Unsupervised Identification of Materials with Hyperspectral Images

Mira Welner

University of California, Davis, CA
mewelner@ucdavis.edu

Abstract

We introduce a novel technique to identify three spectra representing the three primary materials in a hyperspectral image of a scene. We accomplish this using a modified autoencoder. Further research will be conducted to verify the accuracy of these spectra.

Introduction

Hyperspectral images are used to identify materials, objects, and chemical processes in a scene. Each pixel in a hyperspectral image is a spectral profile that characterizes the intensity of light at that pixel as a function of wavelength. Since material has unique spectra, hyperspectral images are often used to identify the presence and relative abundances of materials in a scene. However, identifying which materials can be found in a scene is often a slow and difficult process. The goal of this project is to use an autoencoder to identify the dominant materials found in a hyperspectral image with in an unsupervised manner.

Our method relies on the interesting observation that the spectral complexity of most scenes is not arbitrary; in that, it is well known that most scenes are well approximated with three material and two lighting spectra (Finlayson, Drew, and Funt 1993). As a result it was assumed that the hyperspectral images had three main materials, with each material being represented by a unique spectra, and a material map (Kokaly, King, and Hoefen 2011). The light sources were not taken into account in this initial study. Given this, the hyperspectral image $h(x, y, \lambda)$ can be represented by the following equation:

$$h(x, y, \lambda) = \sum_{r=0}^{2} \pi_r(\lambda)\alpha_r(x, y), \qquad (1)$$

where π_r is is the spectra corresponding to the r-th material map, and $\alpha_r \geq 0$ being the rth material map.

Since most datasets do not have ground truth material spectra and maps, we design an unsupervised approach for estimating them. Specifically, we train an autoencoder for hyperspectral images that is endowed with some key properties. First, the decoder is designed to implement (1), which

attributes physical meaning to the outputs of the encoder. Second, the encoder is designed to output the three spectral profiles, $\{\pi_r(\lambda), r = 0, 1, 2\}$; subsequently, the material map is estimated at each pixel under the assumption that the spectrum at any given pixel lies in the conic hull of the material spectra. Third, we train the autoencoder end-to-end and the network learns to recognize the three dominant materials such that the spectra in any given pixel lies in the conic hull of these materials.

Materials and Methods

Dataset

The dataset used to train our autoencoder was sourced from the Harvard Real World Hyperspectral Images Database (Chakrabarti and Zickler 2011). The database was selected due to the high quality of the images as well as the relative ease of information extraction and inputting the images into a tensorflow pipeline. The dataset consisted of hyperspectral images 1040 pixels long and 1392 pixels wide with 31 spectral bands. 50 images were of indoor and outdoor scenes with daylight illumination, and 27 were of indoor scenes under artificial and mixed illumination. The 77 original images were split into 20 test images and 57 training images. Images in both the training and testing set were separated into four smaller images of size 512 by 512 to augment the dataset as well as to increase processing speed. Images in the training set were duplicated and flipped on the x and y axis to further augment the dataset. These images contained a variety of small scenes photographed at varying distances that would likely be found around a university campus such as desks, trees, walls, and computers.

