

Hyperspectral video restoration using optical flow and sparse coding

Ajmal Mian^{1,*} and Richard Hartley²

¹Computer Science and Software Engineering, The University of Western Australia
35 Stirling Highway, Crawley, WA 6009, Australia

²Research School of Engineering, Australian National University
Canberra, ACT 0200, Australia

*ajmal@csse.uwa.edu.au

<http://www.csse.uwa.edu.au/~ajmal/>

Abstract: Hyperspectral video acquisition is a trade-off between spectral and temporal resolution. We present an algorithm for recovering dense hyperspectral video of dynamic scenes from a few measured multispectral bands per frame using optical flow and sparse coding. Different set of bands are measured in each video frame and optical flow is used to register them. Optical flow errors are corrected by exploiting sparsity in the spectra and the spatial correlation between images of a scene at different wavelengths. A redundant dictionary of atoms is learned that can sparsely approximate training spectra. The restoration of correct spectra is formulated as an ℓ_1 convex optimization problem that minimizes a Mahalanobis-like weighted distance between the restored and corrupt signals as well as the restored signal and the median of the eight connected neighbours of the corrupt signal such that the restored signal is a sparse linear combination of the dictionary atoms. Spectral restoration is followed by spatial restoration using a guided dictionary approach where one dictionary is learned for measured bands and another for a band that is to be spatially restored. By constraining the sparse coding coefficients of both dictionaries to be the same, the restoration of corrupt band is guided by the more reliable measured bands. Experiments on real data and comparison with an existing volumetric image denoising technique shows the superiority of our algorithm.

© 2012 Optical Society of America

OCIS codes: (100.4145) Motion, hyperspectral image processing; (100.3020) Image reconstruction-restoration.

References and links

1. D. Kittle, K. Choi, A. Wagadarikar, and D. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Appl. Opt.* **49**, 6824–6833 (2010).
2. M. Shankar, N. Pitsianis, and D. Brady, "Compressive video sensors using multichannel imagers," *Appl. Opt.* **49**, B9–B17 (2010).
3. A. Wagadarikar, N. Pitsianis, X. Sun, and D. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Opt. Express* **17**, 6368–6388 (2009).
4. Y. Pati, R. Rexaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27th Asilomar Conference on Signals, Systems, and Computers* (IEEE, 1993), 40–44.
5. R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996).
6. M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.* **54**, 4311–4322 (2006).

7. M. Elad, and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Process. **15**, 3736–3745 (2006).
8. M. Elad, and G. Sapiro, "Sparse representation for color image restoration," IEEE Trans. Image Process. **17**, 53–69 (2008).
9. H. Othman, and S. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," IEEE Trans. Geosci. Remote Sens. **44**, 397–408 (2006).
10. S. Bourguignon, D. Mary, and E. Slezak, "Sparsity-based denoising of hyperspectral astrophysical data with colored noise: Application to the MUSE instrument," in 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (IEEE, 2010), 1–4.
11. G. Farneback, "Two-frame motion estimation based on polynomial expansion," in Proceedings of the 13th Scandinavian Conference on Image Analysis (Springer, 2003), 363–370.
12. J. Parkkinen, J. Hallikainen, and T. Jaaskelainen, "Characteristic spectra of munsell colors," J. Opt. Soc. Am. A **6**, 318–322 (1989).
13. J. Mairal, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," J. Mach. Learn. Res. **11**, 19–60 (2010).
14. B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," Ann. Stat. **32**, 407–499 (2004).
15. J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," IEEE Trans. Image Process. **19**, 2861–2873 (2010).
16. P. Ndajah, H. Kikuchi, M. Yukawa, H. Watanabe, and S. Muramatsu, "An investigation on the quality of denoised images," Int. J. Circuits, Systems and Signal Process. **5**, 423–434 (2011).

1. Introduction

Spectroscopy is the measurement and analysis of electro-optical spectra emitted or reflected by an object or transmitted through a medium. When spectral information is measured at multiple spatial points (for example using a rectangular grid), it is known as imaging spectroscopy. Imaging spectroscopy is also referred to as hyperspectral imaging. A hyperspectral image is a data cube with two spatial and one spectral dimension. Measuring this data cube is generally a sequential process. Either 2D spatial images are sequentially acquired at the desired wavelengths (see Fig. 1) or a 1D hyperspectral sensor, simultaneously measuring all wavelengths of interest along a 1D line, is scanned over a scene. In the latter case the sensor, known as a push-broom sensor, is usually mounted on a moving platform like a satellite or aircraft. In this paper, we focus on hyperspectral video (multiple cubes) acquisition using the former technique.

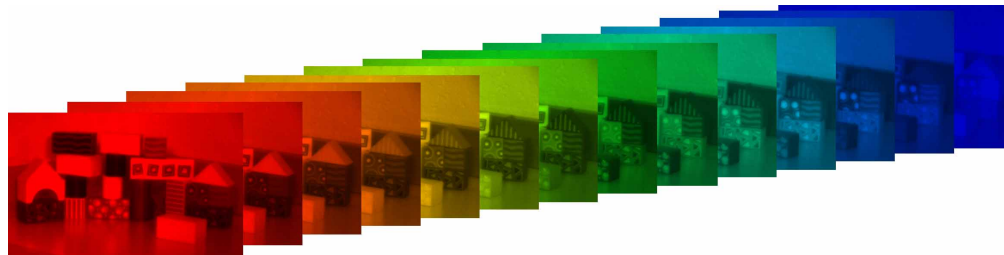


Fig. 1. Sample bands at 690, 650, 620, 610, 600, 590, 580, 540, 520, 510, 500, 490, 480 and 440 nm of a hyperspectral image cube. Each band is rendered as it would be seen by a human eye.

The following terminology is adopted in this paper. A *hyperspectral frame* refers to a hyperspectral data cube acquired at a time instant t and a *band* refers to a single 2D image of a scene within a hyperspectral frame measured at a particular wavelength (see Fig. 2). Hyperspectral video acquisition is a tradeoff between spectral and temporal resolution. We consider the case where only a few bands are measured in each frame so that there is no motion within the bands of a frame and a higher frame rate is maintained. To increase the spectral resolution, the next hyperspectral frame acquires bands with a wavelength offset from the previous frame. Figure 3 illustrates the process. Throughout this paper, it will be assumed that exactly five bands are

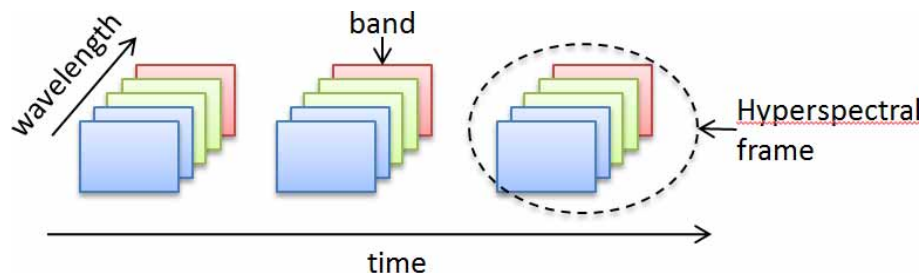


Fig. 2. An illustration of hyperspectral video. There are three hyperspectral frames each with five bands.

