First-order fusion of volumetric medical imagery

C. Wang and Z. Ye

Abstract: A novel fusion scheme for volumetric medical imagery based on first-order local variational information is presented. The authors first define the contrast of a volumetric image with an arbitrary number of bands, which corresponds with the 3D gradient in the special case of a single-band image with a Euclidean metric. This contrast of multi-band image is regarded as the target contrast field. The next step is to look for a single-band volumetric image as the fusion result, which will have the closest gradient field to the contrast of the input multi-band image. It is a functional extremum problem. Using the variational approach, it leads directly to its Euler–Lagrange equation. By iteration of gradient descent, the final result can be obtained. Experimental results are presented to support the performance of the method.

1 Introduction

The purpose of image fusion is to extract and synthesise information from multiple images in order to produce a more accurate, complete and reliable composite image of the same scene or target, so that the fused image is more suitable for human or machine interpretation. It is useful for analysis, detection, recognition and tracing of targets of interest.

Medical imagery can be obtained by many approaches such as magnetic resonance imaging (MRI), CT, SPECT and so on. In MRI, there are different parameters (i.e. T1-weighted, T2-weighted and proton density (PD)). Different medical imagery techniques reflect different features of the human body. MRI-PD conveys structural information. T1- and T2-weighted MRI measure the relaxation to parallel and perpendicular axes, respectively. T1- and T2-weighted modalities are capable of measuring such characteristics as fat, melanin content, blood flow, calcification and so on [1]. CT is sensitive to bone. Because of these different characteristics, combining different information into one image can help the medical analysis and diagnosis and can make the 3D modelling of an organ easy. This capability has significant scientific and medical applications.

Techniques for 2D image fusion have been widely studied. One current area of intense research is discrete wavelet transform (DWT). The common procedures of DWT-based schemes include DWT decomposition of every source band, coefficient combination and inverse DWT (IDWT) of the combined coefficients. In such a framework, the key steps are: the selection of the different wavelet types, determining the maximal decomposition level and the rule for the coefficient combination.

Numerous scientists have done work in this area. Li et al. [2] first advanced the 2D image fusion technique using DWT; the DWT coefficients of each pixel in the low-frequency sub-band are combined using the arithmetic average of corresponding pixels in all input source bands, whereas the high-frequency sub-bands are obtained by the maximum selection rule with consistency verification. Chipman et al. [3] also employed DWT for image fusion, and the corresponding selection rule for both low- and high-frequency coefficients is the same: at each pixel, if the coefficient in some band is much greater than in the others, it will be selected into the combined ones, whereas the average will be calculated if all bands’ coefficients are comparable. Similarly, Wang et al. [4] took an average of the low-frequency sub-band and a maximum absolute amplitude (MAA) in high-frequency sub-bands. Li et al. [5] used the ‘pixel visibility’ to guide the coefficient selection. In [6], Li et al. discussed the issue of optimal decomposition level and also used DWT together with some visual feature of human to perform the fusion task [7]. Kazemi and Moghaddam [8] introduced a multi-wavelet technique for image fusion. Zhang and Blum [9] categorised and summarised all these multi-scale techniques. Nikolov et al. [10] have adopted these DWT-based schemes to address the 3D case; a block diagram is shown in Fig. 1.

DWT-based image fusion schemes have achieved quite good performance; however, as stated in [11], any non-linear operation following compactly supported wavelets will cause many fluctuations near strong edges (‘ringing’ as described in [12]), and with an increasing decomposition level, the fluctuations will be much more severe and thus present an insurmountable difficulty. What’s more, multi-dimensional wavelets derived from a 1D case by a separable 1D basis are not very suitable for representing multi-dimensional singularity [13].

Intelligent algorithms such as artificial neural networks are introduced to aid image fusion [14]. The source images are partitioned into blocks and the ‘clearer’ block is selected in constructing the final image to keep the consistency of pixels. However, because Li et al. [14] did not provide a reasonable method for partitioning, it will cause some unavoidable problems, if one block has more than one feature and the features are in-focus in different bands. Bloch and Maitre [15] have reviewed the image
fusion algorithms focusing on Bayesian approach, Dempster–Shafer evidence theory and fuzzy theory. Aguilar and Garrett [1] and Aguilar and New [16] discussed 2D and 3D image fusion using a so-called ‘shunting’ operator; however, this operator is designed for a specific type of source image and there is no unified framework for the algorithm.