Autoencoder

An autoencoder was programmed in keras to separate the three primary materials (Chollet 2015). Autoencoders are neural networks designed to encode an image into a compressed format, and later decompress the image into its original state (Liou et al. 2014). The portion of the autoencoder that encodes the hyperspectral image into a compressed format is quite similar to a standard autoencoder. First, the image is reshaped into a two dimensional image with the x axis representing the pixel and the y axis representing the spectra. The image is then fed through a series of convolutional
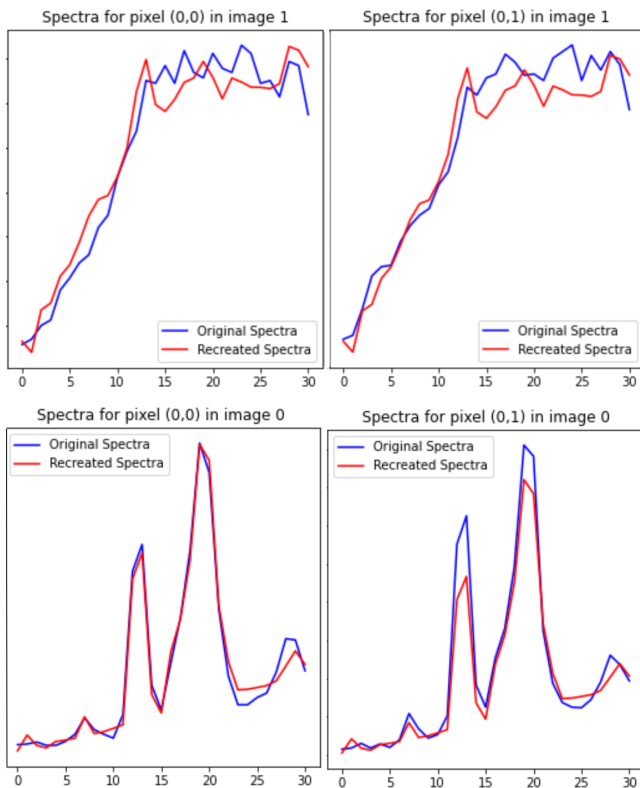
Figure 1: Original spectra compared to recreated spectra for two different pixels in two different test images
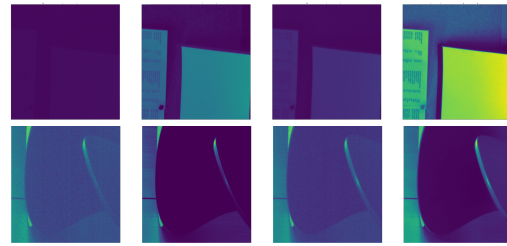

Figure 2: Four material maps, with the corresponding recreated material map to right of the original.

Figure 3: The three main spectra of a wooden floor

and max pooling layers followed by a series of fully connected layers with dropout. The resulting compressed image was represented by a 3×31 matrix.

The decoder, however, was structured differently from a standard autoencoder. Rather than decoding the image through a neural network it was decoded by utilizing the underlying physical principles of a hyperspectral image. The 3×31 matrix was treated as though it represented the spectra of the 3 primary materials making up the scene, each of which had 31 different spectral bands. Material maps were then generated by multiplying the pseudo-inverse of the spectra by the original hyperspectral image.

The three dimensional matrix containing three primary material maps was run through a relu activation layer. The relu activation forced the material maps to become positive, ensuring that the spectra in any given pixel lies in the conic hull of the three primary materials. After being run through the relu, the material maps were multiplied by the spectra to reconstruct the original hyperspectral image. Loss was determined by the normalized mean squared error between the original and reconstructed hyperspectral image.

Results

The autoencoder was trained on the dataset for 25 epochs and a validation loss of 0.0130 was achieved. As seen in figure 2 the recreated spectra for many of the pixels were quite similar to the original spectra, likely due to the spectra

being optimized for over the material maps. This results in lowered accuracy for material maps (figure 3).

The trained model was then split into a decoder and and encoder. The encoder portion compressed a hyperspectral image into the spectra of its three primary materials. An image of a wooden floor generated the three spectra seen in figure 4.

Future research will investigate whether the spectra displayed are similar to materials found in wood, improve the accuracy of the material maps, and improve the ability of the model to generalize on scenes different from the Harvard dataset.

References

Chakrabarti, A.; and Zickler, T. 2011. Statistics of Real-World Hyperspectral Images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 193–200.

Chollet, e. a., Francois. 2015. Keras. https://keras.io. Accessed: 2022-02-09.

Finlayson, G.; Drew, M.; and Funt, B. 1993. Diagonal transforms suffice for color constancy. In *1993 (4th) International Conference on Computer Vision*, 164–171.

Kokaly, R. F.; King, T. V.; and Hoefen, T. M. 2011. Mapping the distribution of materials in hyperspectral data using the USGS Material Identification and Characterization Algorithm (MICA). In *2011 IEEE International Geoscience and Remote Sensing Symposium*, 1569–1572.

Liou, C.-Y.; Cheng, W.-C.; Liou, J.-W.; and Liou, D.-R. 2014. Autoencoder for words. *Neurocomputing*, 139: 84–96.