30540 - Mapping from Aerial and Satellite Images
PCA and MAF

Anders Richard Pankoke Holm
s144125

Hand-in date: 06/03-2019

Principal Component Analysis (PCA)

The `igalmss` imgae supplied for this exercise contains 30 bandwidths. When we are given an image with that many layers, it is a lot of data for our computer to process, and a lot of it may not even be that useful. We can therefore use PCA in order to produce a linear combination of the image data, which will tell us a lot about the image, but with much less information. The PCA is performed by arranging the pixels of each image into a column, and place these side by side. We then subtract the mean from each column (we column center them). Thus each column has a mean of 0 in the new matrix we call $\overline{\mathbf{X}}$. We then compute the dispersion matrix of $\overline{\mathbf{X}}$ by: $\Sigma = cov(\overline{\mathbf{X}})$. We find the eigenvalues λ and -vectors $\mathbf{V}$ from the dispersion matrix, and sort them from highest to lowest. With the eigenvalues, we can determine how much of the variance is represented from the original image.

Eigenvalues	Percentage [%] of total variance
12619.85	89.62
1022.85	96.88
216.59	98.42
38.33	98.69
35.22	98.94

Table 1: First 5 eigenvalues of dispersion matrix Σ

We see that already at the first 2 eigenvalues, we have represented more than 95% of the variance of the data set. Thus we could tell a lot about the image from the first 2 principle components. These can be computed $\mathbf{Q} = \overline{\mathbf{X}}\mathbf{V}$. If the PCA is done properly $cov(\mathbf{Q})$ should have the eigenvalues down it's diagonal and 0's elsewhere. The first 2 PC's can be seen on following figure.

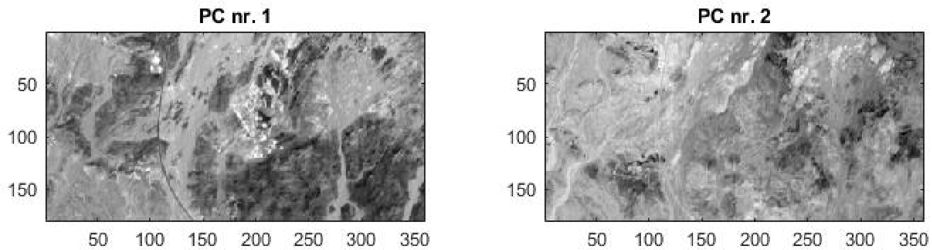

Figure 1: First 2 principal components

Maximum Autocorrelation Factor (MAF)

Much like PCA, with MAF we would like to produce some linear combinations, that can represent our data with less information. We produce two new sets of data, where we shift the respective band by one in the x- and y- direction in order to form a new layer, which is then subtracted from the original band (dimensions are kept by moving the end column/row to the front). We once more arrange all bands into a matrix and column center them to get $\overline{\boldsymbol{X}}_{\Delta x}$ and $\overline{\boldsymbol{X}}_{\Delta y}$, from which we compute dispersion matrices $\Sigma_{\Delta x}$ and $\Sigma_{\Delta y}$. Taking the mean of these, we get Σ_{Δ}. We then solve the problem the problem by maximizing the Rayleigh quotient:

$$\lambda_i = \frac{\mathbf{u}_i^T \Sigma \mathbf{u}_i}{\mathbf{u}_i^T \Sigma_\Delta \mathbf{u}_i} \tag{1}$$

where λ_i is the i'th eigenvalue and $\mathbf{u}_i$ the i'th eigenvector. This can be solved via the general eigenvalue problem $\Sigma\mathbf{U} = \Sigma_\Delta \mathbf{U}\Lambda$, where Λ is a matrix with the eigenvalues sorted in descending order in the diagonal, and $\mathbf{U}$ is the eigenvectors sorted accordingly. We then compute the factors by $\mathbf{F} = \overline{\mathbf{X}}\mathbf{U}$. If the MAF is done properly $corr(\mathbf{F})$ should have the 1's down it's diagonal and 0's elsewhere. All the factors can be seen on following page.

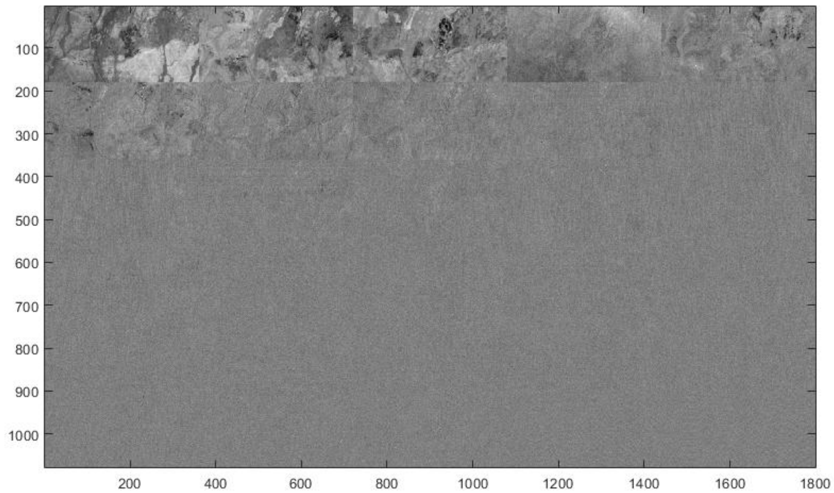

Figure 2: All factors from the MAF

In figure 2 we see all the linear combinations we get from projecting $\overline{\mathbf{X}}$ onto $\mathbf{U}$. The first couple of images clearly showcases some of the features we recognize from the original bands. They are arranged such that we have 5 images per row and 6 in each column. Seemingly only the first 5 images contains useful information, hereafter it is mostly or entirely noise.

MAF is very similar to Minimum Noise Fraction (MNF), where you instead of Σ_Δ have Σ_N, which is the dispersion matrix of the noise in the dataset at all the bandwidth.