Multi-spectral Material Classification in Landscape Scenes Using Commodity Hardware

Gwyneth Bradbury[1]   Kenny Mitchell[2]   Tim Weyrich[1]

[1]University College London   [2]Disney Research

Abstract. We investigate the advantages of a stereo, multi-spectral acquisition system for material classification in ground-level landscape images. Our novel system allows us to acquire high-resolution, multi-spectral stereo pairs using commodity photographic equipment. Given additional spectral information we obtain better classification of vegetation classes than the standard RGB case. We test the system in two modes: splitting the visible spectrum into six bands; and extending the recorded spectrum to near infra-red. Our six-band design is more practical than standard multi-spectral techniques and foliage classification using acquired images compares favourably to simply using a standard camera.

Keywords: Material classification, multi-spectral, vegetation

1   Introduction

Material classification from images is an important task across machine vision and computational photography, with applications in a wide range of domains. One established use is in the classification of foliage and other land cover in remote, multi-spectral, LandSat data. In this case, multi-spectral techniques are key to accurate classification. However, they are also slow, costly and less practical to apply to ground-level image acquisition. Importantly, no mechanism exists to capture multi-spectral images with a hand-held camera in a single shot.

Previously spectral reconstruction has been achieved with a stereo consumer compact camera [10]. Such a system is faster, more practical, and cheaper than a true multi-spectral imaging system and results in six-band, stereo image data. Output is high resolution and images can be captured at high speed or as video.

Several systems are available for multi-spectral 3D imaging spectroscopy for close-range settings [7]. Such implementations often require a laboratory setting making them impractical for use in an uncontrolled setting.

Additionally, spectral multiplexing can be achieved using triple band-pass, dichroic filters, such as those used in Dolby's more recent cinema technology [3].

We explore a novel camera design that aids classification by increasing the spectral information known about a scene. We evaluate the design on the example of foliage classification where, additional colour information is extremely important. In such a case, multi-spectral image acquisition is expensive and impractical while standard RGB imaging lacks accuracy. We create a single-shot,

six-channel, multi-spectral acquisition system. We use this system to address the problem of material classification in ground-level, landscape imagery.

Our system is built using off-the-shelf, commodity hardware: a standard SLR camera, stereo lens and selected filters. We exploit triple band-pass filters extracted from commercial 3D cinema glasses. Each spectral band passed through one of these filters falls into one of the red, green and blue camera response curves. The transmission spectra of the filters themselves are non-overlapping in the visible spectrum, with little information lost between them. This allows capture of additional spectral information in the visible range, splitting the visible spectrum into six bands (rather than the standard three, RGB).

As we demonstrate, foliage classification especially benefits from additional spectral channels due to the very distinct 'spectral footprints' shown by different plant species. Whilst these spectra are dominant in the green region of the visible spectrum (around 550nm), they become much more pronounced in the near infra-red region (upwards of 700nm). This drives us to investigate a second configuration of the system investigating the near infra-red (NIR) spectrum alongside the visible. We implement this using band-pass filters for the regions of the spectrum above and below 720nm. The system allows for reduced error rates for foliage classification in ground-level landscape scenes in both cases.

A significant advantage to our system is its portability and ease-of-use in outdoor environments. As such, we compare our multi-spectral implementation to using a normal SLR camera for image acquisition.

2   Related Work

Multi-spectral imaging typically captures several non-overlapping spectral bands. This non-overlapping property distinguishes it from standard RGB imaging. Whilst hyper-spectral techniques are mostly applied to remote sensing and require specialised equipment such as tunable band-pass filters, multi-spectral systems are both more practical and cheaper, often using exchangeable filters or multiple band-pass filters which can be combined with a standard RGB camera.

Multi-spectral imagery has applications in a number of fields, from recovering the reflectance spectrum at a given point to recording and analysing the Earth's surface vegetation. However, whilst remote sensing systems often incorporate classification methods for analysis of the images, this has been a much lesser used technique for ground-level multi-spectral systems.

