EPSRC

Engineering and Physical Sciences Research Council

InFoMM

Industrially Focused Mathematical Modelling

EPSRC Centre for Doctoral Training in Industrially Focused Mathematical Modelling

Compressed sensing and matrix completion algorithms for demosaicing of spectral imagery

Giancarlo Antonucci

UNIVERSITY OF OXFORD

[dstl]

Contents

1 Introduction

Multispectral imaging is the acquisition of a scene by taking multiple images at different frequency bands. DSTL want to apply it for target identification purposes.

Multispectral imaging (MSI) is the acquisition of a scene by taking multiple images at different frequency bands. It has numerous applications in defence; it is used, for example, for the automated detection of targets based on the materials they are made of.

Cameras can capture such images using either a different sensor for each spectral band or just one sensor for the full spectrum. In the first case, we often have to make severe compromises on the frame rate (time resolution) in order to achieve the desired spectral resolution. In the second case, we can achieve higher frame rates at moderate spatial and low spectral resolutions, which is ideal for certain applications of interest for the Defence Science and Technology Laboratory (DSTL), such as identifying targets.

In order to get the full scene in one exposure, the sensor is covered with a multispectral filter array (MSFA), which allows only one frequency band to be measured at each pixel. Usually, an MSFA consists of a repeated pattern of pixels - which we call a *moxel* - arranged in a checkerboard. Hence, the scene is undersampled according to the location of the MSFA's pixels, and the image needs to be recovered. This process is known as *demosaicing* and is similar in principle to the estimation of RGB components in single-sensor colour cameras fitted with a Bayer colour filter array (CFA). However, whereas CFA demosaicing is a well-studied problem [1], MSFA demosaicing is still in its primitive stages, as it seems that simply extending CFA demosaicing methods does not always give the best results.

Our aim is to find a good MSFA demosaicing algorithm that can be used by DSTL as the main image reconstruction procedure underlying their target identification routines.

Glossary of terms

- **multispectral image:** a 3D cube, each slide corresponding to a different frequency band.
- **channel:** a 2D slide of a multispectral image that corresponds to one frequency band.
- **demosaicing:** the process of reconstructing a multispectral image from partial information obtained fitting a camera with an MSFA.
- **wavelet:** a wave-like oscillation that can be used to extract information from data.
- **sparsifying transform:** an operator that transforms a multispectral image into an array with few non-zero entries.
- **rank:** an integer number describing how much redundant information is contained in a matrix; the lower the rank is, the higher the redundancy is.

2 Conventional methods

Although highly intuitive, most demosaicing methods are heuristics that do not allow much flexibility in their design.

The majority of the most widely used demosaicing methods proposed in the literature are inspired from popular reconstruction techniques for colour images. Being heuristic in nature, they are highly intuitive, but they do not allow much flexibility in their design. The main demosaicing methods are:

Weighted bilinear interpolation (WB), in which every channel is interpolated independently, taking into account only the spatial correlation, so as to estimate the missing values at each pixel thanks to a 2D interpolation of the neighbouring values. A 3D cube is obtained, which is our estimated multispectral image of the scene.

Spectral difference (SD), which relies on WB but takes into account the spectral correlation, based on the assumption that all frequency bands are correlated at every pixel.

Intensity difference (ID), in which we define the intensity at each pixel as the average value over all the frequency bands. We then use a procedure similar to SD to obtain the multispectral image.

Discrete wavelet transform (DWT), which is an approach that uses similar Fourier-based ideas to the CFA demosaicing. We start by using WB to get an initial estimate of the multispectral image, which we then transform via DWT using some wavelet basis in a cube containing the walevet coefficients relative to certain spatial frequency sub-bands. Next, we follow the widely used *replace rule* by assuming that the wavelet coefficients in the high frequency sub-bands are the same for every band, and thus assign the same high frequency coefficients to each band. Finally, we use these coefficients to get our final multispectral image.

