Image Feature Extraction Using Entropy extraction

Brajraj Panwar (0801CS171021)

Sumit Pachaha(0801EC171082)

Pritesh Makwana (0801CS171059)

Bspanwar22010@gmail.com

December 13, 2020

Contents

Chapter 01

Introduction

1.1 Feature extraction

Feature plays a very important role in the area of image processing. Before getting features, various image preprocessing techniques like binarization, thresholding, resizing, normalization etc. are applied on the sampled image. After that, feature extraction techniques are applied to get features that will be useful in classifying and recognition of images. Feature extraction techniques are helpful in various image processing applications e.g. character recognition. As features define the behavior of an image, they show its place in terms of storage taken, efficiency in classification and obviously in time consumption.

1.2 Entropy

In information theory, information entropy is the log-base-2 of the number of possible outcomes for a message.

For an image, local entropy is related to the complexity contained in a given neighborhood, typically defined by a structuring element. The entropy filter can detect subtle variations in thelocal gray level distribution.

The entropy has been used as feature in the proposed methodology. We have extracted entropy from the decomposed components of 2D EWT. The entropy is explained as follows. Entropy- It is a statistical measure of randomness that can be used to differentiate the texture of the input image [. The larger entropy will have more information and vice versa.

Chapter 02

Mathematical Formulation

2.1 Image Formulation

The main innovation of the proposed algorithm is that spatial information is exploited using entropy, while spectral information is exploited using convex set optimization. The concept of entropy [40] has been used in many image processing techniques, including registration, reconstruction, segmentation, classification, and compression. Few researchers have also applied entropy in hyperspectral compression, band-selection and unmixing . Bayliss et. al. developed an ICA-based algorithm for unmixing based on the entropy between spectral signatures. ICA was selected under the assumption that components are statistically independent. However, in the data acquisition process of a hyperspectral sensor, the sum of abundance fractions associated with each pixel adds to one under the abundance sum-to-one constraint. As a result, the sources (endmembers) are not statistically independent . ICA uses the entropy of various spectra as a spectral feature, while the proposed algorithm uses the entropy of each band as a spatial feature.

Let us denote a mixed pixel in the hyperspectral image as an $L \times 1$-dimensional vector:

$$\mathbf{y} = \mathbf{M}\alpha + \mathbf{n}, \qquad (1)$$

where $\mathbf{M}$ is an $L \times Q$ matrix, with Q and L respectively denoting the number of endmembers and the number of bands in the original hyperspectral image. In Eq. (1), $\mathbf{n}$ is a noise vector of size $L \times 1$, which is assumed to be Gaussian in nature. $\alpha = [\alpha_1,\alpha_2,\dots,\alpha_Q]^T$ denotes the abundance vector of size $Q \times 1$, which satisfies the following two constraints:

- Abundance Non-negativity Constraint (ANC):

$$\alpha_i \geq 0, \qquad i = 1,2,\dots,Q. \qquad (2)$$

- Abundance Sum-to-one Constraint (ASC):

$$\sum_{i=1}^{Q} \alpha_i = 1. \qquad (3)$$

Let us denote the hyperspectral image as $\mathbf{Y}$ $[\mathbf{y}_1,\mathbf{y}_2,\dots,\mathbf{y}_L]^T \equiv [\mathbf{y}^1,\mathbf{y}^2,\dots,\mathbf{y}^Z]$, which contains Z mixed pixels of length L. $\mathbf{Y}$ is defined as follows:

$$\mathbf{Y} \equiv \begin{bmatrix} y_{11} & y_{12} & y_{13} & \dots & y_{1Z} \\ y_{21} & y_{22} & y_{23} & \dots & y_{2Z} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{L1} & y_{L2} & y_{L3} & \dots & y_{LZ} \end{bmatrix}, \qquad (4)$$

$y_L 1$

where each band $\mathbf{y}_i$ of size $1 \times Z$ is represented as a row vector $\mathbf{y}_i \equiv [y_i^1, y_i^2, y_i^3, ..., y_i^Z]$, and each mixed pixel vector $\mathbf{y}^i$ of size $L \times 1$ is represented as a column vector $\mathbf{y}^i \equiv$

$[y_1^i, y_2^i, y_3^i, ..., y_L^i]^T$. Here $Z = U \times V$ is the number of pixels in the original image. The height and width of each band are, respectively, U and V. Various notations used in the paper are shown in Table I.

