscale image. This is based on the idea that the output image should maintain the same image structure, and luminance and colour details as the input colour image. It is calculated as

$$\text{rwms}(i) = \sqrt{\frac{1}{||K||}\sum_{j \in K}\frac{1}{\delta_{ij}^2}(\delta_{ij} - |\text{lum}(\mathbf{c}_i) - \text{lum}(\mathbf{c}_j)|)^2} \tag{2.36}$$

where $\text{rwms}(i)$ is the RWMS error at pixel i of the input colour image I, K is the set of all pixels in I, $||K||$ is the number of pixels in I, $\delta_{ij}^2 = (\text{Grey}_{range}/\text{Colour}_{range})||c_i - c_j||$ is the target difference in grey levels for a pair of colours c_i and c_j, and lum is the function that returns the component L^* of a colour. This form of the equation measures the error for the L^* luminance channel, to measure the error of an output greyscale image, it is substituted into the equation instead of the lum function.

We have seen that there are a variety of methods used for image fusion, applications for these methods, and ways of measuring their levels of success. In the next chapter, we will focus specifically on derivative domain image fusion in detail, and then go on to propose new methods based on image derivatives.

Chapter 3

Derivative Domain Image Fusion

Derivative domain image fusion is the set of image fusion methods based on image gradients (derivatives). These gradients are typically calculated from an image using a $[1\ -1]$ filter (for $\partial I/\partial x$ gradients, where I is the image, for $\partial I/\partial y$ gradients a $[1\ -1]^T$ filter would be used). In this thesis we use this filter, as it provides the derivative with the smallest scale (a single pixel), but other filter types are possible, such as $[1\ \ 0\ -1]$.

An early image fusion method using gradients was that of Burt and Kolczynski [12], which used a fusion of the gradients (including in this case diagonal gradients) of the levels of a pyramid transform. The gradients are combined based on salience and match measures - in regions where the input images are similar, the pyramid coefficients are averaged, and where they are different, the most salient input image coefficients are used in the output image. The combined gradients are then used to produce an output pyramid transform, which is then inverted to produce an output fused image.

In this thesis, we focus on gradient-based image fusion methods which are not pyramidal, but instead use only the first-order x and y gradients at a single scale. This scale is generally the scale of the input image, but some of the methods described in this thesis can operate at a smaller (thumbnail) scale to reduce the computational complexity and

therefore speed up the fusion process. The starting point for these methods is the structure tensor, which encodes gradient information from any number of input image channels in a 2×2 inner product matrix.

Structure tensor based methods have many applications in computer vision [6], including in image segmentation [38], edge and junction detection [44], evaluating the liveness of face images [42], corner detection [96], denoising [24] and, relevant to this paper, for image fusion [57].

3.1 The Structure Tensor

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 the $N \times 2$ matrix of derivatives:

$$J = \begin{bmatrix} \frac{\partial I_1}{\partial x} & \frac{\partial I_1}{\partial y} \\ \frac{\partial I_2}{\partial x} & \frac{\partial I_2}{\partial y} \\ \dots & \dots \\ \frac{\partial I_N}{\partial x} & \frac{\partial I_N}{\partial y} \end{bmatrix} \tag{3.1}$$

The Di Zenzo structure tensor[22], which in differential geometry is known as the First Fundamental Form[15], is defined as the inner product of the Jacobian:

$$Z = J^T J \tag{3.2}$$

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

summarizes the combined derivative structure of the multichannel image.

3.2 Socolinsky and Wolff

Socolinsky and Wolff present a method of image fusion based on first-order image derivatives [83]. They focus on visualization of high-dimensional images in a single dimension - greyscale.

Figure 3.1 illustrates the meaning of the direction of maximal contrast. It is the line in the high-dimensional image space of $s(p)$ (also known as the spectral map) along which there is maximal contrast (maximum image gradient magnitude). This is done by solving for $\mathbf{c}$ in the equation

$$J^T J - \lambda I \mathbf{c} = 0 \tag{3.3}$$

meaning we must solve

$$\det(J^T J - \lambda I) = 0 \tag{3.4}$$

so if we define

$$J^T J = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} \tag{3.5}$$

then the two eigenvalues are

$$\lambda^{\pm}(J^T J) = \frac{1}{2}\left(g_{11} + g_{12} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}\right) \tag{3.6}$$

The corresponding eigenvector to $\lambda^{\pm}$ is the line of maximal contrast in the image at that

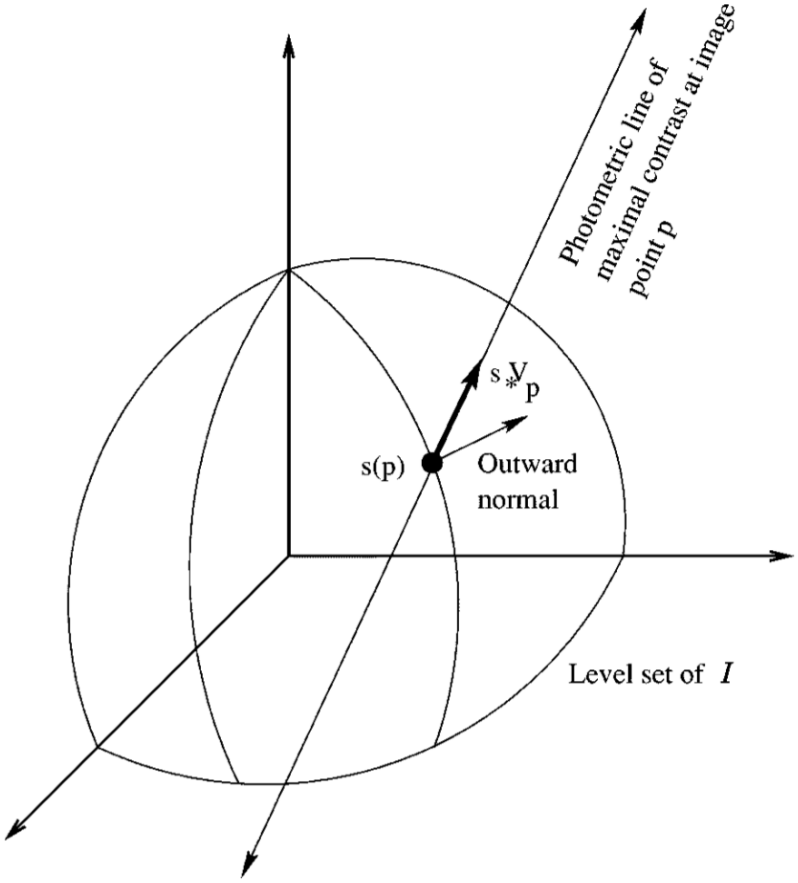


Figure 3.1: Derivative domain image fusion: Socolinsky and Wolff first-order image fusion - photometric line of maximal contrast (from [83]).

pixel, and choosing a positive or negative sign for the eigenvalue corresponds to choosing a direction to face along that line.

In their work, Socolinsky and Wolff use the eigendecomposition of the inner product $J^T J$ to find the direction of maximal gradient contrast, but the singular value decomposition (SVD) of J itself uncovers structure that is useful both for the understanding of Socolinsky and Wolff's image fusion method and also the POP variant of the local linear combination image fusion method presented in section 5.2.1 of this thesis.

$$J = USV^T \tag{3.7}$$

In Eq. 3.7, 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:

$$Z = VS^2V^T \tag{3.8}$$

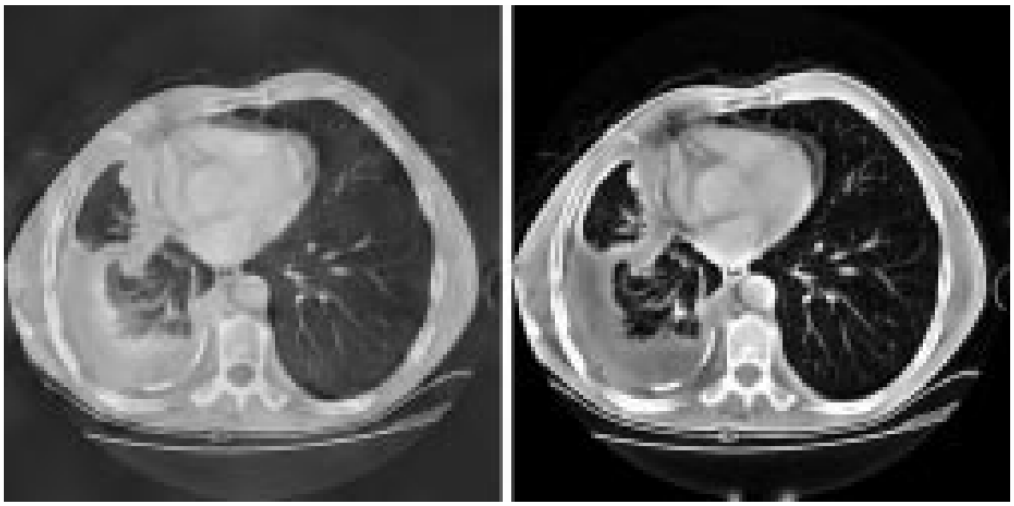
The most significant eigenvalue of Z is S_{11}^2 and the corresponding eigenvector is V_1. This eigenvector defines the direction of maximal gradient contrast in the image plane and S_{11} is the magnitude of this gradient.

In the Socolinsky and Wolff method [83], 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. 3.2). The per-pixel gradient field is written:

$$G(\mathbf{x}) = S_{11}^{\mathbf{x}}V_1^{\mathbf{x}} \tag{3.9}$$

In eq. 3.9 the superscript $^{\mathbf{x}}$ also denotes the x,y image location. We adopt this notation 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. 3.9 is ambiguous in its sign. Socolinsky and Wolff set the



(a) (b)

Figure 3.2: Derivative domain image fusion: Socolinsky and Wolff first-order image fusion - fusion of two contrast windows of chest CT scan - a) DWT fusion, b) SW (from [83]).

sign to match the brightness gradient (i.e. (R+G+B)/3). The sign can also be optimized to maximize the integrability of the derived gradient field [23]. Once we fix the sign, we write

$$\tilde{G}(\mathbf{x}) = sign(\mathbf{x}) S_{11}^{\mathbf{x}} V_1^{\mathbf{x}} \tag{3.10}$$

Figure 3.2 shows an example of their algorithm's output, for the task of fusing two contrast windows of a CT scan, in the field of medical image fusion. The SW fusion result displays clearer details and more contrast than the DWT fusion output.

One limitation of this first-order gradient-based fusion, is that pixels in the image which are not immediately adjacent cannot affect each other. The example they give is of an Ishihara plate, in which the coloured circles are separated by white space. The first-order fusion method may assign the same greyscale value to differently-coloured circles because

they are not adjacent. Our proposed method in chapter 5 solves this problem through large-scale coefficient diffusion.

3.3 Gradient Reintegration and Integrability

In general the derived gradient field $\tilde{G}(\mathbf{x})$ is not integrable (the curl of the field is not everywhere 0). In a generic vector field $V = (V_x, V_y)$, V is integrable if and only if its curl is zero:

$$\text{curl}(V) = \frac{\partial V_x}{\partial y} - \frac{\partial V_y}{\partial x} = 0. \tag{3.11}$$

So, Socolinsky and Wolff calculate the output image $O(\mathbf{x})$ in a least-squares sense by solving the discrete Poisson equation:

$$\tilde{G}_{xx} + \tilde{G}_{yy} = \nabla^2 O(\mathbf{x}) \tag{3.12}$$

where $[\tilde{G}_{xx}\ \tilde{G}_{yy}]$ denotes the divergence of the gradient field.

The iterative solution to this equation is

$$O^{t+1}_{\mathbf{x}} = O^t_{\mathbf{x}} + \frac{1}{4}\left[\nabla O^t_{\mathbf{x}} - (div\ V)_{\mathbf{x}}\right] \tag{3.13}$$

Where O^0 is any initial estimated output image, and $div\ V$ is the divergence of the vector field.

This Poisson gradient field reintegration is used for many other image processing tasks, as well as for image fusion. Pérez *et al.* present many uses for what they describe as Poisson image editing, which involves combining gradient fields and reintegrating [67]. These include seamless object insertion, as shown in fig. 3.3. It has also been used for high

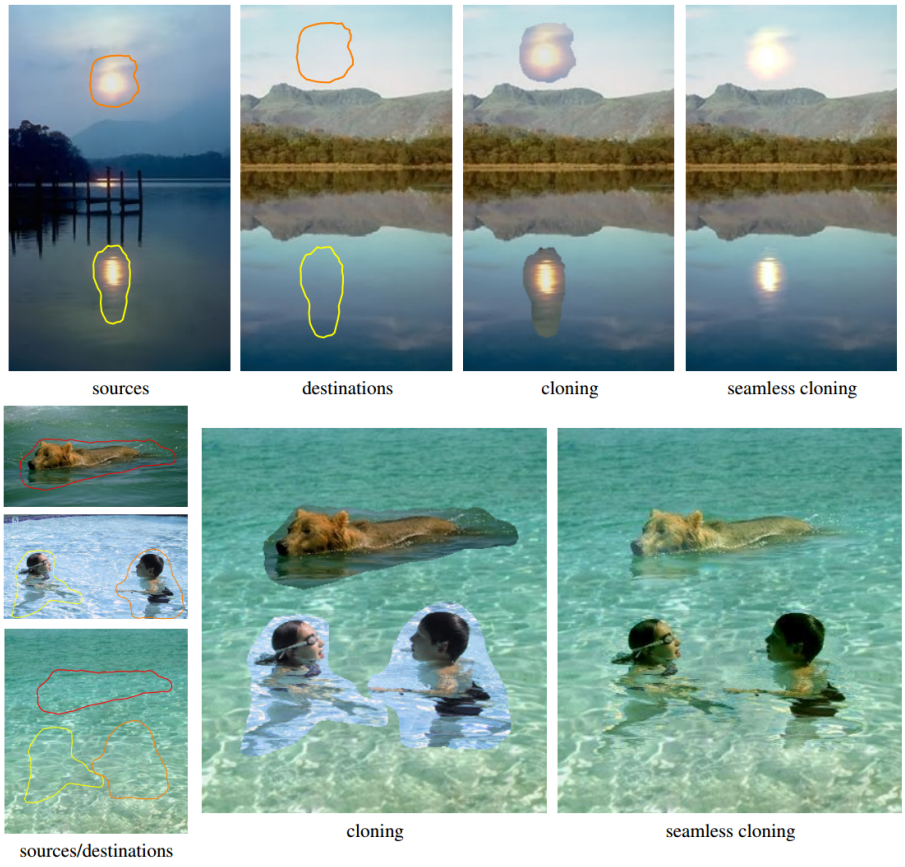


Figure 3.3: Derivative domain image fusion: seamless object insertion using Poisson image editing (from [67]).

dynamic range (HDR) image compression - the HDR gradients are reduced, then reintegrated to produce a low dynamic range (LDR) image citefattal2002gradient. Figure 3.4 is an example of how the HDR image gradients are attenuated and reintegrated to produce an output LDR image, shown using one-dimensional signal comparisons.

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.

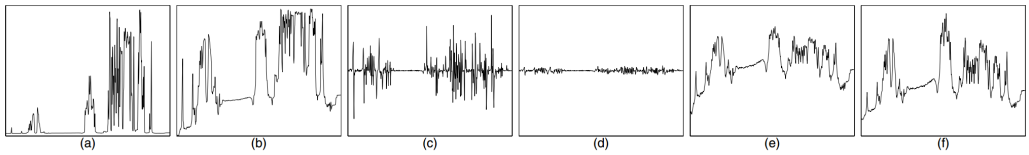


Figure 3.4: Derivative domain image fusion: high dynamic range (HDR) image compression using Poisson reintegration - gradient attenuation, a) scanline of input HDR signal, b) $H(x) = log(scanline)$, c) derivatives $H'(x)$, d) attenuated derivatives $G(x)$, e) reconstructed output LDR signal $O(x)$, f) output scanline $exp(O(x))$ (from [28]).

Figure 3.5 shows an example of an non-integrable colour image. It is clear that no greyscale image can maintain all gradient changes present in this colour image. In cases like this, the gradient reintegration techniques previously mentioned will try to find a least-squares optimal approximation. This approximation introduces artifacts, which visually look like haloes or bending (as shown in figure 3.6e).

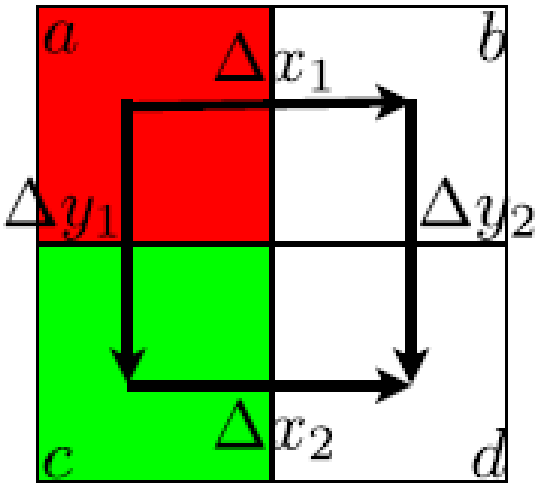


Figure 3.5: Derivative domain image fusion: gradient reintegration and non-integrability - an example of non-integrability (from [63]). It is clear that a greyscale representation of this image can not preserve all of the colour gradients, as no set of scalar values can match these colour gradients.

Another illustrative example of non-integrability that demonstrates these artifacts, using different image fusion methods, is shown in figure 3.6, 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 (as described in chapter 2). 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 [66]. The input images are small so there is only a 7 level wavelet decomposition. In 3.6c and 3.6d we show the outputs using Daubechies 4 and Biorthogonal 1.3 wavelets, the best wavelet types as found in [66]. Clearly neither the basic wavelet method nor the Socolinsky and Wolff method (3.6e) work on this image fusion example. However the POP variant of the LLC image fusion method (3.6f) - explained in chapter 5 of this thesis - succeeds in fusing the images without artifact. The intensity profile of the green line in 3.6f, shown in 3.6h has the desired equiluminant white values, whereas the Socolinsky and Wolff intensity profile 3.6g shows substantial hallucinated intensity variation.

A further example of non-integrability is shown in figure 3.7. Here the fusion task is that of horizontal and vertical white bars, on a black background. The central portion of the images, where the bars overlap in different directions, is particularly problematic. As shown in 3.7c, the SW result has clear bending artefacts around the overlapping corners, where the target gradients are highly non-integrable. The look-up-table reintegration result, shown in 3.7d, produces no artefacts but does not transfer all input details (the two output intensity levels become one). The POP variant of the LLC method in 3.7e (explained in chapter 5 of this thesis), produces an output image which transfers all relevant details while avoiding artefacts.

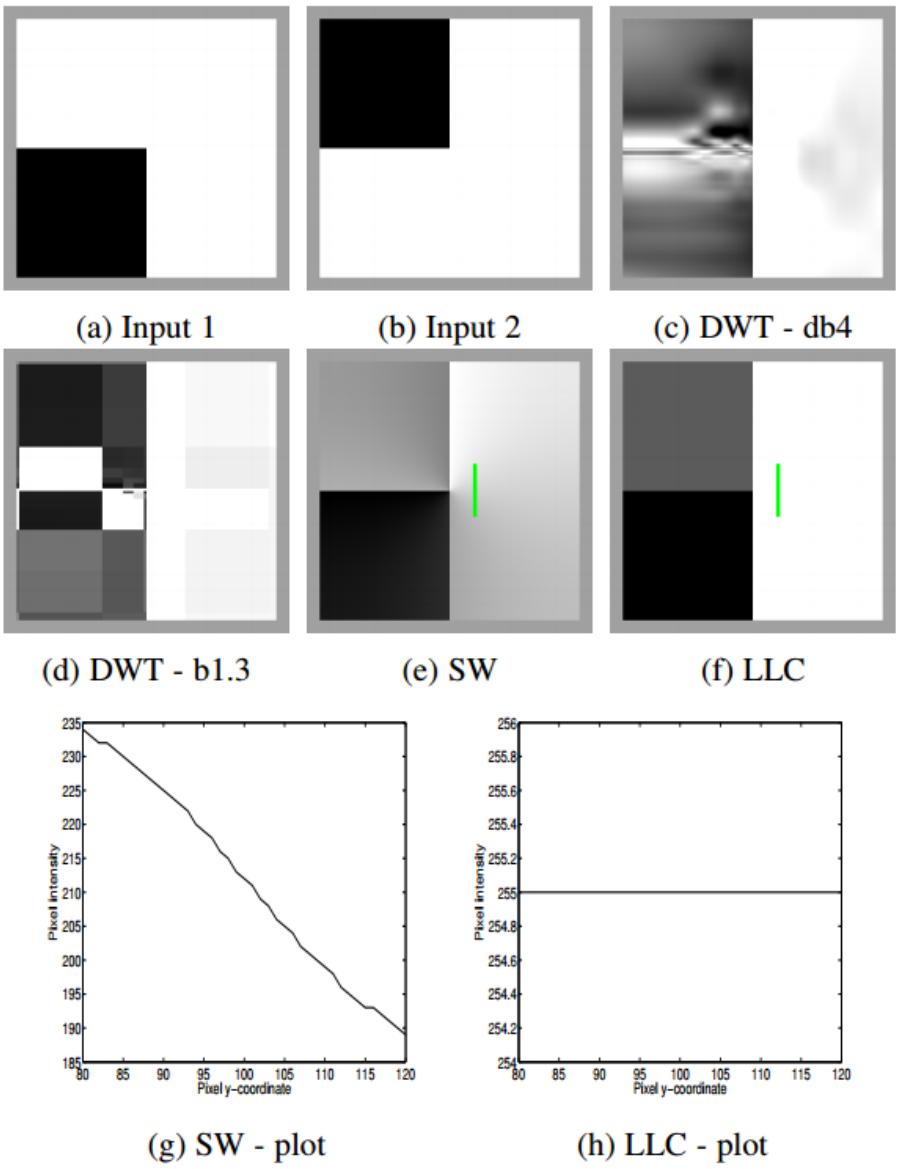


Figure 3.6: Derivative domain image fusion: gradient reintegration and non-integrability - image fusion non-integrability example 1. (a) and (b) are fused by wavelet-based methods (c) and (d), resulting in severe image artifacts. The Socolinsky and Wolff gradient-based method (e) works better, but intensity gradients are hallucinated (g) where none appear in the input images. The LLC method (f) captures all input detail with no artifacts or hallucinated detail (see chapter 5).

Montagna and Finlayson suggest reducing image saturation as a way to decrease the unintegrability of the resulting gradient field, and therefore improving the quality of the output image. They use contrast enhancement to compensate for detail loss due to this saturation reduction. This method shows improved results, but the fundamental problem of integrability remains [63]. Other recent techniques which apply additional constraints to the reintegration problem can sometimes mitigate but not remove these artifacts include using anisotropic weights to reduce the effect of target gradient noise [5], changing the SW sign assignment to reduce non-integrability using a Markov relaxation approach [23], error correction using $L1$ minimization [76] and [71]. In other work [82], the fused image is post processed so that connected components - defined as regions of the input multi-spectral 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.

Another approach to reducing non-integrability artefacts is that of Fattal *et al.* [28]. This is used by Raskar and Yu [75] to stitch together the foreground of one scene with the background of another. The claim is that this alternative reintegration approach will avoid reintegration artefacts, but it is clear from the fused output, shown in the bottom-right of figure 3.8, that severe halo and bending artefacts are still present, particularly above the book and above the man's head.

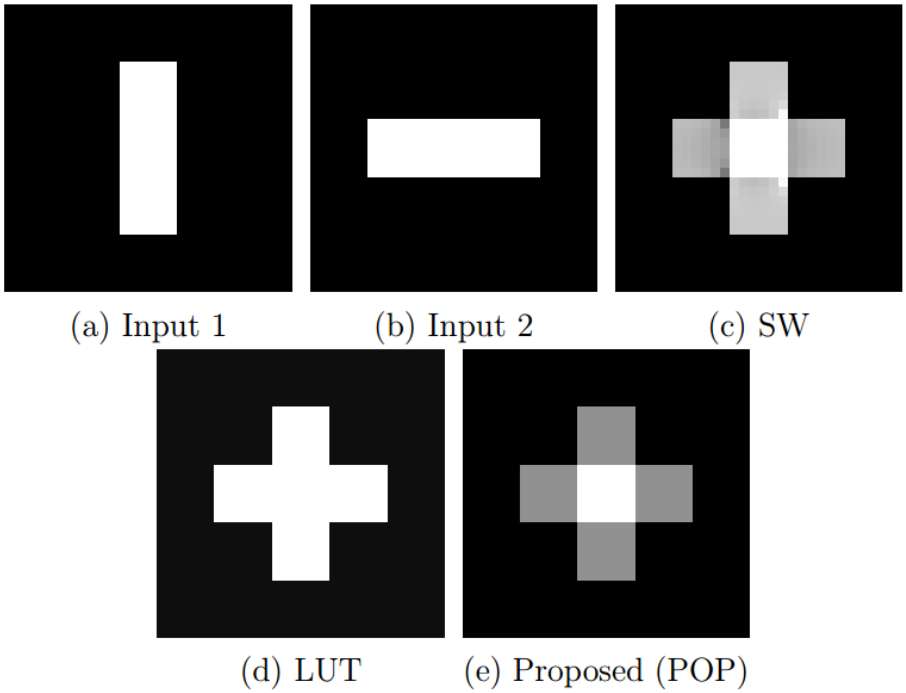


Figure 3.7: Derivative domain image fusion: gradient reintegration and non-integrability - image fusion unintegrability example 2. SW gradients are calculated from (a) and (b) (30 x 30 pixels)[83]. (c) Poisson reintegration result, (d) LUT reintegration result[31], (e) POP variant of proposed local reintegration method.

