IMAGE CLASSIFICATION IN NATURAL SCENES: ARE A FEW SELECTIVE SPECTRAL CHANNELS SUFFICIENT?

Jason Holloway[†]   *Tanu Priya*[⋆]   *Ashok Veeraraghavan*[†]   *Saurabh Prasad*[⋆]

[⋆] University of Houston, ECE Department, Houston, TX
[†] Rice University, ECE Department, Houston, TX

ABSTRACT

A tenet of object classification is that accuracy improves with an increasing number (and variety) of spectral channels available to the classifier. Hyperspectral images provide hundreds of narrowband measurements over a wide spectral range, and offer superior classification performance over color images. However, hyperspectral data is highly redundant. In this paper we suggest that only 6 measurements are needed to obtain classification results comparable to those realized using hyperspectral data. We present classification results for a natural scene using three imaging modalities: 1) using three broadband color filters (RGB) and three narrowband samples, 2) using six narrowband samples, and 3) using six commonly available optical filters. If these results hold for larger datasets of natural images, recently proposed multispectral image sensors [1, 2] can be used to offer material classification results equal to that of hyperspectral data.

Index Terms— Natural Scene Classification, Hyperspectral Imaging, Multispectral Imaging

1. INTRODUCTION

In recent years scene understanding and object classification has received intensive attention as it benefits many practical applications such as autonomous driving, robot vision and content-based image retrieval. The input for many classification tasks are images taken using conventional cameras containing three broadband spectral measurements (the red, green, and blue channels of the image). RGB cameras are bountiful, cheap, and easy to use; however, the coarse sampling of the visible spectrum limits classification accuracy, especially in the presence of metameric scene elements.

Hyperspectral imaging (HSI) systems, on the other hand, record hundreds of measurements and provide fine spectral resolution over a wide range of the electromagnetic spectrum. HSI is able to capture material specific information which greatly improves classification performance. The improved performance comes at a cost. HSI camera systems require specialized processing units, are expensive and bulky, have long acquisition times, and suffer from a low signal-to-noise ratio. Moreover, the spectral profile of elements in natural scenes vary slowly such that neighboring narrowband spectral measurements are highly correlated. This inherent spectral redundancy suggests that far fewer measurements can be used to produce classification fidelity approaching that of HSI.

Recent work in camera and sensor design has led to two imaging systems which can capture a handful of co-located spectral measurements. By manipulating vertical silicon nanowires, [1] is able to create multi-layer image sensors tuned to specific wavelengths while [2] extends the typical Bayer filter found on color images to capture multispectral images. A camera add-on proposed by [3] can be used to increase spectral resolution for consumer cameras. These systems increase spectral measurements on the order of ones, far fewer than the hundreds capable with HSI. It is unclear how much benefit such systems would pose for material classification, or even which measurements should be taken.

In this paper we present the effectiveness of using six channels in natural scene analysis tasks. In particular, we compare the classification performance of using the entire hyperspectral data to a select set of 6 multispectral channels.

We evaluate our approach on 2 scenes with ground truth labeling of 7 different material categories. We hope that these results will motivate more comprehensive research into alternative imaging modalities for scene analysis.

1.1. Related Work

Two well-established techniques for improving upon color image classification performance are to perform classification using (1) joint measurements of RGB and side-band information and (2) dense spectral sampling from HSI.

Augmenting the visible spectrum with a broadband near-infrared channel has been shown to improve material classification [4], agricultural foodstuff discrimination [5], and image segmentation tasks [6]. Multi- and hyperspectral images have been used extensively to improve the performance of image segmentation and classification in specific application areas such as remote sensing (see [7] for a survey of the field), medical diagnosis and bioinformatics [8, 9], and military surveillance [10]. A comprehensive overview of classification using HSI is offered in [11].

Unfortunately, no such effort has been made to improve

classification of natural scenes using HSI data. Work by [12] and [13] shows that natural scenes exhibit spectral profiles which vary slowly and are inherently of a lower dimension than the hundreds of samples captured in HSI data. We offer a foundation for further exploration into using only a few spectral measurements for natural scene classification.