The proposed algorithm takes as an input $\mathbf{Y}$ and the subspace dimension (number of endmembers) The result of the proposed algorithm is a matrix of endmembers $\mathbf{M}$

Since different bands have a different dynamic range (max value - min value), our band normalization strategy is intended to make all bands similar in terms of dynamic range. In our case, band normalization for each i^{th} band $(\mathbf{y}_i)$ is conducted as follows:

$$\bar{y}_i^j = \frac{\left(y_i^j - min(\mathbf{y}_i)\right)}{\left(max(\mathbf{y}_i) - min(\mathbf{y}_i)\right)}, \quad \forall\, y_i^j \in \mathbf{y}_i \quad . \qquad (5)$$

Each pixel value y_i^j of band $(\mathbf{y}_i)$ is normalized as per Eq. (5) and converted to a new normalized value $\bar{y}_i^j$, which is in the range of $[0,1]$. The aforementioned band normalization process [shown in Eq. (5)] is repeated for all the bands of the hyperspectral image. A band-normalized hyperspectral image $\bar{\mathbf{Y}} \equiv [\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, ..., \bar{\mathbf{y}}_L]^T \equiv [\bar{\mathbf{y}}^1, \bar{\mathbf{y}}^2, ..., \bar{\mathbf{y}}^Z]$ is defined as:

$$\bar{\mathbf{Y}} \equiv \begin{matrix} \bar{y}_{11} & \bar{y}_{12} & \bar{y}_{13} & \dots & \bar{y}_{1Z} \\ \bar{y}_{21} & \bar{y}_{22} & \bar{y}_{23} & \dots & \bar{y}_{2Z} \\ \dots & \dots & \dots & \dots & \dots \\ \bar{y}_{L1} & \bar{y}_{L2} & \bar{y}_{L3} & \dots & \bar{y}_{LZ} \end{matrix} \qquad (6)$$

2.2 Entropy Formulation

For a k-state system, Shannon defined the entropy as:

$$H = -\sum_{i=1}^{k} p_i \log p_i \quad , \qquad (7)$$

where p_i is the probability of occurrence for the i^{th} event, and $\sum p_i = 1, 0 \leq p_i \leq 1$. Shannon's entropy is very popular in the field of communications. Many researchers have extended the concept of entropy for image processing purposes [47].

Let $\mathbf{I}$ be a grayscale image of size $U \times V$ and $S_G \in$

$\{0,1,...,G-1\}$ be the set of associated greyscale values. Let G be maximum shade value in the image. Image $\mathbf{I}$ contains $Z = U \times V$ pixels. Let W_i be the frequency of the i^{th} grayscale value, where $i \in S_G$. The entropy for $\mathbf{I}$ (grayscale image) is defined as:

$$H = -\sum_{i=0}^{G-1} p_i \log p_i;\ p_i = W_i/Z \qquad (8)$$

We can extend the entropy definition in Eq. (8) for hyperspectral images and define the k^{th} band entropy (H_k) as:

$$H_k = -\sum_{i=0}^{G-1} p^k_i \log p^k_i;\ p^k_i = W_i^k/Z,\ k = 1,2,...,L \qquad (9)$$

Here, p^k_i is probability of having an i^{th} grey shade in the k^{th} band. The entropy of each band in $\bar{\mathbf{Y}}$ is calculated using Eq. (9), and denoted as $\{H_1,H_2,...,H_L\}$. Entropy can be interpreted as a measure of order (or randomness), or as a measure of homogeneity [48]. Instead of looking at various interpretations, we can look at it as an expression of the number of states of a system. Jianhua [49], Lua [50] and other researchers have used Shannon's entropy concept for information-theoretic divergence between two probability distributions. In the proposed algorithm, this entropy concept is used to measure the divergence between the probability distribution of two bands. A system with many states has high entropy and a system with few states has low entropy. A band with low entropy exhibits fewer variations, and a band with high entropy exhibits more variations. Bands can be rearranged in ascending order based on their values of entropy. A new matrix $\mathbf{S}$ can be thus obtained from $\bar{\mathbf{Y}}$ in such a manner that low entropy bands come first, and high entropy bands come last. The matrix $\mathbf{S}$ is generated such that $H_k < H_{k+1}$ for each value of k.

$$\mathbf{S} \equiv [\mathbf{s}_1,\mathbf{s}_2,...,\mathbf{s}_L]^T \equiv [\mathbf{s}^1,\mathbf{s}^2,...,\mathbf{s}^Z]$$

Chapter 03

algorithms

3.1 Entropy Algorithms

Inputs : $\mathbf{Y}, Q$. Inputs to the algorithm

Step-1: Entropy as spatial information

For k=1 to L

$\bar{\mathbf{y}}_k \leftarrow \mathbf{y}_k$. Band bormalization for k^{th} band

H_k = entropy ($\bar{\mathbf{y}}_k$) . Entropy of k^{th} band

EndFor

$[\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L]^T \leftarrow [\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \ldots, \bar{\mathbf{y}}_L]^T$. Entropy-based band sorting

$\mathbf{S} \leftarrow \bar{\mathbf{Y}}$. Creation of new matrix from normalized hyperspectral image

Step-2: Spectral information for convex set optimization

For i = 1 to $L/2$

$C_i = conv(\mathbf{s}_i, \mathbf{s}_{L-i-1})$. Calculate convex set for two-band data