Socolinsky and Wolff [17, 18] and Socolinsky [19] have proposed a new framework for 2D multispectral image fusion; Socolinsky [19, 20] has also addressed the output’s dynamic range problem. Such a paradigm seems attractive both from the theoretical view and from the practical results. This paper will use a technique similar to Socolinsky’s, focusing on the 3D medical imagery fusion problem and will derive the corresponding model and expressions.

2 Contrast of multi-band 3D image

In [18, 19], the 2D problem of multi-band images has been considered using first-order local variation information. Here, we extend this to a 3D case. We do not consider here the registration problem; that is, we assume the images from different sensors or different time have already been well registered.

2.1 Contrast form

A multi-band volumetric image is defined as a cube $\Omega \subset \mathbb{R}^3$ together with a spectral map $s: \Omega \rightarrow \mathbb{R}^n$ where $\mathbb{R}^n$ denotes an $n$-dimensional photometric space. Let $\mathbb{R}^n$ have an arbitrary Riemannian metric $g$ that can address the sensor property.

Taking a slight abuse of notation for the sake of simplicity for the remainder, let $s: \Omega \rightarrow \mathbb{R}^n$ be a multi-band 3D image. Let $p$ be a point in $\Omega$ and $v$ an arbitrary unit vector in $\mathbb{R}^3$. Analogous to the single-band 2D grayscale image, we define the contrast in $s$ at $p$ in the direction $v$ as the rate of spectral variation at $p$ in that direction. Let $\gamma: [-\epsilon_1, \epsilon_2] \rightarrow \Omega$ be a curve defined on a small interval, with $\epsilon_1 > 0$, $\epsilon_2 > 0$, such that $\gamma(0) = p$ and $\gamma'(0) = v$. The rate of spectral variation at $p$ in the direction of $v$ is given by the magnitude of the vector $s_\gamma(v) = (d/dl)(s \circ \gamma(l))|_{l=0}$, as evaluated by the metric $g$ on $\mathbb{R}^n$ where the composition operator $\circ$ denotes that $(s \circ \gamma)(l) = s(\gamma(l))$, $l \in [-\epsilon_1, \epsilon_2]$. Thus, we can easily obtain

$$s_\gamma(v) = J_p v$$

where $J_p$ is the Jacobian matrix of $s$ at the point $p$. Then the contrast at $p$ in the direction of $v$ is given by the quantity

$$(J_p v)^t g_{s,p}(J_p v) = v^t (J_p^t g_{s,p} J_p) v$$

where the superscript $(\cdot)^t$ denotes the transposition of a vector/matrix.

Let $g_p^s = (J_p^t g_{s,p} J_p)$, $s_k$ denote the $k$th band of the source, and $x_i$ be the $i$th coordinate. From (1) and (2) it follows that

$$(g_p^s)_{ij} = \sum_{k,l=1}^n (g_{s,p})_{kl} \frac{\partial s_k}{\partial x_i} \frac{\partial s_l}{\partial x_j}, \quad 1 \leq i, j \leq 3$$

We define $\chi^2(p) = g_p^s$ to be the image contrast form of $s$ at $p$. Obviously, $\chi^2(p)$ is a symmetric matrix with real elements; therefore its eigenvalues are real and non-negative. Let $\lambda_p$ denote the largest eigenvalue of $\chi^2(p)$. We define the absolute contrast of $s$ at $p \in \Omega$ to be equal to $\sqrt[3]{\lambda_p}$ and we say that the eigenspace of $\chi^2(p)$ corresponding to $\lambda_p$ is the direction of maximal contrast at $p$. It is worthy to note that the eigenspace is a line without a preassigned orientation.