Binary tree-based edge-sensing (BTES), in which we determine the missing values iteratively using both known and previously estimated values, interpolating each channel independently like in WB, but using a tree-like structure to exploit the different amount of information, that is the *probability of assignment* (POA), contained in the different channels.

It is important to note that not all the MSFAs are compatible with BTES, strongly limiting the applicability of this method.

3 Compressed sensing

MSFA demosaicing belongs to the more general category of techniques known as *signal processing*, in which we try to infer quantities of interest from measured information. In particular, we want to reconstruct a signal x from a vector of measured data y via $Ax = y$, where A models the linear measurement (information) process.

When the length of the signal is larger than the amount of data we have, classical linear algebra tells us that the system is undetermined and that there are infinitely many solutions (provided that at least one exists). In other words, it is impossible to recover x from y without additional information. This fact relates to the Nyquist-Shannon sampling theorem [2], which states that the sampling rate of a continuous-time signal must be twice its highest frequency in order to ensure reconstruction.

However, under the assumption that the signal is *sparse*, i.e. most of its components are zeros, the reconstruction is possible. The research area associated to this phenomenon is known as compressed sensing [3]. We aim to employ compressed sensing techniques to tackle demosaicing.

Demosaicing can be obtained via compressed sensing by transforming the image in some sparse basis. This is equivalent to reconstructing the image using a few strokes of a brush.

Natural images are not usually sparse in the canonical pixel basis. However, they often consist of smooth areas, except for curve-like discontinuities that appear over the edges in the picture. So, instead of using pixels, one could describe such images in terms of "strokes of a brush" [4] and paint them using a few strokes. In other words, natural images are often sparse after a suitable transformation.

We represent a multispectral image as a vector, say z, built by stacking together all the columns of each channel. In this way we can write $x = \Psi z$, where Ψ is the sparsifying transform and x the corresponding sparse vector of coefficients. Additionally, we define a subsampling operator P_Ω that selects only those pixels and bands of the full image for which we have information. Since both the sparsifying transform and the subsampling operator we choose are linear, we can rewrite the measurement matrix as $A\Psi = P_\Omega$, and $y = P_\Omega z$.

Hence, we try to find that vector of coefficients x with the fewest possible number of non-zero entries such that $Ax = y$ is satisfied. Unfortunately, this problem is hard to solve in general. Nonetheless, research on compressed sensing has been conducted extensively. The most popular reconstruction methods are usually divided into three categories [3]: convex optimisation methods solving a relaxation of the problem, greedy methods such as orthogonal matching pursuit, and thresholding-based methods like iterative hard thresholding.

Moreover, we need to make sure that the reconstruction error stays under control when the vectors are not exactly sparse and when the measurements y are slightly inaccurate. To this aim, we solve our compressed sensing problem using a recovery procedure called **normalised iterative hard thresholding (NIHT)** [5]. In every iteration, we update the solution using two alternating steps: (i) a line search that minimises the goal function but may depart from a sparse vector x and (ii) a hard-thresholding projection onto a set of sparse vectors.

In matrix completion, we aim to reconstruct an image by assuming it has a low-rank structure.

4 Matrix completion

Rather than recovering a sparse vector x, another possibility is to recover a matrix X from its known elements. This is called *matrix completion*, in which sparsity is replaced by the assumption that X has low rank. Indeed, the small complexity of the set of matrices with given low rank compared to the set of all matrices makes the recovery of such matrices plausible.

For a subsampling operator P_Ω that selects only those elements of X for which we actually have some measurements, suppose that we are given a measurement vector $y = P_\Omega(X)$ of size smaller than the total number of elements in the matrix X. The task is to reconstruct X from y. To stand a chance of success, we assume that X has a rank much smaller than its number of elements, and try to find that matrix X with low rank that satisfies $y = P_\Omega(X)$. Again, this problem is really hard to solve in general.