2.1   Multi-spectral Image Acquisition

3D multi-spectral image capture can be broadly categorised by three main methods: dispersive elements, diffractive media (diffraction grating) or by filtering with exchangeable or tunable bandpass filters.

Habel et al. [5] present a low-budget solution for multi-spectral photography. By creating a custom lens element using off-the-shelf parts, and with careful calibration, they create a low-cost Computed Tomography Image Spectrometer

(CTIS). Whilst this diffraction-grating implementation yields a high spectral resolution in a single shot, the spatial resolution is extremely limited.

Tsuchida and Tanaka [12] present a 6-band, 2-image multi-spectral system. Given the intensity spectrum of the light source, they estimate spectral reflectance using a large-format camera and interference filters. Image capture is slow, using a scanning sensor to scan the scene line by line. This necessitates a static scene and although high resolution is achieved, the system is costly and impractical for dynamic scenes.

Fyffe et al. [3] demonstrate the use of dichroic, triple-band-pass, interference filters for spectral multiplexing. By altering the light sources for a scene, they obtain dense surface normals and surface colour reflectance from a single shot. This is easily extended to dynamic scenes but is not practical for outdoor scenes where the lighting may not be modified. The use of a beam splitter means that the problem of correspondence between shots is avoided, assuming a negligible disparity.

Finally, spectral reconstruction can be achieved with a stereo consumer compact camera as shown by Shrestha et al. [10]. They present a stereo, multi-spectral configuration focusing on spectral reconstruction rather than the stereo implications of the system.

We take a similar approach, using a stereo lens with tri-band-pass interference filters, as well as investigating the visible and NIR case.

2.2   Imaging for Classification of Vegetation

In the vast majority of vegetation classification tasks, hyper-spectral images are captured remotely and processed using texture filtering across the spectral bands. As plant species have unique spectral footprints in both the green and NIR parts of the spectrum, spectral images which focus on these ranges are often used. Hernandez et al. [6], for example, find that red and short-wave infra-red bands produce the best classification results in satellite imagery.

Also in remote sensing, Yu et al. [14] attain a detailed, object-based classification of simple vegetation classes using images with a high spatial resolution. In this case, blue, green, red and NIR comprise the spectral features used for classification along with band-ratio and spectral-derivative features and other textural and topographic features.

Shading and atmospheric interference effects can be very detrimental to vegetation classification, leading to high variation of the spectral distribution within a class, which cannot be modelled linearly [14]. Hue, however, is dependent on the spectral range but independent of illumination intensity changes [9], making it an important feature for classification [14].

Tanser and Palmer [8] compute a spectral reconstruction of vegetation classes based on ten spectral bands using mean, variation and co-variance parameters, and comparing with a tri-band approach. Their results show that, although more detailed spectral information considerably improves the classification result, it is still insufficient for high quality classification. They then use texture, particularly a Moving Standard Deviation Index, to supplement the classification.

At ground level, laser scanners can be used to create 3D maps of an environment [13] (with particular application in simultaneous localisation and mapping problems) which can be used for classification. Similarly, a ground-level survey can be carried out. This can be more accurate, but is also incredibly inefficient. Other ground-level approaches involving multi-spectral cameras are often impractical because of the less-portable nature of these systems and the need to take many sequential, filtered images of a potentially moving scene.