2. HYPOTHESIS AND SOLUTION

We seek to test the following hypothesis: A few (say for example 6) carefully selected spectral channels can be used to achieve classification accuracy comparable to that of classification using HSI. We consider the following three imaging scenarios, we first augment the three bands of a conventional color camera with narrowband measurements, we select a subset of narrowband measurements captured in HSI, and finally we use spectral channels inspired by available optical filters.

Let the true spectral profile of every pixel p in the hyperspectral image be represented by $H(p)$, a length–N column vector where N is the number of spectral bands. We wish to represent each pixel as a length–M vector, $x(p)$, which maximizes discriminatory information between material classes. Each entry in $x(p)$ is a linear combination of the spectral measurements in $H(p)$

$$x(p) = AH(p), \qquad (1)$$

where A is a positive $M \times N$ matrix whose rows contain the weights for each entry in $x(p)$; the rows of A sum to 1.

Material classification is an exercise in labeling where a labeling function $\ell(\cdot)$ maps observed data to an object class $(x \to Y)$. The goal is to reduce error when assigning labels. In particular, we seek to minimize the labeling problem

$$\min \sum_{s \in S} \ell(x(s)) \uparrow y(s), \qquad (2)$$

where S is the set of pixels to be classified, $y(s) \in Y$ is the true object class, and $\uparrow$ returns 1 if $\ell(x(s))$ is not equal to $y(s)$.

To minimize (2) is to find the optimal set of weights A to produce x. Substituting (1) into (2) gives

$$\underset{A}{\operatorname{argmin}} \sum_{s \in S} \ell(AH(s)) \uparrow y(s). \qquad (3)$$

Explicitly solving (3) is intractable, therefore we employ a greedy selection algorithm [14] to find the rows of A sequentially. That is, the first row of A is found which minimizes (3) for $M = 1$. The second row is then computed by minimizing (3) for $M = 2$ and by fixing the first row using the result from the previous step. The procedure repeats until all M rows are found.

Depending on the imaging scenario, further restrictions are placed upon the rows of A. Each row of A corresponding to a narrowband measurement is required to have a single

Fig. 1. **Scenes used for testing:** HSI was collected for the two outdoor scenes shown above. Color images were generated using the spectral profile of the e2v EV76C560 camera sensor to provide realistic measurements for broadband color filters. Ground truth labeling of top scene was used to train the classifier while the bottom scene was reserved for testing. Gamma correction has been applied for display purposes.

non-zero entry. Weights for the RGB color filters are known *a priori* and are fixed before searching for the remaining 3 narrowband measurements. The weights for the common spectral filters are also known *a prioi* which are then used to restrict the search space for each row of A.

In order to ensure consistent inputs to the labeling function we preprocess pixels prior to assigning a label. Each pixel is normalized by dividing the value of each spectral sample by the sum of all spectral samples,

$$x(p) = \frac{x(p)}{\sum_{i=1}^{N} x_i(p)} \qquad (4)$$

where i is an index of the N spectral samples.

Hyperspectral images are spatially redundant as well as spectrally redundant. In an effort to reduce computational overhead, we compute superpixels for the input HSI data to group neighboring pixels. We implement a straightforward extension of the entropy-based superpixel segmentation approach proposed in [15]. Spectral measurements for individual pixels within a superpixel are averaged to form the spectral response for that superpixel.

2.1. Classifier

Our goal in this paper is not to develop a new classifier but to show that a few spectral channels perform surprisingly well.

So, we choose the most popular and common classifier–support vector machines [16]–as the labeling function in (2).

SVM formulations can be constructed to create a non-linear classifier [17] using a kernel to transform input data into a higher dimensional feature space. We employ a Gaussian radial basis function (defined entirely by one parameter, γ) as the transformation kernel. SVM classifiers are highly sensitive to the choice of kernel parameter γ and the soft-margin cost C. We use exponentially increasing sequences to perform a grid-search for the best combination of cost and kernel parameters. Parameter estimation is completed independently for each classification task.