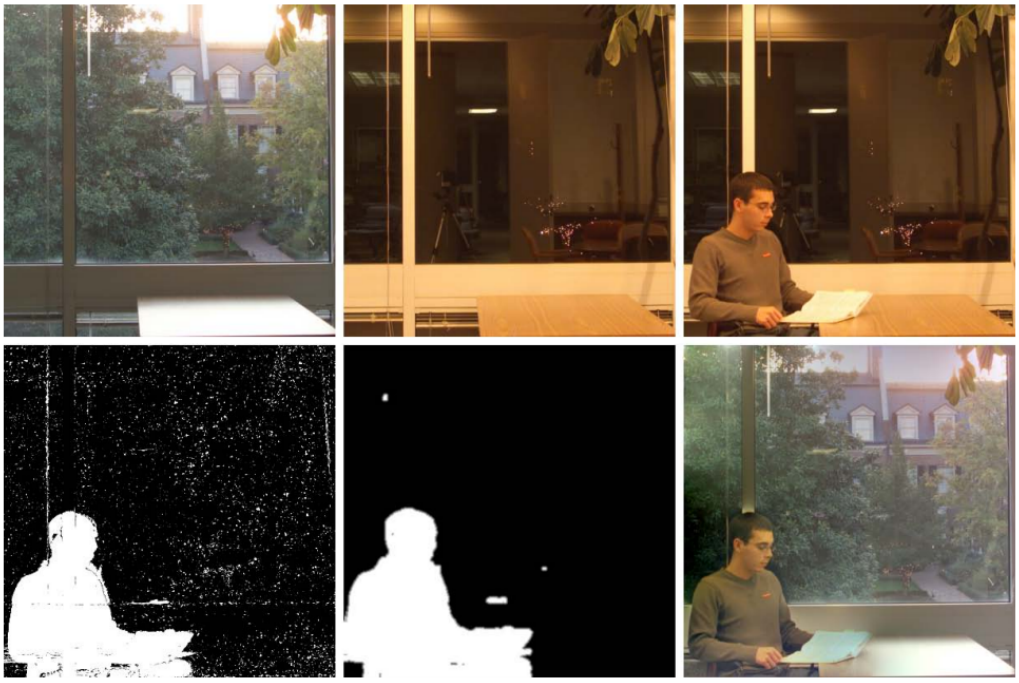


Figure 3.8: Derivative domain image fusion: gradient reintegration and non-integrability - image fusion of different scenes using the alternative to Poisson reintegration of [28] (comparison from [75]).

Lookup-table-based reintegration is an alternative reintegration method, which finds a global lookup-table (LUT) mapping from the input multichannel derivatives to the desired derivatives, and then applies this LUT to the input image channels to form an output image [31]. This is guaranteed to be free of artifacts, but may not provide maximal detail due to its global nature. This reintegration method is used in Spectral Edge (SE) image fusion [16]. The SE image fusion theorem is a way to create desired output RGB image gradients which adds the extra constraint of a desired output color for each pixel. The SE theorem allows gradients to be found which simultaneously capture the detail of all input image channels, and maintain the desired color. 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. The proposed LLC method in chapter 5 has the same desirable property of avoiding the hallucinated artifacts inherent in traditional gradient reintegration as the global look-up-table mapping of Finlayson *et al.* [31], but has far higher levels of detail transfer due to its locality.

Chapter 4

Spectral Edge Image Fusion: Experiments and Applications

Spectral Edge (SE) image fusion is a derivative domain image fusion method proposed by Connah *et al.* [16], which extends gradient-based image fusion into the realm of colour, and underpins one of the variants of LLC fusion set out in chapter 5 of this thesis. It finds equivalent gradients similar to those of Socolinsky and Wolff [83], but instead of solving for a 1-dimensional (greyscale) set of gradients, it finds a 3-dimensional (colour) set of gradients, which simultaneously transfer the gradient detail of any number of input image channels while remaining as close as possible to a guide colour image (often a visible spectrum RGB image) - the mathematics holds for any number of input and output image channels, but we focus on the instances with a 3-D colour output. These gradients are reintegrated using lookup-table-based gradient reintegration [31], which finds a global least-squares optimal mapping from N to 3 image channels.

Sections 4.1 and 4.2 review the previous Spectral Edge image fusion theorem used to calculate equivalent gradients and the lookup-table-based gradient reintegration technique used to create an output fused image from them.

A psychophysical preference experiment is conducted in section 4.3, comparing the Spectral Edge image fusion method to naive luminance channel replacement, the ratio of low-pass pyramid (ROLP) method (as explained in section 2.3 of this thesis), and the method of Schaul *et al.*. The results of the experiment show that the SE method is clearly superior to the input visible RGB image, the naive luminance channel replacement method, and the ROLP method, but the SE and Schaul *et al.* methods are not different to a statistically significant margin.

Section 4.3 also introduces new image fusion metrics, which are extensions of existing 2 to 1 channel image fusion metrics for a higher number of input and output channels. Metrics based on output image colourfulness and contrast are also proposed. The results of these are combined and compared to the results of the psychophysical experiment - a combination of gradient-based and colourfulness metrics gives the closest predicted quality ordering to that of the psychophysical experiment.

Next, section 4.4 introduces a new iterative extension to Spectral Edge image fusion, in which the equivalent gradients calculated by the SE theorem are used as the colour guide gradients for another iteration of gradient calculation - this increases the strength of the detail transfer of the method, as shown with example results for RGB-NIR image fusion, and a psychophysical experiment - more experiments are needed for a conclusive result, but there are tentative indications that using more than one iteration of the Spectral Edge image fusion method may produce superior results, as judged by objective and subjective metrics.

Finally, section 4.4 presents example results of new applications of the Spectral Edge image fusion method: image fusion from a single sensor with an RGB-NIR Bayer pattern, and both natural and false colour RGB-thermal image fusion using the FLIR ONE

smartphone thermal camera.

4.1 Spectral Edge Theorem

Spectral Edge image fusion is a derivative domain image fusion method. It uses the structure tensor, but instead of finding a one-dimensional set of desired gradients from a high-dimensional set of input gradients, to produce a single output channel with maximum detail (as in SW), it finds output derivatives which simultaneously have maximum detail transfer while also remaining as close as possible in color to the input RGB image (or other color guide image) [16]. This allows, instead of N to 1 channel image fusion with maximum gradient transfer, N to 3 channel image fusion which simultaneously has maximum gradient transfer while preserving the desired image colours.

The Spectral Edge theorem, as defined by Connah *et al.*, is as follows (the notation has been changed to match that used in this thesis):

> Given a multi-dimensional image H and a putative RGB "guiding" image R, we can generate a new RGB gradient matrix ∇D that is as close as possible to the gradient of the RGB image, and whose contrast matches that of H exactly.

If we define J as the image Jacobian at a pixel (see equation 3.1), and the inner product of the Jacobian (the structure tensor) as Z, the aim of the SE method is to find an image whose structure tensor exactly matches that of the high-dimensional input image, termed Z_H, while simultaneously keeping the gradients as close as possible to the input RGB gradients J_R. We define the final derived image gradient as ∇D, and is of the form

$$\nabla D = \begin{bmatrix} \frac{\partial D_1}{\partial x} & \frac{\partial D_1}{\partial y} \\ \frac{\partial D_2}{\partial x} & \frac{\partial D_2}{\partial y} \\ \frac{\partial D_3}{\partial x} & \frac{\partial D_3}{\partial y} \end{bmatrix} \tag{4.1}$$

To accomplish this, ∇D must lie within the span of J_R, therefore

$$\nabla D = J_R \, A \tag{4.2}$$

Now $Z_R \equiv Z_H$, therefore $A^T Z_R A \equiv Z_H$. The complete set of A which satisfies this condition is given by

$$A = \left(\sqrt{Z_R}\right)^+ \, B \, \sqrt{Z_H}, \text{ s.t. } B^T B = I_2 \tag{4.3}$$

Where the matrix square root is the unique symmetric root of the real positive semi-definite symmetric matrices Z_R and Z_H, $+$ indicates the Moore-Penrose pseudo-inverse, and I_2 is the 2×2 identity matrix.

Thus far the matrix A will produce a result with equivalent high-dimensional detail, but the second condition is to remain as close to the colour guide image gradients as possible, so

$$J_D \simeq J_R \tag{4.4}$$

$$\implies J_D\, A \simeq J_R \tag{4.5}$$

$$\implies A \simeq I_2 \tag{4.6}$$

$$\implies \sqrt{Z_R}^{+}\, B\, \sqrt{Z_H} \simeq I_2 \tag{4.7}$$

$$\implies B\, \sqrt{Z_H} \simeq \sqrt{Z_R} \tag{4.8}$$

The last line of 4.4 means that B must rotate $\sqrt{Z_H}$ so that it is as close as possible to $\sqrt{Z_R}$. This is the Orthogonal Procrustes Problem - the solution in the least-squares sense is to firstly use a singular value decomposition to express the product of the square roots of Z_R and Z_H:

$$\sqrt{Z_R}\left(\sqrt{Z_H}\right)^{T} = D\,\Gamma\,E^{T} \tag{4.9}$$

Then the solution B is given by

$$B = D\,E^{T} \tag{4.10}$$

This B is then substituted into equation 4.3 to produce A, which is applied to the guide image gradients J_R to produce our final SE equivalent gradients J_D. In this chapter, the output image is created from these gradients using lookup-table-based gradient reintegration, as explained in the following section - in the following chapter, we will show an alternative method of reintegrating these gradients.

4.2 Look-up-table Gradient Reintegration

Let us denote the derived gradient field, which could be calculated via Socolinsky and Wolff[83], or via the Spectral Edge theorem, as ∇D. It could be that there is no image that has derivatives exactly equal to the ones we seek. After all, for every pixel we have an x and y derivative yet the reintegrated image has a single pixel value - the gradient field may be non-integrable, as described in section 3.3 of this thesis. Thus, the typical away to solve this reintegration problem is to solve the Poisson equation to find an output image O with gradients as close as possible to the derived gradients in a least-squares sense:

$$\arg\min_{O} ||\nabla O - \nabla D|| \tag{4.11}$$

As explained in more detail in chapter 3, in finding the image O it is often the case that the Poisson reintegrated image has details not in any of the original N-image planes H. Indeed O will typically have halos and/or bending artifacts. If the gradient field is not integrable, in solving for O (in a least-squares sense) the error manifests itself in these visible artifacts.

One way to remove artifacts from the reintegrated image is to place a constraint on O. Let us denote all images that are a global linear combination of H as

$$O \in P_1(H) \tag{4.12}$$

Or, if we also allow second order polynomial terms (i.e. for an RGB image this would be R^2, G^2, B^2 and RG, GB and BG) we write

$$O \in P_2(H) \tag{4.13}$$

where P_n denotes the order of the polynomial expansion. Finlayson *et al.* [31] proposed solving for O as

$$\underset{O \in P_2(H)}{\arg\min} ||\nabla O - \nabla D|| \tag{4.14}$$

The advantage of ensuring that the output image is a function of the input is that bending and halo artifacts cannot occur (a unique N-pixel in H maps to a unique greyscale in O). Further, adopting a low order polynomial ensures the function is smooth and that the computational process is rapid. The disadvantage of this reintegration method is that a global function of the input image channels can only approximate the target gradients (there will be considerable error, more so than with Poisson reintegration), so a sub-optimal level of detail will be transferred.

4.3 Image Fusion Quality Assessment

In this section, we use a forced choice pairwise psychophysical experiment to compare the perceived subjective image quality of the Spectral Edge image fusion method with other image fusion methods, for the purpose of RGB-NIR image fusion. We then try to find an objective image fusion quality metric which will give the same quality ranking of the methods.

4.3.1 Psychophysical Experiment

In the psychophysical experiment, five classes of image are compared: the original RGB input image, the Spectral Edge image fusion method output, and the results of three other methods based on fusing the RGB luminance channel and the NIR image (using a simple average, ratio of low-pass pyramid (ROLP) image fusion [87], and the dehazing method

of Schaul *et al.*[79]) and substituting the result as a new luminance channel (i.e. replacing the V channel in HSV colour space, or the L channel in CIELAB). As it is a forced choice pairwise comparison, there are 10 comparisons per test image - with 10 test images, and 2 repetitions, this adds up to a total of 200 comparisons per observer. A total of 8 observers, naive to the research, took part in the experiment. Figure 4.1 shows a comparison of all of the methods compared in the experiment used on one of the test images.

The psychophysical experiment is a forced choice pairwise comparison test (Thurstones law case V), as used by Connah et al. [17]. In each comparison test subjects have to select the image they prefer according to personal taste (through forced-choice, i.e., there is no "I dont know" option). All pairs of images (for the different algorithms and the same scene) are presented twice (each pair is presented as a left-right pair and as a right-left pair), in a random order. 10 images from the EPFL RGB-NIR data set are used [8]. We also adopt ISO 3664:2009 recommendations for carrying out image preference experiments. The pairwise preferences for multiple observers are counted in a score matrix and then using Thurstone's method we convert the scores into algorithm ranks and then to preference scores (as explained in section 2.5 of this thesis). Significantly, the Thurstone method also returns confidence intervals (and so it is possible to conclude whether one algorithm is significantly better than another). As the number of comparisons is relatively high (200), each experiment is split into two sessions of 100 comparisons each, to reduce fatigue in volunteers.

The results, with 8 observers naive about the experiment, are shown in figure 4.2. According to this figure we can conclude that the ranking of mean perceived image quality is that the Spectral Edge method is the leading algorithm, second is the dehazing method, third is the ROLP fusion, fourth is the original RGB image, and the luminance channel and

Figure 4.1: Spectral Edge image fusion: psychophysical experiment - example image comparison (top row: RGB, NIR. Middle row: luminance average, ROLP. Bottom row: Schaul *et al*, SE.

NIR average is in last place. Although this is the indicated ranking, the only statistically significant difference (as indicated by the error bars not overlapping), is between the SE method and the original RGB image. More observers are required to confirm the rest of the ranking.

The Spectral Edge method - which provides less detail transfer than some of the competing methods - is, it can be theorized, preferred because of the lack of artefacts, because it is closer to what is expected (from a normal photographic diet), and because the color aspect of image fusion is integral to the method (it is not based on luminance fusion). The dehazing method is also indicated to be possibly preferred over the original RGB image, as it includes large amounts of detail, but with fewer artifacts than the ROLP method, due to its use of edge-preserving filters. The top two methods, SE and Schaul *et al.*, do not differ in subjective quality by a statistically significant margin, but the mean quality result is higher for SE. The other three methods, including the original RGB image, are not significantly different in their preference results, but are significantly lower in subjective image quality than those of SE.

4.3.2 Objective Image Fusion Quality Metrics

Image fusion quality metrics can be divided into two main types - metrics with a reference image and non-reference metrics. As we do not have a reference image, showing the ideal fusion, we examine non-reference metrics here. These typically consist of some measure of the image similarity between each input image and the fused image, often weighted by a measure of image salience.

In our metric testing, we calculate the ranking that each metric gives to our RGB-NIR image fusion algorithms based on a Borda count, which has been used for classifier fusion

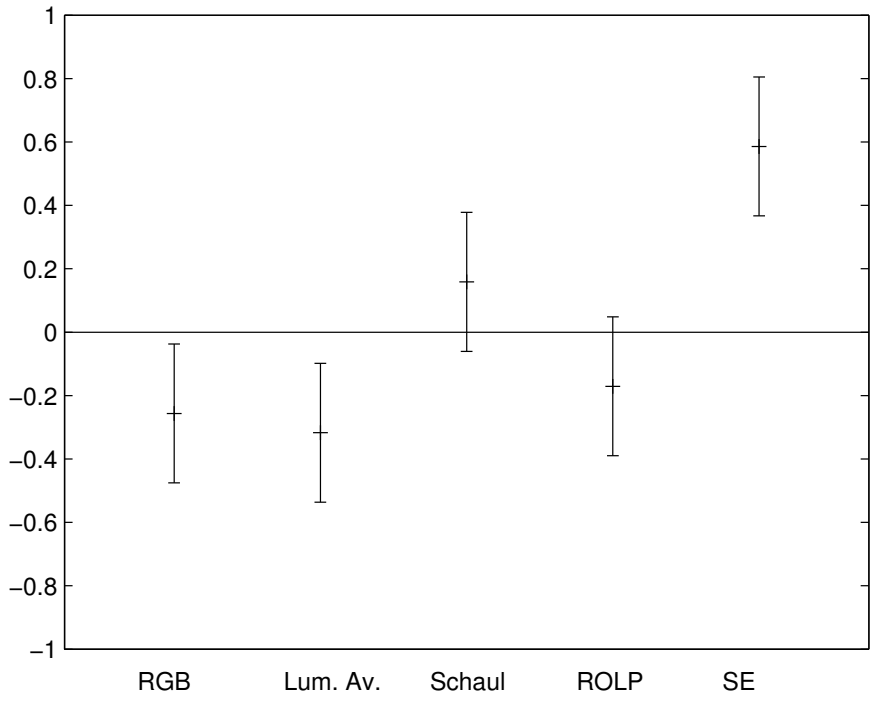


Figure 4.2: Spectral Edge image fusion: psychophysical preference ranking

[78], and in the metric comparison of Liu et al. [56] - the metric scores for the fusion result of one image, with each algorithm, are ranked. The lowest performing algorithm receives 1 point, which increases up to 5 points for the best performing. These points are added up over the 10 images, to obtain a final point total for each algorithm. These point totals are used to rank the algorithms - we can then compare this ranking to the one obtained from our psychophysical experiment. This scoring method is purely based on the ranking of the metric results, not their absolute values, which are often meaningless - if one algorithm has a score of double that of another algorithm, it is not necessarily twice as effective.

Existing metrics are often based on measuring the performance of 2 to 1 channel image fusion, as described in section 2.3.2. To use these metrics for RGB-NIR image fusion, we

could use the RGB luminance channel and NIR image as two greyscale inputs, and the output RGB luminance channel as the single greyscale output, but this would ignore the colour aspects of the input and output images. Therefore we propose extensions of these metrics to M to N channel image fusion, and then compare their results to those of our psychophysical experiment.

The first metric we extend is based on mutual information, a measure of the mutual dependence of two random variables, in this case two images. Hossny et al. defined an image fusion metric based on mutual information [40], and we extend this metric to M to N channel image fusion. We propose

$$Q_{MI} = \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{MI(I_m, O_n)}{E(I_m) + E(O_n)} \tag{4.15}$$

where I_m is the mth input channel, O_n is the nth output channel, $MI(X, Y)$ is mutual information, and $E(X)$ is image entropy. The normalisation step of dividing by the entropies of the current channels is important, to avoid channels with higher entropy dominating the result.

Another measure of image similarity is the structural similarity image measure (SSIM), which is based on similarity in luminance and contrast, and the correlation between images [90]. Piella proposed several similar image fusion metrics based on the SSIM [70]. We did preliminary testing using extended versions of each variation, and the best performing was an extension of the Q metric. We propose

$$Q_{SSIM} = \frac{1}{N} \frac{1}{|W|} \sum_{w \in W} \sum_{m=1}^{M} \sum_{n=1}^{N} \lambda_m(w) SSIM(I_m, O_n | w) \tag{4.16}$$

where I_m is the mth input channel and O_n is the nth output channel. A sliding window

w is moved across the images, and the local regions' SSIM is calculated. The weight λ_m is calculated based on a measure of local image saliency $s(I_m|w)$ such as variance:

$$\lambda_m(w) = \frac{s(I_m|w)}{\sum s(I|w)} \tag{4.17}$$

Image gradients are a natural way of representing image detail, and their similarity is used in the image fusion metric of Xydeas and Petrovic [93]. Our extended version of their metric is defined as:

$$Q_G(x,y) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} Q_n^m(x,y)G^m(x,y)}{\sum_{m=1}^{M} G^m(x,y)} \tag{4.18}$$

where Q_n^m is the edge information preservation value between channels m and n at the pixel location of x and y as defined by Xydeas and Petrovic, which measures how much gradient information has been transferred, and G^m is the gradient magnitude of the input channel at the same pixel location, which acts as a weighting. If we collect the results for each pixel into a vector $\mathbf{Q_G}$, the final metric result is the median value of this vector - we have replaced the mean of the original metric with a median, as the median is more stable to asymmetric and/or skew distributions.

$$Q_G = \text{median}(\mathbf{Q_G}) \tag{4.19}$$

Table 4.1 shows the algorithm ranking given by the three proposed metrics, compared to the ranking of the psychophysical experiment. Note here that the SE and dehazing rankings may be swapped in order in the prediction, as well as the rankings between the bottom three methods, as they are not statistically different in the psychophysical experiment. The metric based on mutual information gives a ranking uncorrelated to the psychophysical ranking,

while the SSIM-based metric gives a ranking that is completely the reverse. The gradient-based metric gives a ranking the most similar, with only the SE result as an anomaly. We are not certain why the SE method is such an anomaly with this metric, but our conjecture is that the SE method creates gradients which are not equal to either the RGB or NIR gradients, but which nonetheless represent a detail and color synthesis. As the Q_G metric measures gradient similarity, it would not score these new synthesized gradients highly. The lookup-table-based gradient reintegration technique used in the SE method may also be a factor, as its global mapping may not transfer maximum detail.

The other three RGB-NIR fusion methods, which are ranked correctly by the gradient-based metric, are all based on luminance fusion. Therefore we suggest that this metric may be useful for RGB-NIR fusion based on luminance channel fusion and replacement.

It may be the case that something other than just detail is causing the SE method to rank so highly in our experiment. The SE method is the only method to try to integrate the NIR information into the colour of the image, therefore colour is a likely area to explain its result. Colourfulness, as measured by chroma in CIELUV color space (as defined in [86]), has been linked to observer preference [29], as has image contrast [14], as measured by root mean square (RMS) contrast. Furthermore, after completing the psychophysical experiment, we interviewed the participants, and they explained that colorfulness was a major factor in their preference decisions, with increased colourfulness being generally preferred.

It may not be true that higher colourfulness and contrast are always preferred (in some cases the reverse could be true), but for the purposes of calculating metrics, we make the assumption that our input RGB images are below the optimum for these values, so we create metrics Q_{Col} and Q_{Con}, which consist of a sinusoidal function, for which a value

of 50% more than the original colorfulness or contrast (defined as $\sum(u^2 + v^2)$ and pixel intensity standard deviation, respectively) is the optimum, giving a result of 1, and which drops to 0 when the colorfulness or contrast drops to 0 or reaches three times its original value. Note we are not saying images should be always have more colorfulness, but that some increased colorfulness is preferred. We recognize that this is a somewhat arbitrary assumption, but as we are conducting initial experiments we feel it is justified - in future research it would be useful to examine this question further.

Table 4.2 shows the algorithm ranking, calculated from a Borda count of the ranking of the algorithms for each image, using the raw colorfulness and contrast values of the fused images for each method. Using raw colorfulness and contrast values, the SE method is accurately predicted as the best method by these metrics, but the rest of the ranking does not match the psychophysical experiment. However, the colorfulness metric gives a ranking quite close to that of the experiment, with only the original RGB and the dehazing method as anomalies.

The metric results so far imply that the SE method does not transfer a maximal amount of detail from the NIR into the output color image, but that instead observer preferences are linked to the increased colorfulness and contrast of its output images.

As none of the measures tested so far gives results entirely consistent with the psychophysical experiment, our next step is to combine the results of several metrics. The metric Q_G gives good results except for the ranking for SE, and the colorfulness metric gives an opposite result, accurately placing SE as the leading algorithm, but misranking two others. Therefore we combine the two metrics, with a weighted combination:

$$Q = \alpha Q_G + (1 - \alpha) Q_{Col} \tag{4.20}$$

Metric	SE	Dehazing	ROLP	RGB	Lum. Av.
Psychophysical experiment	1	2	3	4	5
Q_{MI}	2	4	3	1	5
Q_{SSIM}	5	3.5	3.5	2	1
Q_G	5	1	2	3	4

Table 4.1: Spectral Edge image fusion: comparison of metric rankings

Metric	SE	Dehazing	ROLP	RGB	Lum. Av.
Psychophysical experiment	1	2	3	4	5
Colorfulness	1	5	4	2	3
Contrast	1	4.5	4.5	2	3
Q_{Col}	1	3	4	2	5
Q_{Con}	1	4.5	4.5	2	3

Table 4.2: Spectral Edge image fusion: comparison of rankings by colorfulness and contrast

Table 4.3 shows the results of combining the two metrics, with several values of α. Through preliminary testing we have found the optimal value of α to be approximately 0.6 - with this value the correct ranking is achieved, with the exception of the RGB and ROLP rankings being swapped - however, the three bottom methods are very close in the experiment, with highly overlapping error bars. The ranking of the Spectral Edge and dehazing methods are the most important, and these two are correctly ranked.

Metric	SE	Dehazing	ROLP	RGB	Lum. Av.
Psychophysical experiment	1	2	3	4	5
$\alpha = 0.5$	1	2.5	3	2.5	5
$\alpha = 0.25$	1	3	4	2	5
$\alpha = 0.75$	4	1.5	3	1.5	5
$\alpha = 0.6$	1	2	4	3	5

Table 4.3: Spectral Edge image fusion: comparison of rankings by combined gradient and colorfulness metrics

4.4 Iterative Spectral Edge Image Fusion

As described previously, Spectral Edge image fusion finds an RGB image O from an n-dimensional image H for which a guide RGB image R is known. The nature of the global function used to calculate the output image in this previous version of the method places limits on the levels of detail transfer that can be achieved, and therefore how closely the output image can approximate the desired structure tensor at all pixel locations. In this section, we propose an iterative version of the algorithm, which can come closer to maximum detail transfer and to the desired structure tensor values.

