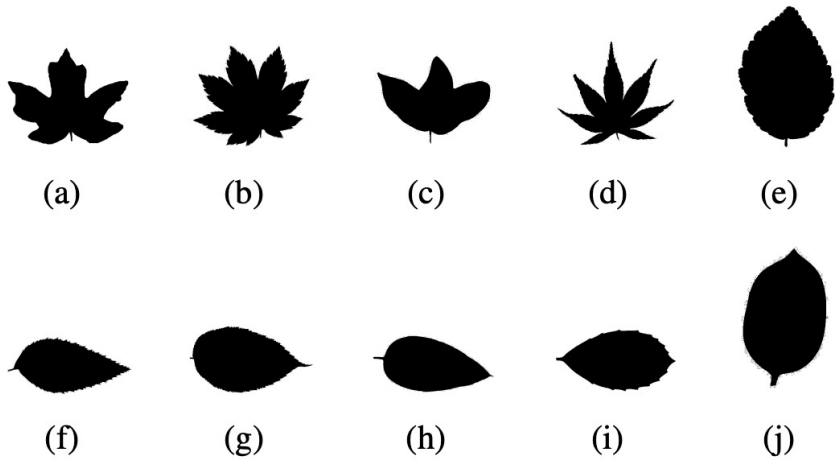
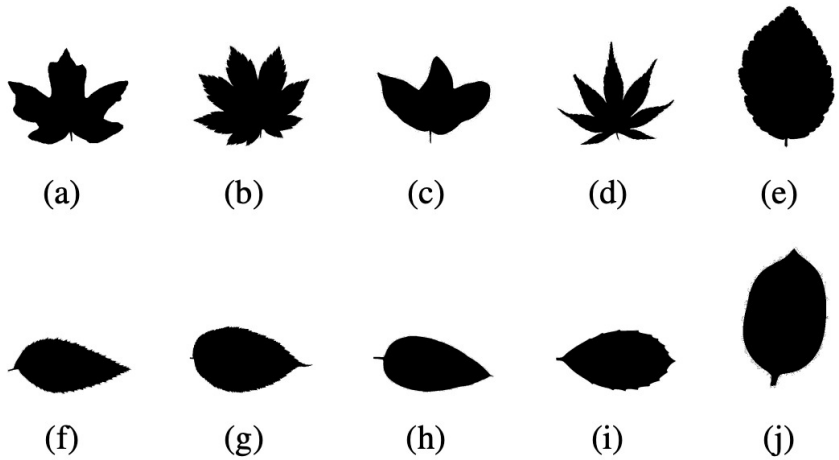
Comparing Multivariate Embedding Methods for Plant Species Identification