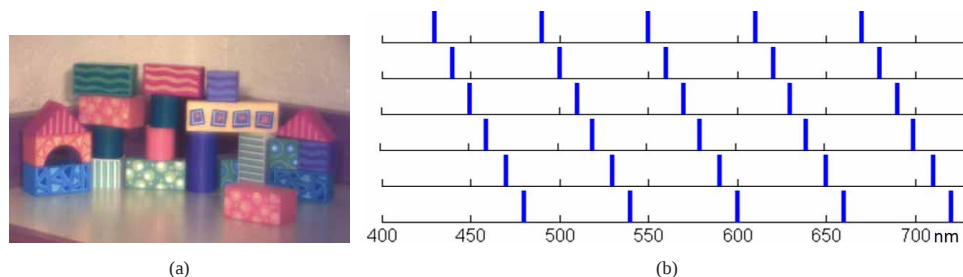


Fig. 3. (a) RGB image of the scene in Fig. 1. (b) Five bands (60nm apart) are sensed in each frame with a between-frame offset of 10nm. Six consecutive frames cover 30 bands (430-720nm) for example the first frame (row) comprises 430, 490, 550, 610, 670nm bands.

measured in each frame and the bands of the next frame are offset by 10nm from the previous ones so that after six frames, 30 bands are sensed covering the range from 430 to 720nm at 10nm resolution. Note that video is useful only for capturing a dynamic scene. A static scene can be fully acquired as a single hyperspectral cube at full spectral resolution without any time constraints.

In this paper, we use the model presented in Fig. 3(b) to restore the dense (30 band) hyperspectral cubes of all the frames. We assume that a frame is acquired instantly hence there is no motion between the bands of a frame. We do not make any further assumptions such as constant velocity, constant acceleration or minimal motion between frames because motion between adjacent frames can be significant. Since most objects (static or moving) are sensed at all wavelengths (bands), in theory their full spectral response can be recovered if correspondences are known between the frames. However, there are three main challenges in achieving this. Firstly, dense correspondence techniques such as optical flow are sensitive to intensity variations between frames. Unlike traditional video, where adjacent frames have similar illumination, the bands of adjacent frames have a wavelength offset in our case, which causes intensity or texture variations between them and makes optical flow more challenging. Secondly, sequential registration of bands that are many frames apart, accumulates optical flow errors. Thus, the resultant spectral response is corrupted at many pixels. Finally, the spectral response of objects that are occluded in some frames or that enter or exit the field of view of the imager is not measured at all bands or wavelengths. In this paper, we address these challenges and propose an algorithm for hyperspectral video restoration. Experiments on real data and comparison with an existing volumetric image denoising technique shows the superiority of our method.

2. Prior work

To the best of our knowledge, this is the first algorithm proposed for hyperspectral video restoration. However, prior work exists on image denoising, volumetric image denoising and RGB color image restoration. From one perspective, our work falls into the category of sparse data acquisition and recovery (or compressive sensing, see e.g. [1–3]) since we acquire only a sparse number of bands at a given instant and the remaining bands are acquired after the scene has changed. However, since correspondences can be established between consecutive frames, we believe that our work is more relevant to image restoration and denoising. This section presents a brief survey of techniques that are most relevant in order to put our work into perspective. We avoid surveying optical flow techniques as our primary contribution is in denoising spectral signals and bands to restore hyperspectral video. Our secondary contribution is the construction of spatio-spectral images, which is likely to improve the accuracy of any optical flow technique.

A set of signals $\mathbf{x} \in \mathbb{R}^n$ is said to exhibit a sparse structure if each signal can be approximated as a linear combination of a few atoms from a dictionary $\mathbf{D} \in \mathbb{R}^{n \times M}$, where M is the number of atoms in the dictionary. The dictionary $\mathbf{D}$ contains prototype signal atoms and is usually overcomplete i.e. the number of atoms is greater than the dimensionality of the signals ($M > n$). An input signal $\mathbf{x}$ is approximated as

$$\hat{\mathbf{x}} \approx \mathbf{D}\hat{\alpha} \quad \text{where} \quad (1)$$

$$\hat{\alpha} = \min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \|\mathbf{D}\alpha - \mathbf{x}\|_2^2 \leq \varepsilon, \quad (2)$$

$$\text{or alternatively } \hat{\alpha} = \arg \min_{\alpha} \{\|\mathbf{D}\alpha - \mathbf{x}\|_2^2 + \gamma\|\alpha\|_0\}. \quad (3)$$

Here $\alpha \in \mathbb{R}^M$ is sparse i.e. it has only a few non-zero elements. The parameter γ sets the trade-off between the approximation and the sparsity of α . The sparsity of α is ensured by minimizing its ℓ_0 pseudo-norm (number of non-zero entries) in the constrained optimization problem of Eq. (2) or its unconstrained version in Eq. (3). Computing ℓ_0 is NP-hard and greedy algorithms such as Orthogonal Matching Pursuit (OMP) [4] are used to solve the above equations. Since the ℓ_1 regularization also results in a sparse solution for α , it is frequently used to formulate sparse coding as a convex optimization problem

$$\hat{\alpha} = \arg \min_{\alpha} \{\|\mathbf{D}\alpha - \mathbf{x}\|_2^2 + \gamma_1\|\alpha\|_1\}, \quad (4)$$

commonly known as the Lasso [5]. Here γ_1 is a regularization parameter similar to Eq. (3).

The choice of dictionary $\mathbf{D}$ is critical, especially for the task of signal denoising and restoration. Aharon et al. [6] proposed the K-SVD algorithm to learn a dictionary in an iterative process. K-SVD alternates between sparse coding of training data based on a current dictionary and updating the dictionary atoms using SVD to better fit the data. Elad and Aharon [7] used the dictionary learned through the K-SVD algorithm to denoise grayscale images. For computational efficiency, the image was divided into smaller overlapping patches and the results were averaged. The main idea was to constrain the denoised image to be close to the original noisy image and to update the dictionary from the noisy image itself. In addition to self-regularization, ℓ_0 regularization was used for sparsity

$$\{\hat{\mathbf{A}}, \hat{\mathbf{X}}\} = \arg \min_{\mathbf{A}, \mathbf{X}} \left\{ \sum_{ij} \|\mathbf{D}\alpha_{ij} - \mathbf{R}_{ij}\mathbf{X}\|_2^2 + \sum_{ij} \gamma_1 \|\alpha_{ij}\|_0 + \gamma_2 \|\mathbf{X} - \mathbf{Y}\|_2^2 \right\}. \quad (5)$$

In this expression, $\mathbf{A}$ represents the set of all α_{ij} where ij runs over all pixels of the image; $\mathbf{Y}$ is the noisy image, $\mathbf{X}$ is its unknown denoised version and $\mathbf{R}_{ij}\mathbf{X}$ is a patch around pixel ij

extracted from image $\mathbf{X}$. This is similar to Eq. (3) except for the last self regularization term which forces the denoised image to be close to the original noisy image. The parameter γ_{ij} controls the relative importance of the sparsity of patch ij and γ_2 controls the relative importance of self-similarity of the complete reconstructed image. This approach worked well for lower noise levels but the results deteriorated rapidly at higher levels of noise. For denoising high dimensional signals such as volumetric images, the same technique [7] can be extended by taking smaller overlapping volume patches. In Section 6, we provide a comparison of our proposed approach with the volumetric image denoising of hyperspectral image cubes and show that we achieve superior results.