We can think of the SE fusion process as a function,

$$O = SE(R, H) \tag{4.21}$$

The output image O has a gradient structure similar to H but has colours similar to R. The main idea of this section is to apply the Spectral Edge algorithm in iteration

$$O_i = \text{SE}\left(O_{i-1}, H\right) \tag{4.22}$$

where

$$O_0 = \text{SE}\left(R, H\right) \tag{4.23}$$

By construction the Spectral Edge algorithm forces the integrated edge to be a function of the global polynomial expansion of the input images. This constraint is applied to avoid the reintegrated gradient field having artifacts and it also makes the whole gradient field reintegration very rapid. However, the gradient field of O_0 (∇O_0) may be quite far from the SE derived gradient field ∇D (especially compared to Poisson reintegration). But, if

we run the algorithm in iteration we should move to a better approximation, because the need to be close to the original RGB image is relaxed.

We produced 15 iterations of outputs from 16 RGB-NIR image pairs from the EPFL RGB-NIR data set [8]. Figure 4.3 shows the mean structure tensor error across the 16 images, measured as the L2 norm of the difference between the structure tensor of the high-dimensional input and the structure tensor of the output image, with the mean taken across all pixel locations.

$$\frac{1}{|X||Y|}\sum_{x\in X}\sum_{y\in Y}||Z_H(x,y)-Z_O(x,y)||_2 \tag{4.24}$$

Where X and Y are the sets of possible x and y coordinates in the image.

Figure 4.4 shows the RGB error, measured as the L2 norm of the difference between each guide RGB pixel value R and that of the output image O,

$$\frac{1}{|X||Y|}\sum_{x\in X}\sum_{y\in Y}||R(x,y)-O(x,y)||_2 \tag{4.25}$$

Where $R(x,y)$ and $O(x,y)$ are length 3 vectors of the R, G and B pixel values at that x and y location.

As the number of iterations increases, the result has a structure tensor closer and closer to that of the high-dimensional input, meaning more of the gradient information is present in the output image. This trend continues up to 6 iterations, after which the tensor error slowly increases. However, the result also differs more and more from the original RGB image, and appears to be approaching an asymptote. This may at some point lead to a less natural and pleasing output image - the images also appear more colourful with more iterations and there should be a point after which subjective preference decreases with

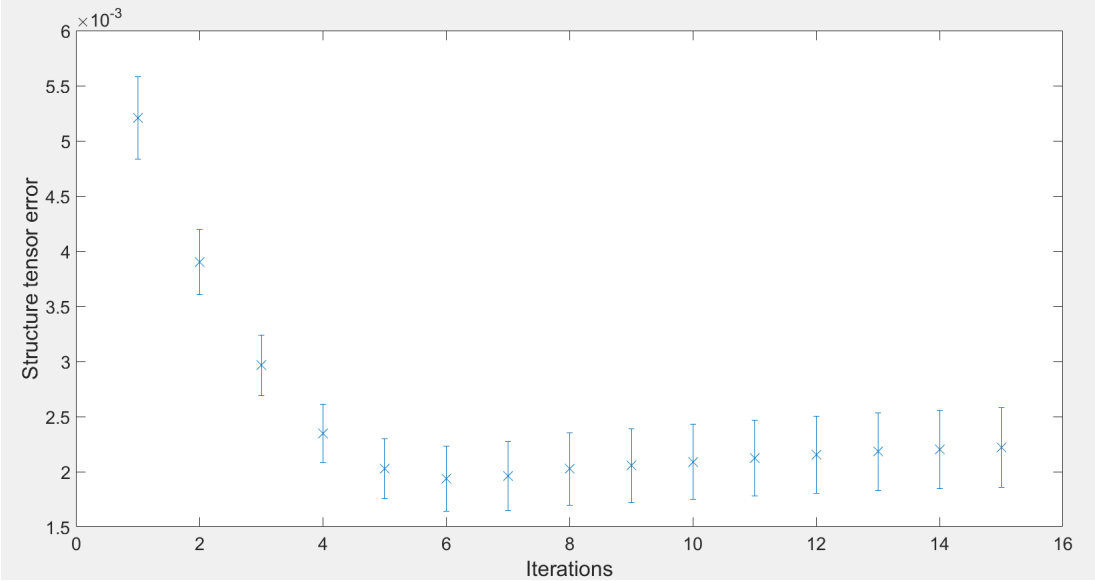


Figure 4.3: Spectral Edge image fusion (iterative): structure tensor error by iteration

increasing colourfulness[29] and detail.

We performed a psychophysical experiment, comparing the standard SE method to the result after 2, 4 and 8 iterations, for RGB-NIR colour image fusion. 16 test images from the EPFL RGB-NIR data set were used [8].

In the previous section of this thesis, we compared the standard SE method to the original RGB image several image fusion methods - in this section we only compare the SE method with its iterative version.

Psychophysical results with 8 observers indicate a preference ranking, as shown in fig. 4.5, where Q is perceived image quality. All the error bars are overlapping, which means that the differences are not statistically significant with this number of observers - however, the order of the mean quality values indicates that the SE method with 2 iterations may be the most preferred, followed by 4 iterations, then the original method, then 8 iterations. Results were not obtained for 3 iterations, but from the ranking it looks like either 2 or 3

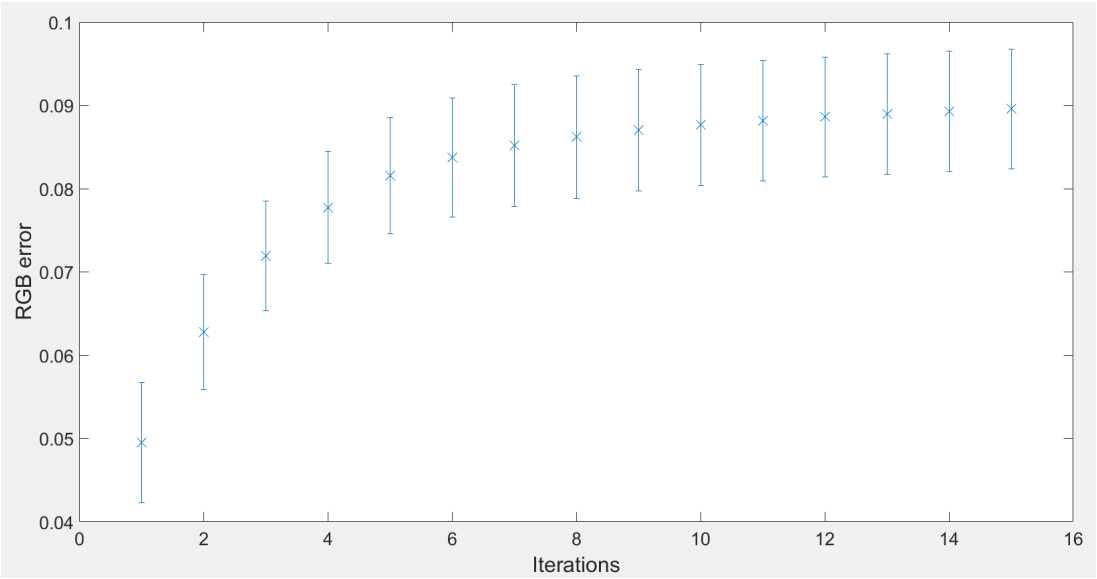


Figure 4.4: Spectral Edge image fusion (iterative): RGB error by iteration

iterations may be the peak for preference.

If the experiment ranking is correct (more observers would be required to prove this), then it would seem that the first few iterations produce extra detail and colour which is beneficial for subjective preference, but then with successive iterations, the output image becomes too extreme and unnatural. Figures 4.6, 4.7 and 4.8 demonstrate this tendency. In figure 4.6, the vegetation becomes increasingly green, and the water increasingly blue with each iteration. At first this is an enhancement, as the colour vividness and contrast is increased, but eventually it becomes too much and unnatural. Figures 4.7 and 4.8 follow a similar pattern with regards to vegetation, and also in figure 4.8 the road's contrast is at first improved, but then it becomes unnaturally purple after 4-8 iterations.

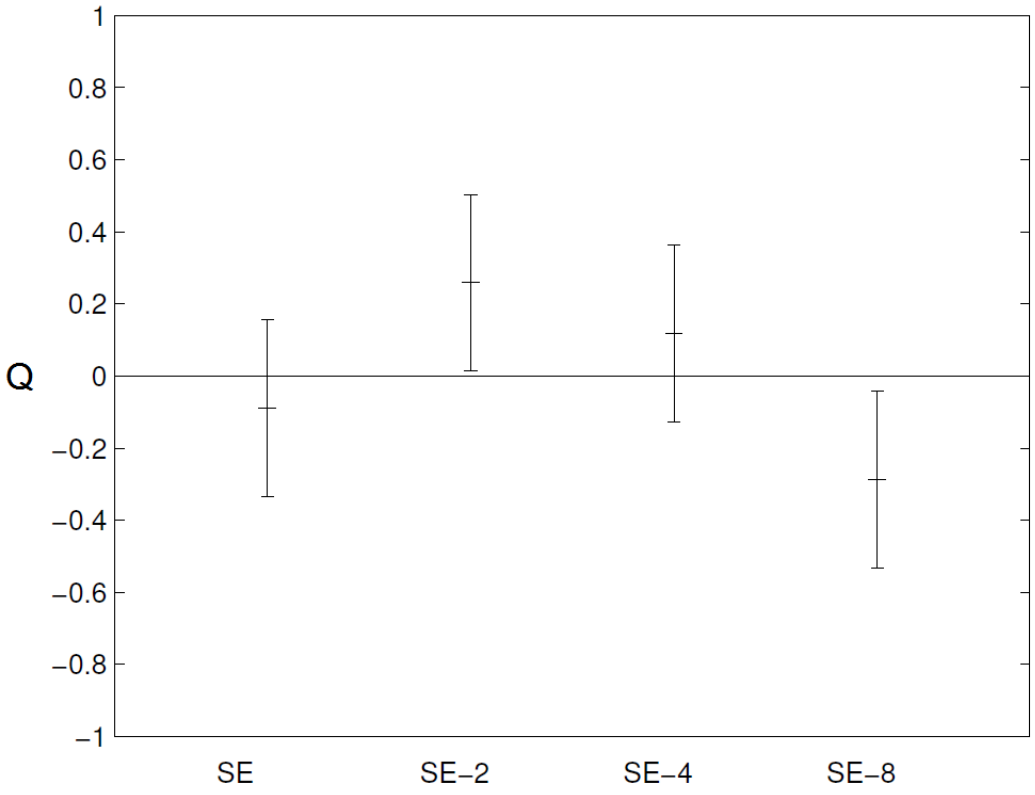


Figure 4.5: Spectral Edge image fusion (iterative): psychophysical experiment results

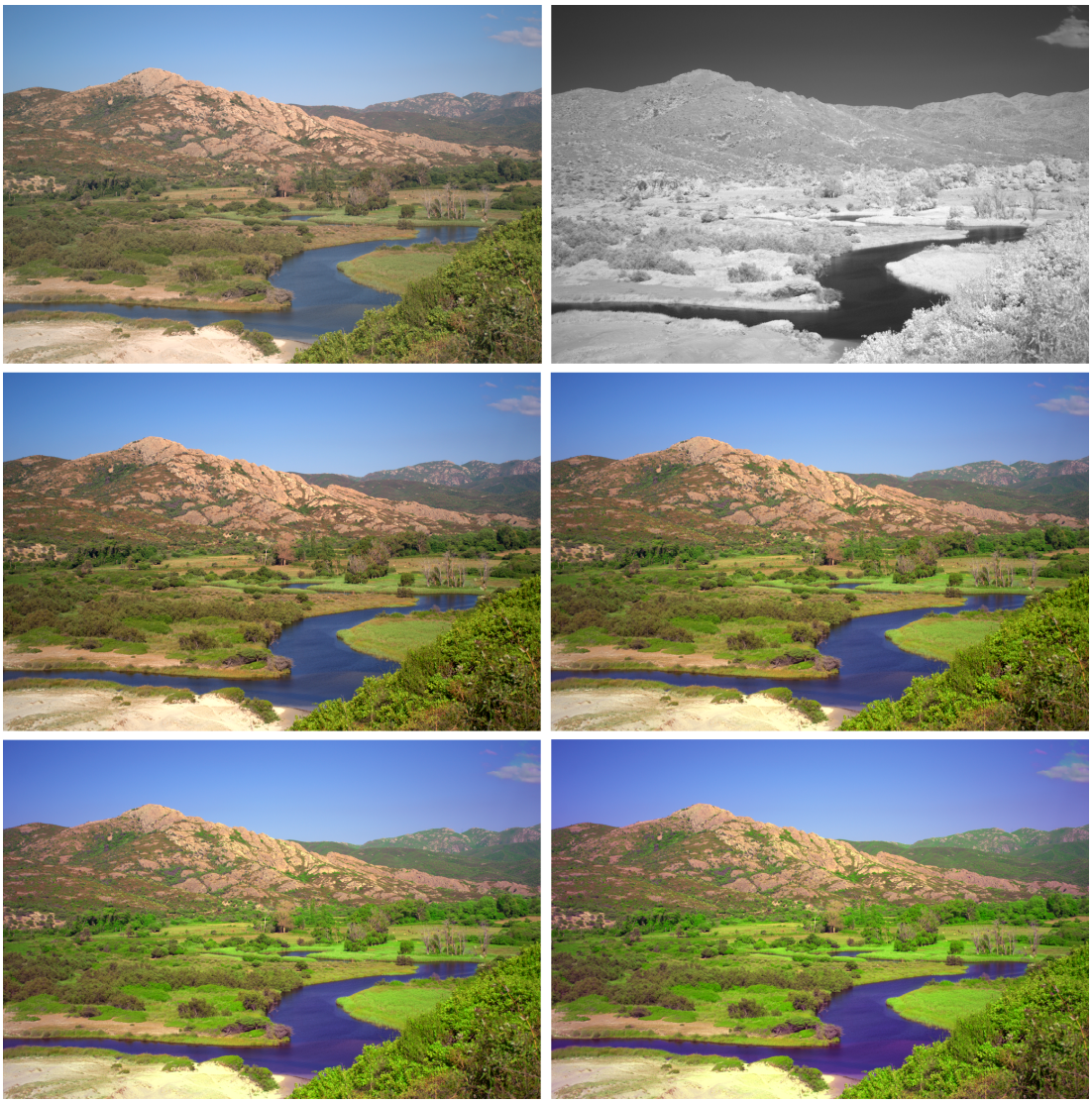


Figure 4.6: Spectral Edge image fusion (iterative): RGB-NIR Image Fusion - 'Country04' comparison (top row: RGB, NIR. Middle row: SE, SE-2. Bottom row: SE-4, SE-8)

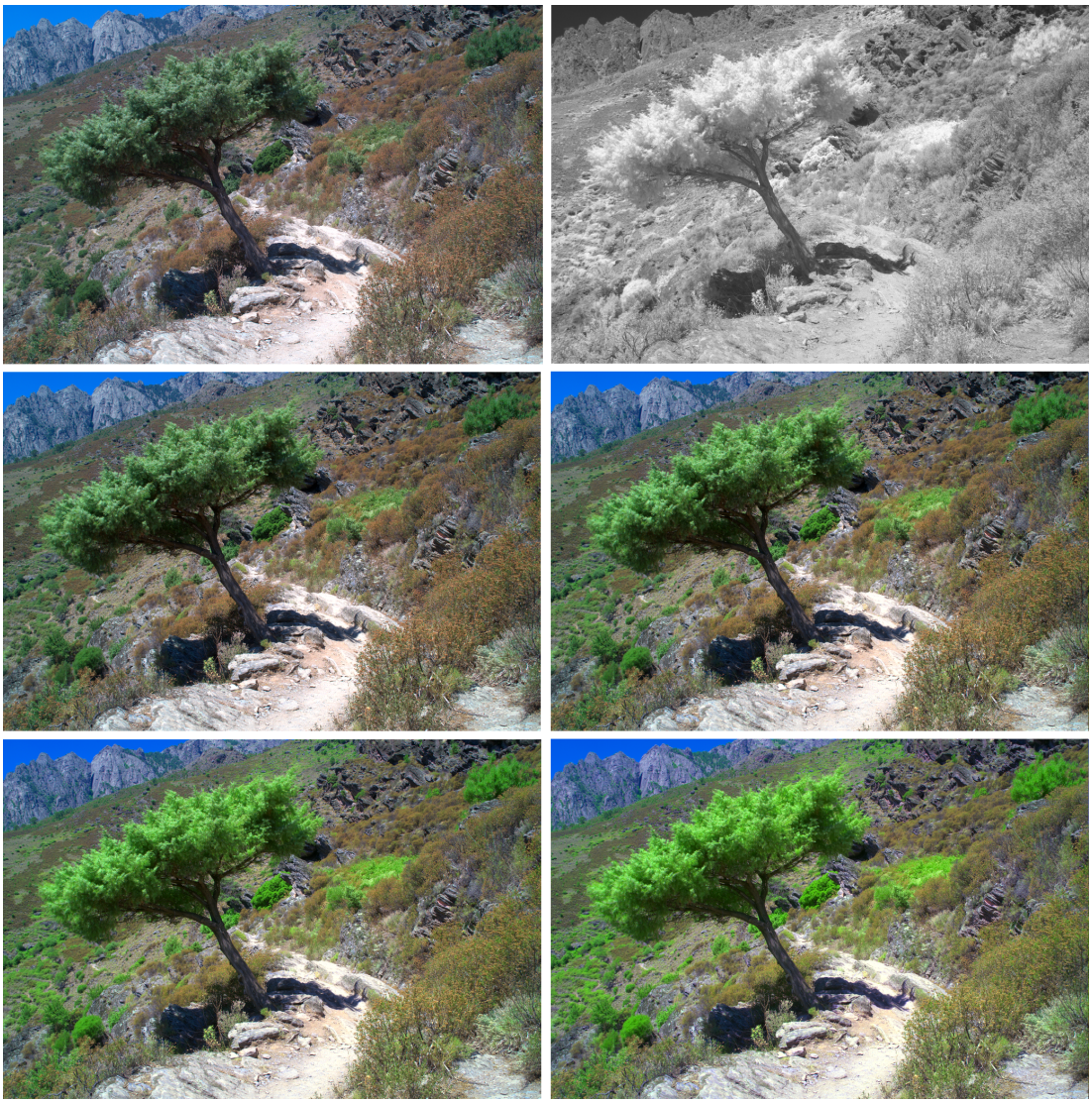


Figure 4.7: Spectral Edge image fusion (iterative): RGB-NIR Image Fusion - ‘Country08’ comparison (top row: RGB, NIR. Middle row: SE, SE-2. Bottom row: SE-4, SE-8)

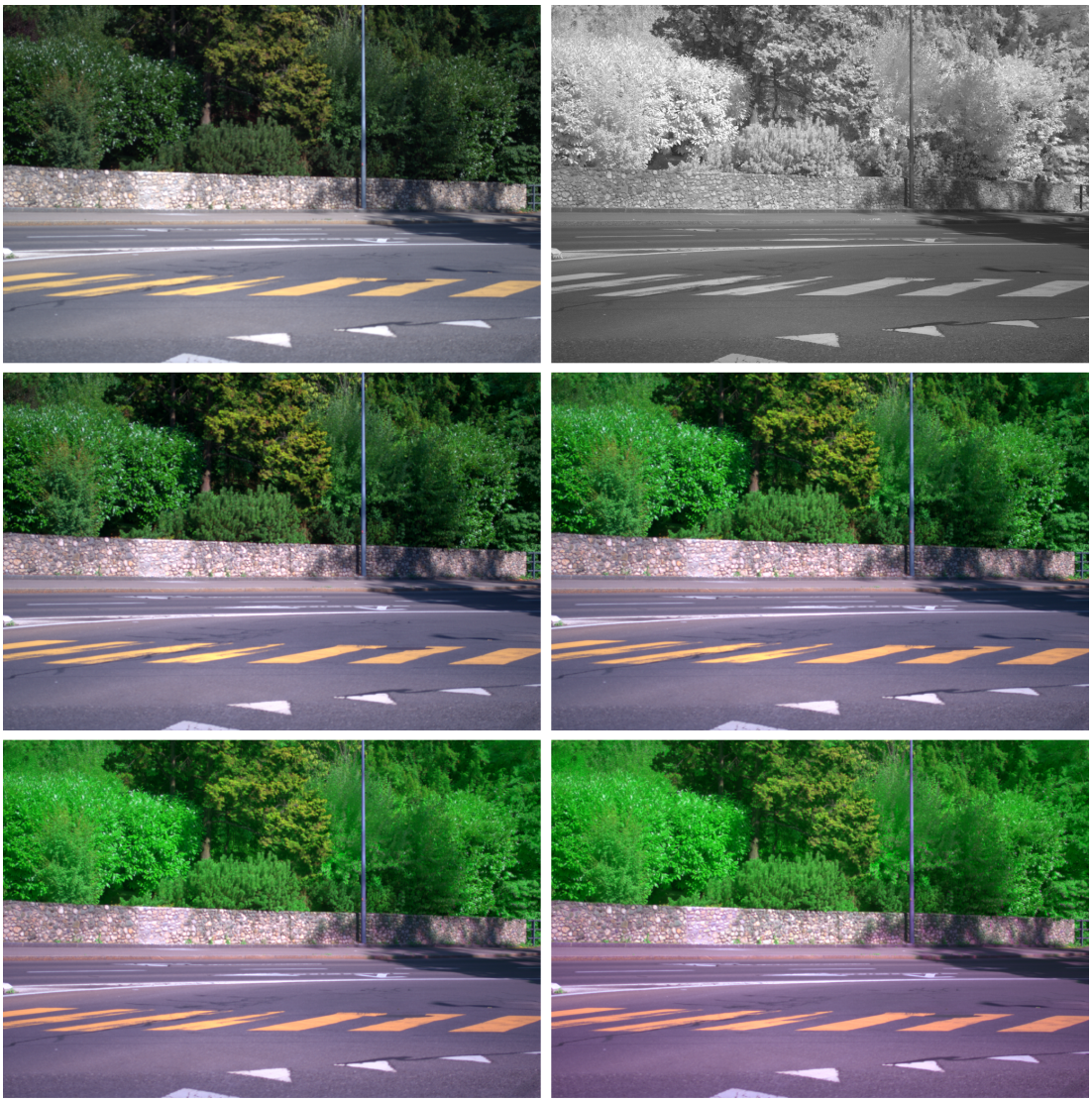


Figure 4.8: Spectral Edge image fusion (iterative): RGB-NIR Image Fusion - 'Street42' comparison (top row: RGB, NIR. Middle row: SE, SE-2. Bottom row: SE-4, SE-8)

4.5 New Applications of Spectral Edge Image Fusion

4.5.1 Image Fusion Using An RGB-NIR Bayer pattern

We have implemented Spectral Edge image fusion using raw sensor data, captured from an Omnivision OV4682 sensor [4], as shown in figure 4.9. This 4–megapixel sensor has a modified Bayer pattern, with one of the green pixels in each 2 x 2 region replaced with a near–infrared pixel (creating the pattern:

$$\begin{bmatrix} R & G \\ NIR & B \end{bmatrix} \tag{4.26}$$

This sensor allows the acquisition of perfectly-registered RGB and NIR image data – previous RGB–NIR image fusion research has used images captured using a standard camera with the hot mirror removed, and different filters placed in front of the camera (the largest data set of this kind is the EPFL RGB–NIR data set [8]). Taking non-simultaneous RGB and NIR images has the problem of objects and the camera moving between the separate acquisitions, resulting in problems with registration, leading to artifacts in the fused images. A single RGB–NIR sensor avoids these problems. However, it poses additional challenges.

The Omnivision sensor only provides RAW sensor data or an output RGB image, so to perform image fusion we created our own custom image pipeline. We first created a demosaicing algorithm based on Pixel Grouping[54] (one of the demosaicing methods available in the open source raw image reader 'dcraw'), but customized for the different RGB–IR Bayer pattern.