Similar to the case of iterative hard thresholding algorithms for compressed sensing, we alternate between the minimisation of the goal function and the projection of the matrix onto the space of low-rank matrices. However, note that we wish to reconstruct multispectral images, which are 3D cubes. Hence, our task is more complicated than just directly applying a matrix completion algorithm, which can deal only with 2D matrices: contrary to the matrix case, the notion of low-rank 3D cube is ambiguous, especially because a decomposition like the singular value decomposition does not exist. Also, the higher dimensionality is an issue.

Therefore, in order to simplify our problem, we vectorise every channel and concatenate the resulting vectors horizontally, so as to represent every multispectral image by a 2D matrix X. Our aim is then to obtain X, which we call the *spectral unfolding* of a multispectral image. The main drawback of this formulation is that we lose some of the intra-channel correlations, but at least we can apply matrix completion.

The dominant computational cost of iterative thresholding algorithms for matrix completion comes from the need to compute the singular value decomposition in every iteration. However, this can be avoided if we ensure that the optimisation stays on the space of low-rank matrices. One way to do this is to consider those low-rank matrices of the form $X = WZ$. Solving our matrix completion problem in terms of W and Z is a least squares problem, which we solve employing the **alternating steepest descent (ASD)** method proposed in [6].

5 Numerical experiments

In this section, we present and explain the numerical results we obtained using the methods described in Sections 2-4. All the algorithms are implemented in a Python package.

Data

The data were collected in a trial of the utility of airborne hyperspectral imaging for target detection that was conducted at DSTL Porton Down in August 2014. From these, we select 140 radiance images, each with 3752×1600 pixels and 44 bands in the VNIR range (400 nm to 1000 nm).

Here, we focus on a single image on which we implement an IMEC snapshot mosaic sensor which contains a 4×4 MSFA with moxel

4	5	12	13
3	6	11	14
2	7	10	15
1	8	9	16

. (1)

It detects 16 spectral bands using Fabry-Perot filters [7] with sensitivity ranging from 470 nm to 630 nm and a spectral bandwidth (FWHM) of approximately 15 nm for each band. The active area of the sensor is composed of 2048×1088 pixels of 5.5 µm across corresponding to 512 × 256 moxels. Results for this case are consistent with the remaining images in our dataset of 140 VNIR images.

Correlation coefficients

Most MSFA demosaicing schemes make some assumptions on the correlation in and between bands. To assess the spatial correlation within a channel, we use the Pearson correlation coefficient [8] between the channel value of every pixel and that of its right neighbour at some distance, for every channel. Similarly, we measure the spectral correlation between two channels using the Pearson coefficient.

The Pearson correlation metric can be interpreted in several ways: from a statistical point of view, it is the covariance of the two variables divided by the product of their standard deviations; from a geometric perspective, it is the cosine of the angle between the two observed quantities in multidimensional space. For a large sample size, the Pearson correlation coefficient is (approximately) unbiased, and thus it provides an accurate estimate of the correlation. Moreover, it is closely related to the inverse of the Mahalanobis distance [9], a quantity that plays a major role in the design and evaluation of target detection algorithms.

The dataset images have a low spatial correlation and a high spectral correlation.

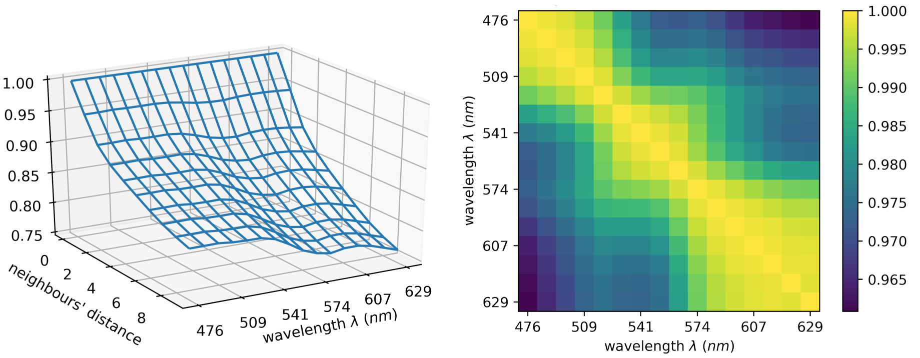

Figure 1 – Spatial (left) and spectral (right) correlation coefficients.

