Análise de componentes principais

Agostinho Brito

2020

- A análise de componentes principais, ou transformada de Karhunen-Loève, ou transformada de Hotelling, realiza uma transformação ortogonal em um conjunto de variáveis correlacionadas (interdependentes) em um outro conjunto de variáveis linearmente descorrelacionadas.
- Dá-se o nome de “componentes principais” porque no espaço transformado, as primeiras componentes concentram a maior parte da informação presente nos dados originais.

Formulação

- Seja

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

um vetor representante de uma população de dados com n-dimensional.

- Define-se o vetor de média e a matriz de covariância dessa população como

$$\bar{m}_x = E\{x\} \qquad C_x = E\{(x - \bar{m}_x)(x - \bar{m}_x)^T\}$$

- Para uma amostra de M elementos, a média e covariância da população podem ser aproximadas por

$$\bar{m}_x = \frac{1}{M}\sum_{k=1}^{M} x_k \qquad C_x = \frac{1}{M}\sum_{k=1}^{M}(\bar{x}_k\bar{x}_k^T) - \bar{m}_k\bar{m}_k^T$$

- Ex: $\bar{x}_1 = (0,0,0)^T$, $\bar{x}_2 = (1,0,0)^T$, $\bar{x}_2 = (1,1,0)^T$ e $\bar{x}_4 = (1,0,1)^T$.

$$\bar{m}_x = \frac{1}{4}\begin{bmatrix} 3 & 1 & 1 \end{bmatrix}^T \quad C_x = \frac{1}{16}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

- A matriz C_x é real e simétrica. Logo, possui autovalores reais e distintos $\lambda_i, i = 1, 2, \cdots, n$, com seus respectivos autovetores $\bar{e}_i$ associados.
- Assuma que estes autovalores são arranjados na ordem decrescente de valor, tais que $\lambda_j \geqslant \lambda_{j+1}$ para $j = 1, 2, \cdots, n-1$.
- Construa uma matriz A, onde cada linha contenha autovetores unitários associados com os autovalores da matriz C_x. A primeira linha de C_x deve conter o autovetor associado com o maior autovalor, restando na última linha o autovetor associado com o menor autovalor.
- A transformação

$$\bar{y} = A(\bar{x} - \bar{m})$$

é denominada de transformada de Hotelling. A população de vetores transformados possui média igual a zero. Sua matriz de covariância é diagonal, preenchida pelos autovalores da matriz C_x na ordem decrescente de seus valores.

Formulação

- É possível reconstruir a população original usando a transformação

$$\bar{x} = A^T \bar{y} + \bar{m}_x$$

- Se algumas componentes do espaço transformado forem dispensadas, utilizando apenas os autovetores da matriz C_x com os maiores autovalores, usando uma matriz A_K, a transformação