We took images of an X–rite ColorChecker Digital SG (140 colour patches) with the OV4682 sensor at different exposure levels, and then acquired rendered sRGB images of

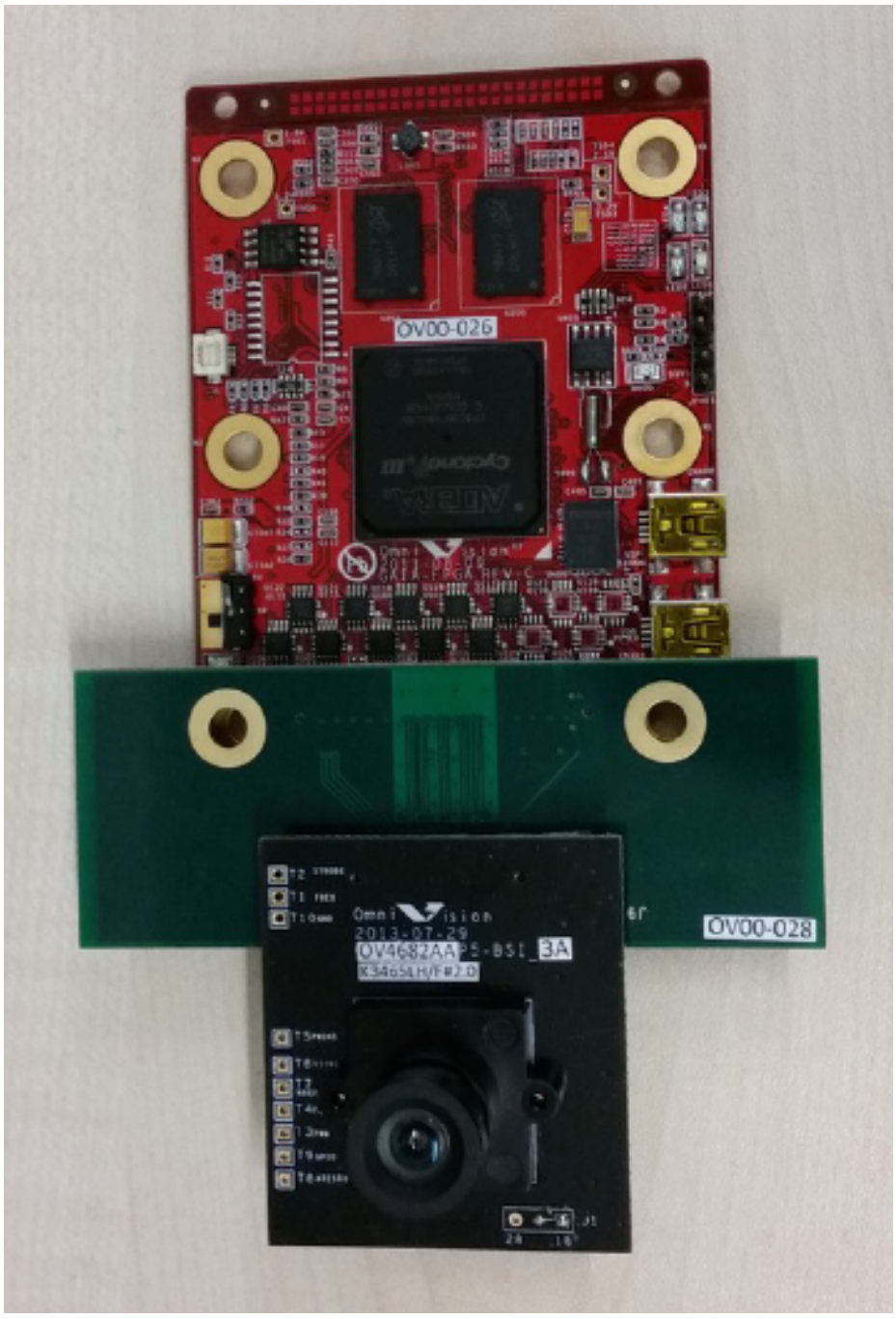
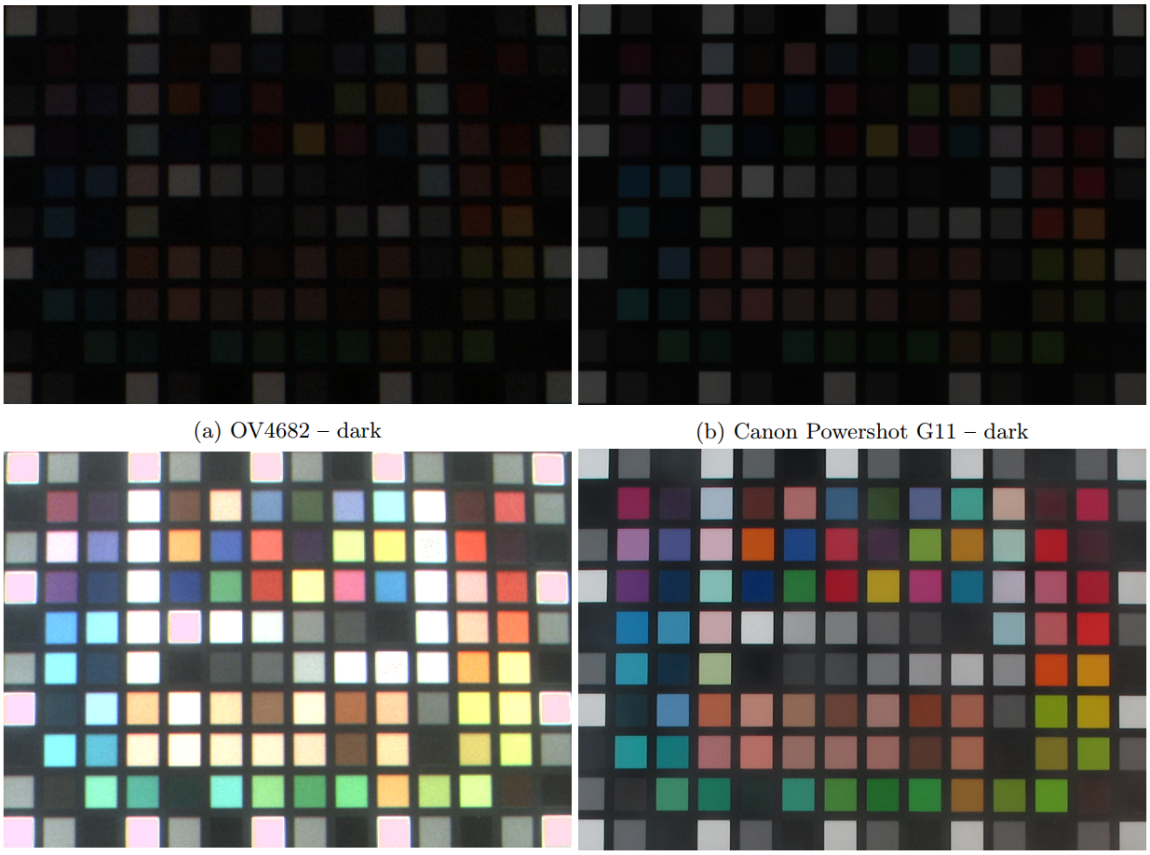


Figure 4.9: Spectral Edge image fusion (new applications): Omnivision OV4682 sensor



(a) OV4682 – dark

(b) Canon Powershot G11 – dark

(c) OV4682 – bright

(d) Canon Powershot G11 – bright

Figure 4.10: Spectral Edge image fusion (new applications): X–rite ColorChecker Digital SG – colour correction images

the same scene using a Canon Powershot G11 camera, two examples of which are shown in figure 4.10 to serve as a reference for correct colour sRGB rendering. We registered these images, and used them to create a custom colour correction matrix which simulates the processing done inside the Canon camera. We recognise that this procedure was somewhat arbitrary, but it is a fast way to simulate the effects of a full image signal processor (ISP).

White balance is an important element of an image processing pipeline, which has been the subject of much research. The key difficulty of white balancing an image is estimating the illuminant colour, which is then used to create a matrix to correct the image

colours. Two of the most well–known algorithms for white balance are the Max–RGB [46] and Grey–World [9] algorithms. In Max–RGB, the maximum RGB values are assumed to correspond to the scene white, which is used to derive the estimated illuminant. In Grey–World, the average RGB values (which should represent grey), provide the estimated illuminant.

In our pipeline, we used the Shades of Grey algorithm [32], which combines the Max–RGB and Grey–World algorithms to find an optimal mid–point between the two extremes. Another possibility would be to use the NIR to help illuminant estimation, as in [35], but that is left to future work.

Finally, we create a colour RGB image, which uses only the visible spectrum information, and a greyscale NIR image, which only uses the near–infrared sensor data. Once we form full–resolution RGB and NIR images, we then apply the Spectral Edge image fusion algorithm to fuse them, and produce a new RGB image with additional detail and superior image quality.

We understand that in real digital camera pipelines the RGB and NIR images may be combined at an earlier stage, in order to combine unprocessed and therefore physically predictable image information, but we chose to fuse them at the end for the sake of demonstrating the fusion process clearly.

Figures 4.11 and 4.12 shows example outputs of our image pipeline. The RGB image (a) is constructed using only visible spectrum information, and can be considered an approximation of the image a typical camera would produce of the scene, while the NIR image (b) only uses the near–infrared intensity to construct the image. The central bush in figure 4.11 appears dark and lacking in detail in the RGB image, but additional details

Figure 4.11: Spectral Edge image fusion (new applications): image fusion using an RGB–IR Bayer pattern - Cambridge street scene 1 (top to bottom: RGB, NIR, SE fusion)

Figure 4.12: Spectral Edge image fusion (new applications): image fusion using an RGB–IR Bayer pattern - Cambridge street scene 2 (top to bottom: RGB, NIR, SE fusion)

are visible in the near–infrared – the chlorophyll present in vegetation has a far higher reflectance in the near–infrared than in the visible spectrum. A similar effect is visible in figure 4.12. The SE fusion result (c) is superior in both cases to the original RGB image, as the near–infrared details are transferred, while maintaining natural colours. These examples show quite a typical image scenario in which SE fusion can dramatically improve image quality.

We can easily see this being used to improve images from digital cameras which have RGB–NIR capability, which have been previously primarily used for machine vision applications.

Further work could be done in this area to create a more sophisticated image signal processor (ISP) for RGB-NIR sensors, and to integrate image fusion earlier in the image pipeline.

4.5.2 RGB–thermal Image Fusion Using the FLIR ONE

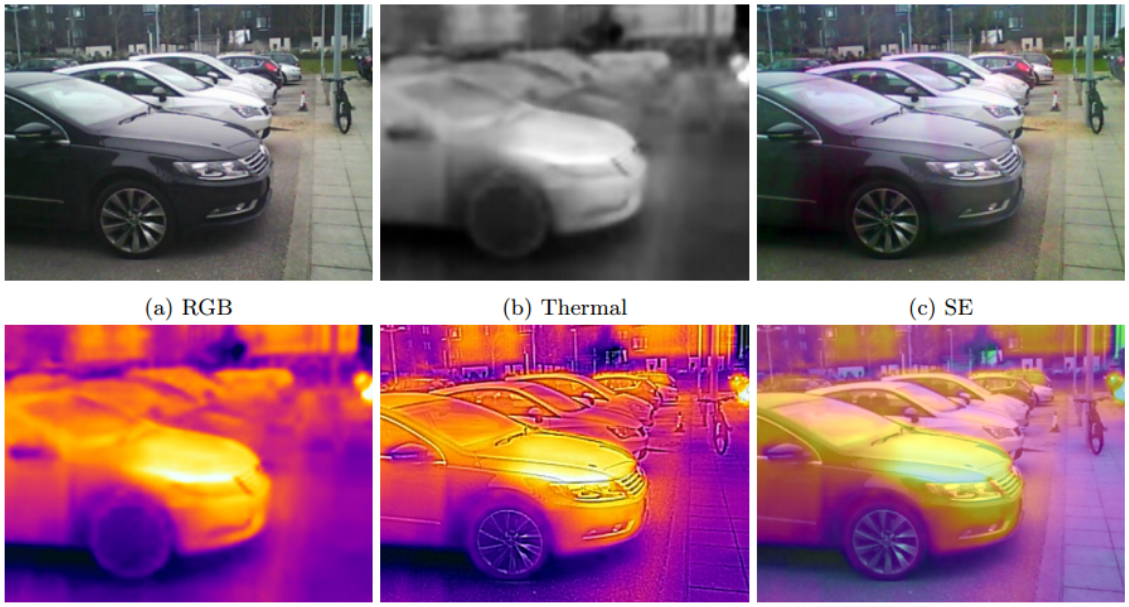
The FLIR ONE is a thermal camera accessory for smartphones, with 160 x 120 thermal resolution[1]. It has both visible RGB and thermal cameras, and is capable of exporting both modalities separately as well as fusing them with its own patented method [84].

We used the FLIR ONE to capture visible and thermal images, and then applied the Spectral Edge algorithm to produce a colour output image. For this application we used the iterative Spectral Edge variant from section 4.4, which produces stronger results. In the FLIR fusion patent, they assert that standard fusion methods such as SE are not preferred because "results are generally difficult to interpret and can be confusing to a user since temperature data from the IR image, displayed as different colours from a palette or different greyscale levels, are blended with colour data of the visual image", but we show here that

the results of combining visible colours and thermal detail can interestingly portray thermal details with natural colours. As an alternative fusion result, one more similar to the MSX technology used by FLIR, we take the false colour from the thermal image, and use this as the colour input for SE fusion, with the luminance channel of the RGB image used as an additional detail input.

In figures 4.13, 4.14, and 4.15, we show three example scenes. In each scene, we show the RGB image taken by the FLIR ONE visible spectrum camera in (a), the greyscale thermal image in (b), and the SE fusion result, the RGB image enhanced with the thermal image information in (c). We then show the false color thermal image in (d), the fused false colour image produced by FLIR MSX technology in (e), and our alternative fused false color image in (f), with the false colour thermal image used as our RGB input, and the visible spectrum image used to enhance its detail.

The first result, figure 4.13, shows a scene of several parked cars. The nearest car is considerably warmer than the other cars, perhaps having been recently used, and this heat is transferred into the natural colour SE fusion result (c) as extra brightness compared to the original. There is a chromatic artefact caused by brightening very dark pixels, which makes the car appear purple - this could be solved by desaturating dark areas before performing image fusion, so they would become grey instead of creating artificial colours. The water cooler in figure 4.14 shows high thermal readings in the center of the cooler, due to the heat of the cooling mechanism. This heat is effectively shown in the natural colour fusion result (c), as a warm glow. The third scene is a night scene, with a boat full of rowers hidden in the darkness in the visible image, but their body heat is visible in the thermal image. The natural colour fusion result shown in figure 4.15c shows somewhat unnatural colours, due to the extremely dark visible RGB image, lacking colour information, but nevertheless



(a) RGB (b) Thermal (c) SE

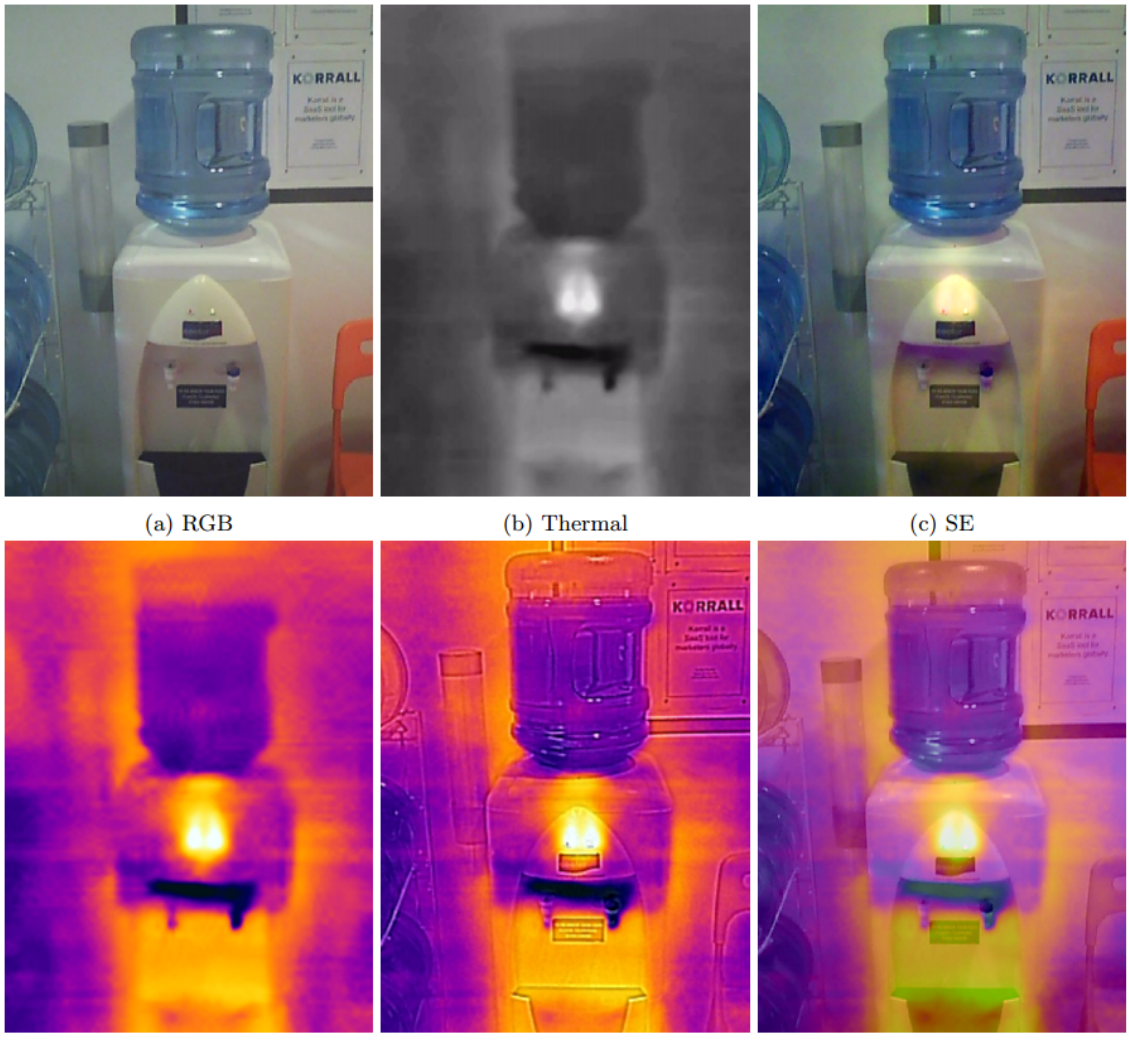
(d) Thermal (false color) (e) FLIR MSX fusion (f) SE (false color)

Figure 4.13: Spectral Edge image fusion (new applications): RGB-thermal image fusion using the FLIR ONE - scene 1 (cars)

effectively transfers the thermal detail of the rowers in the center of the image.

The false colour SE fusion results in (f) of each figure transfer virtually all RGB details while keeping the false color intact. The details are more natural and subtle than the FLIR MSX fusion results of (e), which appear to use direct edge transfer and possible edge sharpening, in comparison with the milder lookup–table–based gradient reintegration used in the SE fusion method. Each of the two methods has their merits, and a judgment of the preferred method would have to be made depending on the specific application.

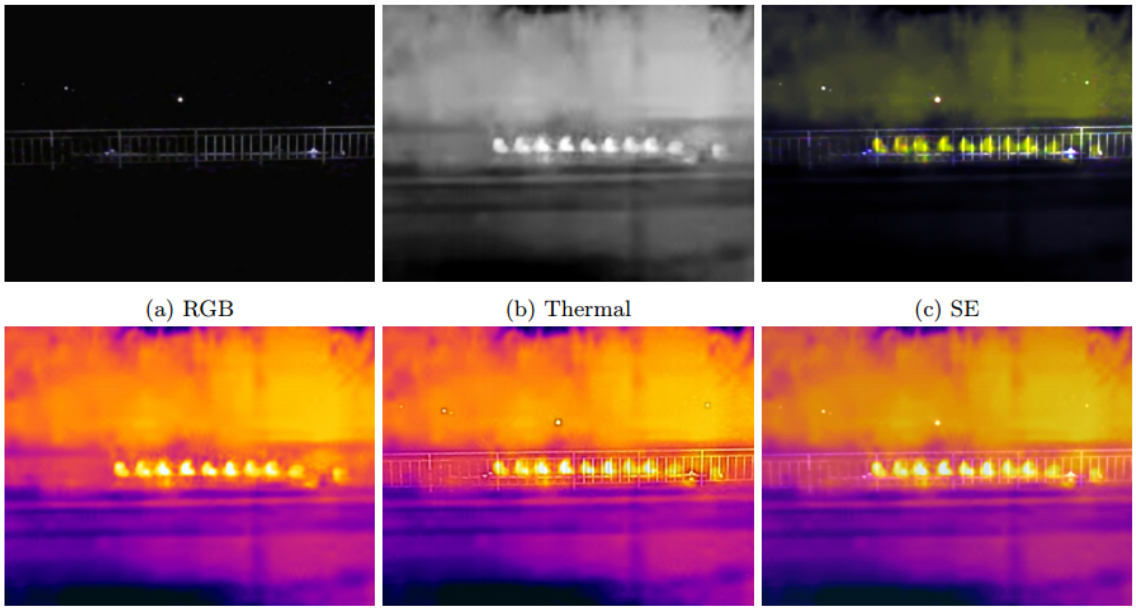
The RGB–thermal fusion shown in (c) of these figures could be integrated into a security camera for a surveillance application. A single fused image could simultaneously give a human observer both visible and thermal details, possibly requiring less attention and leading to faster object or person detection. The false color fusion shown in (f) of these figures may be a possible alternative to the current FLIR MSX fusion method used in the



(a) RGB (b) Thermal (c) SE

(d) Thermal (false color) (e) FLIR MSX fusion (f) SE (false color)

Figure 4.14: Spectral Edge image fusion (new applications): RGB–thermal image fusion using the FLIR ONE - scene 2 (water cooler)



(a) RGB (b) Thermal (c) SE

(d) Thermal (false color) (e) FLIR MSX fusion (f) SE (false color)

Figure 4.15: Spectral Edge image fusion (new applications): RGB–thermal image fusion using the FLIR ONE - scene 3 (rowers at night)

FLIR ONE.

Chapter 5

Local Linear Combination Image Fusion

In this chapter, local linear combination (LLC) image fusion is introduced. First, a generalized derivative domain image fusion scheme based on local linear combination coefficients is described, then two instantiations of the model: the first calculates LLC coefficients using the principal characteristic vector of the outer product (POP) of the Jacobian matrix of derivatives, and the second finds a least-norm regression from input derivatives to Spectral Edge (SE) derivatives (with regularization). Finally, optimization steps and experiments are conducted with the image fusion model.

5.1 Local Linear Combination Image Fusion Model

Unlike previous image fusion methods, which use complex minimizations and nonlinear optimization [81] to produce good results, but are computationally cumbersome, or previous gradient reintegration methods which produce artifacts (see chapter 3), we aim to propose a simple but effective framework for the problem of image fusion to produce state-of-the-art results at high speed.

We reduce the problem to finding a smoothly-varying local linear mapping from N to M image channels (dimensions), in other words a local linear combination of the input

channels with different coefficients for each pixel. We can formulate this in a similar way to that of Ma *et al.* [58],

$$O(\mathbf{x}) = \sum_{n=1}^{N} \mathcal{P}_n(\mathbf{x}) H_n(\mathbf{x}) \tag{5.1}$$

Where O is the fused image, N is the number of input images, $H_n(\mathbf{x})$ is the input image intensity value at the $\mathbf{x}$-th pixel in the n-th input image and $\mathcal{P}_n(\mathbf{x})$ the weight at the $\mathbf{x}$-th pixel in the n-th input image. This formulation will do for N to 1 channel image fusion, for which at each pixel location there will be an $N \times 1$ vector of weights - for N to M channel fusion the set of weights will become an $N \times M$ matrix at each pixel location.

This formulation has two benefits: firstly, providing we maintain a smooth mapping, we guarantee that no new artefacts will be introduced to the output fused image (if artefacts or sensor noise are present in the input images they may be transferred) - secondly, the model's simplicity leads to large performance benefits compared to many other image fusion methods. The process of applying the coefficients is a simple per-pixel dot product, while the calculation of coefficients may be reduced in complexity by operating on a thumbnail resolution and upsampling (see section 5.3).

We propose two methods for calculating these weights, for two different image fusion tasks. For N to 1 channel image fusion, we propose using the Principal characteristic vector of the Outer Product (POP) of the image Jacobian to calculate a projection which provides our local linear combination coefficients $\mathcal{P}$, and for N to 3 channel image fusion, we propose using the Spectral Edge (SE) image fusion theorem to calculate target gradients at each pixel, and then using a least-norm regression from the input gradients (with regularization) to find the local linear combination coefficients $\mathcal{P}$ that we require.

Our two fusion variants can be thought of as instances of a general gradient-based local

image fusion model:

1. For all input image locations $\mathbf{x}$ calculate the Jacobian $J^{\mathbf{x}}$.
2. Calculate per-pixel linear combination coefficients $\mathcal{P}(\mathbf{x})$ using $J^{\mathbf{x}}$.
3. *diffuse*$(\mathcal{P}(\mathbf{x}))$.
4. Apply $\mathcal{P}(\mathbf{x})$ to produce output image.

The first step of calculating the input image derivatives, and the last step of applying the linear combination coefficients to produce an output image, are always the same. The POP variant and the SE variant manifest step 2 in different ways. Whereas the POP variant uses the POP image fusion theorem to calculate an N to 1 set of linear combination coefficients based on the projection of the input derivatives which matches the Socolinsky and Wolff equivalent gradient, regularized by assuming the projection vector must be positive, the SE variant first calculates target RGB gradients using the Spectral Edge image fusion theorem, and then finds an N to 3 linear combination using a least norm regression, regularized by the constraint that the output image should not diverge too far from the input RGB image. Step 3, the diffusion of linear combination coefficients, can be implemented in a variety of ways, as explained later in this chapter.

Our image fusion method makes assumptions common to most image fusion methods: firstly, we assume that the input images are well registered, and secondly that they contain no significant noise or artifacts - any noise or artifacts present in the input images will be detected as gradient information and transferred into the fused output image.

5.2 Calculating Linear Combination Coefficients

5.2.1 POP variant

The derived gradient in the Socolinsky and Wolff (SW) method is a mathematically well founded fusion of the all the available gradient information from a multichannel image. However, reintegrating these gradients is an unsolved problem, which leads to hallucinated artefacts in the output fused images. The POP variant of our method avoids these problems by avoiding gradient reintegration entirely - instead we find a local linear combination of the input images which has the desired gradients.

The basic premise of the POP variant of our method is that we can find a per-pixel linear combination of the input channels such that if we differentiated the output image we would generate the equivalent gradients as found by SW, by finding a projection direction from the principal characteristic vector of the outer product of the Jacobian.