3 System Design and Image Acquisition

We build two systems with the following hardware. Both feature a twin-aperture, stereo lens (see Figure 1), then either of the following combinations are used: complementary triple band-pass filters (also shown in Figure 1) and standard DSLR; or 730 nm infrared long-pass filter and 400-720 nm band-pass filter and an infrared-modified DSLR.
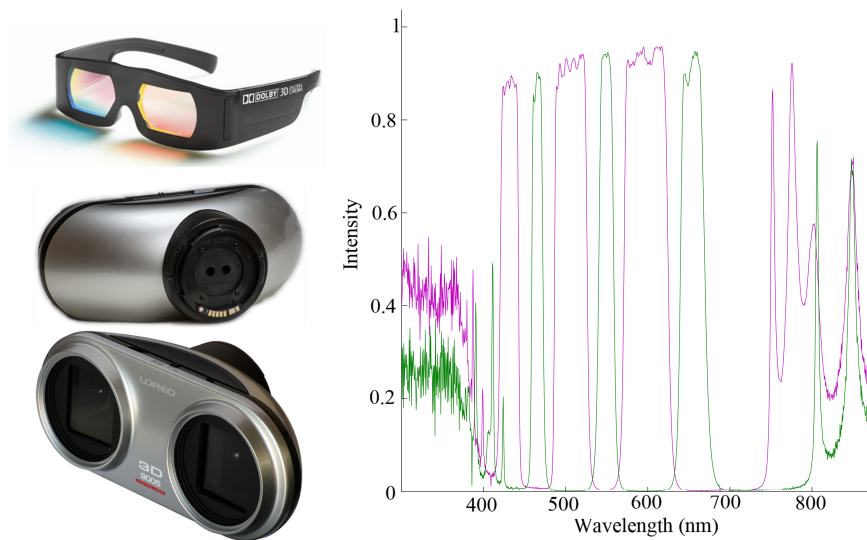

Fig. 1. Left, from top: commercial stereo cinema glasses using dichroic, tri-band-pass, interference filters; twin-aperture stereo lens. Right: Transmission spectra for filters (left).

Filtering the images implies a significant proportion of the available light is lost. This means a compromise must be made between a fast shutter speed, narrow aperture and low sensor gain. Faster shutter speeds ensure that motion is frozen whilst a narrow aperture ensures that the depth of field is large enough to capture the scene. Subr et al. [11] investigate the effects that a low light-budget can have on stereo reconstruction, finding that gain, exposure time and aperture

restrictions lead to significant degradation of the resulting depth image but that often a compromise can be reached, determined by the type of scene.

For landscape scenes, we use a narrow aperture to obtain the greatest depth of field; however, as the filters block a significant proportion of available light, we find this is limited to f/11. We use an ISO of 1250 and a tripod to compensate for any camera shake caused by the shutter speeds required - $^1/_{40}$ s to $^1/_{60}$ s.

The stereo nature of the lens is simultaneously a drawback and benefit to our system. Whilst using filtered stereo images makes stereo matching more challenging, we also gain information about the depth of the materials in the scene, a feature which can be used to augment a classifier. We process this pair using SURF features for stereo rectification, allowing us to compute a good disparity mapping between the two images using LibELAS [4].

In the tri-band case, disparity maps are much harder to compute when both stereo images are filtered due to the difference in spectrum between the images. Filtering just one image leaves greater overlap between the recorded spectra, that is, greater correlation between the images. Depth recovery is then more successful and additional spectral information is retained. We use a neutral density filter (ND4) for the unfiltered image to ensure similar exposure between the pair.

We find that the radiometric distortion is generally less of a problem for feature matching in the visible-NIR case and that good depth maps can be recovered using SURF features for rectification followed by LibELAS.

4 Supervised Classification in Six-Band, Stereo Images

The system aims to aid classification by increasing the spectral information known about a scene. Using the raw captured image pairs, we label several distinct classes in the RGB image and apply Random Forest classification [2].

We label each image with the classes: grass (two types), tree (two types), sky, and man-made, extracting colour and texture features for classification.

Previous work [1] has shown that colour, hue in particular, and entropy are the most successful features for vegetation classification in landscape images. Hence we use the 6-colour RGB bands and their transformed values in HSV space (Hue, Saturation and Value) along with local entropy. Colour features alone are not sufficient for material classification [8] and so we apply 2-dimensional discrete stationary wavelet response filters to enhance the classifier. As texture features change with the scale of the material, we also use the disparity map as a feature.

5 Results

We evaluate our system on the task of foliage classification first testing the classifier on monocular image data, using the visible-spectrum images from each dataset. This gives us a point of comparison for the datasets we generate.