March 26, 2025

Introduction

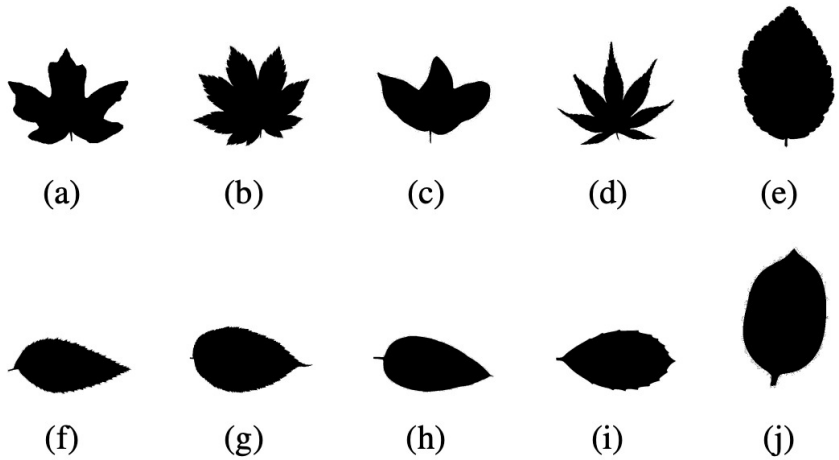


Introduction



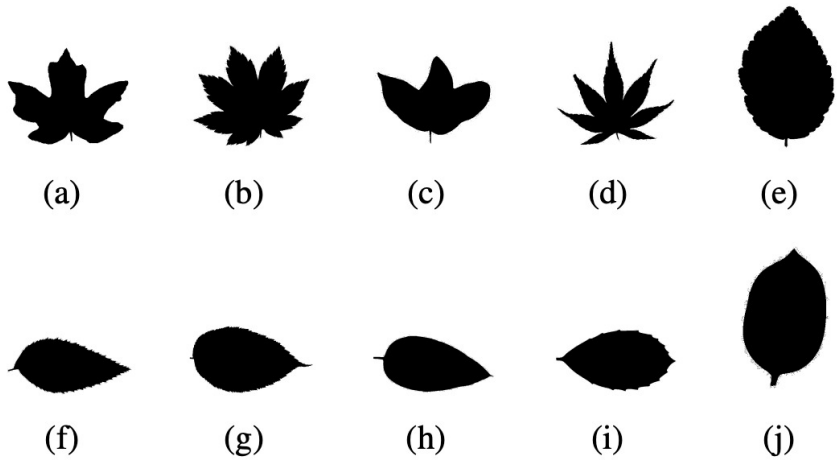
- 98 species, each with 16 images ($N = 1568$)

Introduction



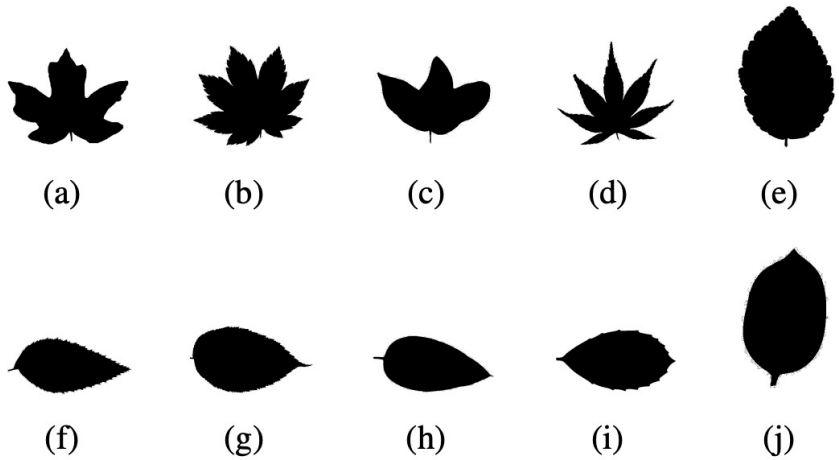
- 98 species, each with 16 images ($N = 1568$)
- 192 features: 64 each of shape, margin and pattern

Introduction



- 98 species, each with 16 images ($N = 1568$)
- 192 features: 64 each of shape, margin and pattern
- Extension- apply dimension reduction before clustering:

Introduction



- 98 species, each with 16 images ($N = 1568$)
- 192 features: 64 each of shape, margin and pattern
- Extension- apply dimension reduction before clustering:
  1. Principal Component Analysis (PCA)
  2. Isometric Mapping (Isomap)
  3. Locally Linear Embedding (LLE)

Principal Component Analysis

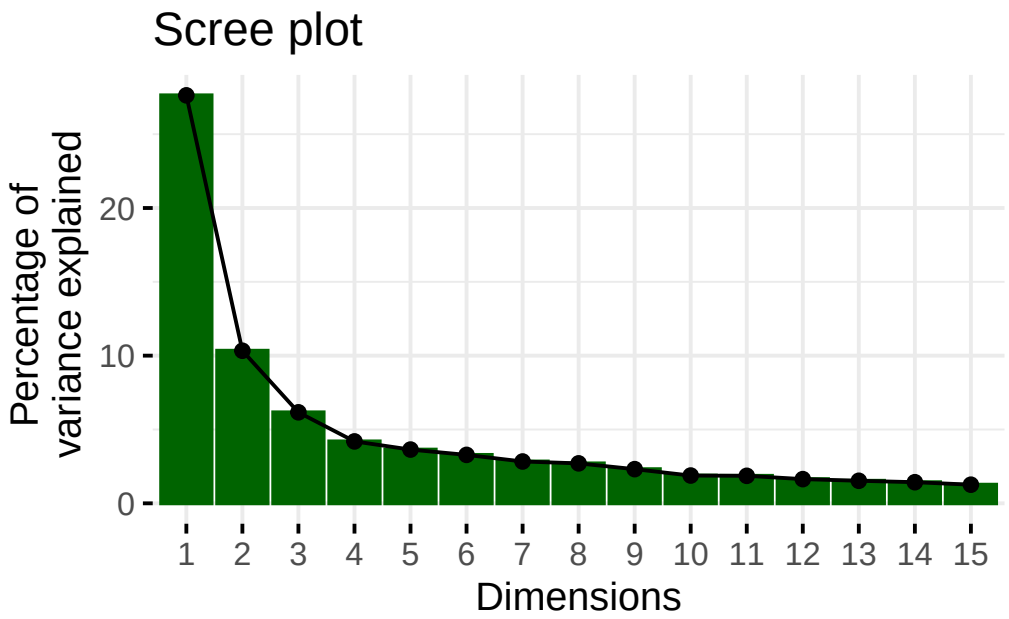
Steps:

1. Calculate the Principal Components (PC's) from the correlation matrix
2. Determine the number of PC's to keep using a Scree plot

Principal Component Analysis

Steps:

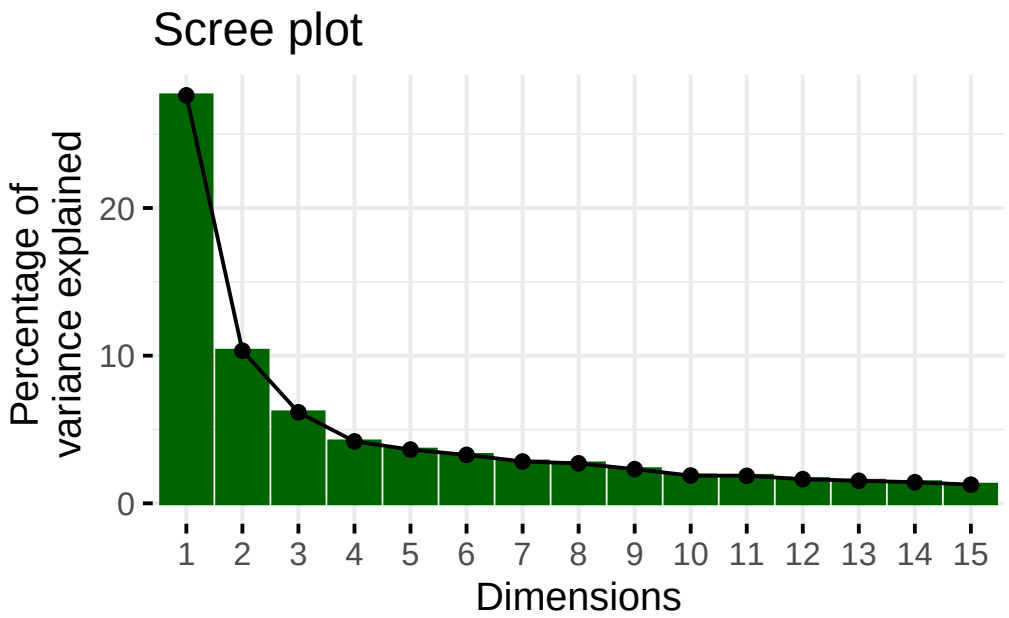
1. Calculate the Principal Components (PC's) from the correlation matrix
2. Determine the number of PC's to keep using a Scree plot



Principal Component Analysis

Steps:

1. Calculate the Principal Components (PC's) from the correlation matrix
2. Determine the number of PC's to keep using a Scree plot



- Lower dimensional embedding to be used: {4 PCs}, {10 PCs}, {30 PCs}

Isometric Mapping

Steps:

1. Construct neighborhood graph
2. (Find k_{min})
3. Calculate geodesic distances
4. Apply MDS
5. Calculate residual variance

Isometric Mapping

Residual variance

$$\text{Residual Variance} = 1 - R^2 = 1 - \text{cor}^2(D_G, D_Y) \quad (1)$$

1. D_G: matrix of geodesic distances between all pairs of observations in $\mathbb{R}^D$
2. D_Y: matrix of Euclidean distances between all pairs of observations in $\mathbb{R}^d$

