FACE RECOGNITION – EIGENFACES

Danny Styvens Cardona Pineda, Juan Camilo Ceballos Arias, Juan Esteban Torres Marulanda, Miguel Ángel Mejía Muñoz

Master in Data Sciences and Analytics

Medellin
June - 2019

About the project:

- The aim of the project is to recognize a person's face by making a comparison between its characteristics and those of other people's faces which are already known. To perform this task, Eigenfaces method is implemented together with a previous preparation of the images.

- Application of distances for outlier detection and image classification by mathematical calculations.

First section: Outliers identification

- **Datasets used:**
  - Faces94:
    - Directory: male
  - Landscapes

- **Number of images:**
  - Faces: 2260
  - Natural landscapes: 210
  - Total: 2470

- **Initial calculations:**
  - Mean image
  - Median image
  - Real median image
  - Atypical data distances (Manhattan, Euclidean, Chebyshev, Minkowsky):
    - Based on percentile value
    - Based on interquartile range

Second section: Eigenfaces method

- **Datasets used:**
  - Faces94:
    - Directories: male, female, malestaff
  - Landscapes

- **Number of images:**
  - Faces: 3059
  - Natural landscapes: 210
  - Total: 3269

- **Calculations made:**
  - Covariance matrix (centralizing with respect to the mean image)
  - Singular value decomposition (for PCA)
  - Creation of subspace - facespace (from eigenvectors)
  - Projection of images on subspace.

Second section: Eigenfaces method

- **Calculations made:**
  - Distances and outliers:
    - Manhattan, Euclidean, Chebyshev
    - Three standard deviations away
  - Face recognition from image projection on subspace

Third section: Image classification

- **Datasets used:**
  - Faces94:
    - Directories: male, female, malestaff
  - Landscapes

- **Number of images:**
  - Faces: 3059
  - Natural landscapes: 210
  - Total: 3269

- **Calculations made and models used:**
  - PCA from real median face and PCA assisted by libraries
  - Unsupervised models:
    - K-means for clustering
    - T – Distributed stochastic neighbor embedding (TDSNE)
  - Supervised models:
    - Linear discriminat analysis (LDA)
    - Logistic regression

References

- Jain AK. Data clustering: 50 years beyond K-means.Pattern Recognit Lett, 31(8):651–66, 2010
- P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using classspecific linear projection).IEEE Xplore, 19(7):711–20, 1997
- K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images).J. Opt. Soc.Am. A, 14(8):1724–33, 1997
- Sturdivant R Hosmer D., Lemeshow S. Logistic regression models for multinomial and ordinal outcomes.InApplied Logistic Regression, pages 269–310, Hoboken, New Jersey, 2013
- Sturdivant R Hosmer D., Lemeshow S. Discrimination and classification. InApplied Multivariate StatisticalAnalysis RichardJohnson DeanWichern, pages 575–670, United States of America, 2014
- Sima L Karl W. Principal components analysis. InApplied Multivariate Statistical Analysis, pages 319–57,Berlin, Germany, 2014
- Xu Z. Xiuming M Lan G, Wenya L. Improved PCA + LDA Applies to Gastric Cancer Image ClassificationProcess.Physics Procedia, 24:1689–95, 2012
- M. Macho.Topología de Espacios Métricos.2010.Retrieved fromhttp://www.ehu.eus/ mtwmastm/TEM0910.pdf

References

- Somers K Peres P., Jackson D. How many principal components? stopping rules for determining the numberof non-trivial axes revisited.).Computational Statistics Data Analysis, 49(4):974–97, 2005
- M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. 1991 IEEE ComputerSociety Conference on Computer Vision and Pattern Recognition Proceedings, pp.586-591, 1991.https://doi.org/10.1109/CVPR.1991.139758
- W. Zhao, R. Chellappa, P. J. Phillips, and A Rosenfeld. Face Recognition: A Literature Survey.).ACMComput. Surv., 35(4):399–458, 2003
- L.v.d. Maaten, G. Hinton. Visualizing data using t-SNE. Journal of Machine Learning Research, Vol 9(Nov), pp. 2579—2605. 2008.

Thank you!