Characteristics and statistics of digital remote sensing imagery

Digital Images:
**Digital Image**

- With raster data structure, each image is treated as an array of values of the pixels.
- Image data is organized as rows and columns (or lines and pixels) start from the upper left corner of the image.
- Each pixel (picture element) is treated as a separate unite.

**Digital Image Terminology**

<table>
<thead>
<tr>
<th>Rows (i)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>15</td>
<td>12</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>16</td>
<td>18</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>22</td>
<td>24</td>
<td>22</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Columns (j)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>15</td>
<td>12</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>16</td>
<td>18</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>22</td>
<td>24</td>
<td>22</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Brightness value range (often 8-bit)</th>
<th>255</th>
<th>127</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associated grayscale</td>
<td>white</td>
<td>gray</td>
<td>black</td>
</tr>
</tbody>
</table>

Picture element (pixel) at location row 4, column 4, band 1 has a brightness value of 24, i.e., $B_{4,4,4} = 24$

**Statistics of Digital Images Help:**

- Look at the frequency of occurrence of individual brightness values in the image displayed
- View individual pixel brightness values at specific locations or within a geographic area;
- Compute univariate descriptive statistics to determine if there are unusual anomalies in the image data; and
- Compute multivariate statistics to determine the amount of between-band correlation (e.g., to identify redundancy).
Statistics of Digital Images

It is necessary to calculate fundamental univariate and multivariate statistics of the multispectral remote sensor data.

This involves identification and calculation of
- maximum and minimum value
- the range, mean, standard deviation
- between-band variance-covariance matrix
- correlation matrix, and
- frequencies of brightness values

The results of the above can be used to produce histograms.

Such statistics provide information necessary for processing and analyzing remote sensing data.

A “population” is an infinite or finite set of elements.

A “sample” is a subset of the elements taken from a population used to make inferences about certain characteristics of the population. (e.g., training signatures)
Large samples drawn randomly from natural populations usually produce a symmetrical frequency distribution.

Most values are clustered around the central value, and the frequency of occurrence declines away from this central point.

A graph of the distribution appears bell shaped is called a normal distribution.

Histogram and Its Significance to Digital Remote Sensing Image Processing

The histogram is a useful graphic representation of the information content of a remote sensing image indicating the quality of the original data, e.g. whether it is low in contrast, high in contrast, or multimodal in nature.

Many statistical tests used in the analysis of remote sensing data assume that the brightness values recorded in a scene are normally distributed.
Histogram and Its Significance to Digital Remote Sensing Image Processing

Unfortunately, remotely sensed data may not be normally distributed. In such instance, nonparametric statistical theory may be preferred.
Univariate Descriptive Image Statistics

Measures of Central Tendency in Remote Sensor Data

- **Mode**: is the value that occurs most frequently in a distribution and is usually the highest point on the curve. Multiple modes are common in image datasets.
- **Median**: is the value midway in the frequency distribution.
- **Mean**: is the arithmetic average and if defined as the sum of all observations divided by the number of observations.

Measures of Central Tendency in Remote Sensor Data

The *mean* is the arithmetic average and is defined as the sum of all brightness value observations divided by the number of observations. It is the most commonly used measure of central tendency.

The mean ($\mu_k$) of a single band of imagery composed of $n$ brightness values ($BV_{ik}$) is computed using the formula:

$$\mu_k = \frac{\sum_{i=1}^{n} BV_{ik}}{n}$$
Sample **mean** is a poor measure of central tendency when the set of observations is skewed or contains an outlier.

Further measurements are needed.

<table>
<thead>
<tr>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>19</td>
</tr>
</tbody>
</table>

Mean = \( \frac{27}{9} = 3 \)

(Does not represent the dataset well)

[Image of sample data with outlier]

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Mean = \( \frac{9}{9} = 1 \)

[Image of skewed data]

**Skewed**

**Measures of Dispersion**

Measures of the dispersion about the mean of a distribution provide valuable information about the image.