Mairal et al. [8] extended the K-SVD grayscale image denoising algorithm [7] to restore RGB color images. The color denoising algorithm follows the original K-SVD algorithm applied to $p \times p \times 3$ RGB patches except for a new projection method in the OMP step. The inner product $\mathbf{y}^T \mathbf{x}$ in the original OMP [4] is replaced with $\mathbf{y}^T (\mathbf{I} + \gamma \mathbf{K}/p) \mathbf{x}$ where γ is an empirically selected control parameter and

$$\mathbf{K} = \begin{pmatrix} \mathbf{J}_p & 0 & 0 \\ 0 & \mathbf{J}_p & 0 \\ 0 & 0 & \mathbf{J}_p \end{pmatrix}, \text{ where } \mathbf{J}_p \text{ is a } p \times p \text{ matrix of ones.} \quad (6)$$

Othman and Qian [9] proposed wavelet shrinkage based hyperspectral image denoising. First, noise is removed in the spatial domain followed by noise removal in the spectral domain which also corrects artifacts resulting from spatial denoising. The algorithm operates on the spectral derivative of the hyperspectral cube. Bourguignon et al. [10] used sparse representations in redundant transformation spaces for denoising astrophysical spectra. They model astrophysical data as the sum of line and continuous spectra and use the ℓ_1 -norm regularization to impose sparsity constraints on their respective canonical and DCT bases. Results are reported on *simulated* data from the MUSE (Multi Unit Spectroscopic Explorer) consortium.

In general, the above algorithms assume that the noise is white Gaussian with zero mean. However, optical flow based registration of hyperspectral frames does not introduce Gaussian noise but artifacts due to non-linear mixing of multiple spectra. Thus, assumptions such as noise being mostly concentrated in the high frequency components of the signal do not hold.

3. Hyperspectral frame registration using optical flow

We use the model proposed in Fig. 3(b) where 5 bands are instantly acquired in a frame. Hence, there is no motion between the bands of a frame. The 5 bands of the next frame are acquired at 10nm offsets and there is significant motion from frame to frame. The aim of optical flow is to find the horizontal and vertical pixel displacements (δ_x, δ_y) such that the error between two consecutive video frames is minimized

$$(\delta_x, \delta_y) = \min_{x,y} \|I_{(x,y)} - I_{(x+\delta_x, y+\delta_y)}\|_2^2. \quad (7)$$

We calculated the pixel displacements with Farneback's two frame optic flow estimation [11] based on polynomial expansion. Each pixel neighbourhood is approximated by a quadratic polynomial $f(\mathbf{x}) \approx \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$ and the displacement between two pixels is estimated as $\mathbf{d} = -\frac{1}{2} \mathbf{A}^{-1} (\mathbf{b}_2 - \mathbf{b}_1)$ where $\mathbf{A}$ is the symmetric matrix of coefficients common to both polynomials and $\mathbf{b}_1$ and $\mathbf{b}_2$ are the vectors of coefficients of their respective polynomials. Farneback also incorporates integration over a neighbourhood, multiscale analysis and an iterative process to increase optical flow accuracy [11]. However, as with other optical flow algorithms, image regions with minimal and ambiguous texture remain the main sources of errors.

Estimating the optical flow between hyperspectral frames with heterogeneous bands introduces an additional challenge because of the different wavelength dependent spectral response

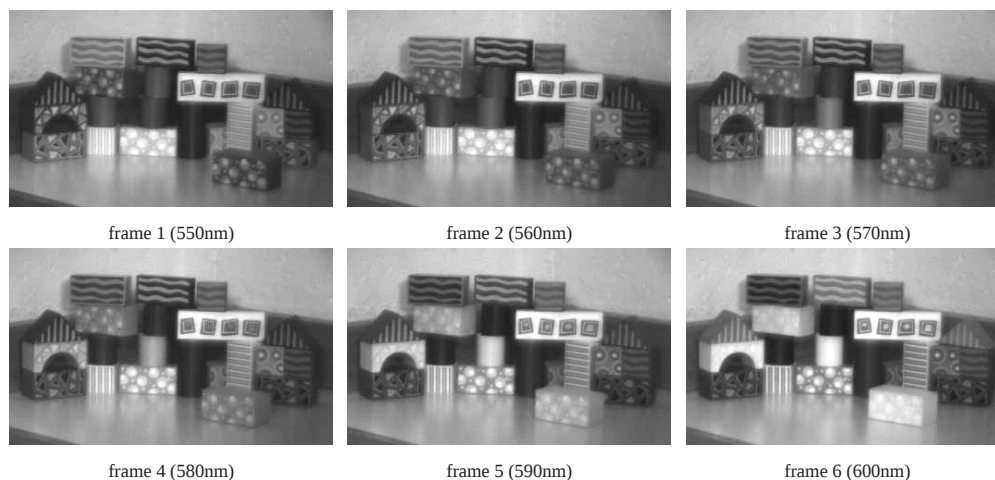


Fig. 4. Third band (rendered as gray scale images) from six consecutive frames of a dynamic scene with static background and a moving block in the foreground. Notice the varying texture which makes it challenging to calculate optical flow between frames/bands.

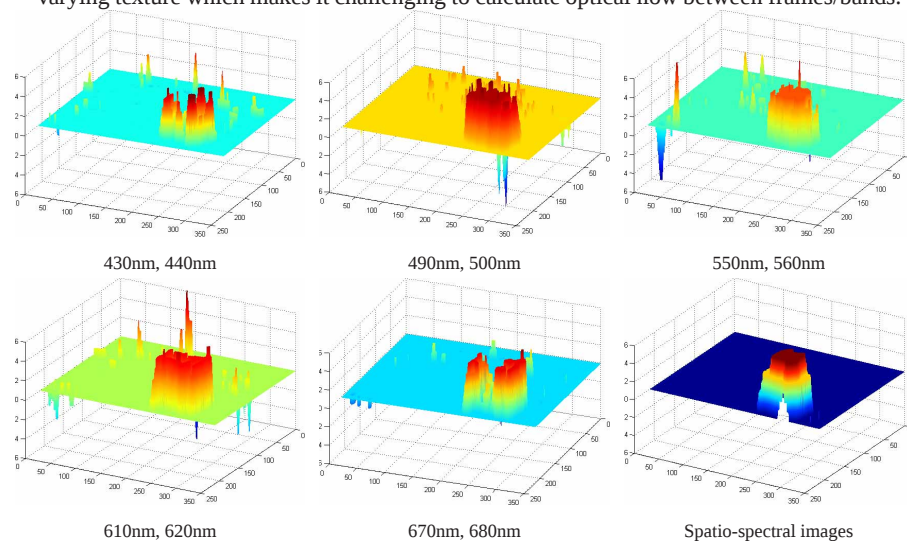


Fig. 5. Optical flow in the horizontal direction for the scene in Fig. 4 represented as 3D plots. The x, y directions correspond to the image dimensions and the vertical direction corresponds to horizontal displacement between frames. The frame bands used to calculate the optical flow are written under each plot. The bottom right plot shows optical flow calculated between five band spatio-spectral images of consecutive hyperspectral frames.