Multi-class labeling is achieved by creating multiple binary classification problems. We employ a one-versus-one scheme where the binary classifier discriminate between two classes. Test samples are assigned to the label which performs the best in the most head-to-head comparisons.

3. EXPERIMENTS

We validate our hypothesis using two scenes captured outdoors on the University of Houston campus. The scenes were acquired using a Headwall Photonics hyperspectral imager which provided measurements in 325 spectral bands with a spatial resolution of 1004×2500. The hyperspectral data uniformly spanned the visible and near-infrared spectrum from 400nm-1000nm. Each scene includes objects in 7 different categories–vegetation, metal, concrete, pathway, skin, fabric, and rubber. Ground truth labels were marked by hand for both scenes.

Color image representations of the two scenes were synthesized using the spectral profile of the e2v EV76C560 camera sensor to provide a real-world RGB profile. The camera sensor is used in machine vision cameras and is representative of the spectral response of many cameras. Each scene includes a diffuse Spectralon calibration target with 99% reflectance in the wavelength range used in this study which was used to apply white balancing to the color images. A color representation of the scenes is shown in Figure 1; the top scene was used for classifier training while the bottom scene was reserved for testing the classifiers. Complex lighting in the scene resulted in acquiring spectral information for scene elements under varying lighting conditions, i.e. in direct sunlight and in shadow.

Benchmark accuracy was computed using the full HSI data, RGB color filters, and RGB coupled with a broadband near-infrared channel (as realized using the same camera sensor with an IR pass filter). In addition, we compute classification results using RGB and 3 narrowband channels, 6 narrowband channels, and 6 optical filters. All of the use cases were synthesized using the HSI data.
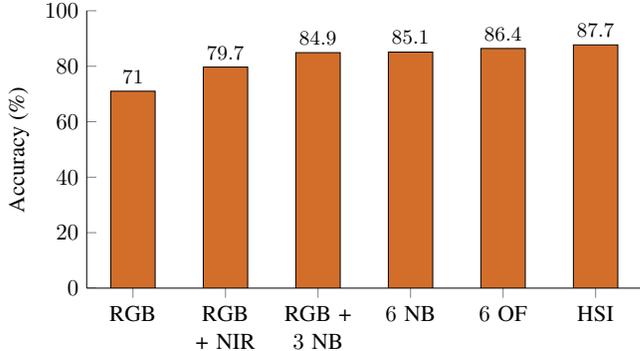

Fig. 2. **Classifier results:** Object classification accuracy using 6 different imaging modalities and SVM classifier. Classification accuracy increases with the number of measurements but faces diminishing returns. The three test cases using 6 measurements are within 3 percentage points of the hyperspectral classifier despite using only 2% of the measurements.

3.1. Classification Results

Training of the SVM classifier was conducted using a 5-fold scheme where the training samples were randomly divided into 5 groups. Training was then conducted using a leave one out methodology, 4 groups are used to learn the support vectors and the fifth group is used to test against. Parameter estimation was completed using a grid-search with exponentially increasing values for the cost and kernel parameter, $C \in \{2^0, 2^1, \ldots, 2^8\}$, $\gamma \in \{2^0, 2^1, \ldots, 2^8\}$. SVM classification tasks were executed using the LIBSVM library [18]. 916 samples were extracted from the top scene in Figure 1 to train the classifier while 649 samples from the bottom scene were used for testing.

Overall accuracy results are shown for the six use cases in Figure 2. The results show a clear trend in the data—incorporating more measurements improves overall accuracy. As expected, using only three RGB broadband filters results in the lowest overall accuracy, while having access to all 325 narrowband samples in the HSI data yields the best results.