We begin by defining $\acute{U}^{\mathbf{x}}$ as the principal (i.e. first) characteristic vector of the outer product of the Jacobian matrix of derivatives (equation 3.1) at location x, where the superscript x is shorthand for a certain x, y pixel location.

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 $[\frac{\delta}{\delta_x}(P(\mathbf{x}))\ \frac{\delta}{\delta_y}(P(\mathbf{x}))]^T = s^{\mathbf{x}} G(\mathbf{x})$ (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,

$$\begin{aligned} \tfrac{\delta}{\delta_x}(P(\mathbf{x})) &= \textstyle\sum_{k=1}^{N} \acute{U}_k^{\mathbf{x}} \tfrac{\delta}{\delta_x}(I_k(\mathbf{x})) \\ \tfrac{\delta}{\delta_y}(P(\mathbf{x})) &= \textstyle\sum_{k=1}^{N} \acute{U}_k^{\mathbf{x}} \tfrac{\delta}{\delta_y}(I_k(\mathbf{x})) \end{aligned} \tag{5.2}$$

Remembering that $U^{\mathbf{x}}$ is part of the singular value decomposition of the Jacobian - see Equation 3.7 - 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

$$[\frac{\delta}{\delta_x}(P(\mathbf{x}))\ \frac{\delta}{\delta_y}(P(\mathbf{x}))] = [S^{\mathbf{x}}_{11}V^{\mathbf{x}}_1] \tag{5.3}$$

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 [83] heuristic method. Rather we choose the sign so that the projected image is positive (a necessary property of any fused image):

$$s^{\mathbf{x}} = sign(\acute{U}^{\mathbf{x}}.I(\mathbf{x})) \tag{5.4}$$

Equation 5.4 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

$$U^{\mathbf{x}} = J^{\mathbf{x}} V^{\mathbf{x}} [S^{\mathbf{x}}]^{-1} \tag{5.5}$$

that is, $U^{\mathbf{x}}$ is the product of the Jacobian and the square root of the Di Zenzo structure tensor. 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.

If we divide the image plane into 2×2 pixel sections, and calculate a single projection for each section at the upper-left pixel, the output fused gradients at that pixel will equal the Socolinsky and Wolff equivalent gradients exactly.

From this we can imagine we can sample $\mathcal{P}$ at every second pixel exactly, which means we could interpolate between these sampled coefficients and reconstruct a band limited version of the optimal set of $\mathcal{P}$.

However, we wish to extract the maximum frequency of projection information. Also we must deal with image areas with discontinuities or which lack meaningful gradient information, therefore we use the coefficient diffusion set out in the previous section to combine projection information between neighboring similar pixels.

Aside: There is another side effect of coefficient diffusion, which can eliminate a potential mathematical problem. Previously, we proved that the derivative of the output image

$$O = \mathcal{P} I \tag{5.6}$$

equals the SW gradient for a single pixel. However, when dealing with a continuous surface, the Product Rule must be used, giving

$$\partial_{\mathbf{x}} O = (\partial_{\mathbf{x}} \mathcal{P}) I + \mathcal{P}(\partial_{\mathbf{x}} I) \tag{5.7}$$

This means that the derivative of O will not match exactly the desired derivative, because of the extra term $(\partial_{\mathbf{x}} \mathcal{P}) I$ created from the Product Rule.

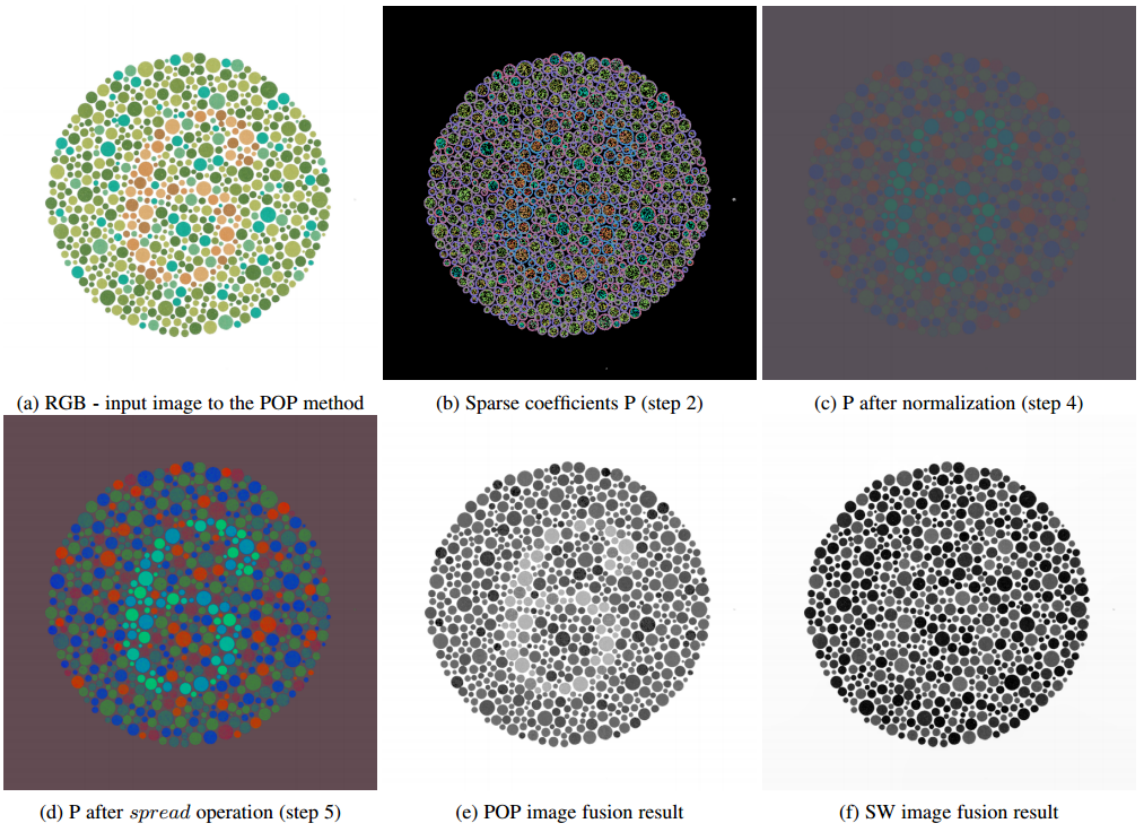
We can think of coefficient diffusion as (in part) eliminating this extra and unwanted result, through combining and averaging results from a large number of pixels.

Algorithm for calculating POP variant coefficients

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

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

The POP variant of our method is an instance of the general scheme defined in section 3. It has an extra conditional element in step 2, by which only coefficients in image locations with significant gradients and projection structure are used (typically we set θ_1 to 0.01, and θ_2 to 0.8 - these values have been chosen by experimentation), and has two extra steps: after the bilateral filtering, $\mathcal{P}$ is dense, but each linear combination coefficient vector is not a unit vector. This is remedied by normalizing the length of each vector. Finally, we



(a) RGB - input image to the POP method (b) Sparse coefficients P (step 2) (c) P after normalization (step 4)

(d) P after *spread* operation (step 5) (e) POP image fusion result (f) SW image fusion result

Figure 5.1: Local linear combination image fusion: illustration of the POP variant of our method - in (a) we show an Ishihara plate. The initial linear combination coefficient image derived in the POP variant of our method 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 coefficients 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.

apply a spreading function $spread()$ to move each of the coefficient vectors a fixed multiple of angular degrees away from the mean (the diffusion step pulls in the opposite direction and results in coefficient vectors 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.

Fig. 5.1 shows an example of this coefficient calculation process - in 5.1b-d the 3-dimensional coefficients are visualised by assigning their values to the corresponding RGB channels. In this example, the previous method of Socolinsky and Wolff fails to create an output greyscale image of this Ishihara plate with a readable number, because its Poisson equation-based gradient reintegration does not consider details which are not immediately adjacent to each other - our proposed method solves this problem thanks to its large-scale coefficient diffusion.

Global POP Variant

Instead of the local linear combination previously explained, the POP image fusion theorem may also be used to implement a global image fusion scheme. We project the input image Jacobian J by the sign-normalized first characteristic vector U at each pixel,

$$G(\mathbf{x}) = \acute{U}^{\mathbf{x}}.J(\mathbf{x}) \tag{5.8}$$

This creates a set of target gradients G without the sign problems inherent in previous structure tensor methods. From these gradients we form the target Laplacian (∇O) from their second derivatives - from here we could solve the Poisson equation to find an output image, but instead we use the LUT-based reintegration method from Finlayson *et al.* [31]. To do this, we find a least-squares regression from a polynomial function (the $poly$ function) of the Laplacian of the input channels (∇I) to the target Laplacians,

$$Z = (poly(\nabla I)^T poly(\nabla I))^{-1} poly(\nabla I)^T \nabla O \tag{5.9}$$

This set of weights is equivalent to a lookup-table, and can be applied to a polynomial function of the input images to produce an output fused result ($O = poly(I) * Z$).

This has the advantage of guaranteeing a lack of artifacts, and dramatically increasing the algorithm's efficiency.

5.2.2 SE variant

In the Spectral Edge (SE) variant of our proposed method, we use the SE image fusion theorem to calculate 3-dimensional target derivatives, and use these to calculate an N x 3 matrix of linear combination coefficients, where N is the number of input image channels. The principle of our method is to find the per-pixel linear combination of the input derivatives which best matches the target derivatives at each pixel, and then apply this to the input image pixels to form an output image. In this section we assume we are working at a single pixel location $\mathbf{x}$, but the minimizations will be calculated $\forall\, \mathbf{x} \in I$, where I is the image plane.

Algorithm for calculating SE variant coefficients

1. For all input image locations $\mathbf{x}$ calculate the Jacobian $J^{\mathbf{x}}$.
2. Calculate target Jacobian $\hat{J}^{\mathbf{x}}$ using Spectral Edge image fusion theorem (see section 4.1).
3. Calculate per-pixel linear combination coefficients $\mathcal{P}(\mathbf{x})$ to match target derivatives (eq. 5.10).
4. $\mathcal{P}(\mathbf{x}) = \textit{diffuse}(\mathcal{P}(\mathbf{x}))$.

For the three-dimensional RGB gradients provided by the SE theorem and four input image channels (e.g. RGB-NIR), $\mathcal{P}$ becomes a 4 x 3 projection matrix (this can also be thought of as three vectors of length 4, 3 simultaneous regressions of the previous type), and the target is the $M^2 * 2$ x 3 matrix of gradients ∇R in a window of width M around the current pixel, calculated from the SE theorem. We also add regularization to the minimization, constraining the solution so that the projection matrix $\mathcal{P}$, when applied to the original input image channels I, should produce pixel values close to the RGB values of the guide RGB image $\tilde{R}$. The parameter λ controls the strength of the regularization, our default value is $||\hat{J}|| * 10^1$.

$$\arg\min_{\mathcal{P}} ||(\nabla(\hat{J}^T\mathcal{P}) - \nabla R) + \lambda(I^T\mathcal{P} - \tilde{R})|| \tag{5.10}$$

With this regularization (assuming a nonzero λ value), in smooth image regions with zero derivatives, the chosen projection coefficients will produce an output image equal to the input guide image.

After this regression is calculated at each pixel (or on a thumbnail as explained in section 3.2), the 12 values (in the case of 4 to 3 channel image fusion) of $\mathcal{P}$ are known at every location. The *diffuse*() function (in our implementation using a cross bilateral filter) is then applied to ensure a continuous linear mapping across the image plane. Unlike the POP variant, normalization and spreading of the coefficients are not required, due to the greater degree of stability caused by calculating the regression in a window and regularization. The output image is a simple per-pixel matrix multiplication:

$$O(\mathbf{x}) = I(\mathbf{x}) \,.\, \mathcal{P}(\mathbf{x}) \tag{5.11}$$

Fig. 5.2 shows the entire image fusion process of the SE variant of our method. Gradients are calculated from the high-dimensional and guide images, and from these target SE gradients are calculated. Using a per-pixel least-norm regression, linear combination coefficients are found which best map the input gradients to the target SE gradients (with regularization). The upper-right output image is that produced the previous global lookup-table-based image fusion reintegration method (see section 4.2), and the lower-right output image is the fused image of the proposed method.

Structure tensor error is a meaningful measure of detail transfer in derivative domain image fusion, measured as the L2 norm of the difference between the structure tensor of the high-dimensional input and the structure tensor of the output image, summed across the image, normalized by the L2 norm of the high-dimensional tensor,

$$\sum_{x \in X} \sum_{y \in Y} ||(Z_H(x,y) - Z_{R_i}(x,y) + \Omega)||_2 \; / \; (||Z_H(x,y)||_2 + \Omega) \tag{5.12}$$

Where X and Y are the sets of possible x and y coordinates in the image, and Ω is set to 0.01 to stabilize the division.

Table 5.1 shows the mean structure tensor errors for the original RGB images, the previous SE method, and the new local reintegration applied to the SE method, averaged over 10 images from the EPFL RGB-NIR data set [8]. Our method has a lower error, meaning it transfers detail more effectively than the previous SE method, while simultaneously being preferred by human perception, as shown in section 6.1.

RGB	SE	Proposed
0.7960	0.7784	0.7693

Table 5.1: Local linear combination image fusion: structure tensor error - image fusion detail transfer error averaged over 10 RGB-NIR image pairs.

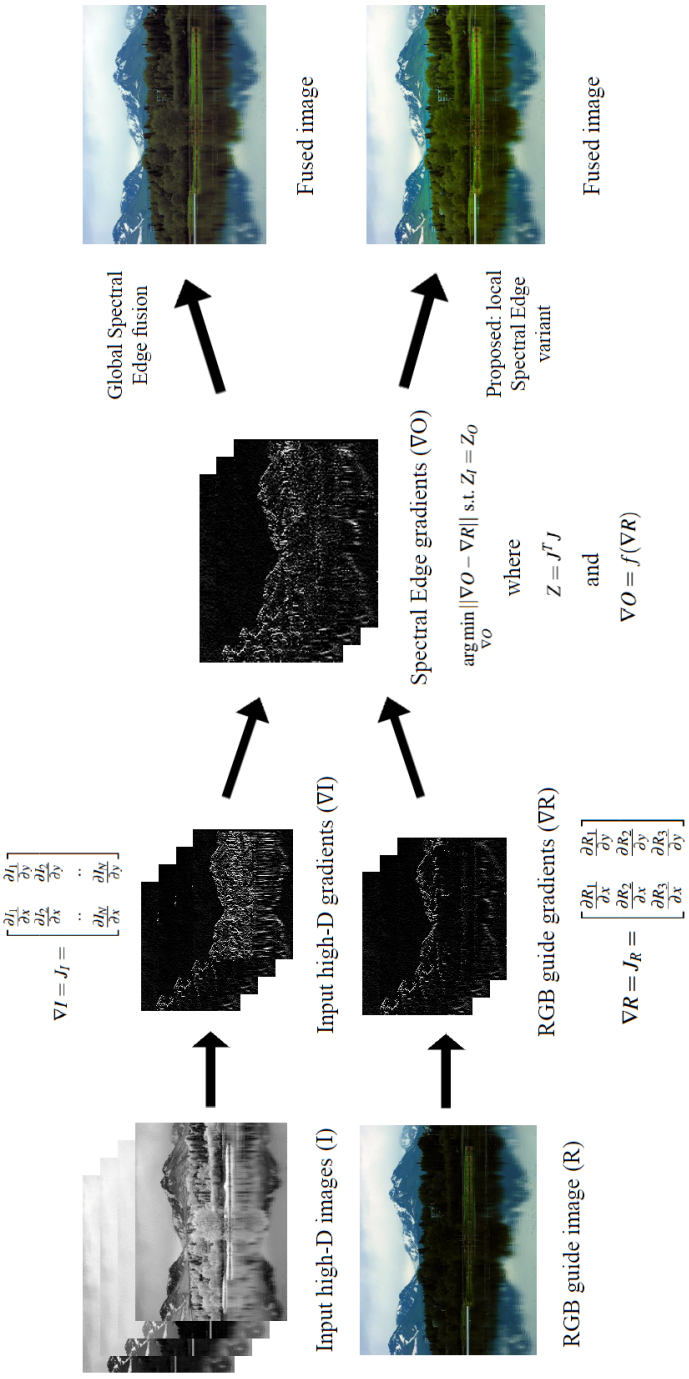


Figure 5.2: Local linear combination image fusion: Spectral Edge variant of the proposed method - flow diagram.

5.2.3 Diffusion of linear combination coefficients

In image regions with zero derivatives or where the image Jacobian has coincident eigenvalues (e.g. corners) there may be no meaningful gradient information, or a large change in the linear mapping direction found at one image location compared to another (discontinuity). We can generalize this to say that the coefficients will always have a certain level of error in real-world conditions. It follows then that we must interpolate, or diffuse, the linear combination coefficients that are well defined across the image. We can achieve this in a number of ways, as shown in fig. 5.3.

The image fusion task in figure 5.3 is colour to greyscale conversion, in this case Monet's 'Impression, soleil levant' - this image is often chosen to demonstrate colour to greyscale conversion, because the sunrise, easily visible in the colour image, has much lower contrast in standard luminance channel conversions such as 5.3b. We show the CIELAB luminance channel in figure 5.3b, and the Socolinsky and Wolff fused output in figure 5.3c. Figure 5.3d shows the POP variant of the LLC method without any coefficient diffusion - clearly there are artefacts and gaps in the projection, in regions with discontinuities or without meaningful gradient information. Figure 5.3e and figure 5.3f use Gaussian filtering to diffuse the projection coefficients, with a Gaussian kernel of size and standard deviation 20 and 160 respectively, while figure 5.3g and figure 5.3h use cross bilateral filtering for coefficient diffusion (the guide edge image is the corresponding input image channel) - the domain standard deviation is 20 and 160 pixels, and the range standard deviation is fixed at 0.4. We have also experimented with using guided filtering to diffuse our coefficients, as in [50], but we found this to work less effectively.

Table 5.2 shows the results of comparing these diffusion methods with the results of Socolinsky and Wolff's approach for colour to greyscale conversion [83]. The metric used

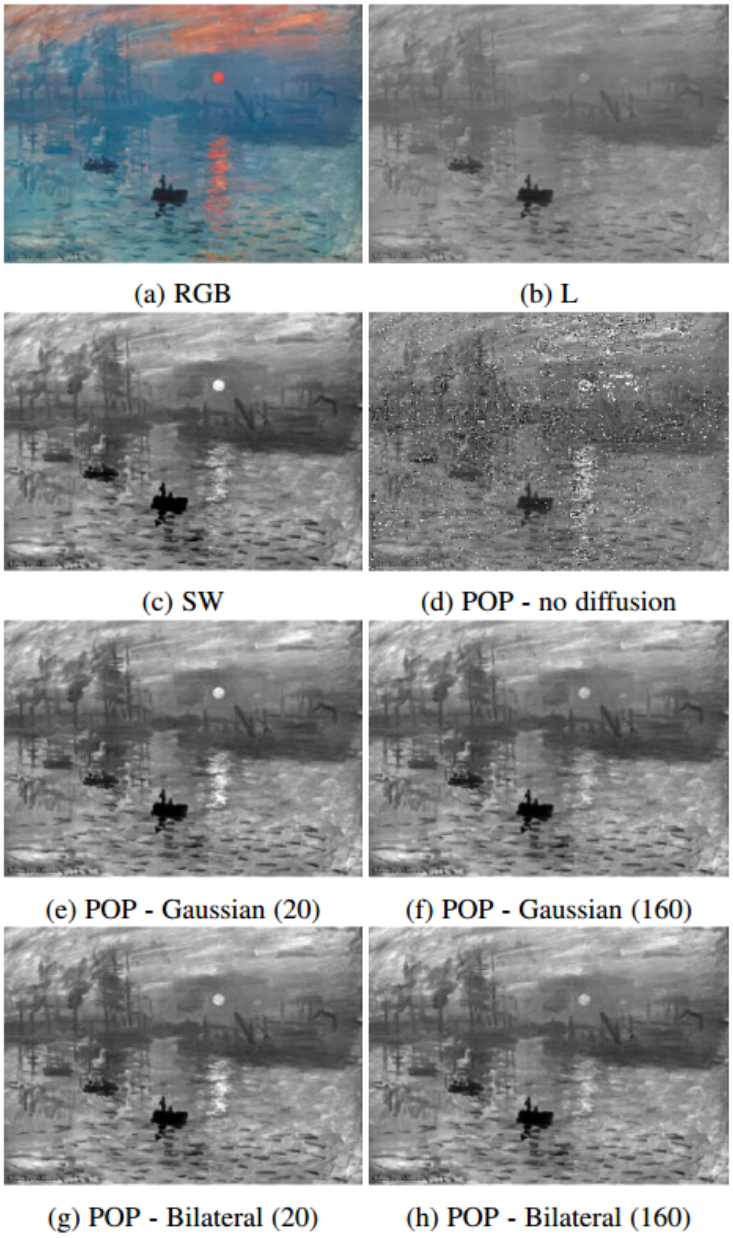


Figure 5.3: Local linear combination image fusion: comparison of coefficient diffusion methods - colour to greyscale conversion of 'Impression, soleil levant' image (a), LAB luminance (b), Socolinsky and Wolff result (c), POP without coefficient diffusion (d), (e-h) POP with various diffusion methods.

is structure tensor error - the norm of the difference between the RGB structure tensor and the output greyscale structure tensor is averaged across the image (therefore lower is better). The results are averaged across the 25 images of the Ĉadík data set [13]. With no diffusion, the error is higher than that of Socolinsky and Wolff, as there remain some pixels where the image does not contain meaningful gradient information, or discontinuities, so the projection information is incorrect or not present. The other errors are all lower than Socolinsky and Wolff, indicating that the output image's structure tensor (and therefore detail) is closer to the input colour image. A larger scale diffusion produces lower error, and bilateral filtering produces lower error than Gaussian filtering. Our assumption is that cross bilateral filtering works better for coefficient diffusion because it is sensitive to the input image structure - the coefficients are blended more between pixels with similar intensity values.

Method:	SW	No diffusion	Bilateral (20)
Error:	0.1392	0.1762	0.1051
Method:	Bilateral (160)	Gaussian (20)	Gaussian (160)
Error:	0.0865	0.1163	0.0940

Table 5.2: Local linear combination image fusion: structure tensor error - colour to greyscale conversion detail transfer error averaged over 25 RGB images from the Ĉadík data set[13].

Therefore we define a *diffuse*() function, which represents the diffusion of linear combination coefficients to infill missing values where edge information is not significant. In our default implementation this uses 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 $\sigma_d \to \infty$ and $\sigma_r \to \infty$, the diffusion becomes a global mean, and the linear combination tends to a global weighted

sum of the input channels. If $\sigma_d \to \infty$ and $\sigma_r = 0$ each distinct vector of values in the image will be associated with the same set of coefficients and so the bilateral filtering step defines surjective mapping which could be implemented as a look-up table [31]. Excepting these boundary cases the 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.

5.3 Optimization

To speed up the technique, the input images may be downsampled and $\mathcal{P}$ calculated only for the thumbnail image. In this case, a method must be chosen to upsample the resulting $\mathcal{P}$ values to provide a linear combination at every pixel of the full-size image plane. The cross bilateral filter used in the case of full-resolution linear combination coefficient images becomes joint bilateral upsampling, and is used on the thumbnail coefficient image to upsample it, using the corresponding input image channel as a full-resolution guide image [43]. The global version of the POP variant also works using a thumbnail version of the input images. The linear combination coefficient vectors, target gradients and Laplacians are first calculated at the small scale. The set of weights Z is calculated, and applied to a polynomial function of the full-resolution input images to produce the fusion result.

With this thumbnail implementation, the POP variant of our method becomes extremely fast. Using the example of color to greyscale conversion of one of the images from the Kodak data set, a 3 to 1 channel fusion problem on an image of 768 x 512 pixels, the global and local variants take 5.13s and 5.16s respectively at full resolution (using a MATLAB implementation). If we use thumbnails of a quarter resolution ($1/2$ in each dimension) this drops to 1.31s and 1.41s, and to 0.35s and 0.49s at a $1/16$ downsampling level ($1/4$

in each dimension). Thumbnails of an even smaller size can be used, with corresponding performance gains, while maintaining a high-quality output image. We remark that this thumbnail computation also has the advantage that the coefficient image can be computed in tiles i.e. we never need to calculate the full resolution coefficient image.

5.4 Experiments

5.4.1 POP variant

Figures 3.6 and 3.7, shown in chapter 3 of this thesis, show experiments using the POP variant of the LLC method on extreme image fusion cases that present problems to many existing image fusion techniques. Unlike the previous derivative domain method of Socolinsky and Wolff (SW), and the discrete wavelet transform (DWT) method, our proposed method produces output images which transfer all salient image details without artefacts.

The POP variant of the LLC method provides several parameters which may be tuned for optimal image fusion performance. The domain and range standard deviations of the cross bilateral filter used for coefficient diffusion can be varied. To test this, we used 10 images from the EPFL RGB-NIR data set, and fused the luminance channel of the RGB image with the NIR image. We then assessed the fusion quality using the standard 2 to 1 metrics Q_G (based on image gradients), Q_{MI} (based on mutual information) and Q_Y (based on SSIM) defined in section 4.3. The mean metric result over 10 images is presented in the following graphs.

Figure 5.4 shows the results of varying the domain (spatial) standard deviation, on a logarithmic scale, from 2^0 to 2^{14} pixels, with the range standard deviation fixed at 0.2. The performance improves with a greater domain standard deviation for the Q_{MI} metric, as well as the mean of the metric results (although the meaning of this mean metric result is

questionable), as the local linear mapping becomes smoother, and reaches its maximum at a standard deviation of 2^8 pixels - above this the graph is almost flat, and appears to be reaching an asymptote. However, the Q_G and Q_Y metrics show peak results at lower domain standard deviations, of 2^2 and 2^3 respectively. As the metrics are in disagreement on the optimal domain standard deviation, we use the larger result of 2^8 pixels in the rest of our experiments, as a larger domain standard deviation is more likely to guarantee fused outputs free of artefacts.