of the scene. Figure 4 shows the third band from six consecutive frames of a hyperspectral video. In addition to changes due to motion (the moving block in the foreground), the intensity or texture of the images is also varying significantly even though there is only 10nm difference between the bands of consecutive frames. The 3D plots in Fig. 5 represent optical flow (horizontal direction only) calculated with Farneback's algorithm [11] for the scene in Fig. 4. Notice that optical flow calculated from pairs of heterogeneous bands contain numerous errors (vertical spikes). Motion is incorrectly found in static regions of the scene and sometimes in the wrong direction on the moving block.

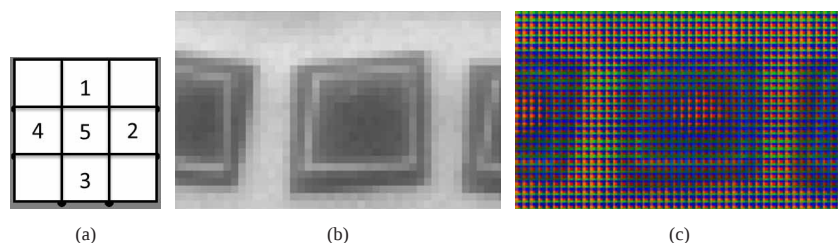


Fig. 6. (a) Ordering of bands in a spatio-spectral image. Each set of 3×3 pixels is formed by ordering the corresponding pixels of the 5 bands. The corner pixels are interpolated from the nearest pixels in the same set excluding the center pixel. (b) A scene patch at 550nm rendered as a gray scale image. (c) A spatio-spectral image of the same patch constructed from 5 bands i.e. 430, 490, 550, 610 and 670nm.

A naive approach is to use a common band between frames for registration. However, this approach decreases the efficiency (frame rate or spectral resolution) by 20% and offers no improvement in optical flow since a single narrow band cannot capture all the texture in the scene. Thus deciding on a common band is a problem in itself as the common band should ideally be scene-specific.

To address the above challenges, we construct a spatio-spectral image by ordering the five measured bands of a hyperspectral frame as shown in Fig. 6(a). Each set of 3×3 pixels is formed by ordering the corresponding pixels of the 5 bands (similar to the Bayer pattern). The corner pixels are interpolated from the nearest pixels in the same set excluding the center pixel. Thus, the spatio-spectral image is nine times larger compared to any single band image. Notice that in Fig. 6(b), the inner orange patch of the square is not distinguishable from its blue boundary at 550nm whereas it is visible in the spatio-spectral image (rendered as RGB in Fig. 6(c)).

Optical flow calculated from the spatio-spectral images is more accurate and has almost no incorrect motion in static regions of the scene (see Fig. 5 bottom-right). Note that an offset of 10nm still exists between the spatio-spectral images constructed from consecutive hyperspectral frames. However, due to their increased textural information, they result in more acute optical flow. Figure 7 shows an example of sequentially registered 540nm bands five frames apart. The left-most image is the measured 540nm band at frame 6, the middle one is sequentially registered using optical flow between five successive pairs of single heterogeneous bands and the right-most is sequentially registered using optical flow calculated from five pairs of spatio-spectral images. Registration based on spatio-spectral optical flow is significantly better. Notice that the block as well as its reflection from the table is distorted in the middle image whereas the distortions of the block and its reflection are minimal in the right-most image.

4. Hyperspectral video restoration

Although, the use of spatio-spectral images increases the accuracy of optical flow, some errors still exist. These errors accumulate and become more obvious after sequential registration of bands that are many frames apart. Notice that some distortions still exist in Fig. 7 when the spatio-spectral images are used for optical flow. The spectral curves at the distorted image pixels are also distorted. Since six frames are registered to complete a hyperspectral cube, each spectral response could be a mixture of up to six different spectra. Pure spectral response will be obtained at pixels with no optical flow errors whereas mixtures of six spectra will be obtained at pixels where optical flow errors exist between all pairs of frames.

Parkkinen et al. [12] measured the visible-range reflectance spectra of the 1257 chips in the



Fig. 7. A 540nm band (left) is sequentially registered from frame 6 to 1 using inter-band optical flow (center) and spatio-spectral image-based optical flow (right).

Munsell Book of Color and reported that the spectra can be well approximated as a linear combination of eight characteristic spectra. This indicates sparsity in the reflectance spectra and the possibility of restoring the spectral response at corrupted pixels as a sparse linear combination of an overcomplete dictionary of spectral atoms. We used training data to learn the dictionary and tested two dictionary-learning algorithms for this purpose, namely the K-SVD algorithm [6] and the online learning algorithm of Mairal et al. [13]. We report results for the later technique, as it performed better.

4.1. Spectral restoration

We propose a spectral restoration model that capitalizes on the fact that the spectral response is *measured* at five out of 30 wavelengths in each hyperspectral frame. Therefore, these five measurements are more reliable compared to the remaining 25, which come from optical flow based registration and may contain errors. Let $\mathbf{s}_{ij} \in \mathbb{R}^{30}$ be the spectral response at pixel i, j of the 30 band hyperspectral frame (cube) obtained from optical flow and $\mathbf{D}_\lambda \in \mathbb{R}^{30 \times M_\lambda}$ be the overcomplete (spectral) dictionary (where M_λ is the size of the spectral dictionary) learned from the static pixels of the hyperspectral frame. Then, according to sparse coding theory the denoised hyperspectral frame can be recovered as

$$\hat{H}_{ij} = \mathbf{D}_\lambda \hat{\alpha}_{ij} \quad (8)$$

where $\hat{\alpha}_{ij}$ are sparse coefficients given by

$$\hat{\alpha}_{ij} = \arg \min_{\alpha_{ij}} \{ \|\mathbf{D}_\lambda \alpha_{ij} - \mathbf{s}_{ij}\|_2^2 + \gamma_1 \|\alpha_{ij}\|_1 \}, \quad (9)$$

and γ_1 is the sparsity regularizer. Note that each vector $\hat{\alpha}_{ij}$ is computed separately. The restored multispectral image $\hat{\mathbf{H}}$ is the ensemble of spectra $\hat{H}_{ij}$ defined at each pixel position ij . Note that Eq. (9) minimizes the ℓ_2 error between the input signal $\mathbf{s}_{ij}$ and its sparse approximation giving equal weights to all dimensions (bands/wavelengths). However, $\mathbf{s}_{ij}$ is more reliable at the measured bands/wavelengths. Moreover, optical flow between neighbouring frames is less likely to have errors compared to optical flow accumulated over five consecutive frames. Therefore, we introduce a weighting term such that

$$\hat{\alpha}_{ij} = \arg \min_{\alpha_{ij}} \{ \|\mathbf{W}(\mathbf{D}_\lambda \alpha_{ij} - \mathbf{s}_{ij})\|_2^2 + \gamma_1 \|\alpha_{ij}\|_1 \}, \quad (10)$$