In both implementations, the left stereo image always represent the visible spectrum whereas the right image is filtered. We rely on relevance detection

Classifier input data	Tri-band Pixel	Tri-band Super-pixel	Infrared Pixel	Infrared Super-pixel
Right image	**0.0233**	0.1967	0.0391	0.2786
Left image	**0.0240**	0.1977	0.0384	0.2790
Right image and disparity	**0.0212**	0.1966	0.0375	0.2789
Left image and disparity	**0.0315**	0.2228	0.0418	0.2594
Right and left images	**0.0198**	0.1988	0.0227	0.2426
Right and left images with disparity	**0.0199**	0.1941	0.0221	0.2419

Table 1. Classification error results, showing classification error as different inputs are used with either pixel-based or super-pixel classification (right image filtered as detailed).

in the random forest implementation and comparing classification error with different combinations of left and right images and disparity as input.

We also compare using a pixel classifier to a super-pixel classifier, noting that the wavelet features we use incorporate a sense of spatial extent and that the super-pixel approach may render this superfluous.

Table 1 details the error achieved against the data set used. The lowest error occurs when both images from the multi-spectral pair are used. Use of the disparity map has little influence, and using tri-band filters to split the visible spectrum results in lower error than extending the spectrum to NIR light.

6 Discussion

There are several conclusions we can draw from our results.

- Firstly, imaging with tri-band filtering performs better than imaging over the visible and infrared parts of the spectrum, despite the strong infrared reflectance of foliage. The NIR implementation only gives us one effective additional colour channel so the information gained is greater in the tri-band case than the NIR case.
- Secondly, pixel-based classification performs better than super-pixel classification – probably because texture filters already contain some spatial information which makes a super-pixels unnecessary.
- Finally, the most favourable results are obtained using a filtered pair of images in the visible spectrum, that is, splitting the visible spectrum into six bands rather than the standard three.

Given that we know vegetation to have a more pronounced spectral footprint in the NIR region of the spectrum, it is surprising that the visible-NIR test case does not out-perform the tri-band-filtered case. There are, however, several explanations for this. Most likely is that the NIR images, due to the nature of the image sensor, only yielded one additional grey-scale channel representing the NIR spectrum and therefore there was a limited information gain in this case. It
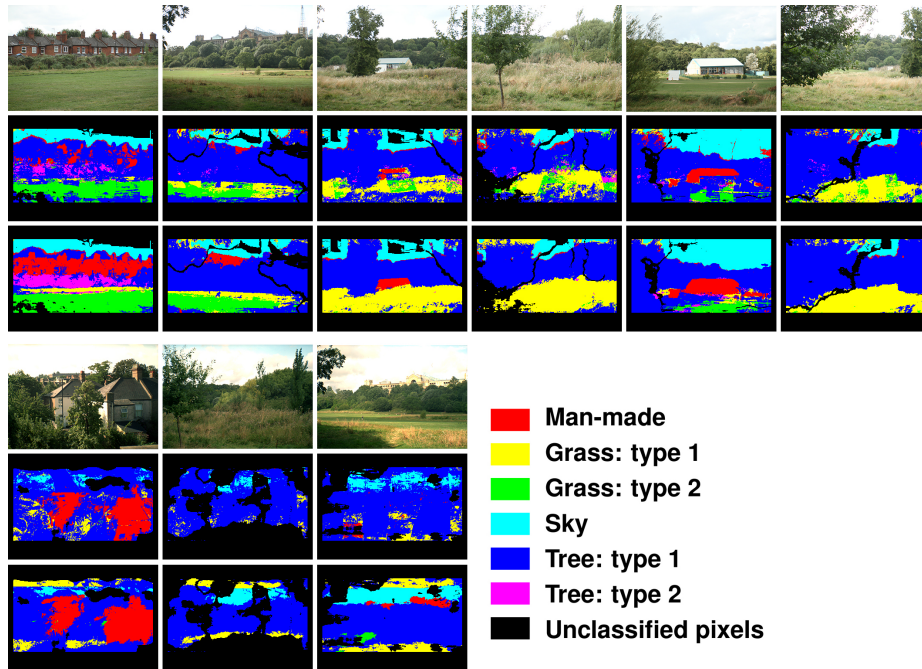
Fig. 2. Sample classification results showing an improvement when both images from each pair are used for classification. Top set, tri-band filtering results; bottom set, visible-infrared filtering results. For each set: Top,original unfiltered image; middle, classification results using only the unfiltered image; bottom, classification results using both images from the pair (filtered and unfiltered).

is clear, however, in both cases that the additional spectral information improves the classification result.