Figure 5.5 shows the results of varying the range standard deviation, on a logarithmic scale from $1/2^0$ to $1/2^{14}$ with the domain standard deviation fixed at 2^8 pixels. Again, the metrics disagree in the optimal standard deviation - Q_G sets it at $1/2$, Q_Y at $1/8$, and Q_{MI} at $1/2^8$. The mean metric result peaks at a range standard deviation of $1/4$, and we use this in our further experiments; again we err on the side of greater diffusion (with a higher standard deviation), as this is more likely to guarantee a lack of artefacts.

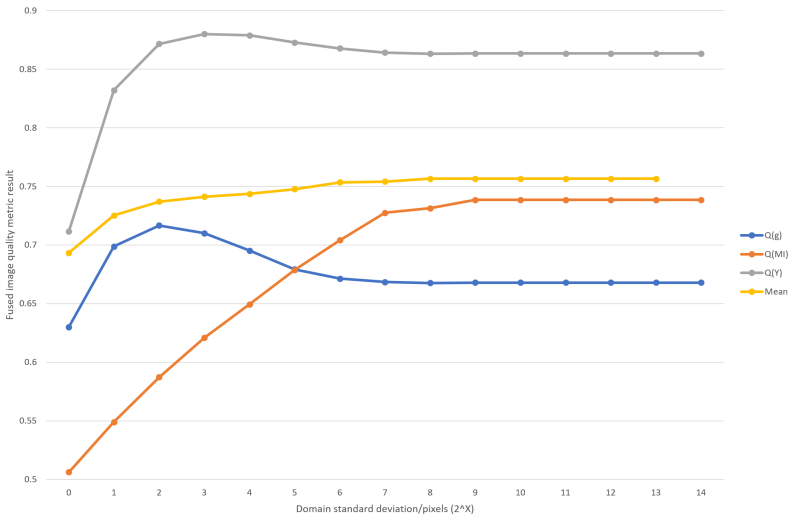


Figure 5.4: Local linear combination image fusion (POP variant): Q_G, Q_{MI}, Q_Y metrics, and mean of the three metrics - results with varying domain standard deviation.

5.4.2 SE variant

The SE variant of the LLC method provides several parameters which may be tuned for optimal image fusion performance, depending on the image fusion task. The window size M on which the least-squares regression is calculated can be varied, as well as the regularization parameter λ. The domain and range standard deviation of the cross bilateral filter used for coefficient diffusion can also be varied.

To tune these parameters, we tested the fusion results on RGB-NIR image fusion, using 10 images from the EPFL RGB-NIR data set [8]. The quality of the fusion results is calculated using the M to N channel Q_G metric defined in section 4.3 of this thesis.

First, we tested various values for the regularization parameter λ, with a fixed window size of 1 pixel, and found that increasing this value gives higher quality output images, and that at all values equal to or above 10^{-10}, the output is the same. The bilateral filter parameters were fixed, with a domain standard deviation of 40 pixels, and a range standard deviation of $0.1 * (max(I) - min(I))$.

Next, we tested different window sizes from 1 to 15 for the least-squares regression, with a fixed λ value of 10^{-10}, and the bilateral filter parameters fixed as in the previous test. Q_G metric results are averaged over the 10 images, and the peak result is at a window size of 9. However, all the results are within a narrow range, the difference between the highest and lowest results is just over 0.02.

The same procedure is repeated, with the variable this time being the domain standard deviation. Figure 5.6 shows the results - domain standard deviations of 8 and 16 had the joint best metric results, and the performance decreases with increasing standard deviation above this, although the difference is not dramatic. The graph for range standard deviation looks similar, with a peak metric result with a standard deviation of $0.25 * (max(I) -$

$min(I)$).

In this chapter, we have proposed a new framework for image fusion, based on a local linear combination of the input images. We have proposed two ways of calculating local linear combination weights, explained how the method can be optimized, and experimented with coefficient diffusion parameters.

In the next chapter, we will show the versatility of the proposed method, demonstrating state-of-the-art results on a wide variety of applications - most image fusion methods are designed specifically for a single application.

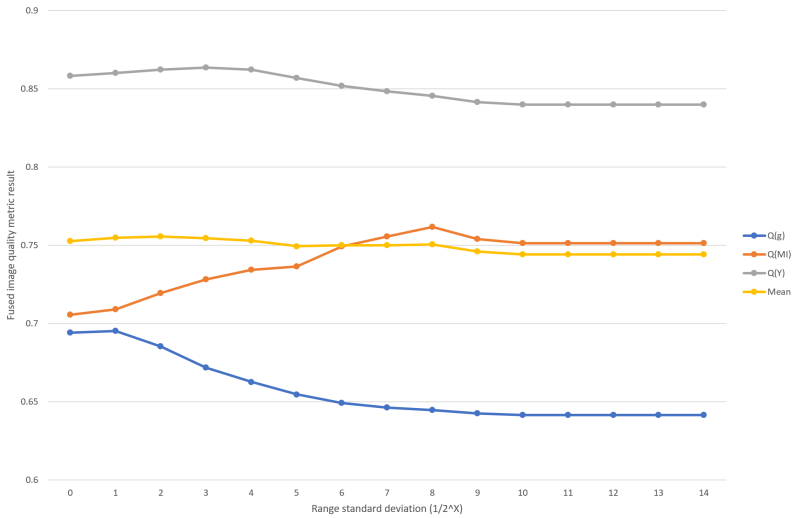


Figure 5.5: Local linear combination image fusion (POP variant): Q_G, Q_{MI}, Q_Y metrics, and mean of the three metrics - results with varying range standard deviation.

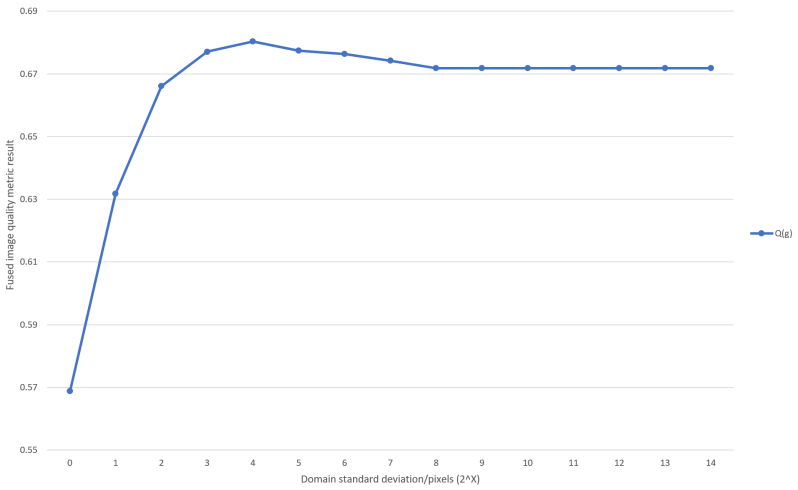


Figure 5.6: Local linear combination image fusion (SE variant): mean Q_G metric result with varying domain standard deviation.

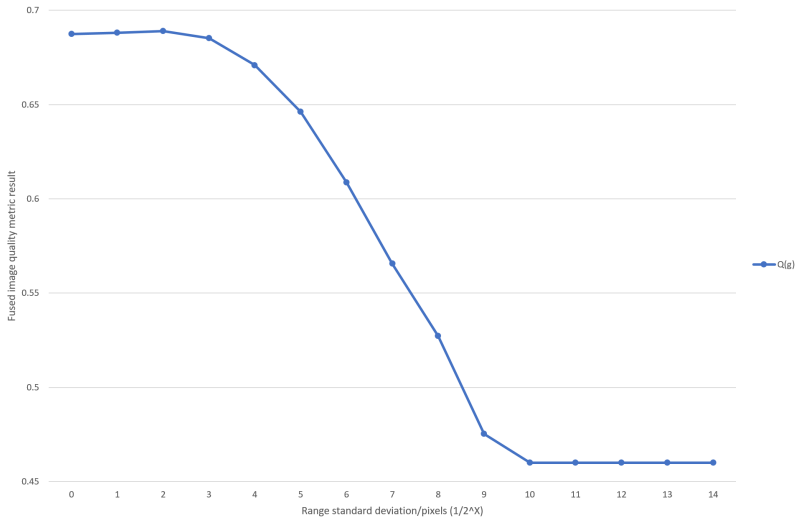


Figure 5.7: Local linear combination image fusion (SE variant): mean Q_G metric result with varying range standard deviation.

Chapter 6

Local Linear Combination Image Fusion: Applications

In this chapter, various applications of the local linear combination image fusion method are presented - these disparate applications are chosen to show the flexibility of the proposed method. The POP or SE variants of the method are used depending on the individual application. In all cases, the input images are assumed to be registered.

6.1 RGB-NIR Image Fusion

In figure 6.1 we wish to fuse the conventional visible spectrum colour RGB image (6.1a) with a greyscale near-infrared (NIR) image (6.1b), to produce a colour output image with details from both input images, but also with natural colours. NIR images can often see through haze, correct for overexposure, or reveal greater detail in vegetation, among other properties [60].

Of course, this fusion can be done with the goal of maximum detail (for technical applications such as machine vision, security, surveillance etc.), or for photographic enhancement for human perception (i.e. creating a pleasing, subjectively improved image). In this section, we are attempting the latter. In chapter 4 of this thesis, it was found that the fusion results of the Spectral Edge method are clearly preferred to the input RGB image by

observers in a psychophysical experiment.

We compare the results of RGB-NIR fusion using both variants of our LLC image fusion model with the previous Spectral Edge method using global look-up-table reintegration [31], using images from the EPFL RGB-NIR data set [8], as well as some provided by Spectral Edge Ltd. Figures 6.1, 6.2 and 6.3 show examples of the RGB and NIR input images, and the fusion results for each method.

In figure 6.1, the NIR image shows much more detail in the sky, which is overexposed and hazy in the RGB image. All three methods transfer this detail effectively into the fused image, but the LLC (SE) method most effectively combines detail transfer and colour enhancement. Figure 6.2 is an example of the NIR-reflecting properties of vegetation (chlorophyll). Vegetation appears very bright in the near-infrared due to its high NIR reflectance, and therefore more detailed and less noisy in areas which are dark in the visible spectrum. Again the LLC(SE) method provides the best balance of detail transfer and colour faithfulness. Figure 6.3 similarly shows the vegetation properties of NIR imaging, but also displays its dehazing abilities. In the background of the landscape, the NIR image captures much more detail, whereas the RGB image is hazy in these areas. In this case, the LLC(POP) method transfers more of this haze detail, but produces an inferior overall colour vividness and image quality.

Table 6.1 shows the results of a psychophysical experiment, with 6 observers. The SE variant of the LLC method is by far the most preferred, followed by the POP variant of the LLC method, then the previous SE method, and finally the original RGB image is the least preferred. This shows that all of the image fusion methods tested are a useful form of image enhancement, but the SE variant of the LLC method is the best performing on this task.

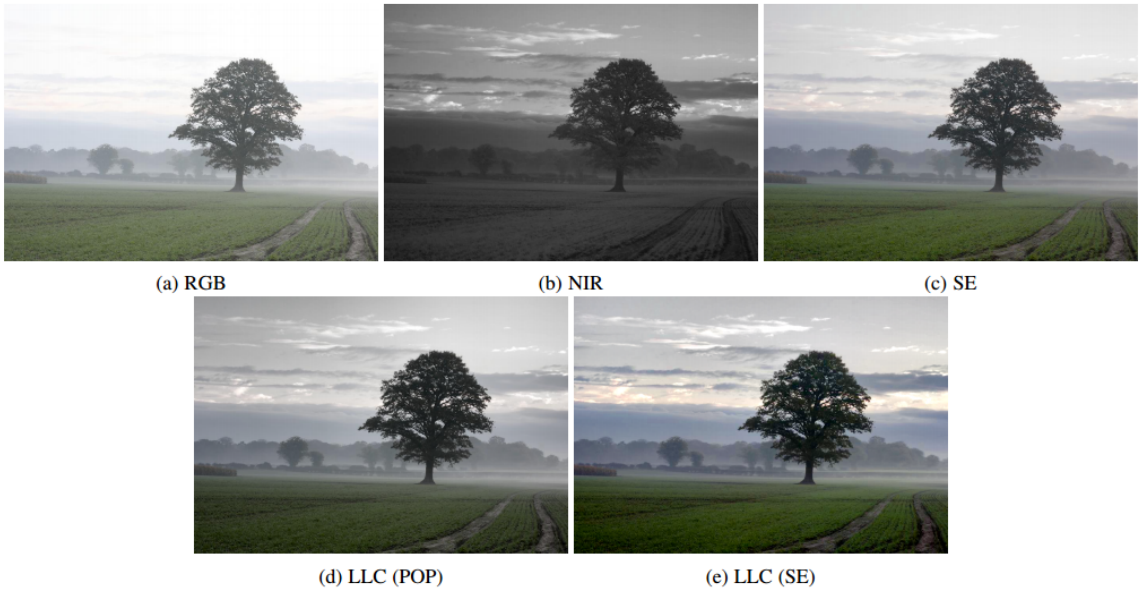


Figure 6.1: Local linear combination image fusion applications: RGB-NIR image fusion (image courtesy of Spectral Edge Ltd.) - original RGB and near-infrared input images, fusion result of Spectral Edge [16], and proposed results of the SE and POP variants of our LLC method.

Method:	RGB	SE	LLC(POP)	LLC(SE)
zscore:	-0.733	0.142	0.139	0.452

Table 6.1: Local linear combination image fusion: RGB-NIR image fusion - psychophysical experiment results.

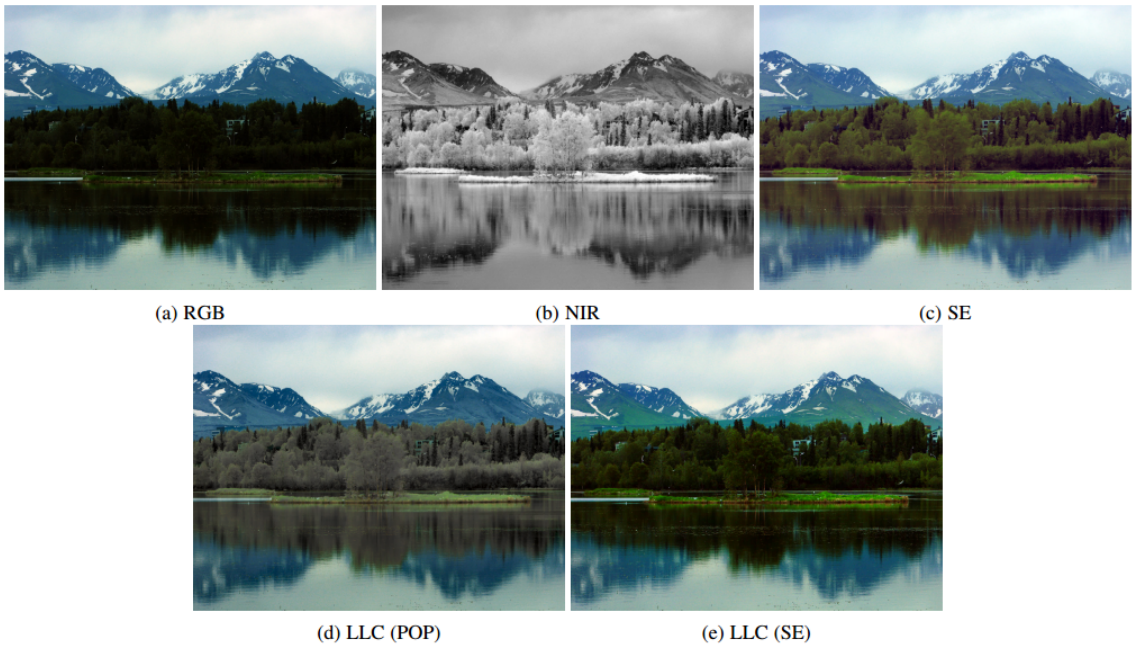
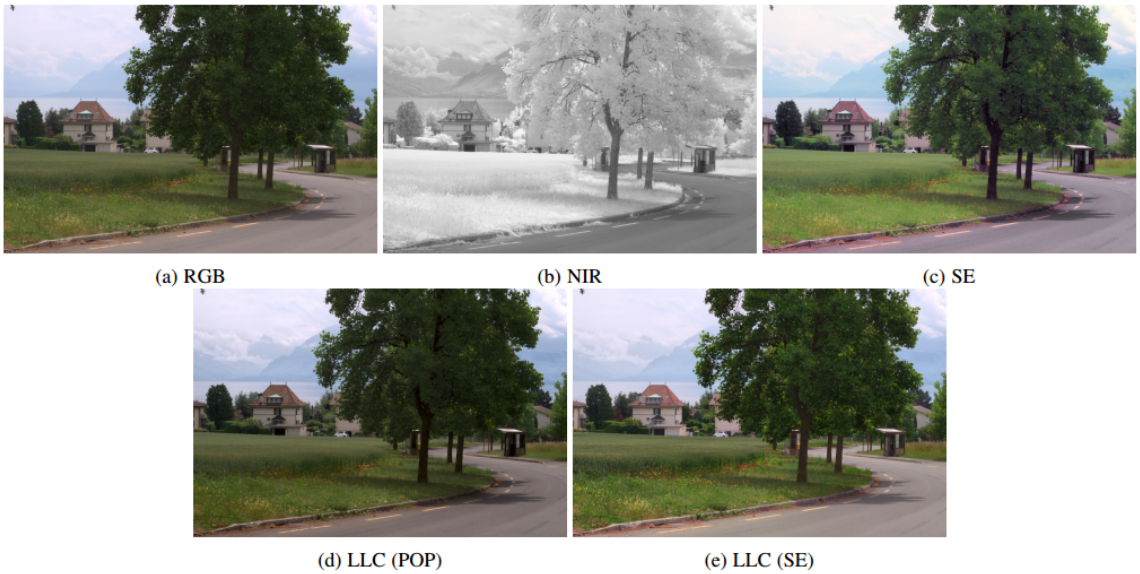


Figure 6.2: Local linear combination image fusion applications: RGB-NIR image fusion (image from Zhang *et al.* [97]) - original RGB and near-infrared input images, fusion result of Spectral Edge [16], and proposed results of the SE and POP variants of our LLC method.



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

(d) LLC (POP) (e) LLC (SE)

Figure 6.3: Local linear combination image fusion applications: RGB-NIR image fusion (image from EPFL RGB-NIR data set [8]) - original RGB and near-infrared input images, fusion result of Spectral Edge [16], and proposed results of the SE and POP variants of our LLC method.

6.2 RGB-thermal Image Fusion

Often, greyscale visible spectrum images are fused with thermal images to produce a greyscale output, as seen in fig. 1.1. It is increasingly common for security and surveillance cameras to capture a colour visible spectrum image, as well as the thermal infrared image of the same scene, for applications such as security and surveillance imaging. The RGB image often shows more detail in bright parts of the scene, but may miss details (particularly people) in dark areas. Image fusion can be used to create an image with all salient details, as well as realistic colours. Figure 6.4 shows an example of this fusion, operating on a frame taken from registered visible and thermal videos from the OTCVBS data set [21]. The RGB luminance channel and the thermal image are fused using the POP variant of the LLC method, and then the RGB luminance channel is replaced with the fused output. The figure in the shadows is clearly visible in the fused output image, and this may prove useful for the human observers of a security camera system.

In chapter 4, the FLIR ONE smartphone thermal camera was used to capture visible and thermal images for image fusion. We can also fuse these results using the POP variant of the LLC fusion method. Figure 6.5 shows the results, compared to the iterative Spectral Edge image fusion method used in the previous chapter. The bright heat in the centre of the water cooler is visible in both results, but the POP result avoids artefacts caused by thermal detail transferred into the lower portion of the cooler.

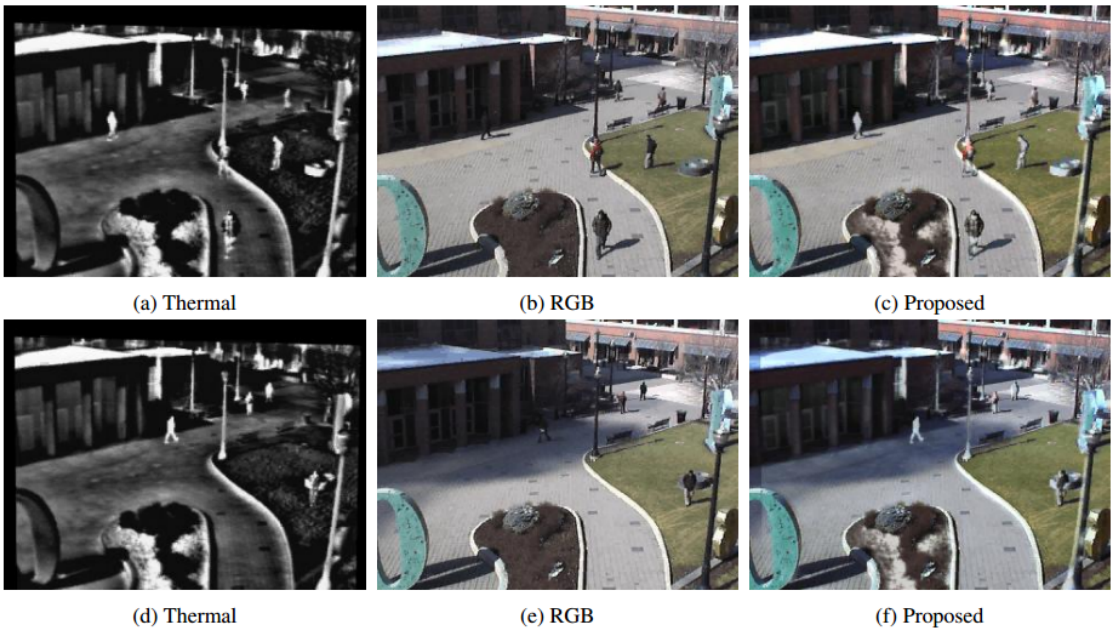


Figure 6.4: Local linear combination image fusion applications: RGB-thermal image fusion - Thermal (7-14 μm) + RGB fusion - video frame 1 (a-c) and frame 400 (d-f), from OTCVBS data set [21]. Fused using the POP variant of the proposed LLC method.

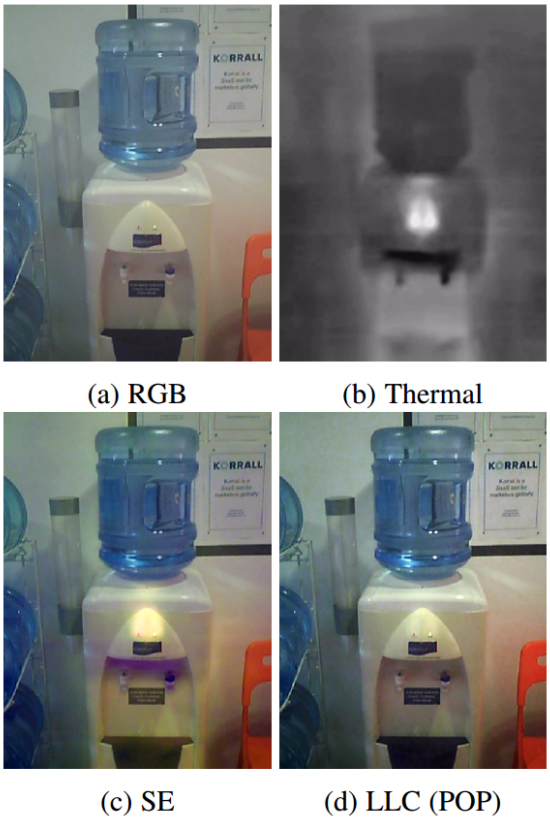


Figure 6.5: Local linear combination image fusion applications: RGB-thermal image fusion using the FLIR ONE camera - a) visible RGB, b) thermal, c) SE result (see chapter 4), d) LLC (POP variant).

6.3 Multifocus Image Fusion

Multifocus image fusion is another application of image fusion - standard multifocus image fusion involves fusing two greyscale input images with different focal settings [51] [52]. 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 6.2 shows a comparison of the performance of the POP variant of the proposed image fusion model on this task, on several standard multifocus image pairs, using standard image fusion quality metrics. The Q_G metric is based on gradient similarity [93], the Q_Y metric is based on the structural similarity image measure (SSIM) [90] [95], and the Q_{MI} metric is based on mutual information [40]. These metrics are reviewed in section 2.5 of this thesis. The results are compared to the method of Zhou and Wang, based on multi-scale weighted gradient-based (MWGF) fusion [98], as well as a standard DWT fusion, using a Daubechies wavelet and CM (choose max) coefficient selection - the POP variant result comes out ahead in the majority of cases.

Figure 6.6 shows the input images and results on the 'Pepsi' image pair - there are visible artifacts around the lettering in the DWT result, while the other two results have no visible artifacts. For this application we use a downsampling ratio of 0.5 and a k stretching parameter of 2.5.