Using coordinates $x, y, z$ on $\Omega$, note that the selection of the largest eigenvalue $\lambda_p$ resonates with principal component analysis if we take $\xi = (s_1, s_2, s_3)^t$ as a random vector with $n$ snapshots, denoted by $\xi_i = (s_{1i}, s_{2i}, s_{3i})^t$, $i = 1, 2, \ldots, n$. Here the subscripts $x, y, z$ denote the partial differentiation and $i$ denotes the $i$th band. Then under the Euclidean metric, the covariance matrix has almost the same form as $\chi^2$ except a constant $1/n$.

$$\text{Var}(\xi) = \frac{1}{n} \begin{pmatrix} s_{x1} & s_{x2} & \cdots & s_{xn} \\ s_{x1} & s_{x2} & \cdots & s_{xn} \\ \vdots & \vdots & \ddots & \vdots \\ s_{xn} & s_{yn} & \cdots & s_{zn} \end{pmatrix} = \frac{1}{n} \chi^2$$

Thus the largest eigenvalue of $\chi^2$ resonates with the power of the principal component of random vector $\xi$. Then the eigenvector corresponding to $\lambda_p$ denotes the major direction of vector $\xi$ and thus the maximal rate of variation in $s$.

For a special case, we consider the single-band volumetric image under the Euclidean metric at each point $p \in \Omega$. It is easy to find

$$\chi^2 = \begin{pmatrix} s_x s_x & s_x s_y & s_x s_z \\ s_x s_y & s_y s_y & s_y s_z \\ s_x s_z & s_y s_z & s_z s_z \end{pmatrix} = \begin{pmatrix} s_x \\ s_y \\ s_z \end{pmatrix} \begin{pmatrix} s_x & s_y & s_z \end{pmatrix}$$

We can compute that the eigenvalues of $\chi^2$ in (5) are $\{0, s_x^2 + s_y^2 + s_z^2\}$; that is, the maximal eigenvalue $\lambda_p = |\nabla s(p)|^2$ and the direction of maximal contrast is spanned by $\nabla s(p)$. This agrees with the definition of gradient in a single-band image.
2.2 Contrast field

Now, the contrast field can be constructed as follows. For each point \( p \in \Omega \), the contrast at \( p \) is a vector \( V(p) \), whose length is the absolute contrast \( \sqrt{\lambda_p} \) and the direction is denoted by the eigenvector \( \mathbf{E}(\lambda_p) \) corresponding to \( \lambda_p \).

However, the corresponding eigenvector may have two opposite directions, denoted by \( \pm \mathbf{E}(\lambda_p) \); they do not have a preassigned priority. We must consider a criteria to use in selecting the value. An auxiliary function \( \psi \) can help. We may choose it as the gradient of the summation of all bands

\[
\psi(x, y, z) = \nabla \sum_{k=1}^{n} s_k(x, y, z)
\]

Thus \( V(p) \) can be modified as

\[
V(p) = \text{sign}(\psi(p)) \cdot V_0(p) \cdot V_0(p)
\]

where \( V_0(p) = \sqrt{\lambda_p} \cdot \mathbf{E}(\lambda_p) \), and \( \langle ., . \rangle \) denotes the inner product. The implications of (6) and (7) are clear: the direction of multi-band image contrast must agree with the variation of the major bands; one simple scheme is to assign each band the same weight. As an example, (6) does just that and works rather well as shown in Section 4. \( V \) derived from (7) is defined as the contrast field of the input multi-band image.

3 Single-band image reconstruction from local contrast

In Section 2, we defined the contrast field \( V \) of a multi-band volumetric image. \( V \) is derived from the first-order local variational information of each band. Our purpose is therefore to preserve as much local image contrast information in \( V \) as possible. The next problem is to find a single-band image that can embody the contrast field \( V \) of multi-band image optimally; that is, we are looking for a function \( f: \Omega \to \mathbb{R} \) whose gradient is closest to \( V \).

3.1 Reconstruction using \( L^2 \) norm

To find a function \( f: \Omega \to \mathbb{R} \), which has the closest gradient field to \( V \), an intuitive scheme would be to solve the equation \( \nabla f = V \). However, this equation in general will have no solution because \( V \) constructed from (3), (6) and (7) does not necessarily satisfy the condition of a gradient; that is, \( V \) may not be integrable. So generally, there does not exist a single-band volumetric image that has the exact contrast information as \( V \).