We compute these coefficients on the image we selected and show the results in Figure 1. We see that the spatial correlation between neighbouring pixels decreases as their distance increases. On the other hand, even if three blocks are clearly distinguishable, the spectral correlation between the channels is extremely high overall (in fact, bigger than 0.96). This confirms the reliability of the assumptions of the conventional methods that we previously described. Furthermore, the results in Figure 1 suggest that the spectral unfolding is likely to be extremely low rank, thus making the matrix completion approach potentially very well-suited.

Measure of the quality of a reconstruction

The most common measure of the quality of an image estimated by demosaicing is the *peak signal-to-noise ratio* (PSNR), an engineering term for the ratio between the maximum possible power of (the channel of) an image and the power of corrupting noise that affects the fidelity of its representation, computed in terms of the average squared difference between the pixel intensities of the reference image and its reconstruction.

Results

We show our results in Figure 2, in which we compare the various methods. We see that DWT performs the worst, possibly because the replace rule is just too simplistic. Better results are achieved by WB and BTES, which rank very similarly: this happens because the moxel (1) associates the same POA to every pixels, and thus the tree structure of BTES does not provide any advantage in the interpolation procedure. ID and SD perform even better: this was expected, since both techniques take into account the spectral correlation between the channels. However, SD is the best performing method: the high spectral correlation between all the frequency bands shown in Figure 1 implies that SD should be extremely accurate in this instance, since SD is based on the assumption that all channels are correlated at every pixel.

Hence, we use the output of SD as the initial guess for both NIHT and ASD. We can see that the former does not improve much over SD - up to 4% on average. The improvement is not large: the high redundancy of the measurement matrix A and the regularity of the mosaic pattern do not let us explore the full structure of the multispectral image. In other words, we need as many points as possible to converge to an accurate reconstruction.

Using the output of the best conventional method, the matrix completion approach seems to be the best demosaicing algorithm here.

On the other hand, the results we show in Figure 2 suggest that ASD is an efficient reconstruction method when employed in combination with SD. Despite the spectral unfolding, images of fields are better represented with low-rank matrices as they possess less directional features, and thus ASD applies well to our image. In particular, we find that rank = 3 seems to give the best reconstruction. This is consistent with Figure 1, as the high correlation between the channels suggests a low-rank structure of our spectral unfolding, as well as with the fact that 95.27% of the full radiance information is contained in the first 3 singular values of the spectral unfolding.