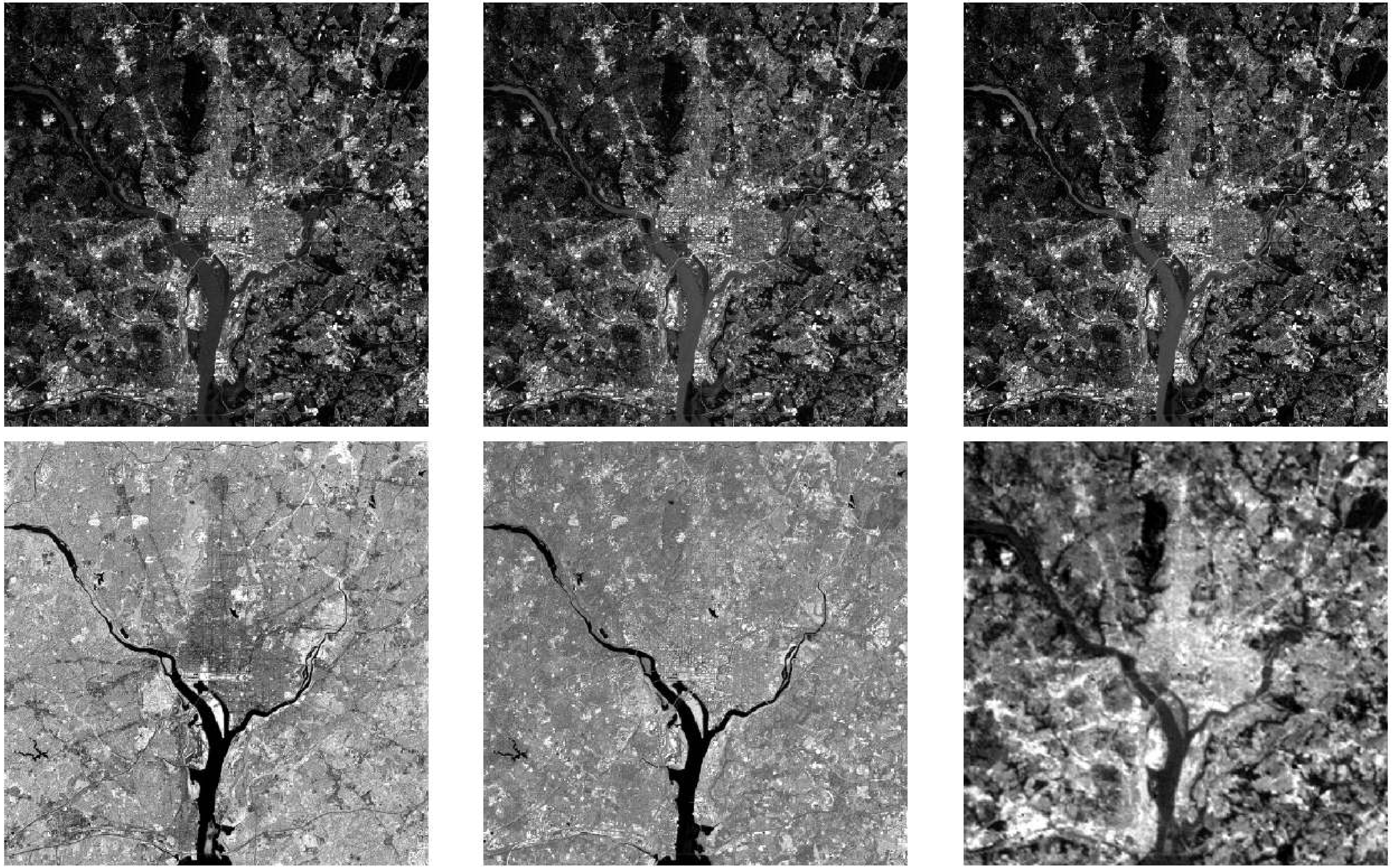
$$\hat{x} = A_K^T \bar{y} + \bar{m}_x$$

  recompõe a população de forma aproximada. Essa ideia possibilita realizar compressão de dados.

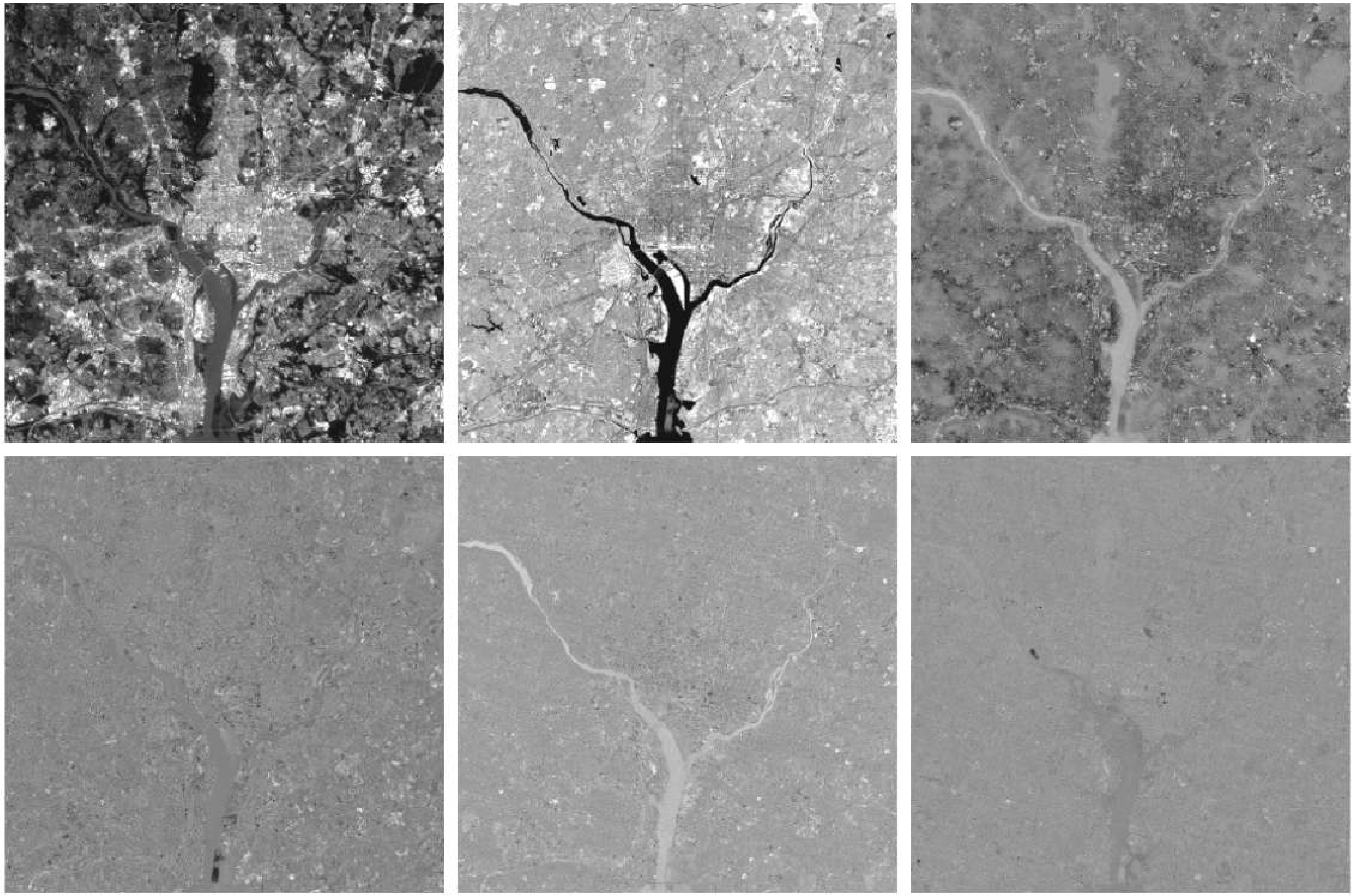
Usos do processo de análise de componentes principais