Isometric Mapping

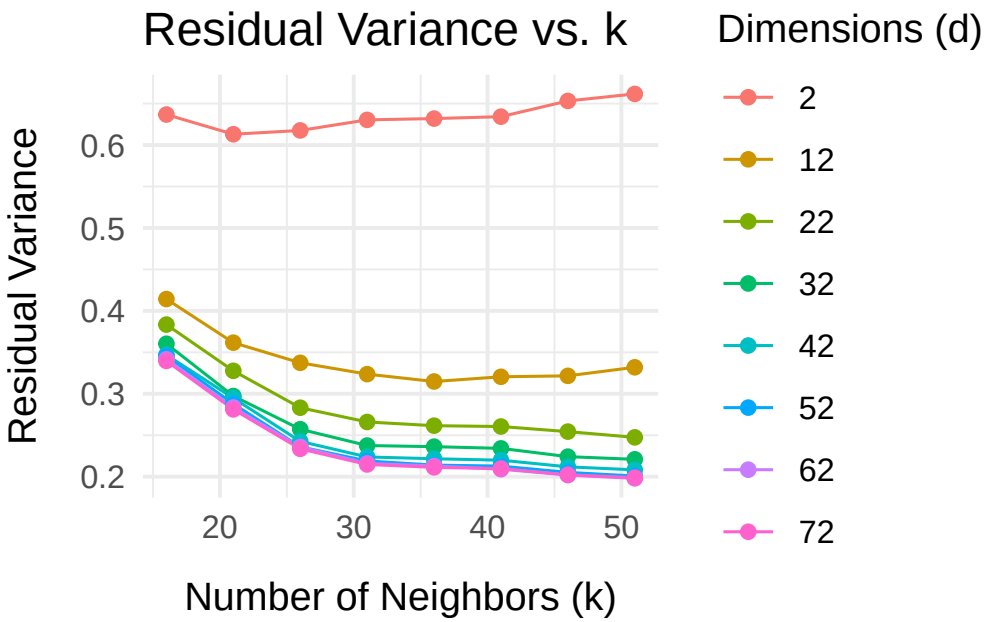
Steps:

1. Construct neighborhood graph
2. (Find k_{min})
3. Calculate geodesic distances
4. Apply MDS
5. Calculate residual variance

Isometric Mapping

Steps:

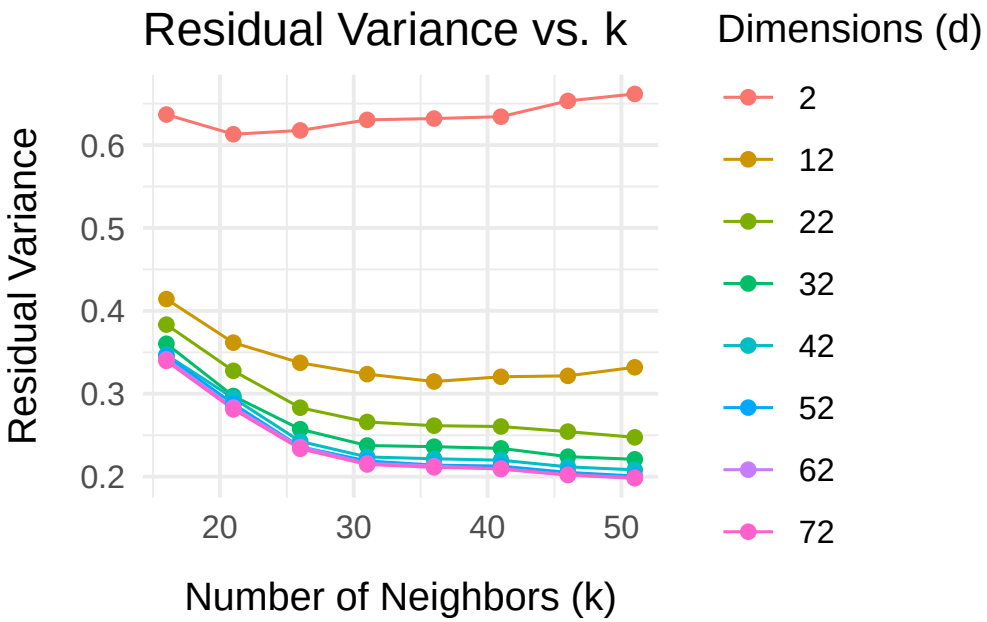
1. Construct neighborhood graph
2. (Find $k_{\min}$)
3. Calculate geodesic distances
4. Apply MDS
5. Calculate residual variance



Isometric Mapping

Steps:

1. Construct neighborhood graph
2. (Find $k_{\min}$)
3. Calculate geodesic distances
4. Apply MDS
5. Calculate residual variance