7   Conclusion

Identifiable sources of error in the experiments include the following: predefined classes covered a large number of potential sub-classes such as tree species, and types of man-made structures such as paths and houses were grouped together.

Data we acquired sampled only a single type of landscape and so there is a possibility that the random forest was over-trained on this data-set and would not produce good results for any other environment.

We conclude that our novel, multi-spectral system leads to improved foliage recognition in ground-level landscape imagery and is more practical and lower cost than current alternatives, and we claim that additional spectral data improves material classification in landscapes.

Although the incorporation of depth information does marginally reduce classification error, the biggest improvement is seen when multi-spectral images are

used. We suggest then, that colour features still have more of an impact than texture features in classification of different foliages classes but note that further testing would be required in order to validate this.

References

1. Gwyneth Bradbury. Material Classification in Outdoor Scenes. *MSc Computer Graphics, Vision and Imaging, University College London*, 2010.
2. Leo Breiman. Random Forests. *Machine Lerarning*, Volume 45(1):29, January 2001.
3. Graham Fyffe. Single-Shot Photometric Stereo by Spectral Multiplexing. *Proceedings ACM SIGGRAPH Asia Sketches*, pages 2–7, 2010.
4. Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient Large-Scale Stereo Matching. *Proceedings of the 10th Asian Conference on Computer Vision*, pages 25–38, 2010.
5. Ralf Habel, Michael Kudenov, and Michael Wimmer. Practical spectral photography. *Computer Graphics Forum (Proceedings EUROGRAPHICS 2012)*, 31(2):449–458, May 2012.
6. Luis Hernandez-Stefanoni and Raul Ponce-Hernandez. Mapping the Spatial Distribution of Plant Diversity Indices in a Tropical Forest Using Multi-Spectral Satellite Image Classification and Field Measurements. *Biodiversity and Conservation*, pages 13(14):2599–2621 , 2004.
7. Min Kim, Todd Harvey, David Kittle, Holly Rushmeier, Julie Dorsey, Richard Prum, and David Brady. 3D Imaging Spectroscopy for Measuring Hyperspectral Patterns on Solid Objects. *ACM Transactions on Graphics*, 31(4):38:1–38:11, July 2012.
8. Anthony Palmer and Frank Tanser. Vegetation Mapping of the Great Fish River Basin, South Africa: Integrating Spatial and Multi-Spectral Remote Sensing Techniques. pages 197–204, 2000.
9. Z. Qi. Extraction of Spectral Reflectance Images From Multi-Spectral Images by the HIS Transformation Model. *International Journal of Remote Sensing*, 17, 1996.
10. Raju Shrestha, Jon Yngve Hardeberg, and Alamin Mansouri. One-Shot Multispectral Color Imaging with a Stereo Camera. *Digital Photography VII. Proceedings of the SPIE*, 7876:787609–787609–11, 2011.
11. Kartic Subr, Gwyneth Bradbury, and Jan Kautz. Binocular-Stereo Photography Under a Light-Budget. *Proceedings of CVMP 2012*, 2012.
12. Masaru Tsuchida, Keiji Yano, and Hiromi T. Tanaka. Development of a High-Definition and Multispectral Image Capturing System for Digital Archiving of Early Modern Tapestries of Kyoto Gion Festival. *2010 20th International Conference on Pattern Recognition*, pages 2828–2831, August 2010.
13. Denis Wolf, Andrew Howard, and Gaurav Sukhatme. Towards Geometric 3D Mapping of Outdoor Environments Using Mobile Robots. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1507–1512, 2005.
14. Qian Yu, Peng Gong, Nick Clinton, Greg Biging, Maggi Kelly, and Dave Schirokauer. Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. *Photogrammetric Engineering and Remote sensing*, 72(7):799–811, 2006.