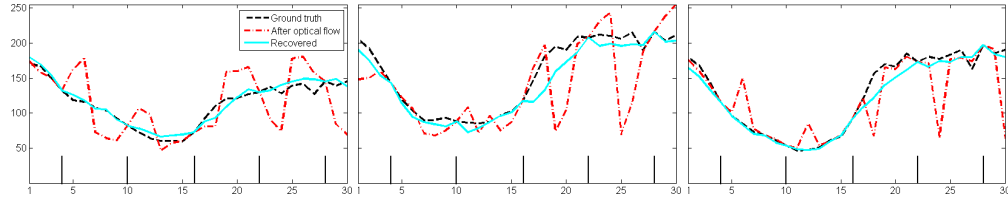


Fig. 8. Optical flow errors lead to incorrect spectral reflectance curves at many pixels (see above examples). Using the proposed technique, the correct spectral reflectance can be recovered.

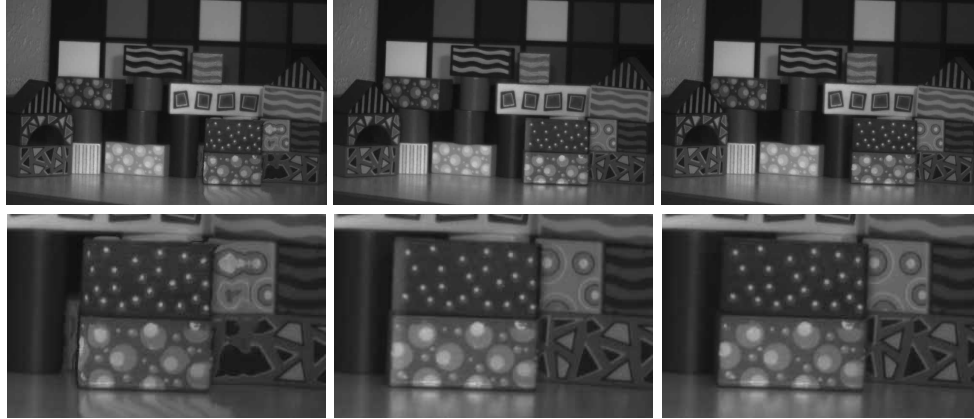


Fig. 9. Left: A 550nm band registered sequentially from frame 1 to frame 6. Center: The errors propagated from optical flow are corrected by spectral restoration. Right: Ground truth 550nm band acquired with frame 6.

where $\mathbf{W} \in \mathbb{R}^{30 \times 30}$ is a diagonal matrix of weights that gives the highest weight to the measured wavelengths followed by those which are estimated from optical flow between nearest frames. Bands/wavelengths that are registered between distant frames get the least weights.

It was observed that improved results were obtained by applying a simple edge-preserving filter to the image prior to the restoration step. The spectral image is mixed with its median-filtered version to remove impulsive noise. Thus, we define

$$\tilde{s}_{ij} = (1 - \gamma_2)s_{ij} + \gamma_2\bar{s}_{ij},$$

where $\bar{s}_{ij}$ is the median of the eight-connected neighbours and γ_2 is a mixing constant that regularizes the relative importance of s_{ij} and $\bar{s}_{ij}$. The median is computed independently in each band. Then, weighting coefficients α_{ij} are computed using this filtered image:

$$\hat{\alpha}_{ij} = \arg \min_{\alpha_{ij}} \{ \|\mathbf{W}(\mathbf{D}_\lambda \alpha_{ij} - \tilde{s}_{ij})\|_2^2 + \gamma_1 \|\alpha_{ij}\|_1 \}, \quad (11)$$

Equation 11 is convex for a known dictionary. The dictionary is learned in the formulation of Eq. (4) using the online dictionary learning algorithm [13]. The spectra in static regions of the scene are used as training data for dictionary learning. Note that $\mathbf{W}\mathbf{D}_\lambda$ needs to be calculated once only. Equation (11) can be solved using the Least Angle Regression (LARS) [14] algorithm. Figure 8 shows examples of recovered spectra using the above model and Fig. 9 shows an example of a band recovered after optical flow based registration from frame 1 to

frame 6.

4.2. Spatial restoration

The bands of a hyperspectral cube have high spatial correlation which is not exploited by the above spectral restoration model. Spectral restoration removes most of the artifacts resulting from optical flow and significantly improves the RMSE. However, minor artifacts still remain especially around the boundaries of moving objects (see Fig. 9 and 10). To remove these artifacts, we propose a guided dictionary learning and sparse representation model that restores each band individually. For each band, we learn two dictionaries, one for the five measured bands and one for the band to be restored, such that the same coefficients can be used to sparsely approximate the measured bands and the band to be restored.

$$\{\hat{\beta}_{ij}, \hat{\mathbf{D}}_{sm}, \hat{\mathbf{D}}_{se}\} = \arg \min_{\beta_{ij}, \mathbf{D}_{sm}, \mathbf{D}_{se}} \{ \|\mathbf{D}_{sm}\beta_{ij} - \mathbf{R}_{ij}\mathbf{H}_m\|_2^2 + \gamma_4 \|\mathbf{D}_{se}\beta_{ij} - \mathbf{R}_{ij}\mathbf{H}_e\|_2^2 + \gamma_3 \|\beta_{ij}\|_1 \}, \quad (12)$$

where $\mathbf{R}_{ij}$ is an operator that extracts $p \times p$ patch(es) from any number of bands (5 measured bands $\mathbf{H}_m$ or a single band to be restored $\mathbf{H}_e$), $\mathbf{D}_{sm} \in \mathbb{R}^{5p^2 \times M_s}$ is the spatial dictionary learned to approximate the $p \times p \times 5$ patches of the measured bands, $\mathbf{D}_{se} \in \mathbb{R}^{p^2 \times M_s}$ is the corresponding dictionary learned to approximate $p \times p$ patches of the band to be restored. By enforcing the sparse coefficients β_{ij} (and hence the size M_s) of both dictionaries to be the same, we establish

a link between the two dictionaries. By combining the two dictionaries $\mathbf{D}_s = \begin{bmatrix} \mathbf{D}_{sm} \\ \delta \mathbf{D}_{se} \end{bmatrix}$ and the

data to be approximated $\mathbf{H} = \begin{bmatrix} \mathbf{H}_m \\ \delta \mathbf{H}_e \end{bmatrix}$ (where δ is a tradeoff parameter), the above expression simplifies to

$$\{\hat{\beta}_{ij}, \hat{\mathbf{D}}_s\} = \arg \min_{\beta_{ij}, \mathbf{D}_s} \{ \|\mathbf{D}_s\beta_{ij} - \mathbf{R}_{ij}\mathbf{H}\|_2^2 + \gamma_3 \|\beta_{ij}\|_1 \}. \quad (13)$$

We set $\delta = 1$ so that the five measured (more reliable) bands get 5/6 weight and the estimated band gets 1/6 weight. Our guided dictionary approach is similar to the work of Yang et al. [15] who used coupled dictionaries for single image based super resolution. However, our formulation is different because we are seeking a sparse representation of data estimated by spectral restoration whereas Yang et al. [15] use a low resolution image patch to predict its high resolution version. Unlike [15], the data to be sparsely approximated is available but contains noise from optical flow in our case and the approximation is guided by the measured bands that are more reliable.