- Compressão de dados
- Determinação de orientação de objetos

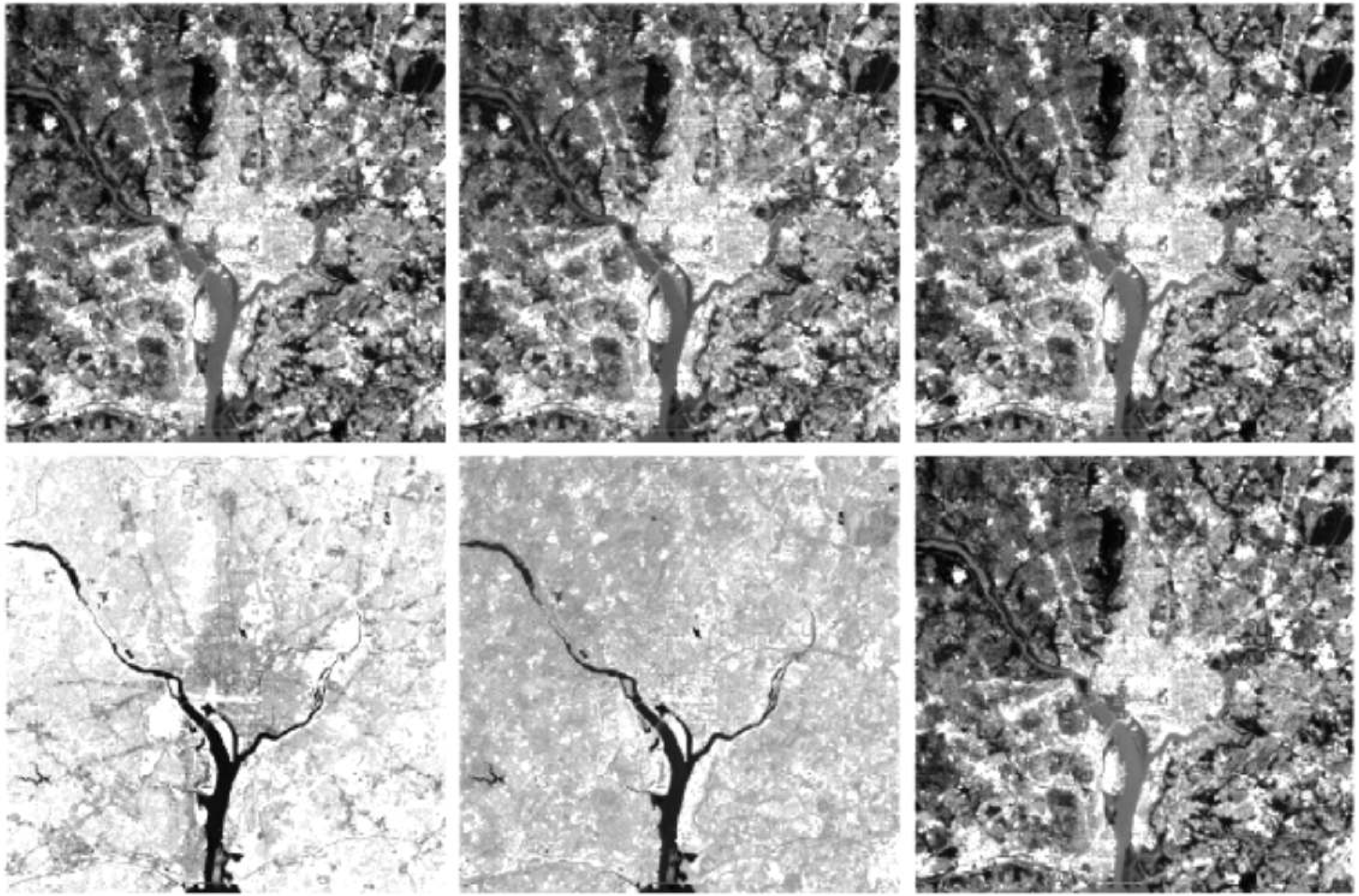
Compressão



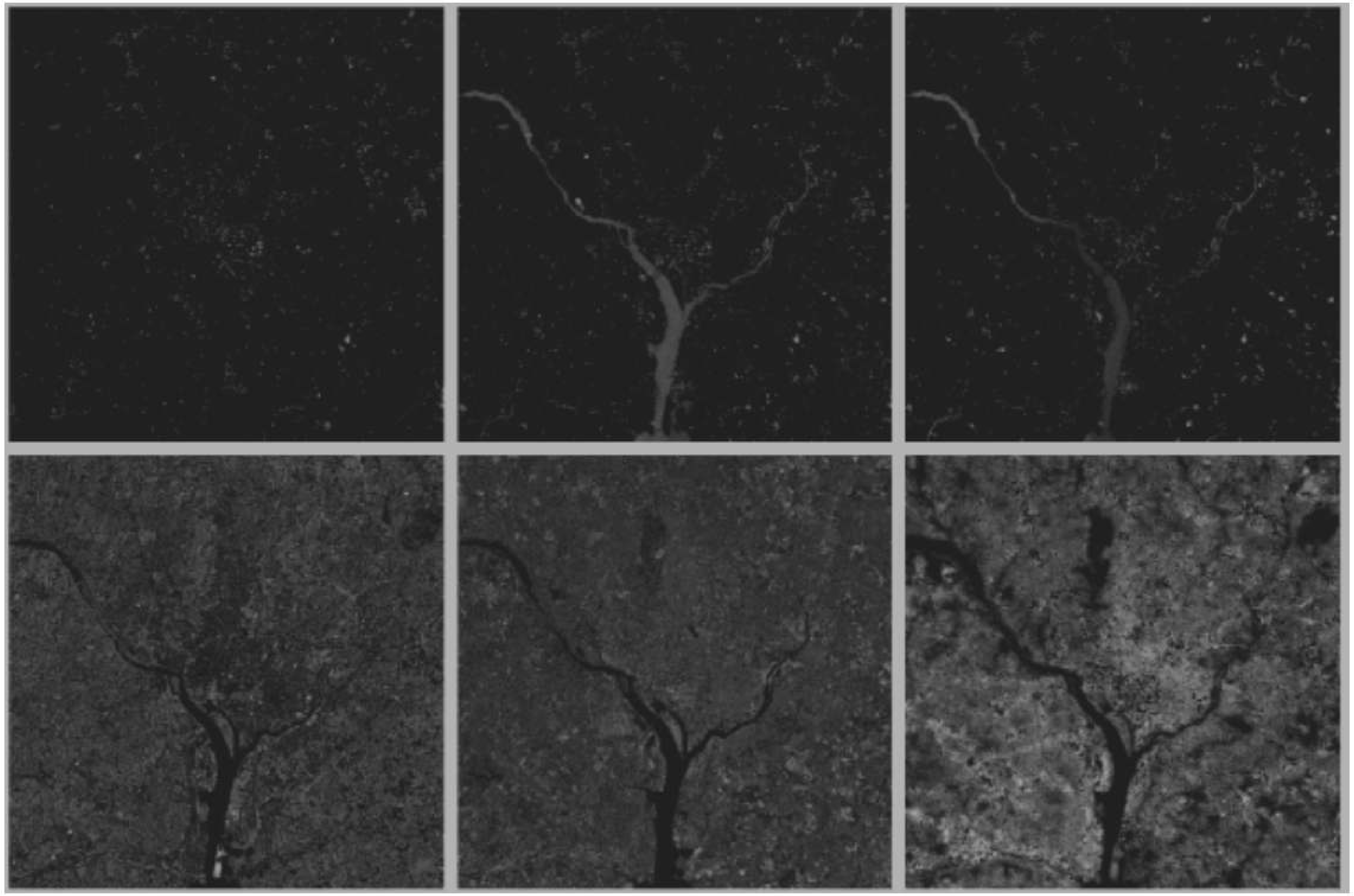
Compressão



Compressão - reconstrução com apenas duas componentes



Compressão - erro produzido



1. Segmentação pelo método de Otsu.
2. Identificação de componentes conectadas.
3. Extração dos contornos.
4. Uso dos pontos do contorno para alimentar a transformada PCA.
5. Uso dos autovetores para indicar a orientação das peças.
6. Cada autovetor está desenhado com tamanho proporcional ao seu autovalor.