Plenoptic photography provides various refocusing options of color images, allowing images with different depths of focus to be created from a single exposure [65]. The POP variant of the proposed 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

Image Pair	Metric	DWT	MWGF	POP
Book	Q_G	0.8208	0.8327	**0.8332**
	Q_Y	**0.8053**	0.8027	0.8008
	Q_{MI}	0.9738	**1.227**	1.057
Clock	Q_G	0.7860	0.7920	**0.7956**
	Q_Y	**0.8008**	0.7955	0.7910
	Q_{MI}	0.7475	1.142	**1.248**
Desk	Q_G	0.7907	**0.8287**	0.8242
	Q_Y	0.7933	0.7978	**0.7979**
	Q_{MI}	0.7261	1.072	**1.248**
Pepsi	Q_G	0.8648	0.8800	**0.8820**
	Q_Y	**0.7950**	0.7725	0.7792
	Q_{MI}	0.8751	1.196	**1.210**

Table 6.2: Local linear combination image fusion applications: multifocus fusion - table of metric results.

(a) Near Focus (b) Far Focus (c) DWT (d) MWGF (e) Proposed

Figure 6.6: Local linear combination image fusion applications: multifocus Fusion - two greyscale input images with different points of focus, and the fusion results of the DWT, MWGF[98] and POP variant of the proposed LLC method.

a crystal clear output image, in focus at every pixel, to be created.

Figure 6.7 shows an image (from Ng *et al.* [65]), in which four different refocused images are created from a single exposure. The POP variant of our method is used to fuse these differently focused images into a single image in focus at every point by finding a replacement luminance channel - in comparison the result of the method of Eynard *et al.* does not show perfect detail in all parts of the image, and has unnatural colour information (we created this output using code provided by Davide Eynard, experiments with this code

Figure 6.7: Local linear combination image fusion applications: 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 variant of our proposed LLC method. The POP variant of our method brings details across the image into sharper focus with natural colour.

often lead to unnaturally coloured results).

6.4 Multi-exposure Image Fusion

Multi-exposure fusion (MEF) fusion is a simple and practical alternative to high-dynamic range (HDR) imaging, which avoids the step of creating an HDR image by going directly from a set of input images with different exposures to an output fused image. This method

assumes all input images are perfectly registered, and is widely used in consumer photography [77].

A comparison of MEF algorithms [58] poses MEF fusion as a weighted average problem:

$$O(\mathbf{x}) = \sum_{n=1}^{N} W_n(\mathbf{x}) I_n(\mathbf{x}) \tag{6.1}$$

Where O is the fused image, N is the number of multi-exposure input images, $I_n(\mathbf{x})$ is the luminance (or other coefficient value) and $W_n(\mathbf{x})$ the weight at the $\mathbf{x}$-th pixel in the n-th exposure image. The weight factor $W_n(\mathbf{x})$ may be spatially varying or global.

In a subjective comparison, based on assessing 8 MEF algorithms by their mean opinion score (MOS), rated from 1 to 10, the best performing algorithm was that of Mertens *et al.* [62]. This is based on a multiscale Laplacian pyramid decomposition of the input images, with the coefficients from each image weighted by a combination of contrast, saturation and well-exposedness, and then reintegrated to produce a fused image.

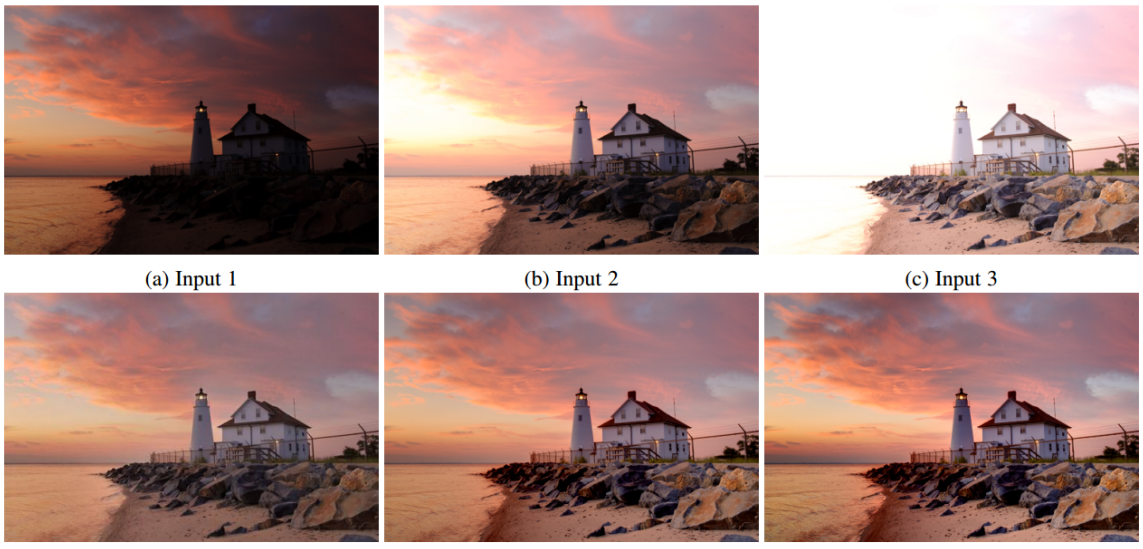
Another powerful method for multi-exposure image fusion is that of Shen *et al.* [80], which aims to achieve an optimal balance between local contrast and color consistency, while combining details from different exposures. A globally optimal solution is calculated subject to the two quality measures by formulating the fusion problem as probability estimation.

Figures 6.8, 6.9 and 6.10 show the results of using the LLC image fusion method on this task, compared to the methods of Mertens *et al.* and Shen *et al.*. We use the Spectral Edge variant of our proposed method to calculate a spatially varying set of linear weights W from the input multi-exposure images. However, our method also requires a guide image, to provide the desired output color at each pixel - we use the output of the method

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

(d) Shen *et al.* (e) Mertens *et al.* (f) LLC(SE)

Figure 6.8: Local linear combination image fusion applications: multi-exposure fusion - 'Balloons' image sequence courtesy of Erik Reinhard.



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

(d) Shen *et al.* (e) Mertens *et al.* (f) LLC(SE)

Figure 6.9: Local linear combination image fusion applications: multi-exposure fusion - 'Lighthouse' image sequence courtesy of HDRsoft.

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

(d) Shen *et al.* (e) Mertens *et al.* (f) LLC(SE)

Figure 6.10: Local linear combination image fusion applications: multi-exposure fusion - 'Cave' image sequence courtesy of Bartlomiej Okonek.

of Mertens *et al.* as our guide image in the results shown here.

6.5 Colour to Greyscale Conversion

Colour to greyscale conversion is the process of converting a colour RGB image to a summary greyscale, which should represent all of the intensity and colour details of the original as closely as possible. There have been various previous methods proposed to accomplish this goal, such as [36], [74] and [37]. These methods and others were presented and compared in the work of Eynard *et al.* [25], and their proposed method was found to be the best performing, based on the RWMS metric [45] and a psychophysical experiment.

The problem of converting from colour to greyscale, although not generally thought of as an image fusion task, is an N to 1 dimensionality reduction problem - exactly what the POP variant of the LLC image fusion method effectively does. Therefore, for this task we can use the POP variant of our proposed method. To produce optimal performance for colour to greyscale conversion, a $spread$ parameter of 2 is used on the LLC vectors to rotate them away from the mean mapping vector - it is important to have highly separated mapping vectors in this task as 3 input dimensions must be compressed into 1 output dimension. Another optimization for this task is to use the hue in CIE LUV colour space as the guide image for cross bilateral filtering/bilateral upsampling - this helps to ensure that image areas with different hues have a different projection, and therefore are more likely to have a different output greyscale value.

Table 6.3 shows a comparison of colour to greyscale conversion performance for all images from the Ĉadík data set [13]. The POP method is compared with CIE L (luminance) and the results of Eynard *et al.* [25]. The metric used is the root mean weighted square (RWMS) error metric of Kuhn *et al.* [45], which compares colour differences between pixels in the input RGB image with differences in intensity in the output greyscale image, and is the metric used in [25]. The POP method is the best performing in a plurality of the

Image	CIE L	Eynard *et al.*	POP
155_5572_jpg	**1.18**	1.45	1.53
25_color	0.853	**0.459**	0.601
34445	0.746	**0.634**	0.701
C8TZ7768	1.36	**1.22**	1.31
ColorWheelEqLum200	5.41	**1.98**	3.12
ColorsPastel	7.70	**4.43**	5.62
DSCN9952	**1.00**	1.36	1.37
IM2-color	4.58	0.918	**0.835**
Ski_TC8-03_sRGB	1.07	1.06	**1.03**
Sunrise312	1.78	**1.32**	1.37
arctichare	0.815	0.599	**0.570**
balls0_color	**1.15**	1.21	1.21
butterfly	0.746	0.578	**0.549**
fruits	1.02	0.865	**0.839**
girl	0.8347	0.8364	**0.8346**
impatient_color	1.116	1.113	**0.945**
kodim03	1.10	1.16	**1.07**
monarch	1.10	1.03	**0.902**
portrait_4v	0.645	0.613	**0.521**
ramp	8.30	**1.30**	3.71
serrano	**1.29**	1.31	1.36
text_4v	0.749	**0.683**	0.968
tree_color	0.672	**0.565**	0.571
tulips	1.13	1.06	**1.04**
watch	0.639	**0.610**	0.714

Table 6.3: Local linear combination image fusion applications: color to greyscale qualitative comparison. Mean RWMS error metric values (to 3 s.f, except where more are necessary) for CIE L (luminance), Eynard *et al.* and the POP method. All values are $\times 10^{-3}$.

test images.

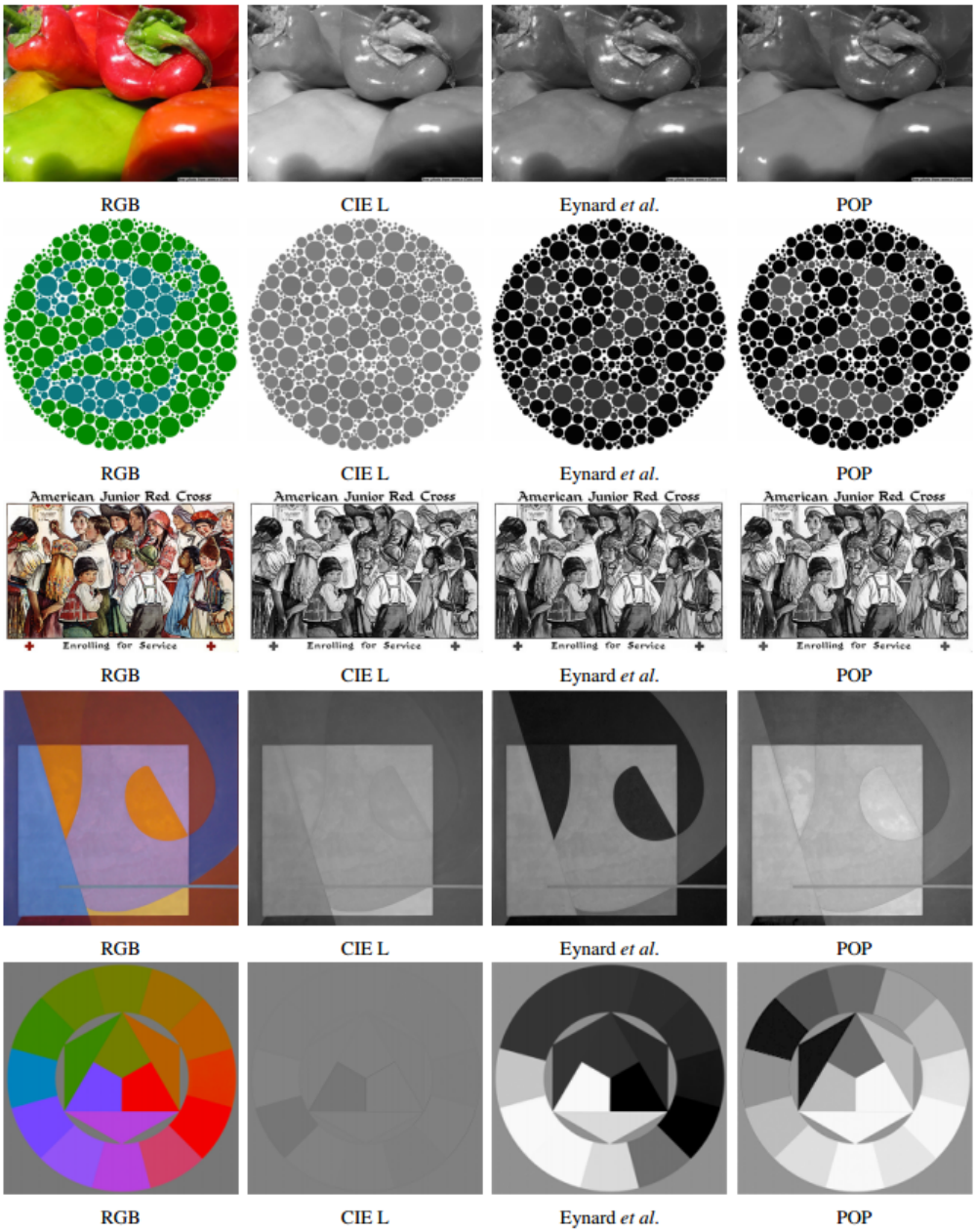


Figure 6.11: Local linear combination image fusion applications: color to greyscale conversion (Čadík data set[13]) - '155_5572_jpg', '25_color', 34445', 'C8TZ7768' and 'ColorWheelEqLum200'. Input RGB, CIE L, results of Eynard *et al.*[25] and POP variant of the proposed LLC method.

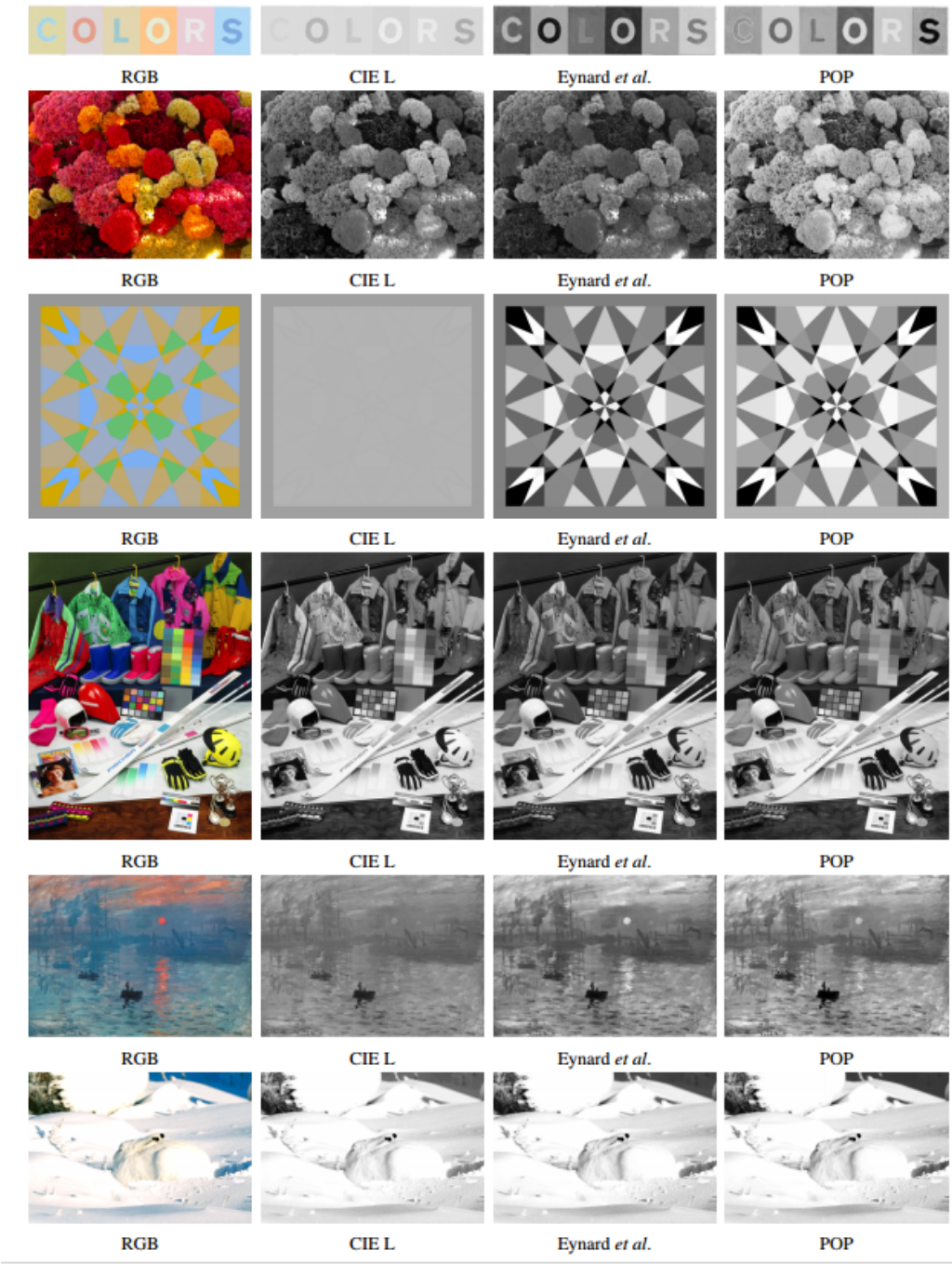


Figure 6.12: Local linear combination image fusion applications: color to greyscale conversion (Ĉadík data set[13]) - 'ColorsPastel', 'DSCN9952', 'IM2-color', 'Ski_TC8-03_sRGB', 'Sunrise312' and 'arctichare'. Input RGB, CIE L, results of Eynard *et al.*[25] and POP variant of the proposed LLC method.

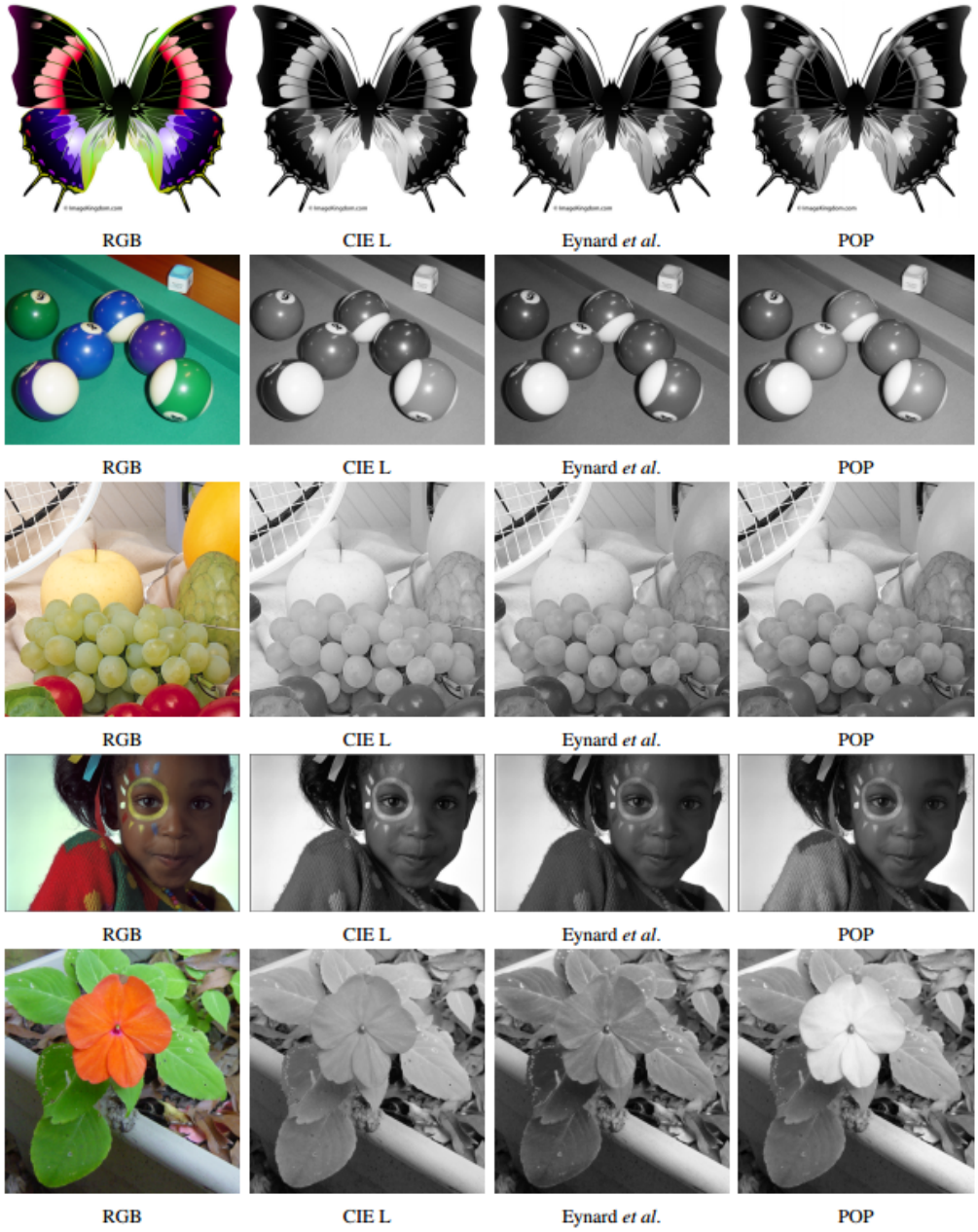


Figure 6.13: Local linear combination image fusion applications: color to greyscale conversion (Ĉadík data set[13]) -'butterfly', 'balls0_color', 'fruits', 'girl' and 'impatient_color'. Input RGB, CIE L, results of Eynard *et al.*[25] and POP variant of the proposed LLC method.

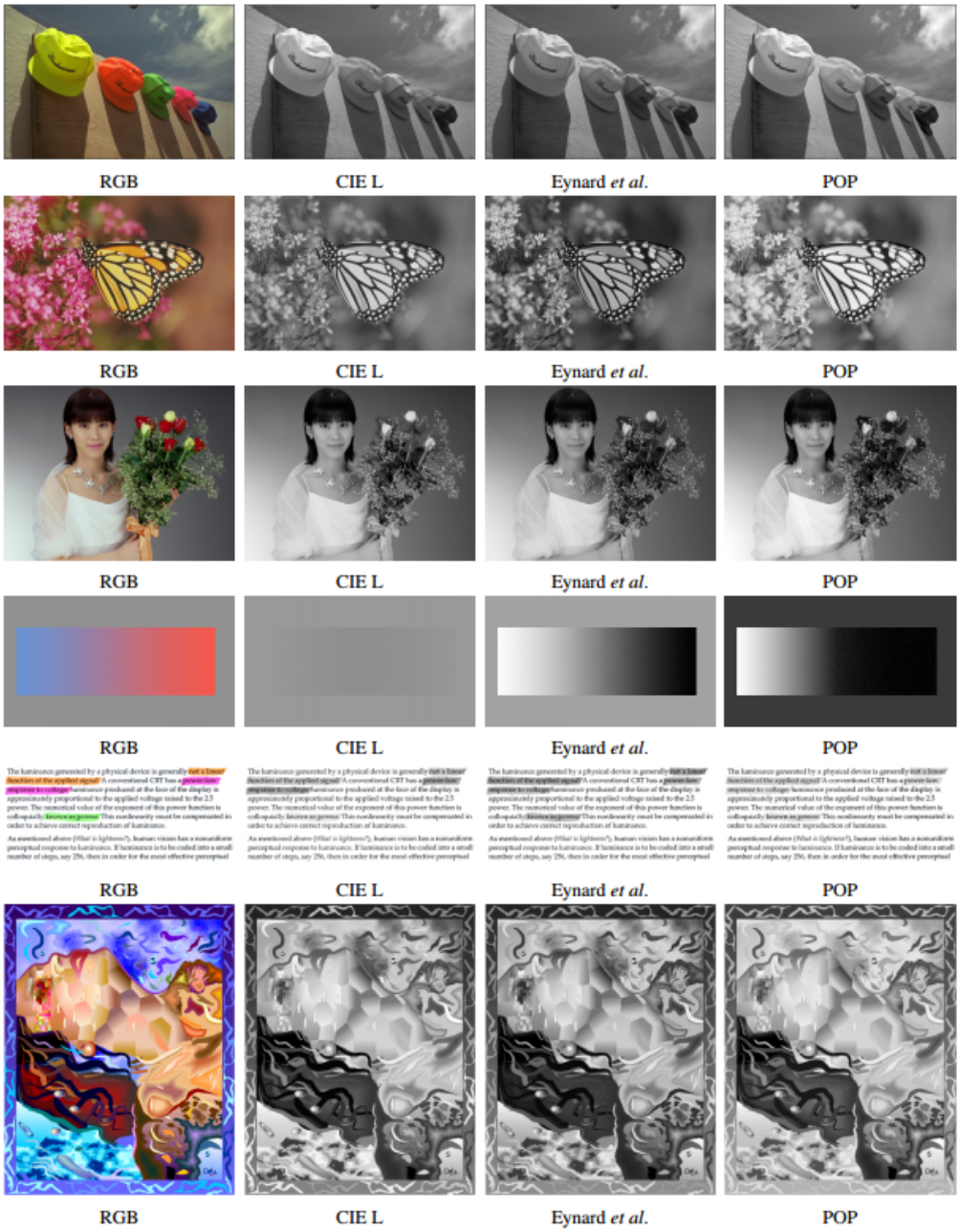


Figure 6.14: Local linear combination image fusion applications: color to greyscale conversion (Ĉadík data set[13]) -'kodim03', 'monarch', 'portrait_4v', 'ramp', 'text' and 'serrano'. Input RGB, CIE L, results of Eynard *et al.*[25] and POP variant of the proposed LLC method.