Augmenting color images with side-channel information greatly improves the classification accuracy. Adding the broadband near-infrared channel accounts for half of the difference in accuracy between RGB images and HSI data. Adding three narrowband measurements to the RGB data offers further improvement in accuracy, resulting in nearly 85% of the samples being correctly identified. The three narrowband samples were selected using the previously mentioned greedy selection algorithm to solve (3) after setting the first three rows of the weighting matrix to match the red, green, and blue spectral response curves of the e2v EV76C560 camera sensor. Interestingly, the solution favored measurements from within the visible spectrum with the narrowband

Table 1. Confusion matrix for 6 optical filters using SVM

Class	V	M	C	P	S	F	R	Accuracy (%)
Vegetation	181	12	3	2	2	0	0	90.5
Metal	0	104	1	16	0	0	5	82.5
Concrete	0	4	32	11	0	0	0	68.0
Pathway	0	13	0	178	0	0	9	89.0
Skin	0	5	0	5	12	0	0	54.5
Fabric	0	0	0	0	0	21	0	100
Rubber	0	5	0	1	0	0	27	81.8
Accuracy (%)	100	72.7	88.9	83.6	85.7	100	65.8	**86.4**

Table 2. Confusion matrix for HSI using SVM

Class	V	M	C	P	S	F	R	Accuracy (%)
Vegetation	183	16	1	0	0	0	0	91.5
Metal	0	109	0	15	0	2	0	86.5
Concrete	0	3	31	13	0	0	0	66.0
Pathway	0	13	0	179	0	0	8	89.5
Skin	0	3	0	0	19	0	0	86.4
Fabric	0	0	0	0	0	21	0	100
Rubber	0	4	0	1	0	0	28	84.8
Accuracy (%)	100	73.6	96.9	86.1	100	91.3	77.8	**87.7**

Fig. 3. **Spectral responses:** The spectral response for the various imaging modalities superimposed upon a representation of the spectrum sampled by the HSI. (a) shows the RGB and NIR spectral responses, (b) shows RGB and three narrowband filter locations (in white), (c) shows the location of the six narrowband filters (in white), and (d) shows the location and bandwidth of the six optical filters. The broadband spectral curves have been normalized for clarity.

samples located at 416nm, 466nm, and 641nm.

As seen in Figure 2, selecting 6 narrowband samples to classify the scene is as accurate as using RGB and 3 narrowband filters and is within three percentage points of the HSI benchmark. Unlike the RGB and 3 narrowband filters, the distribution of measurements spans the entire spectrum and is not concentrated within the visible range. The narrowband samples are centered at 405nm, 419.5nm, 690nm, 767.5nm, 825nm, and 919.5nm.

The highest classification accuracy came from imaging with 6 optical filters. The optical filter specifications were inspired by actual filters commonly available through commercial retailers. We included 4 filter types (shortpass, longpass, bandpass and band reject) with cut-off/cut-on/center wavelengths set every 25nm. The bandpass and band reject filters had variable bandwidths of 25nm, 50nm, and 75nm at each center wavelength.

Despite the plethora of choices available, the selected filters (chosen using a greedy approach) were all bandpass filters and 5 of the 6 had a bandwidth of 25nm. The distribution of the filters seems to follow the same general trend as the six narrowband filters, with center wavelengths of 425nm, 450nm, 650nm, 725nm, 775nm, and 800nm. The filter centered at 650nm had a bandwidth of 50nm. With an overall accuracy of 86.4% the six channels were 98.5% as accurate at classification as HSI while using only 2% of the total number of measurements.

The confusion matrices shown in Table 1 and Table 2 show SVM classification performance using 6 optical filters and HSI data respectively. The imaging modalities tend to have similar strengths and weaknesses. The only significant difference between the two is that HSI classification was able to learn the sparsely represented skin class better than the 6 optical filters.
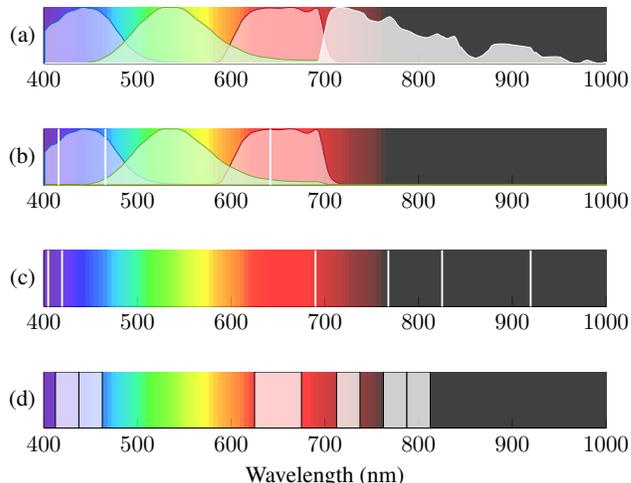
The distribution of spectral samples for the imaging modalities are shown in Figure 3. The broadband red, green, blue, and near-infrared channels have been normalized for clarity. It is interesting to note that no measurements were used between 500nm-600nm when solving (2). This suggests that RGB classification is less likely to succeed since the green channel provides little distinguishing information. In fact, using just the red, blue, and near-infrared channels outperforms the red, green, and blue channels in classification accuracy (77.5% accuracy for RB+NIR vs 71% for RGB).