**Range** of a band of imagery \( range_k \) is computed as the difference between the **maximum** \( max_k \) and **minimum** \( min_k \) values:

\[
range_k = \max_k - \min_k
\]
Measures of Dispersion

When the minimum or maximum values are extreme or unusual, the range could be a misleading measure of dispersion.

When unusual values are not encountered, the range is a very important statistic often used in image enhancement functions such as min–max contrast stretching.

Variance: is the average squared deviation of all possible observations from the sample mean. The variance of a band of imagery, $\text{var}_k$, is computed using the equation:

\[
\text{var}_k = \frac{\sum_{i=1}^{n} (BV_{ik} - \mu_k)^2}{n - 1}
\]
\begin{equation}
\text{var } k = \frac{\sum_{i=1}^{n} (BV_{ik} - \mu_k)^2}{n - 1}
\end{equation}

\begin{tabular}{ccc}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1 \\
\end{tabular}

\begin{equation}
\text{var } _i = \frac{\sum_{k=1}^{K} (1 - 1_i)^2}{9 - 1} = 0
\end{equation}

\begin{tabular}{ccc}
19 & 1 & 12 \\
18 & 1 & 13 \\
1 & 14 & 12 \\
\end{tabular}

\begin{equation}
\text{var } _i = \frac{\sum_{k=1}^{K} (BV_{ik} - \mu_k)^2}{n - 1}
\end{equation}

\begin{align*}
&= \frac{(19 - 9)^2 + (11 - 9)^2 + (12 - 9)^2 + (18 - 9)^2 + (1 - 9)^2 + (13 - 9)^2 + (14 - 9)^2 + (12 - 9)^2}{8} \\
&= \frac{432}{8} = 54
\end{align*}

**Standard Deviation ($s_k$):** is the positive square root of the variance.

\begin{equation}
s_k = \sqrt{\text{var } k}
\end{equation}

A small $s_k$ suggests that observations clustered tightly around a central value.

A large $s_k$ indicates that values are scattered widely about the mean.

The sample having the largest variance or standard deviation has the greater spread among the values of the observations.
**Standard Deviation ($s_k$):** is the positive square root of the variance.

$$s_k = \sqrt{\text{var}_k}$$

<table>
<thead>
<tr>
<th>19</th>
<th>1</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>

$$\text{var}_k = \frac{\sum (BV_{ik} - \mu_k)^2}{n}$$

$$= \frac{(19 - 9)^2 + (11 - 9)^2 + (12 - 9)^2 + (18 - 9)^2 + (11 - 9)^2 + (13 - 9)^2 + (14 - 9)^2 + (12 - 9)^2}{8}$$

$$= \frac{432}{8} = 54$$

$$s_k = \sqrt{\text{var}_k} = \sqrt{54} = 7.35$$

---

**Measures of Distribution (Histogram) Asymmetry and Peak Sharpness**

**Skewness** is a measure of the asymmetry of a histogram and is computed using the formula:

$$skewness_k = \frac{\sum_{i=1}^{n} \left( \frac{BV_{ik} - \mu_k}{s_k} \right)^3}{n}$$

A perfectly symmetric histogram has a skewness value of zero.
Histogram of A Single Band of Landsat Thematic Mapper Data

Max. = 102
Min. = 6
Mean = 27
Median = 25
Mode = 9

Histogram of Thermal Infrared Imagery of a Thermal Plume in the Savannah River

Max. = 188
Min. = 38
Mean = 73
Median = 68
Mode = 75
Remote Sensing Multivariate Statistics

The different remote-sensing-derived spectral measurements for each pixel often change together in some predictable fashion.

i.e., an increase or decrease in one band’s brightness value is accompanied by a predictable change in another band’s brightness value.
Remote Sensing Multivariate Statistics

If there is no relationship between the brightness value in one band and that of another for a given pixel, the values are mutually independent;

i.e., an increase or decrease in one band’s brightness value is not accompanied by a predictable change in another band’s brightness value.

Remote Sensing Multivariate Statistics

Remote sensing is often concerned with the measurement of how much radiant flux is reflected or emitted from an object in more than one band (e.g., in red and near-infrared bands).