T_i = count (C_i) . Calculate the number of points in the convex set

EndFor

$\mathbf{T} = \{T_1, T_2, \ldots, T_{L/2}\}$. Set of number of convex set points

$\underset{N \in \mathbf{T}}{\text{minimize}} \quad N - Q \quad \text{subject to} \quad N \geq Q$. Find N using optimization problem

Step-3: Removing extra $(N - Q)$ points

If $N > Q$

For k=1 to N

$D(\mathbf{x}_k)$. Calculate Euclidean distance for each point $\mathbf{x}_k \in C$

EndFor

Remove $(N - Q)$ points from C. Removal of points having lowest $D(\mathbf{x}(i,j), \mathbf{x}(i_1, j_1))$

EndIf

$\hat{\mathbf{M}}$ = Convex spectral signatures from C

Output : $\hat{\mathbf{M}}$

Chapter 04

4.1 Implementation :

```
import matplotlib.pyplot as plt
import numpy as np
from skimage import data
from skimage.util import img_as_ubyte
from skimage.filters.rank import entropy
from skimage.morphology import disk

noise_mask = np.full((128, 128), 28, dtype=np.uint8)
noise_mask[32:-32, 32:-32] = 30

noise = (noise_mask * np.random.random(noise_mask.shape) - 0.5 *
    noise_mask).astype(np.uint8)
img = noise + 128

entr_img = entropy(img, disk(10))

fig, (ax0, ax1, ax2) = plt.subplots(nrows=1, ncols=3, figsize=(10, 4))

img0 = ax0.imshow(noise_mask, cmap='gray')
ax0.set_title("Object")
ax1.imshow(img, cmap='gray')
ax1.set_title("Noisy image")
ax2.imshow(entr_img, cmap='viridis')
ax2.set_title("Local entropy")

fig.tight_layout()
# Texture detection
image = img_as_ubyte(data.gravel())

fig, (ax0, ax1) = plt.subplots(ncols=2, figsize=(12, 4),
                               sharex=True, sharey=True)
```

```
img0 = ax0.imshow(image, cmap=plt.cm.gray)
ax0.set_title("Image")
ax0.axis("off")
fig.colorbar(img0, ax=ax0)

img1 = ax1.imshow(entropy(image, disk(5)), cmap='gray')
ax1.set_title("Entropy")
ax1.axis("off")
fig.colorbar(img1, ax=ax1)

fig.tight_layout()

plt.show()
```

REFERENCES

[1] A. J. Brown, S. J. Hook, A. M. Baldridge, J. K. Crowley, N. T. Bridges, B. J. Thomson, G. M. Marion, C. R. de Souza Filho, and J. L. Bishop, “Hydrothermal formation of clay-carbonate alteration assemblages in the nili fossae region of mars,” *Earth and Planetary Science Letters*, vol. 297, no. 1-2, pp. 174–182, 2010.

[2] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, “Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches,” *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 5, no. 2, pp. 354–379, 2012.

J. S. Bhatt, “Novel approaches for spectral unmixing of hyperspectral data,” Ph.D. dissertation, Dhirubhai Ambani Institute of Information and Communication Technology, 2014

[3] S. L. Ruixing Li, “Improving hyperspectral subpixel target detection using hybrid detection space,” *Journal of Applied Remote Sensing*, vol. 12, no. 1, pp. 1 – 22 – 22, 2018. [Online]. Available: https://doi.org/10.1117/1.JRS.12.015022

[4] Y. Zhang, X. Huang, X. Hao, J. Wang, W. Wang, and T. Liang, “Fractional snow-cover mapping using an improved endmember extraction algorithm,” *Journal of Applied Remote Sensing*, vol. 8, no. 1, p. 084691, 2014.

[5] F. Huang, Y. Yu, and T. Feng, “Hyperspectral remote sensing image change detection based on tensor and deep learning,” *Journal of Visual Communication and Image Representation*, vol. 58, pp. 233 – 244, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/ pii/S1047320318302773

B. Tu, N. Li, L. Fang, H. Fei, and D. He, “Classification of hyperspectral images via weighted spatial correlation representation,” *Journal of Visual Communication and Image Representation*, vol. 56, pp. 160 – 166, 2018. [Online]. Available: http://www.sciencedirect. com/science/article/pii/S1047320318302256

[6] J.-C. Wu and G.-c. Tsuei, “Comparison of hyperspectral endmember extraction algorithms,” *Journal of Applied Remote Sensing*, vol. 7, no. 1, p. 073525, 2013.

[7] Entropy Based convex set optimization for spatial spectral Endmember Extraction from Hyperspectral image Dharambhai Shah, *Student Member, IEEE*, Tanish Zaveri, *Senior Member, IEEE*, Y N Trivedi and Antonio Plaza *Fellow, IEEE*