A common approach for this type problem is to find a closest solution regarding the \( L^2 \) norm; that is, if \( V \) is the target contrast field, we want the gradient of the solution to have the least-squared-error to \( V \). Thus mathematically, we are to find a function \( f \) for which the following functional is minimised

\[
\iiint_{\Omega} |\nabla f - V|^2 \, dx \, dy \, dz
\]

The following notations are adopted

\[
V = V_1 \hat{x} + V_2 \hat{y} + V_3 \hat{z}
\]

\[
F = |\nabla f - V|^2 = (f_x - V_1)^2 + (f_y - V_2)^2 + (f_z - V_3)^2
\]

Using the variational approach [21], the Euler–Lagrange equation for this functional can be directly found by

\[
\frac{\partial F}{\partial f_x} = \frac{\partial}{\partial x} \left( \frac{\partial F}{\partial f_x} \right) + \frac{\partial}{\partial y} \left( \frac{\partial F}{\partial f_y} \right) + \frac{\partial}{\partial z} \left( \frac{\partial F}{\partial f_z} \right) = 0
\]

After simplifying, it can be converted into the form of following partial differential equation (PDE)

\[
\Delta f = \text{div} \, V \quad \text{on} \quad \Omega
\]

As far as the boundary condition is concerned, because we have no prior knowledge, it is hard to assign Dirichlet conditions. In image processing, the researchers often extend the boundary symmetrically so that a Neumann condition can be assigned to this PDE

\[
\nabla f \cdot n = 0 \quad \text{on} \quad \partial \Omega
\]

Assuming that \( V \) on the boundary \( \partial \Omega \) is calculated from symmetrically extended source images, it follows that \( V \cdot n = 0 \) stands on \( \partial \Omega \) where \( n \) is the outward unit normal to \( \Omega \). Combine (12) and (13)

\[
\{ \Delta f = \text{div} \, V \quad \text{on} \quad \Omega \}
\]

On solving a Poisson equation like (14), there are a lot of methods to apply; here, according to [22, 23], a simple iterative scheme is adopted

\[
f_{i,j,k}^{t+1} = f_{i,j,k}^{t} + \frac{1}{6} \left[ \Delta f_{i,j,k}^{t} - (\text{div} \, V)_{i,j,k} \right]
\]

where \( t \) is the iteration variable and \( f^0 \) denotes an arbitrary initial guess. In (15), we must define a discretised version of each differential operator. Let the volumetric grid have the size of \( \Omega = [0, I-1] \times [0, J-1] \times [0, K-1] \). Thus the possible first-order neighbourhood of any point \((i,j,k) \in \Omega \) is

\[
M_{i,j,k} = \{ (i-1, j, k), (i+1, j, k), (i, j-1, k), (i, j+1, k), (i, j, k+1) \}
\]

And for \( (i,j,k) \in \Omega \), its first-order neighbourhood is defined as

\[
N_{i,j,k} = \{ (x,y,z) | (x,y,z) \in M_{i,j,k} \cap \Omega \}
\]

Then we can discretise the Laplacian operator as

\[
\Delta f_{i,j,k} = \left( \sum_{(x,y,z) \in N_{i,j,k}} f_{x,y,z} \right) - \{ \#N_{i,j,k} \} \cdot f_{i,j,k}
\]

where \( \#N_{i,j,k} \) denotes the number of all the elements in \( N_{i,j,k} \). To ensure that the iterative result does not have a visible space shift, we must guarantee that the combined operation of the gradient operator \( \nabla \) (\( \partial / \partial x \), in producing \( V \)) and the divergence function \( \text{ div } \) must have a symmetric support. Thus we choose a forward difference to realise \( \nabla \) and a backward one for \( \text{div} \).

3.2 Dynamic range constraint

In Section 3.1, we considered the linear case of the solution, using the \( L^2 \) norm. However, using the iteration step (15), it is obvious that the dynamic range of \( f \) is not necessarily within the permitted bounds. Since the result will be either shown on a monitor, printed or plotted using some other device, the dynamic range should be constrained to \([0,255] \). If \( f \) obtained by (15) has a range exceeding
To verify the performance of the proposed methods, we use
3D-DWT method for comparison

However, both re-scaling and clipping operations will
decrease the contrast of \( f \). In a digital field, re-scaling
maps some adjacent intensities in the neighbourhood to the
same intensity level, which will soften these contrasts to
zero. Similarly, clipping will make the contrast zero if the
intensities of a point and its neighbourhood all exceed the
range. A 1D case is shown in Fig. 2, in which we can see that the re-scaling and clipping operation may decrease or discard some important detail information. So some other means of avoiding this must be adopted. In [20], the same problem in 2D is discussed, here we directly generalise it to a 3D case. With a constraint condition introduced to (8), we must thus minimise the functional

\[
\int \int \int_{\Omega} |\nabla f - V|^2 \, dx \, dy \, dz \quad \text{subject to } 0 \leq f(x, y, z) \leq 255
\]