Equation 13 is convex with respect to each variable when the other is fixed. First, the dictionary is learned [13] from part of the hyperspectral frame where motion is not detected and then the region where motion is detected is restored using the same dictionary. Unlike [15], during restoration of a band, the noisy patch as well as the corresponding noise-free patch from the measured bands are both sparsely approximated using the learned dictionary. Thus the restoration process is guided by both five noise-free patches of measured bands and one noisy patch of the band estimated from optical flow followed by spectral restoration. Intuitively this gives better accuracy and was verified experimentally as well by removing the noisy patch during the sparse coding step. Once all the (overlapping) patches have been calculated, the values corresponding to the band to be recovered are averaged. Figure 10 shows a sample band after spectral restoration and spectral + spatial restoration. A magnified view is given in Fig. 11. Notice the improvement around the boundary of the moving blocks.

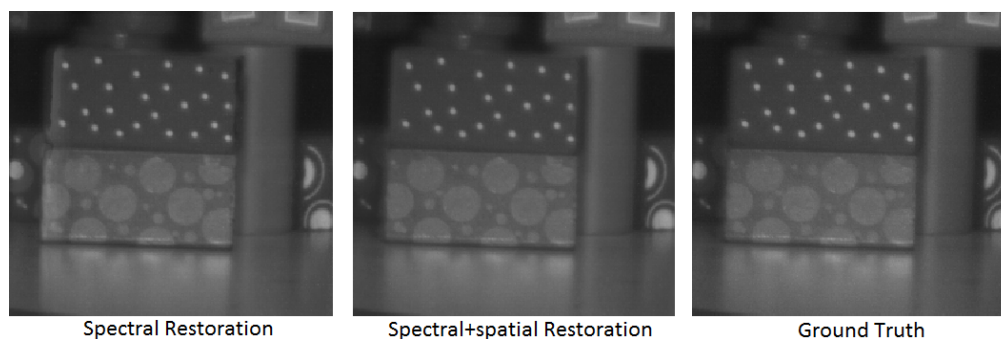


Fig. 10. Spectral restoration of a 490nm band sequentially registered from five frames apart. Some errors can be noticed around the boundaries of the moving blocks which are removed by the spatial restoration.



Fig. 11. Magnified views of the (left most middle part of the) three images in Fig. 10. After spectral+spatial restoration (middle image), the boundaries are better recovered and the image has more resemblance to the ground truth (right image).

5. Experimental setup and data collection

Our hyperspectral imaging system includes a CRi VariSpec Liquid Crystal Tunable Filter (LCTF), a 25mm lens and a Basler sCA750-60fm camera with 752×480 spatial resolution (see Fig. 12). A halogen light was used to illuminate the scene and the Macbeth color checker was used for spectral calibration. The LCTF was tuned and synchronized with the camera using custom software. An image was acquired each time the filter tuned to a different wavelength. The LCTF can be tuned to 33 different wavelengths from 400 to 720nm with 10nm step. Figure 12 shows the transmittance of the LCTF, the camera's CCD sensitivity (both provided by the respective manufacturers) and the spectrum of the halogen light measured with the StellarNet spectroradiometer. The exposure time of the camera was varied during acquisition to cater for the varying LCTF transmittance. Due to low LCTF transmittance, the 400, 410 and 420nm bands were dropped (see Fig. 12(b)) and the last 30 bands were used in our experiments. We collected a sequence of 30 frame hyperspectral video of a static scene with a moving object. A sample image of the scene is given in Fig. 9. The scene was static while the bands of a frame were measured. However, between the frames, an object was moved in the scene. The movement was manually performed and was significant between frames. All 30 bands were

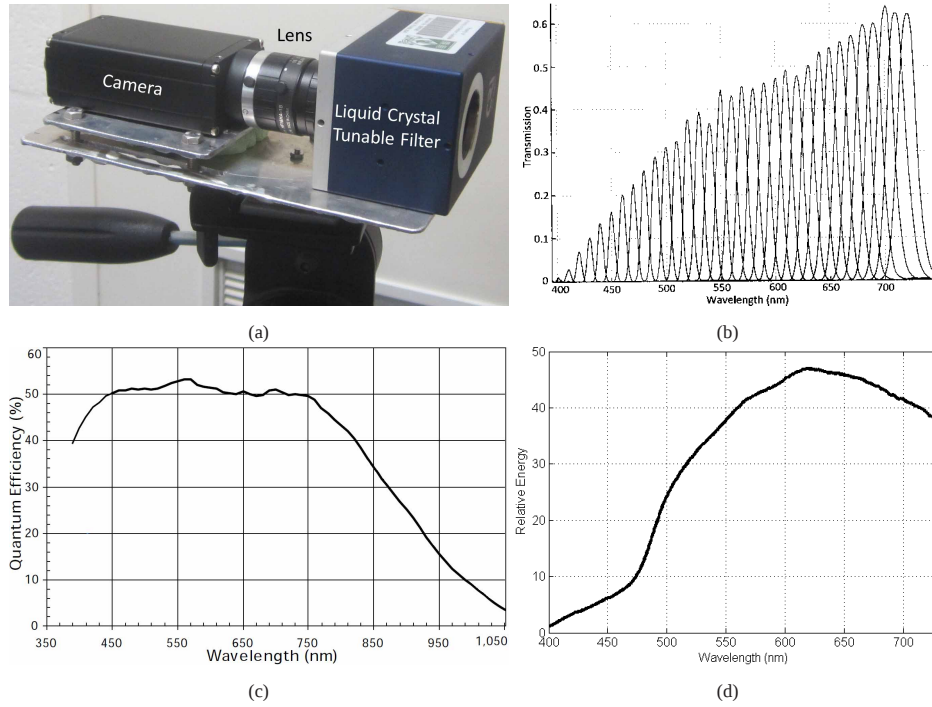


Fig. 12. (a) Hyperspectral camera setup. (b) Transmittance of the LCTF. (c) Quantum efficiency of the camera CCD. (d) Spectral curve of the halogen light.

measured in each hyperspectral frame but as shown in Fig. 3, only five bands per frame were used to recover the full 30 band hyperspectral frames. The remaining 25 bands per frame were used as ground truth for quantitative analysis. This data is available for research purposes on the first author's website.