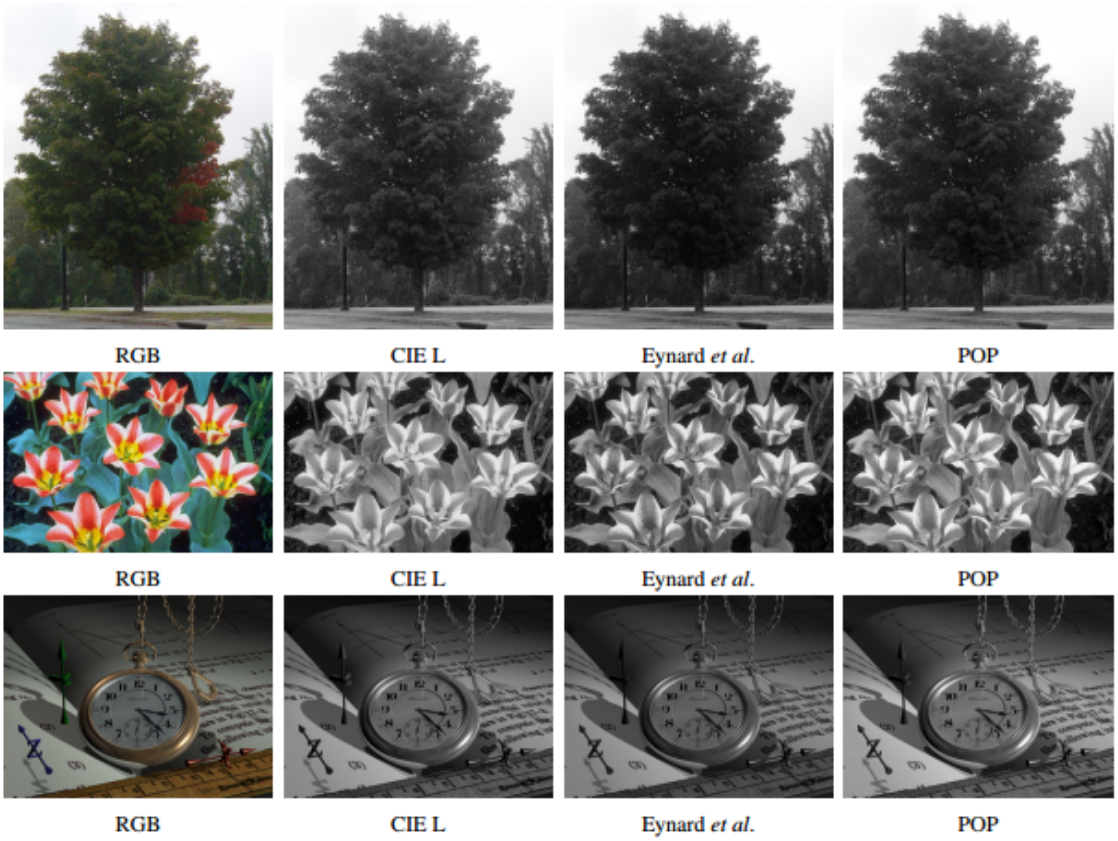


Figure 6.15: Local linear combination image fusion applications: color to greyscale conversion (Ĉadík data set[13]) -'tree_color', 'tulips' and 'watch'. Input RGB, CIE L, results of Eynard *et al.*[25] and POP variant of the proposed LLC method.

6.6 Flash and No-flash Image Enhancement

Flash and no-flash image pairs can be used to create an improved final image. The flash image captures high levels of detail, while the no-flash image captures ambient illumination and natural color. Petschnigg *et al.* use these image pairs to transfer detail from the flash to the no-flash image and denoise the no-flash image[69]. They denoise the no-flash image using a joint bilateral filter, with the flash image used as the edge image. Additional detail is transferred by adding in high-frequency bilateral filter coefficients from the flash image.

(a) Flash (b) No flash (c) Petschnigg *et al.* (d) Proposed

Figure 6.16: Local linear combination image fusion applications: flash and no-flash image enhancement example - flash and no-flash images (a) and (b), (c) result of Petschnigg *et al.*[69], (d) result of SE variant of LLC method.

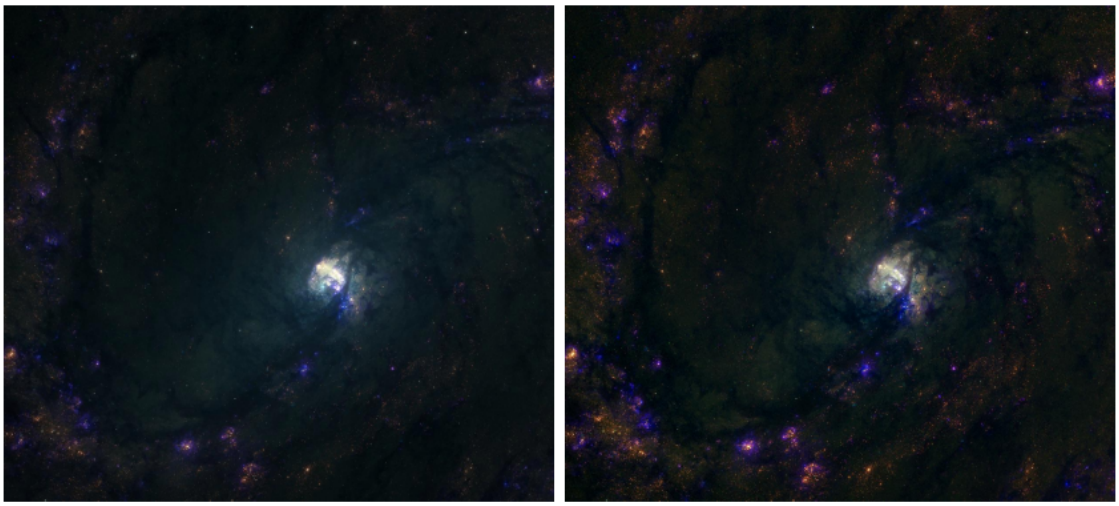
These changes are not applied to all parts of the image - a shadow mask prevents flash artifacts or specular highlights from being transferred into the output image.

We can create similar results in a simpler way using our LLC method. Fig. 6.16 shows an example result using the SE variant of our method. We first use a spatial blur to reduce the noise of the no-flash image, before using it as the RGB guide image in our fusion process, together with the flash image as the high-dimensional input. Compared to the result of Petschnigg *et al.* we transfer more of the derivative information of the flash image for improved contrast and detail. However, we would need to add something (such as a shadow mask) to prevent problems in scenes with greater flash artifacts.

6.7 Image Fusion for Astronomical Visualization

Space and ground-based telescopes provide a rich source of image data. These telescopes often have multi-band imaging capabilities, but often a false color visualization is constructed by a simple assignment of three bands as the red, green and blue image channels, along with manual postprocessing in programs such as Adobe Photoshop [2].

We have used the Spectral Edge variant of our method to improve astronomical visualization. We performed manual postprocessing on each of 8 multiband images of the



(a) False Color RGB (b) Proposed

Figure 6.17: Local linear combination image fusion applications: astronomical fusion for visualization (Hubble image of M83 galaxy) - (a) false color image composed of 3 out of 8 multiband images, (b) output of SE variant of LLC method.

M83 galaxy captured by the Hubble space telescope [3] in MATLAB, and then arbitrarily assigned 3 channels as a false color RGB. We then used this as the guide image for the proposed image fusion method, with all 8 multiband images as the high-dimensional input. Figure 6.17 shows the results, with an increase in detail transfer and contrast in the fused output image.

Chapter 7

Summary and Conclusions

7.1 Summary

This thesis has presented several new contributions:

Firstly, the Spectral Edge image fusion method is compared to other image fusion methods for RGB-NIR image fusion in a psychophysical experiment, and its results shown to be preferred to the input RGB images - a tentative preference order between the other methods ranks the SE method highest, but not to a statistically significant degree. The ranking given by several proposed objective image fusion quality metrics is then compared to that of the experiment.

Secondly, an iterative extension to the Spectral Edge image fusion method has been proposed, and compared to the previous method. Extra iterations give the fusion greater detail and colour vividness, but can also lead to unnatural colours. Using between 2 and 4 image fusion iterations may give the most preferred results, but more investigation is needed for a definitive answer.

Thirdly, new applications of the Spectral Edge image fusion method have been presented: RGB-NIR image fusion from a single sensor which captures both visible and near-infrared image information, and RGB-thermal image fusion using the FLIR ONE smartphone thermal camera. In both cases, useful results are demonstrated from the image fusion process, giving improved image quality and detail as measured by psychophysical and metric experiments.

Finally, a new image fusion model is proposed - local linear combination image fusion. In our model, the problem of image fusion is reduced to finding linear combination coefficients, different for each pixel, which map the input images to an output image. Two methods of calculating local linear combination coefficients are explained: the first is based on the Principal characteristic vector of the Outer Product (POP) of the Jacobian matrix of derivatives, which produces a mapping that creates an output image with derivatives equal to those calculated by Socolinsky and Wolff, and the second is based on finding a mapping by fitting a least-squares regression (with regularization) to the colour derivatives calculated from the Spectral Edge theorem. This produces highly local and detailed image fusion results, without artifacts, with low computational complexity. We show it to produce state of the art results on a wide range of applications.

7.2 Future Work

There is far more potential work to be done on creating reliable image fusion quality metrics which work on any number of input images and a colour output fused image - almost all previous metrics are based on fusing two greyscale images. Psychophysical experiments will remain the best way of measuring subjective image fusion quality, but metrics are useful where time or money is limited. Little previous work has looked at metrics for any

image fusion cases other than 2 to 1 channel image fusion, so there is tremendous potential for research in this area.

More work could be done to compare the Spectral Edge image fusion method (either in its original form, or with iteration) with other methods - some of the psychophysical experiments in this thesis failed to show which method is preferred to a statistically significant degree. The experiments in this thesis focused on RGB-NIR image fusion, but there are a wide variety of image fusion applications which could be tested and compared, and a large number of image fusion methods to compare against. There are recent image fusion methods using deep learning (such as [72]), and it would be interesting to compare these with our method.

Our new local linear combination model of image fusion contains a great deal of scope for future research. Different ways of calculating coefficients for LLC fusion could be considered. Any method of image analysis or decomposition could potentially provide the basis for calculating LLC coefficients - for example the coefficients used in DWT fusion. Other methods for diffusing the coefficients across the image plane could also be considered - any method of image smoothing or filtering could potentially be applicable here. The model is so versatile and open to experimentation that any number of different permutations could be tried to improve its speed and performance. It can also be applied to other image fusion tasks, such as image fusion for medical imaging or manuscript analysis.

More work could be done on the mathematical basis of the LLC image fusion model and the POP and SE coefficient calculation variants proposed. Although proofs have been presented here, it is still not fully clear, at all steps, why the image fusion model works so well. In particular, the mathematics of going from proving that POP gradients are equal to SW gradients at a single pixel up to the gradient field of a whole image could be developed

further.

7.3 Conclusions

Derivative domain image fusion is an area of great research, with various methods using the structure tensor to provide greyscale output gradients, and then using gradient reintegration methods used to produce an output image. However, despite many efforts, the output images suffer from artefacts due to the non-integrability of gradient fields.

The Spectral Edge image fusion method integrates colour into its mathematics in an elegant and effective way, producing colour output gradients which simultaneously transfer maximum detail and retain natural colours based on a guide image. The previous approach used global look-up-table based gradient reintegration to avoid artefacts, but this does not transfer maximum detail. The output images of this method are preferred to other methods in psychophysical experiments and can be applied to various applications. An iterative extension of this method has the potential to create output images with higher levels of detail transfer than the original method.

Our proposed local linear combination method applies different combination coefficients to each pixel, allowing greater detail transfer and output image quality than global methods can achieve. The POP theorem can produce linear combination coefficients for an output greyscale image, or we can reintegrate the colour gradients produced by the Spectral Edge theorem - other ways could also be considered to calculate its coefficients. It is a versatile and powerful method - as shown in Chapter 6 of this thesis, it can be applied to a wide variety of applications producing excellent fusion results. Its computational complexity is also low, making it suitable for real-time implementation.

Bibliography

[1] FLIR ONE. http://www.flir.co.uk/flirone/. Accessed: 2016-03-08.

[2] How Photoshop helps NASA reveal the unseeable. http://blogs.adobe.com/conversations/2015/09/how-photoshop-helps-nasa-reveal-the-unseeable.html. Accessed: 2016-06-30.

[3] Hubble legacy archive. http://hla.stsci.edu/. Accessed: 2016-06-30.

[4] Omnivision OV4682 RGB-IR sensor. http://www.ovt.com/products/sensor.php?id=145. Accessed: 2016-03-08.

[5] Amit Agrawal, Ramesh Raskar, and Rama Chellappa. What is the range of surface reconstructions from a gradient field? *Computer Vision, European Conference on*, pages 578–591, 2006.

[6] Josef Bigun. *Vision with direction*. Springer, 2006.

[7] Rick S Blum and Zheng Liu. *Multi-sensor image fusion and its applications*. CRC press, 2005.

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

[9] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980.

[10] Peter J Burt and Edward H Adelson. The laplacian pyramid as a compact image code. *Communications, IEEE Transactions on*, 31(4):532–540, 1983.

[11] Peter J Burt and Edward H Adelson. Merging images through pattern decomposition. *Applications of Digital Image Processing VIII*, 575:173–181, 1985.

[12] Peter J Burt and Raymond J Kolczynski. Enhanced image capture through fusion. In *Computer Vision, 1993. Proceedings., Fourth International Conference on*, pages 173–182. IEEE, 1993.

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

[14] Anthony J Calabria and Mark D Fairchild. Perceived image contrast and observer preference i. the effects of lightness, chroma, and sharpness manipulations on contrast perception. *Journal of imaging Science and Technology*, 47(6):479–493, 2003.

[15] David Connah, Mark S. Drew, and Graham D. Finlayson. Spectral edge image fusion: Theory and applications. *European Conference on Computer Vision, IEEE Conference on*, pages 65–80, 2014.

[16] David Connah, Mark Samuel Drew, and Graham David Finlayson. Spectral Edge image fusion: Theory and applications. *Computer Vision, European Conference on*, pages 65–80, 2014.

[17] David Connah, Graham D Finlayson, and Marina Bloj. Seeing beyond luminance: A psychophysical comparison of techniques for converting colour images to greyscale. *Color and Imaging Conference*, 2007(1):336–341, 2007.

[18] Nedeljko Cvejic, David Bull, and Nishan Canagarajah. Region-based multimodal image fusion using ica bases. *IEEE Sensors Journal*, 7(5):743–751, 2007.

[19] Nedeljko Cvejic, Artur Łoza, David Bull, and Nishan Canagarajah. A similarity metric for assessment of image fusion algorithms. *International journal of signal processing*, 2(3), 2006.

[20] Nedeljko Cvejic, Artur Loza, David R Bull, and Cedric Nishan Canagarajah. A novel metric for performance evaluation of image fusion algorithms. In *IEC (Prague)*, pages 80–85, 2005.

[21] James W Davis and Vinay Sharma. Background-subtraction using contour-based fusion of thermal and visible imagery. *Computer Vision and Image Understanding*, 106(2):162–182, 2007.

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

[23] Mark S Drew, David Connah, Graham D Finlayson, and Marina Bloj. Improved colour to greyscale via integrability correction. *IS&T/SPIE Electronic Imaging*, pages 72401B–72401B, 2009.

[24] Virginia Estellers, Stefano Soatto, and Xavier Bresson. Adaptive regularization with the structure tensor. *IEEE Transactions on Image Processing*, 24(6):1777–1790, 2015.

[25] 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.

[26] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics*, 27(3):67, 2008.

[27] Raanan Fattal. Single image dehazing. In *ACM Transactions on Graphics (TOG)*, volume 27, page 72. ACM, 2008.

[28] Raanan Fattal, Dani Lischinski, and Michael Werman. Gradient domain high dynamic range compression. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 249–256. ACM, 2002.

[29] Elena A Fedorovskaya, Huib de Ridder, and Frans JJ Blommaert. Chroma variations and perceived quality of color images of natural scenes. *Color Research & Application*, 22(2):96–110, 1997.

[30] Chen Feng, Shaojie Zhuo, Xiaopeng Zhang, Liang Shen, and Sabine Süsstrunk. Near-infrared guided color image dehazing. *Image Processing, IEEE International Conference on*, pages 2363–2367, 2013.

[31] Graham D Finlayson, David Connah, and Mark S Drew. Lookup-table-based gradient field reconstruction. *Image Processing, IEEE Transactions on*, 20(10):2827–2836, 2011.

[32] Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. *Color and Imaging Conference*, 2004(1):37–41, 2004.

[33] Clément Fredembach, Nathalie Barbuscia, and Sabine Süsstrunk. Combining visible and near-infrared images for realistic skin smoothing. In *Color and Imaging Conference*, volume 2009, pages 242–247. Society for Imaging Science and Technology, 2009.

[34] Clément Fredembach and Sabine Süsstrunk. Colouring the near-infrared. *Color and Imaging Conference*, 2008(1):176–182, 2008.

[35] Clément Fredembach and Sabine Susstrunk. Illuminant estimation and detection using near-infrared. In *Digital Photography V*, volume 7250, page 72500E. International Society for Optics and Photonics, 2009.

[36] Amy A Gooch, Sven C Olsen, Jack Tumblin, and Bruce Gooch. Color2gray: salience-preserving color removal. *ACM Transactions on Graphics (TOG)*, 24(3):634–639, 2005.

[37] Mark Grundland and Neil A Dodgson. Decolorize: Fast, contrast enhancing, color to grayscale conversion. *Pattern Recognition*, 40(11):2891–2896, 2007.

[38] Shoudong Han, Wenbing Tao, Desheng Wang, Xue-Cheng Tai, and Xianglin Wu. Image segmentation based on grabcut framework integrating multiscale nonlinear structure tensor. *Image Processing, IEEE Transactions on*, 18(10):2289–2302, 2009.

[39] Paul R Hill, Cedric Nishan Canagarajah, and David R Bull. Image fusion using complex wavelets. In *BMVC*, pages 1–10. Citeseer, 2002.

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

[41] Alex Pappachen James and Belur V Dasarathy. Medical image fusion: A survey of the state of the art. *Information Fusion*, 19:4–19, 2014.

[42] Klaus Kollreider, Hartwig Fronthaler, and Josef Bigun. Evaluating liveness by face images and the structure tensor. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pages 75–80. IEEE, 2005.

[43] Johannes Kopf, Michael F Cohen, Dani Lischinski, and Matt Uyttendaele. Joint bilateral upsampling. *ACM Transactions on Graphics*, 26(3):96, 2007.

[44] Ullrich Köthe. Edge and junction detection with an improved structure tensor. In *Joint Pattern Recognition Symposium*, pages 25–32. Springer, 2003.

[45] Giovane R Kuhn, Manuel M Oliveira, and Leandro AF Fernandes. An improved contrast enhancing approach for color-to-grayscale mappings. *The Visual Computer*, 24(7-9):505–514, 2008.

[46] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971.

[47] Cheryl Lau, Wolfgang Heidrich, and Rafal Mantiuk. Cluster-based color space optimizations. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1172–1179. IEEE, 2011.

[48] Martin Lettner, Markus Diem, Robert Sablatnig, and Heinz Miklas. Registration and enhancing of multispectral manuscript images. In *Signal Processing Conference, 2008 16th European*, pages 1–5. IEEE, 2008.

[49] Hui Li, BS Manjunath, and Sanjit K Mitra. Multisensor image fusion using the wavelet transform. *Graphical models and image processing*, 57(3):235–245, 1995.

[50] Shutao Li, Xudong Kang, and Jianwen Hu. Image fusion with guided filtering. *IEEE Transactions on Image Processing*, 22(7):2864–2875, 2013.

[51] Shutao Li, James T Kwok, and Yaonan Wang. Multifocus image fusion using artificial neural networks. *Pattern Recognition Letters*, 23(8):985–997, 2002.

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

[53] Shutao Li, Haitao Yin, and Leyuan Fang. Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Transactions on Geoscience and Remote Sensing*, 51(9):4779–4789, 2013.

[54] Chuan-kai Lin. Pixel grouping. `https://sites.google.com/site/chklin/demosaic`. Accessed: 2016-03-08.

[55] Tony Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2):225–270, 1994.

[56] Zheng Liu, Erik Blasch, Zhiyun Xue, Jiying Zhao, Robert Laganiere, and Wei Wu. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(1):94–109, 2012.

[57] Bibo Lu, Hui Wang, and Chunli Miao. Medical image fusion with adaptive local geometrical structure and wavelet transform. *Procedia Environmental Sciences*, 8:262–269, 2011.

[58] K. Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *Image Processing, IEEE Transactions on*, 24(11):3345–3356, 2015.

[59] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, 1989.

[60] Klaus Mangold, Joseph A Shaw, and Michael Vollmer. The physics of near-infrared photography. *European Journal of Physics*, 34(6):S51, 2013.

[61] Rafał K Mantiuk, Anna Tomaszewska, and Radosław Mantiuk. Comparison of four subjective methods for image quality assessment. In *Computer Graphics Forum*, volume 31, pages 2478–2491. Wiley Online Library, 2012.

[62] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum*, 28(1):161–171, 2009.

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

[64] Filippo Nencini, Andrea Garzelli, Stefano Baronti, and Luciano Alparone. Remote sensing image fusion using the curvelet transform. *Information Fusion*, 8(2):143–156, 2007.

[65] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report*, 2(11), 2005.

[66] Gonzalo Pajares and Jesus Manuel De La Cruz. A wavelet-based image fusion tutorial. *Pattern recognition*, 37(9):1855–1872, 2004.

[67] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 313–318. ACM, 2003.

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

[69] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM transactions on graphics (TOG)*, 23(3):664–672, 2004.

[70] Gemma Piella. New quality measures for image fusion. *International Conference on Information Fusion*, pages 542–546, 2004.

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

[72] K Ram Prabhakar, V Sai Srikar, and R Venkatesh Babu. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4724–4732. IEEE, 2017.

[73] Guihong Qu, Dali Zhang, and Pingfan Yan. Information measure for performance of image fusion. *Electronics letters*, 38(7):313–315, 2002.

[74] Karl Rasche, Robert Geist, and James Westall. Re-coloring images for gamuts of lower dimension. In *Computer Graphics Forum*, volume 24, pages 423–432. Wiley Online Library, 2005.

[75] Ramesh Raskar, Adrian Ilie, and Jingyi Yu. Image fusion for context enhancement and video surrealism. In *ACM SIGGRAPH 2005 Courses*, page 4. ACM, 2005.

[76] Dikpal Reddy, Amit Agrawal, and Rama Chellappa. Enforcing integrability by error correction using 1-minimization. *Computer Vision and Pattern Recognition, IEEE Conference on*, pages 2350–2357, 2009.

[77] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[78] Dymitr Ruta and Bogdan Gabrys. An overview of classifier fusion methods. *Computing and Information Systems*, 7(1):1–10, 2000.

[79] Lex Schaul, Clément Fredembach, and Sabine Süsstrunk. Color image dehazing using the near-infrared. *Image Processing, IEEE International Conference on*, pages 1629–1632, 2009.

[80] Rui Shen, Irene Cheng, Jianbo Shi, and Anup Basu. Generalized random walks for fusion of multi-exposure images. *IEEE Transactions on Image Processing*, 20(12):3634–3646, 2011.

[81] Takashi Shibata, Masayuki Tanaka, and Masatoshi Okutomi. Gradient-domain image reconstruction framework with intensity-range and base-structure constraints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2745–2753, 2016.

[82] Diego A. Socolinsky. *A Variational Approach to Image Fusion*. PhD thesis, John Hopkins University, 2000.

[83] Diego A Socolinsky and Lawrence B Wolff. Multispectral image visualization through first-order fusion. *Image Processing, IEEE Transactions on*, 11(8):923–931, 2002.

[84] Katrin Strandemar. Infrared resolution and contrast enhancement with fusion. *Patent 9,171,361*, 2015.

[85] Louis L Thurstone. A law of comparative judgment. *Psychological review*, 34(4):273, 1927.

[86] Marko Tkalcic and Jurij F Tasic. *Colour spaces: perceptual, historical and applicational background*, volume 1. IEEE, 2003.

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

[88] Eduard Vazquez, Theo Gevers, Marcel Lucassen, Joost Van De Weijer, and Ramon Baldrich. Saliency of color image derivatives: a comparison between computational models and human perception. *JOSA A*, 27(3):613–621, 2010.

[89] Zhou Wang and Alan C Bovik. Mean squared error: love it or leave it? a new look at signal fidelity measures. *Signal Processing Magazine, IEEE*, 26(1):98–117, 2009.

[90] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[91] Qi Wei, José Bioucas-Dias, Nicolas Dobigeon, and Jean-Yves Tourneret. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7):3658–3668, 2015.

[92] Zhiyun Xue and Rick S Blum. Concealed weapon detection using color image fusion. In *Proceedings of the 6th International Conference on Information Fusion*, volume 1, pages 622–627, 2003.

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

[94] Bin Yang and Shutao Li. Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement*, 59(4):884–892, 2010.

[95] Cui Yang, Jian-Qi Zhang, Xiao-Rui Wang, and Xin Liu. A novel similarity based quality metric for image fusion. *Information Fusion*, 9(2):156–160, 2008.

[96] Lin Zhang, Lei Zhang, and David Zhang. A multi-scale bilateral structure tensor based corner detector. In *Asian Conference on Computer Vision*, pages 618–627. Springer, 2009.

[97] Xiaopeng Zhang, Terence Sim, and Xiaoping Miao. Enhancing photographs with near infra-red images. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[98] Zhiqiang Zhou, Sun Li, and Bo Wang. Multi-scale weighted gradient-based fusion for multi-focus image. *Information Fusion*, 2014.