An iterative operation to solve (21) may be

\[
\begin{align*}
 f_{t+1}(x, y, z) &= f_{t}(x, y, z) + \frac{1}{6} \left[ \Delta f_{t}(x, y, z) - (\text{div} \, V)_t(x, y, z) \right] \\
 f_{t+1}^c(x, y, z) &= \max(0, \min(255, f_{t+1}(x, y, z)))
\end{align*}
\]

It is very clear that each voxel of result \( f \) generated by (22) is within \([0, 255]\), which can be directly represented by a monitor or printed in slices.

4 Experiments

In the preceding sections, the procedures to obtain a fusion result from a multi-band volumetric image are presented. In this section, we will verify our method by experiments. In the following sections, the metric matrix \( g_{i_d(p)} \) at each point \( p \in \Omega \) is selected as the Euclidean metric and we stop the iteration (15) and (22) when \( t \) reaches 800.

4.1 3D-DWT method for comparison

To verify the performance of the proposed methods, we employ the 3D wavelet fusion techniques for comparison.

Here we only adopt one scheme in [10]. The 3D wavelet we choose is 3D separable Daubechies 8 (db8) wavelet. The maximal wavelet decomposition level is three. The coefficients in smooth (low-frequency) component of the highest decomposition level are selected as the arithmetic average of the ones in all input bands. For the remaining seven high-frequency components in each decomposition level, a simple MAA rule is utilised. That is to say, when each corresponding voxel is considered, the coefficient of high-frequency components, in which band it has the MAA, will be preserved in the fused coefficients. The fused result can be obtained by 3D inverse wavelet transform from all the fused coefficients.

4.2 Experimental results

The volumetric medical images we used were all provided by BrainWeb [24], including MRI-T1, MRI-T2 and MIRIPD. The single-band source image is 181 mm \( \times \) 217 mm \( \times \) 181 mm and each voxel stands for 1 mm \( \times \) 1 mm \( \times \) 1 mm. The source and fusion result images are all shown in slices and here we choose slices 42, 92 and 142 out of the total 181 slices. The corresponding source image slices are shown in Fig. 3.

First, T1- and T2-weighted MRI are used as the source images. Three schemes are employed: the proposed methods based on \( L^2 \)-norm and on dynamic range constraint, respectively, and the scheme based on 3D separable DWT. The resulting slices are shown in Fig. 4. The results based on the \( L^2 \)-norm and on the DWT-based scheme are both shown using the ‘re-scaling’ as defined in (19). From the first two rows, it is clear that the fat behind eyes in the T1-weighted and calcification in T2-weighted have been well combined in the fusion results. The contrast of dynamic-range-constraint solution is a little larger than the \( L^2 \)-norm solution. As far as the third row DWT-based solution is concerned, it is quite blurred and the fluctuations, or ‘ringings’, are very serious and the characteristics of the source images are not well combined at all. Both solutions of the first-order fusion scheme are clearly superior to a DWT-based solution.

The results of T2-weighted and PD MRI images being used as the source images are shown in Fig. 5. The structural information in PD and calcification in T2-weighted images are fused into results using the two first-order schemes proposed in this paper. They reflect the respective characteristics of the source images and are much better than DWT-based schemes.
The proposed first-order fusion schemes can also be used when the total number of bands is greater than two. Here the case of three bands is demonstrated. All the bands of T1-, T2-weighted and PD MRI are used as the source images and the results are shown as Fig. 6. The different features that the sources conveyed are well fused into the results. The reflected details may help with medical analysis and diagnosis. However, in the results derived by our proposed schemes, there are still some problems. If the boundary area is quite smooth, it runs short of variational information in these areas as can be seen in Figs. 4–6. According to (16)–(18), we can find that in the boundary areas, the definition of the Laplacian operator uses fewer voxels and this is even more serious in corner areas. This definition will cause fewer constraints in the boundary areas. So the result in the boundary areas is not very satisfying. This corner problem may be ameliorated if we make an appropriate initial guess for \( f^0 \) in (15) and (22). One possible guess for \( f^0 \) is to choose it as the average of all source bands, 
\[
    f^0 = \frac{1}{n} \sum_{k=1}^{n} s_k,
\]
where \( n \) is the number of source bands, and the fusion result must embody the majority information of sources. Fortunately, for the use of medical images, these boundary and corner areas have less useful information in general and this corner problem will not typically decrease the value of the fusion results provided by the proposed method.