There are three free parameters in the proposed restoration model. Their values were set to $\gamma_1 = \gamma_3 = 0.15$ and $\gamma_2 = 0.3$. The spectral dictionary size was set to over three times the dimensionality of spectral signal (the number of bands) i.e. $M_\lambda = 100$ so that the dictionary is overcomplete. The patch size p in spatial restoration was set to 3 since a smaller than 3×3 patch does not contain significant spatial information. Accordingly, the spatial dictionary size was set to be over three times the dimensionality of the patch ($3 \times 3 \times 6$, where 6 corresponds to 5 measured bands + one band to be restored) i.e. $M_s = 256$. We also report results for $p = 5$ and $M_s = 700$.

6. Hyperspectral video restoration results

We report quantitative results using RMSE (Root Mean Squared Error) between the recovered hyperspectral cube and the measured ground truth. The RMSE between a recovered hyperspectral frame $H^r \in \mathbb{R}^{u \times v \times n}$ and its corresponding ground truth $H^g \in \mathbb{R}^{u \times v \times n}$ is given by

$$RMSE = \sqrt{\frac{1}{u v n} \sum_{i=1}^u \sum_{j=1}^v \sum_{k=1}^n (H_{ijk}^r - H_{ijk}^g)^2}, \quad (14)$$

where n is the spectral and $u \times v$ are the spatial dimensions. To avoid bias in the results, RMSE was always measured at only those pixels where motion was detected by optical flow. RMSE of the full frames were much lower due to averaging with more reliable spectra in the static

regions. The input frames were divided into static regions and motion regions using two conservative masks obtained from optical flow. The masks ensured that only static regions were used for learning the dictionaries and that RMSE was calculated at the motion pixels.

In the first experiment, we compare optical flow based registration to cubic interpolation (between measured bands) and use the proposed algorithm to restore the hyperspectral frames in both cases. Since learning a dictionary from interpolated bands leads to incorrect sparse representations, additional training data is required in the form of at least one full 30 band hyperspectral cube. We used an additional hyperspectral cube for learning the dictionary for restoration of the interpolated bands. In the case of optical flow (this and all the remaining experiments), the dictionaries were learned from the static regions of the input frame.

Table 1 shows the results of our first experiment. Optical flow gives a smaller RMSE from the ground truth compared to cubic interpolation. Moreover, the proposed restoration algorithm recovers the hyperspectral cubes more accurately from optical flow based registered bands. Nevertheless, it is interesting to see that our algorithm recovers the dense hyperspectral cube with reasonable accuracy from a cube constructed from interpolation between only five measured bands.

Table 1. RMSE of restored hyperspectral frames from measured ground truth. Frames registered with optic flow give better restoration accuracy compared to cubic interpolation.

Frame #	1	2	3	4	5	6	Average
Interpolation	41.44	35.14	26.73	20.96	17.65	24.64	27.76
Restored	7.65	6.27	8.63	8.75	8.66	9.33	8.22
Optic Flow	15.38	12.16	11.92	16.26	16.26	21.57	14.47
Restored	3.94	4.15	4.75	4.73	4.57	5.11	4.54

In the second experiment, we run the proposed restoration algorithm in a loop to find the minimum number of required iterations. Table 2 shows that maximum improvement in RMSE was achieved in the first iteration and the second iteration improved the RMSE in only a few cases where the maximum frame distance of optical flow was higher. However, while widely accepted, RMSE is not the best quality measure for denoised images since it does not take structural distortions into account [16]. Visual inspection shows that the second iteration improves the structural quality of the images but RMSE drops slightly due to blurring which is caused by the second term in Eq. (11) and the averaging of the overlapping patches during the spatial restoration. The results in Table 2 are given for two different patch sizes used in the spatial restoration stage i.e. patch sizes of $3 \times 3 \times 6$ and $5 \times 5 \times 6$. The third dimension is fixed at 6 because each time there are five measure bands and one additional band to be restored. Note that the smaller patch gives slightly better results and is also computationally more efficient to use since a lower dimensional vector needs to be approximated from a smaller size dictionary.

In the last experiment, we compare the proposed algorithm with the K-SVD volumetric image denoising algorithm [7]. For K-SVD denoising, we used the implementation provided by the authors [7]. The output of the optical flow was used as input to both algorithms. In both cases, the (initial) dictionaries were learned from exactly the same regions of the input hyperspectral frames i.e. where motion was not detected by optical flow. Figure 13 shows a sample band recovered with the proposed approach and K-SVD for qualitative analysis whereas Table 3 provides a quantitative comparison of the two algorithms for all 30 hyperspectral frames. The K-SVD algorithm did not perform well in removing the optical flow artifacts and achieved lower performance than our spectral restoration model alone. For comparison of our spatial restoration model, we combined the K-SVD volumetric denoising with our spectral restoration

Table 2. RMSE of recovered hyperspectral frames w.r.t. the number of iterations.

Frame #	Patch size Opt. Flow	$3 \times 3 \times 6$			$5 \times 5 \times 6$		
		Iter 1	Iter 2	Iter 3	Iter 1	Iter 2	Iter 3
1	15.38	3.94	3.82	3.84	4.31	4.28	4.31
2	12.16	4.15	4.59	4.82	4.46	4.96	5.21
3	10.84	4.75	5.49	5.82	5.04	5.81	6.15
4	11.92	4.73	5.31	5.48	4.81	5.49	5.75
5	16.26	4.57	5.09	5.38	4.78	5.19	5.41
6	21.57	5.11	5.32	5.58	5.18	5.27	5.39
7	18.51	6.32	6.36	6.48	5.32	5.19	5.24
8	15.89	5.98	6.32	6.48	4.90	5.16	5.36
9	14.61	5.37	5.77	6.15	5.28	5.92	6.29
10	16.08	5.00	5.27	5.45	5.24	5.56	5.84
11	19.83	4.75	4.74	5.04	4.97	5.14	5.41
12	23.56	5.22	5.05	5.45	5.43	5.45	5.70
13	16.59	4.89	4.60	4.69	4.99	4.92	5.03
14	15.31	4.76	4.65	4.85	4.99	5.13	5.36
15	13.27	5.32	5.82	6.17	5.49	6.05	6.38
16	14.82	5.21	5.34	5.50	5.32	5.51	5.67
17	17.20	5.40	5.46	5.71	5.48	5.50	5.69
18	20.32	7.11	7.15	7.44	7.19	7.15	7.37
Avg	16.34	5.14	5.34	5.57	5.18	5.43	5.64

model in different configurations. The proposed spectral restoration followed by K-SVD volumetric denoising ($\lambda + K_{SVD}$) did not improve the RMSE except for frame number 7, 8 and 18. Overall, the best performance was achieved by the proposed spectro-spatial restoration model. Note that there is more motion between certain frames causing their RMSE to be greater than others.