It is necessary to compute multivariate statistical measures such as covariance and correlation among the several bands to determine how the measurements covary.
Correlation between Multiple Bands of Remotely Sensed Data

To estimate the degree of interrelation between variables in a manner not influenced by measurement units, the correlation coefficient, $r$, is commonly used.

The correlation between two bands of remotely sensed data, $r_{kl}$, is the ratio of their covariance ($\text{cov}_{kl}$) to the product of their standard deviations ($s_k s_l$); thus:

$$r_{kl} = \frac{\text{cov}_{kl}}{s_k s_l}$$

Correlation Coefficient:

- A correlation coefficient of $+1$ indicates a positive, perfect relationship between the brightness values of the two bands.
- A correlation coefficient of $-1$ indicates that the two bands are inversely related.
- A correlation coefficient of zero suggests that there is no linear relationship between the two bands of data.
### Landsat-7 ETM+ Band 1 (Blue band)

<table>
<thead>
<tr>
<th>Band</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>242</td>
<td>65.163137</td>
<td>10.231356</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>115</td>
<td>25.797593</td>
<td>5.956048</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>131</td>
<td>23.958016</td>
<td>8.469890</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>105</td>
<td>26.530666</td>
<td>13.690054</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>193</td>
<td>32.014001</td>
<td>24.296417</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>128</td>
<td>15.103553</td>
<td>12.738188</td>
</tr>
<tr>
<td>7</td>
<td>102</td>
<td>124</td>
<td>110.734372</td>
<td>4.385063</td>
</tr>
</tbody>
</table>

### Landsat-7 ETM+ Band 2 (Green band)

<table>
<thead>
<tr>
<th>Band</th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
<th>Band 5</th>
<th>Band 6</th>
<th>Band 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000000</td>
<td>0.964874</td>
<td>0.953195</td>
<td>0.433582</td>
<td>0.575042</td>
<td>0.724979</td>
<td>0.555425</td>
</tr>
<tr>
<td>2</td>
<td>0.964874</td>
<td>1.000000</td>
<td>0.964263</td>
<td>0.487311</td>
<td>0.626501</td>
<td>0.762857</td>
<td>0.577699</td>
</tr>
<tr>
<td>3</td>
<td>0.953195</td>
<td>0.964263</td>
<td>1.000000</td>
<td>0.570068</td>
<td>0.726797</td>
<td>0.845615</td>
<td>0.653461</td>
</tr>
<tr>
<td>4</td>
<td>0.433582</td>
<td>0.487311</td>
<td>0.570068</td>
<td>1.000000</td>
<td>0.899511</td>
<td>0.788821</td>
<td>0.693087</td>
</tr>
<tr>
<td>5</td>
<td>0.575042</td>
<td>0.626501</td>
<td>0.726797</td>
<td>0.845615</td>
<td>1.000000</td>
<td>0.850000</td>
<td>0.753462</td>
</tr>
<tr>
<td>6</td>
<td>0.724979</td>
<td>0.762857</td>
<td>0.845615</td>
<td>0.788821</td>
<td>0.850000</td>
<td>1.000000</td>
<td>0.814648</td>
</tr>
<tr>
<td>7</td>
<td>0.555425</td>
<td>0.577699</td>
<td>0.653461</td>
<td>0.693087</td>
<td>0.793462</td>
<td>0.814648</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

### Landsat-7 ETM+ Band 3 (Red band)

<table>
<thead>
<tr>
<th>Band</th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
<th>Band 5</th>
<th>Band 6</th>
<th>Band 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>242</td>
<td>65.163137</td>
<td>10.231356</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>115</td>
<td>25.797593</td>
<td>5.956048</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>131</td>
<td>23.958016</td>
<td>8.469890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>105</td>
<td>26.530666</td>
<td>13.690054</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>193</td>
<td>32.014001</td>
<td>24.296417</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>128</td>
<td>15.103553</td>
<td>12.738188</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>102</td>
<td>124</td>
<td>110.734372</td>
<td>4.385063</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Landsat-7 ETM+ Band 4 (Near IR band)
Because spectral measurements of individual pixels may not be independent, some measure of their mutual interaction is needed.