- Lower dimensional embedding to be used: $\{d = 22,\ k = 16\}$, $\{d = 22,\ k = 21\}$, $\{d = 22,\ k = 26\}$

Locally Linear Embedding

Steps:

1. Find the K-NN of each observation

Locally Linear Embedding

Steps:

1. Find the K-NN of each observation
2. Calculate w_{ij}

Locally Linear Embedding

w_{ij} calculation

$$\mathbf{x}_i \approx \sum_{j \in N(i)} w_{ij} \mathbf{x}_j \quad \forall\ i = 1, \ldots, N \tag{2}$$

subject to

$$\sum_j w_{ij} = 1 \tag{3}$$

$$w_{ij} = 0 \quad \forall\ j \notin N(i) \tag{4}$$

Locally Linear Embedding

Steps:

1. Find the K-NN of each observation
2. Calculate w_{ij}
3. Calculate lower dimensional embedding $\mathbf{Y}$

Locally Linear Embedding

Minimize the Reconstruction Error

$$\Phi(\mathbf{Y}) = \sum_{i=1}^{N} \left\| \mathbf{y}_i - \sum_{j} w_{ij} \mathbf{y}_j \right\|^2 \tag{5}$$

where

$$\mathbf{Y} \in \mathbb{R}^{n \times d} \tag{6}$$

Locally Linear Embedding

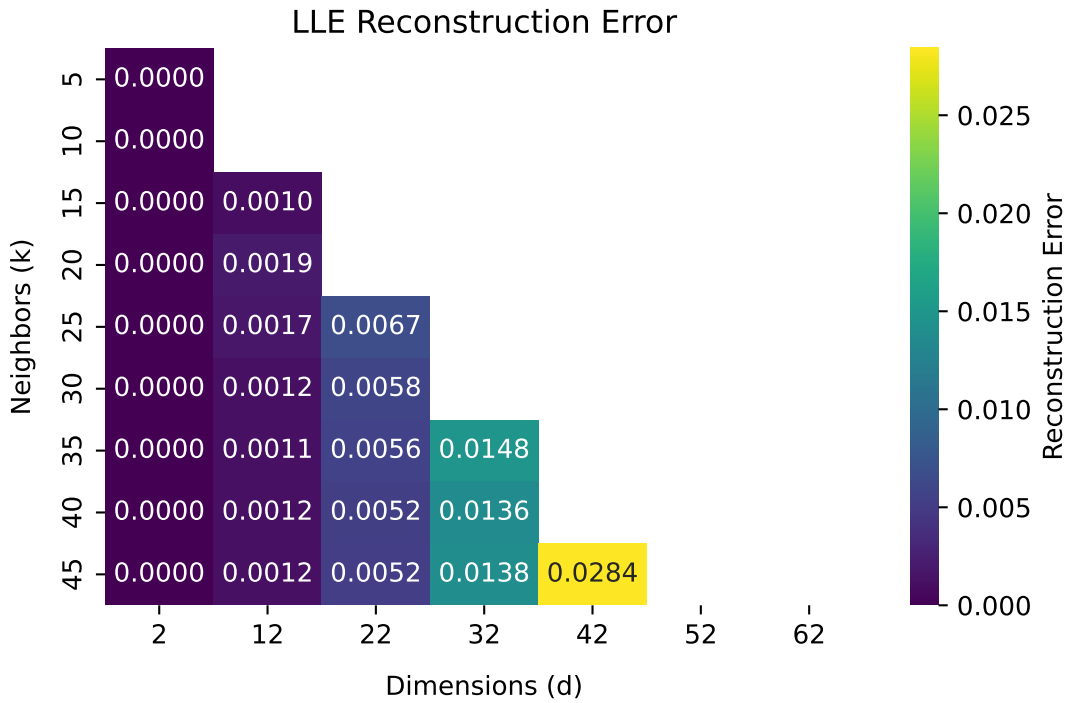
Steps:

1. Find the K-NN of each observation
2. Calculate w_{ij}
3. Calculate lower dimensional embedding $\mathbf{Y}$
4. Calculate the reconstruction error

Locally Linear Embedding

Steps:

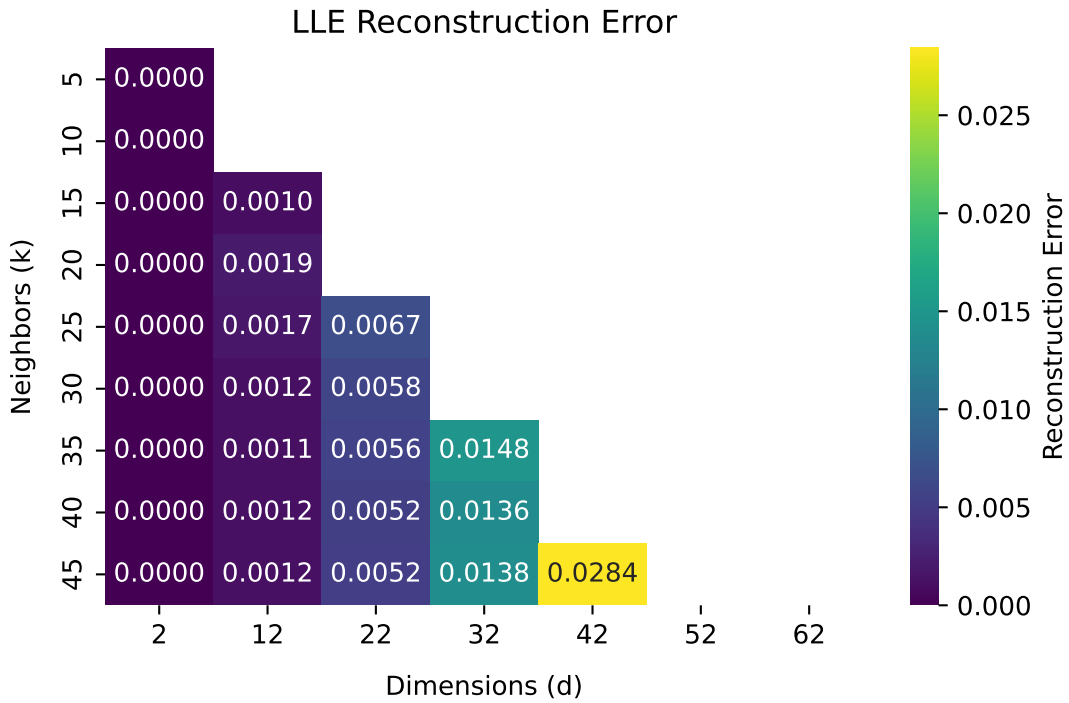
1. Find the K-NN of each observation
2. Calculate w_{ij}
3. Calculate lower dimensional embedding $\mathbf{Y}$
4. Calculate the reconstruction error



Locally Linear Embedding

Steps:

1. Find the K-NN of each observation
2. Calculate w_{ij}
3. Calculate lower dimensional embedding **Y**
4. Calculate the reconstruction error



- Lower dimensional embedding to be used: $\{d = 2,\ k = 15\}$, $\{d = 12,\ k = 15\}$, $\{d = 22,\ k = 25\}$, $\{d = 32,\ k = 35\}$, $\{d = 42,\ k = 45\}$

What's left?

- Apply a clustering algorithm
  1. K-Means
  2. Kernel K-Means
- Determine if dimension reduction improved the results

Comments and discussion

Questions?

References

Beck, M. A., Liu, C. Y., Bidinosti, C. P., Henry, C. J., Godee, C. M., Ajmani, M. (2020). An embedded system for the automated generation of labeled plant images to enable machine learning applications in agriculture. *PLoS One*, 15(12), e0243923.

Tariku, G., Ghiglieno, I., Gilioli, G., Gentilin, F., Armiraglio, S., Serina, I. (2023). Automated Identification and Classification of Plant Species in Heterogeneous Plant Areas Using Unmanned Aerial Vehicle-Collected RGB Images and Transfer Learning. *Drones*, 7(10), 599.

Christenhusz, M. J. M., Byng, J. W. (2016). The number of known plant species in the world and its annual increase. *Phytotaxa*, 261(3), 201–217.

Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., Soares, J. V. B. (2012). Leafsnap: A Computer Vision System for Automatic Plant Species Identification. In *Proc. ECCV*, LNCS 7573, 502–516.

Mallah, C., Cope, J., Orwell, J. (2013). Plant Leaf Classification Using Probabilistic Integration of Shape, Texture and Margin Features. In *Proc. IASTED Int. Conf. SPPRA*, 105–110.