7. Processing time

All algorithms were tested on a 2.4GHz quad core machine with 32bit operating system and 4GB RAM. The code for hyperspectral video acquisition and the optical flow based registration were implemented in Visual C++. Acquisition time for one hyperspectral frame was 0.66 seconds. Optical flow based registration took 4.04 seconds per hyperspectral frame. The spectral and spatial restoration algorithms were implemented in Matlab and the Sparse Modeling Software (SPAMS) [13] was used for dictionary learning and least angle regression [14]. The time required for spectral dictionary learning and restoration was 6.74 seconds whereas the time required for spatial dictionary learning and restoration was 4.70 seconds per band.

8. Conclusion

We presented an algorithm for the restoration of dense hyperspectral video from a few measured bands per frame. The proposed approach increases the frame rate or spectral resolution of imaging systems by many folds. It exploits the sparsity in spectral response of natural objects and the spatial correlation between images acquired at different wavelengths. The measured bands of each frame are arranged in a Bayer like pattern to make spatio-spectral images which offer better optical flow accuracy. Errors from optical flow are first removed using a spectral

Table 3. Comparison with the volumetric K-SVD algorithm under different configurations.
 λ : proposed spectral restoration only, λ +G: proposed spectral+spatial restoration, K_{SVD} :
K-SVD volumetric denoising [6]. The overall best performance is achieved by the proposed
spectral+spatial restoration λ +G.

Frame #	Opt. Flow	λ	λ +G	K_{SVD}	λ + K_{SVD}	λ + K_{SVD} + λ	$K_{SVD} + \lambda$
1	15.38	4.68	3.94	11.68	5.58	4.51	4.87
2	12.16	4.18	4.15	9.18	5.80	4.89	4.85
3	10.84	4.70	4.75	8.02	6.06	5.40	4.75
4	11.92	4.63	4.73	8.56	6.18	5.70	5.06
5	16.26	5.09	4.57	12.89	6.91	6.44	6.40
6	21.57	6.26	5.11	17.93	7.74	6.21	7.04
7	18.51	10.32	6.32	14.79	9.46	11.65	10.44
8	15.89	8.61	5.98	12.30	8.29	9.90	8.60
9	14.61	6.97	5.37	10.45	7.05	6.85	6.55
10	16.08	6.00	5.00	11.12	6.28	5.63	5.64
11	19.83	5.96	4.75	15.07	6.75	5.40	6.50
12	23.56	7.20	5.22	18.48	7.91	5.87	7.26
13	16.59	6.79	4.89	13.20	6.82	5.89	6.15
14	15.31	6.29	4.76	12.42	6.92	5.52	6.05
15	13.27	6.35	5.32	9.62	7.11	6.44	5.96
16	14.82	6.41	5.21	10.79	6.96	5.96	5.71
17	17.20	7.89	5.40	13.53	8.42	6.86	6.80
18	20.32	13.60	7.11	16.50	13.41	16.51	14.26
19	20.54	7.20	5.49	16.42	7.78	6.99	7.33
20	16.34	5.72	4.74	12.55	6.81	5.96	6.15
21	13.22	5.70	5.37	9.15	6.57	6.39	5.59
22	13.62	5.15	5.03	9.35	6.32	5.57	5.16
23	15.17	5.24	4.96	11.11	6.51	5.29	5.43
24	17.70	5.90	5.32	13.55	7.09	5.62	6.05
25	19.96	7.07	5.93	13.62	7.31	6.48	7.04
26	16.29	6.19	5.70	12.22	6.88	5.92	6.21
27	13.39	6.24	6.64	9.18	6.68	6.68	5.77
28	14.34	5.84	6.29	10.10	6.63	6.13	5.88
29	16.61	6.65	6.74	12.63	7.45	6.22	6.62
30	30.39	8.98	7.79	16.48	9.47	8.04	9.11
Avg	16.72	6.59	5.42	12.43	7.31	6.70	6.64

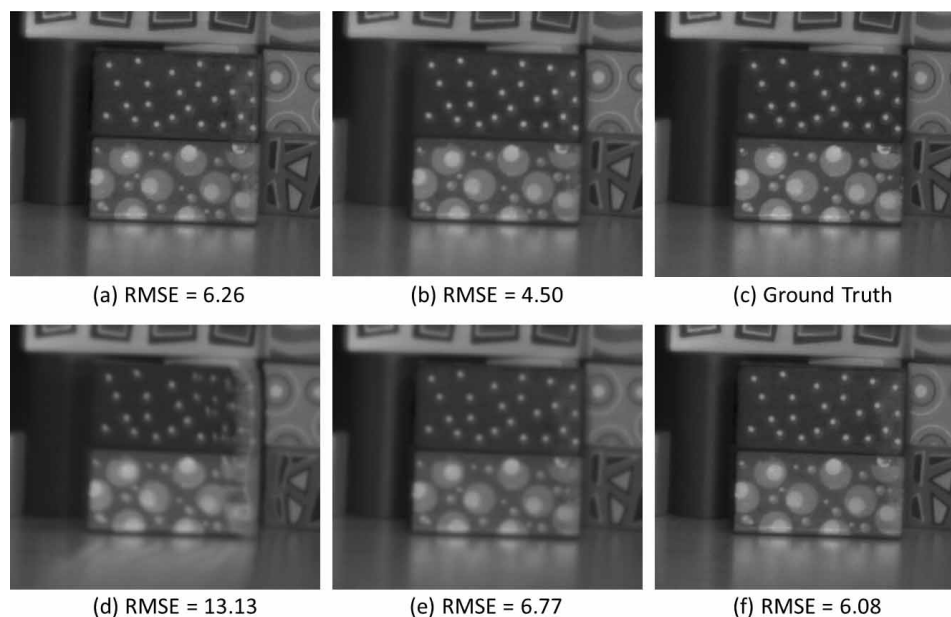


Fig. 13. Comparison with K-SVD denoising. A 560nm band, registered from 5 frames apart is restored with (a) spectral restoration, (b) spectral + spatial restoration, (d) K-SVD volumetric denoising, (e) spectral restoration + KSVD volumetric denoising and (f) K-SVD volumetric denoising + spectral restoration. Measured ground truth is in (c).

restoration model followed by a spatial restoration model. Different formulations of sparse coding are used in both models and the dictionary is learned from regions of the input frame (to be restored) where no motion is detected. Experimental analysis on real data and comparison with an existing state-of-the-art volumetric image denoising technique, under various experimental configurations, shows that the proposed approach consistently achieves higher restoration accuracy. Unlike the majority of image restoration literature, we did not attempt to remove noise or artifacts that had been synthetically introduced in the ground truth images. In fact, we measured the ground truth bands for comparison, which is a more realistic setting.

It is worth mentioning that the number of bands per frame determine a trade-off between accuracy and efficiency. The algorithm will still work with fewer or more number of bands per frame. Fewer bands per frame will increase the hyperspectral video frame rate but will deteriorate the optical flow and restoration accuracy. Similarly, measuring more wavelengths is likely to improve accuracy at the cost of lower frame rate.

Acknowledgment

This research was supported by Australian Research Council grant DP110102399.