This measure, called the covariance, is the joint variation of two variables about their common mean.

To calculate covariance:

First, compute the corrected sum of products (SP) defined by the equation:

$$SP_{kl} = \sum_{j=1}^{n} (BV_{jk} - \mu_j)(BV_{il} - \mu_l)$$

It is computationally more efficient to use the following formula to arrive at the same result:

$$SP_{kl} = \sum_{j=1}^{n} (BV_{jk} \times BV_{il}) - \frac{1}{n} \sum_{i=1}^{n} BV_{ik} \sum_{j=1}^{n} BV_{jl}$$

This quantity is called the uncorrected sum of products.
Remote Sensing Multivariate Statistics

Then, Covariance is calculated by dividing $SP$ by $(n - 1)$.

The covariance between brightness values in bands $k$ and $l$, $\text{cov}_{kl}$, is equal to:

$$
\text{COV}_{kl} = \frac{SP_{kl}}{n - 1}
$$

Covariance: is the joint variation of two variables about their common mean. $SP_{kl}$ is the corrected Sum of Products between bands $k$ and $l$.

$$
\text{cov}_{kl} = \frac{SP_{kl}}{n - 1}
$$

$$
SP_{kl} = \sum_{i=1}^{n} (BV_{ik} - \mu_k)(BV_{il} - \mu_l)
$$

<table>
<thead>
<tr>
<th>Band $k$</th>
<th>19</th>
<th>1</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>1</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Band $l$</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

$SP_{kl} = (19-9)x(1-1)+(1-9)x(1-1)+(12-9)x(1-1)+(18-9)x(1-1)+(1-9)x(1-1)+(13-9)x(1-1)+
(1-9)x(1-1)+(14-9)x(1-1)+(12-9)x(1-1) = 0$

$\text{Cov}_{kl} = 0$
Covariance: A more efficient formula is:

\[ SP_{kl} = \sum_{i=1}^{n} (BV_{ik} \times BV_{il}) - \frac{1}{n} \left( \sum_{i=1}^{n} BV_{ik} \right) \left( \sum_{i=1}^{n} BV_{il} \right) \]

The sums of products (SP) and sums of squares (SS) can be computed for all possible band combinations.

\[ SP_{kl} = [(19x1)+(1x1)+(12x1)+(18x1)+(1x1)+(13x1)+(1x1)+(14x1)+(12x1)] - [91x9/9] = 0 \]

\[ Cov_{kl} = 0 \]

The sums of products (SP) and sums of squares (SS) can be computed for all possible band combinations.

**Format of a Variance-Covariance Matrix**

<table>
<thead>
<tr>
<th></th>
<th>Band 1 (green)</th>
<th>Band 2 (red)</th>
<th>Band 3 (near-infrared)</th>
<th>Band 4 (near-infrared)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 1</td>
<td><strong>SS_1</strong></td>
<td><strong>cov_1,2</strong></td>
<td><strong>cov_1,3</strong></td>
<td><strong>cov_1,4</strong></td>
</tr>
<tr>
<td>Band 2</td>
<td><strong>cov_2,1</strong></td>
<td><strong>SS_2</strong></td>
<td><strong>cov_2,3</strong></td>
<td><strong>cov_2,4</strong></td>
</tr>
<tr>
<td>Band 3</td>
<td><strong>cov_3,1</strong></td>
<td><strong>cov_3,2</strong></td>
<td><strong>SS_3</strong></td>
<td><strong>cov_3,4</strong></td>
</tr>
<tr>
<td>Band 4</td>
<td><strong>cov_4,1</strong></td>
<td><strong>cov_4,2</strong></td>
<td><strong>cov_4,3</strong></td>
<td><strong>SS_4</strong></td>
</tr>
</tbody>
</table>
Variance–covariance and correlation matrices are the key for principal components analysis (PCA), feature selection, image classification and accuracy assessment.

Remote Sensing Multivariate Statistics

Variance–covariance and correlation matrices are the key for principal components analysis (PCA), feature selection, image classification and accuracy assessment.