### 4.3 Future directions for study

There are still quite a few aspects of our proposed method deserving further discussion and improvement for future study as briefly listed below.

- The choice of a more reasonable metric \( g \), which may be different at different voxels, and can reflect the source feature differently and reasonably.
- Regarding the corner problem, the basic solution is to provide additional boundary conditions. What kind of boundary conditions might these be?
- The auxiliary function \( \psi \) in Section 2.2 may be chosen more theoretically, not only practically.
- A multiscale technique can be introduced into this method that will agree with the human visual system more appropriately.

### 5 Conclusion

In this paper, we have presented a definition of contrast in a general multi-band volumetric image and showed how our definition agrees with the standard gradient in the case of a single-band volumetric image under the Euclidean metric. The contrast field is constructed as the target contrast and the ‘optimal’ image to the target contrast is obtained by a mathematical formulation via variational approach. Next, the case of dynamic range constraint was discussed. This first-order scheme showed an obvious advantage over the existing DWT-based one and is capable of producing consistently high-quality visualisation of multi-band volumetric images. The enhanced feature of this fusion is a powerful aid in medical analysis and diagnosis. A number of examples and comparisons supporting our results are also shown.
Fig. 4  Some slices of the volumetric fusion results from T1- and T2-weighted MRI
From top to bottom: $L^2$ norm solution, dynamic-range-constraint solution and 3D wavelet solution [10]. From left to right: slice 42, 92 and 142

Fig. 5  Some slices of the volumetric fusion results from T2-weighted and PD MRI
From top to bottom: $L^2$ norm solution, dynamic-range-constraint solution and 3D wavelet solution [10]. From left to right: slice 42, 92 and 142
Fig. 6 Some slices of the volumetric fusion results from T1-, T2-weighted and PD MRI
From top to bottom: $L^2$ norm solution and dynamic-range-constraint solution. From left to right: slice 42, 92 and 142

6 Acknowledgments

The authors are grateful to the anonymous reviewers and editors for their kind comments and constructive suggestions. They also thank Dr. Yong Su (USTC) for occasional discussion on mathematical issues.

7 References


8 Appendix: 1D example clarifying the origin of ‘ringing effect’

In Section 4.2, from the fusion results based on the 3D-DWT method in [10], as shown in Figs. 4 and 5, the fluctuations, or ‘ringing effects’, are very serious and the result is lacking. Here we employ a 1D case to clarify the origin of such a ‘ringing effect’ as shown in Fig. 7.

The source signals $x$ and $y$ are impulse signals and the locations of the impulses are two points away from each
other. After filtering by lowpass and highpass filters, respectively, corresponding to a wavelet of db8, the source signals $x$ and $y$ are decomposed into four components: $x_{\text{low}}$, $x_{\text{high}}$, $y_{\text{low}}$ and $y_{\text{high}}$. Note that the wavelets are fluctuating and so are both the filters. Thus the low-frequency and high-frequency component of a single impulse are both fluctuating; for example, $x_{\text{low}}$ and $x_{\text{high}}$. The fused low-frequency component and high-frequency one are obtained via the MAA rule. This kind of non-linear operation will omit some components in the decomposed ones (i.e. $x_{\text{low}}$ and $x_{\text{high}}$ could not be exactly preserved into the fused one and neither could $y_{\text{low}}$ and $y_{\text{high}}$). When the fused components are input to IDWT-filters, the omitted component will generate fluctuations because of the fluctuating IDWT filters. This is essentially a Gibbs-like phenomena.

In addition, when the scale becomes coarser, these kind of fluctuations will occur at points further away from the impulse position of $x$ and $y$. This is a difficulty with all DWT-based fusion methods.

**Fig. 7** Origin of ‘ringing effect’ in DWT-based fusion: a 1D case