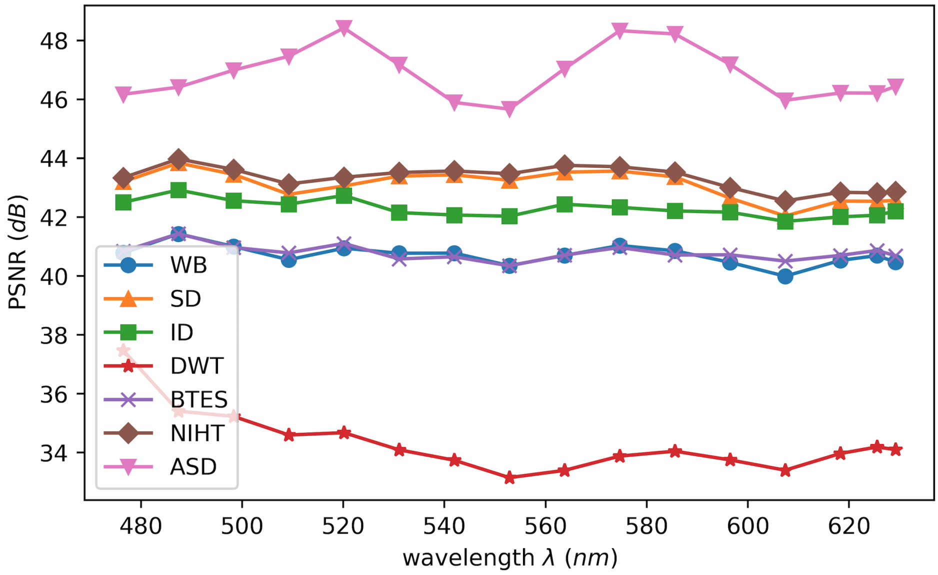

Figure 2 – PSNR of each channel given its band centre.

6 Discussion, conclusions & recommendations

We have described several state-of-the-art methods for demosaicing. We introduced compressed sensing and matrix completion techniques in the multispectral imaging context, and analysed the reconstructions of the subsampled images using different values of the parameters involved. We showed that there is a relevant multidimensional structure that can be used to reconstruct the image from far less observations than one would initially assume to be needed based on the ambient dimensionality of the image tensor. The successful results of matrix completion on the spectral unfolding of the test images give us hope that techniques like ASD can be employed by DSTL to achieve very good reconstruction, which would not otherwise be possible by using conventional methods. Finally, we implemented all the algorithms in a Python package which can be easily used by DSTL for further testing.

Future work should include further improvements of the methods we have studied. In particular, in the compressed sensing case, we need to consider less redundant transforms, e.g. wavelets. For the matrix completion case, a key action is to improve ASD.

Nevertheless, we note that the performance of the algorithms are highly case dependent. Therefore, it is likely that more sophisticated tensor reconstruction techniques that take both the spatial and spectral information into account will be required, as described in [4].

7 Potential impact

Our results will help DSTL choosing a demosaicing algorithm for target identification. Our findings suggest that ASD in combination with SD achieve good performance, showing a competitive edge above state-of-the-art methods.

Jonathan Piper, Senior Scientist at DSTL, said: "*This project is a critical step in developing effective, low-size, weight and power spectral sensors for use in military applications. Mosaic filter-based imagers will enable deployment of spectral sensing on a much wider range of platforms, such as small spacecraft and solar powered aircraft, where previous sensing technologies were impractical. This brings the benefits of spectral sensing, which can detect militarily-significant objects that are indiscernible to the human eye, to a much wider range of applications. The results of this project will be used in data exploitation systems for such sensors, enabling them to meet their full potential for collecting intelligence and performing surveillance or reconnaissance, and enabling users to have confidence in the quality of the information generated.*"

References

[1] B. K. Gunturk, J. Glotzbach, Y. Altunbasak, R. W. Schafer, and R. M. Mersereau. Demosaicking: Color filter array interpolation. *IEEE Signal Processing Magazine*, 22(1), 2005.

[2] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[3] S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. *Applied and Numerical Harmonic Analysis*, 2013.

[4] S. Vary. Sparse approximation and low-rank models of multispectral imaging. Master's thesis, University of Oxford, 2016.

[5] T. Blumensath and M. E. Davies. Normalized iterative hard thresholding: Guaranteed stability and performance. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):298–309, 2010.

[6] J. Tanner and K. Wei. Low rank matrix completion by alternating steepest descent methods. *Applied and Computational Harmonic Analysis*, 40(2):417–429, 2016.

[7] G. Hernandez. *Fabry-Perot Interferometers*. Cambridge University Press, 1986.

[8] S. Mihoubi, O. Losson, B. Mathon, and L. Macaire. Multispectral demosaicing using pseudo-panchromatic image. *IEEE Transactions on Computational Imaging*, 3(4):982–995, 2017.

[9] D. Manolakis, D. Marden, and G. A. Shaw. Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, 14(1):79–116, 2003.