4. DISCUSSION AND FUTURE WORK

In this paper we showed that there is potential to greatly improve the accuracy of material classification in natural scenes by sampling only a few spectral channels. Specifically, using 6 bandpass filters may be sufficient to obtain classification accuracy comparable to that of a hyperspectral imager but without the need to capture hundreds of spectral measurements. A fruitful course for future research is two-fold. First a comprehensive analysis is of narrowband HSI of natural scenes is to determine the large-scale applicability of taking only a few measurements for material classification. The second is to build and test a prototype imaging system which operates in the modalities described in this paper.

5. REFERENCES

[1] Kwanyong Seo, Munib Wober, Paul Steinvurzel, Ethan Schonbrun, Yaping Dan, Tal Ellenbogen, and Kenneth B Crozier, "Multicolored vertical silicon nanowires," *Nano letters*, vol. 11, no. 4, pp. 1851–1856, 2011.

[2] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamice range, and spectrum," *Image Processing, IEEE Transactions on*, vol. 19, no. 9, pp. 2241–2253.

[3] Alkhazur Manakov, John F Restrepo, Oliver Klehm, Ramon Hegedüs, Elmar Eisemann, Hans-Peter Seidel, and Ivo Ihrke, "A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 47, 2013.

[4] N Salamati, C Fredembach, and S Susstrunk, "Material classification using color and nir images," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2009, pp. 216–222.

[5] Phil Williams, Karl Norris, et al., *Near-infrared technology in the agricultural and food industries.*, American Association of Cereal Chemists, Inc., 1987.

[6] Neda Salamati, Diane Larlus, Gabriela Csurka, and Sabine Süsstrunk, "Semantic image segmentation using visible and near-infrared channels," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer, 2012, pp. 461–471.

[7] Dengsheng Lu and Qihao Weng, "A survey of image classification methods and techniques for improving classification performance," *International journal of Remote sensing*, vol. 28, no. 5, pp. 823–870, 2007.

[8] Bahram Khoobehi, James M Beach, and Hiroyuki Kawano, "Hyperspectral imaging for measurement of oxygen saturation in the optic nerve head," *Investigative ophthalmology & visual science*, vol. 45, no. 5, pp. 1464–1472, 2004.

[9] Cong Phuoc Huynh and Antonio Robles-Kelly, "Hyperspectral imaging for skin recognition and biometrics," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 2325–2328.

[10] Dimitris Manolakis, David Marden, and Gary A Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.

[11] Chein-I Chang, *Hyperspectral data exploitation: theory and applications*, John Wiley & Sons, 2007.

[12] Te-Won Lee, Thomas Wachtler, and Terrence J Sejnowski, "The spectral independent components of natural scenes," in *Biologically Motivated Computer Vision*. Springer, 2000, pp. 527–534.

[13] Ayan Chakrabarti and Todd Zickler, "Statistics of real-world hyperspectral images," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 193–200.

[14] George L. Nemhauser and Laurence A. Wolsey, *Integer and Combinatorial Optimization*, Wiley-Interscience, New York, NY, USA, 1988.

[15] Ming-Yu Liu, Oncel Tuzel, Srikumar Ramalingam, and Rama Chellappa, "Entropy rate superpixel segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2097–2104.

[16] Corinna Cortes and Vladimir Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[17] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992, pp. 144–152.

[18] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.