

Data Mining

Versuch 6 Gesichtserkennung

von

Armin Schwarz (as219)

Florian Tatzel (ft020)

Marc Walter (mw136)

Das git Repository mit den Quellcodes findet sich unter <https://github.com/floriant/DataMining>
Alle Dateien zu diesem Versuch befinden sich im Unterordner Versuch_6.

2.5 Vor dem Versuch zu klärende Fragen

Was sind Eigenvektoren und Eigenwerte?

Eigenvektor:

Der Eigenvektor ist ein um einen Betrag λ skaliertes Vektor des Ursprungsvektors. Bei einer Matrixtransformation wird ein Eigenvektor nur skaliert, aber nicht gedreht.

Eigenvektoren sind die Vektoren, über die die Daten die maximale Abweichung haben.

Eigenwerte:

Der Betrag um den der Ursprungsvektor skaliert wird, ist der Eigenwert.

Was versteht man unter Eigenfaces?

Eigenfaces sind die Eigenvektoren (u_i), die für die weitere Berechnung verwendet werden.

Eigenfaces sind im Prinzip Bilder von normalisierten Gesichtern, die durch Eigenvektoren ausgerichtet werden.

Wie kann mit der PCA eine Dimensionalitätsreduktion durchgeführt werden?

Die Dimensionalität wird im Grund dadurch reduziert, dass man die Dimensionen bei der Kovarianzmatrix umkehrt.

So sind die Anzahl der Bildvektoren die vorgegebene Dimension statt der bisherigen.

Grund dafür ist die Annahme, dass nur bestimmte Vektoren relevant sind, während die Meisten nahe null liegen und deswegen für die Eindeutigkeit nicht relevant sind.

Wie werden mit dem Python Modul Image Bilder in ein Python Programm geladen?

```
import Image
#manchmal auch:
#from PIL import Image

im = Image.open("../res/FaceRecogBilder/training/1-2.png")
print im.format, im.size, im.mode
im.show()
```

Hinweis: Des Öffnen funktioniert `im.show()` nicht, sondern wirft nur einen Fehler aus

3.4 Aufgaben

1. Ab welcher Anzahl K von verwendeten Eigenvektoren treten Fehlklassifikationen ein?

Bei unseren Gesichtern (Bilder 5 bis 7) treten zuverlässig Fehlklassifikationen ab einer Anzahl von zwei Eigenvektoren auf, bei anderen Bildern allerdings schon mit vier Eigenvektoren.

2. Wie groß ist dann die Mindestdistanz zwischen Test- und nächstliegendem Trainingsbild?

Distanz bei Flo (5): 61.914 // Marc(6): 8.54 // Armin (7): 62.759

3. Wie ändert sich die Distanz zwischen Bildern, wenn die Anzahl der Eigenvektoren reduziert wird?

Die Distanz wird kleiner, weil die Anzahl der Punkte, zwischen denen die Distanz berechnet wird, abnimmt.

4. Wie könnte dieser Einfluss der Eigenvektor-Anzahl auf die Mindestdistanz reduziert werden?

Wenn nach der Auswahl der Eigenvektoren diese Vektoren normiert werden. Dadurch wirkt sich die Anzahl der Eigenvektoren nicht mehr so stark aus.

5. Nennen Sie zwei Algorithmus-unabhängige Parameter, die starken Einfluss auf Rate korrekter Gesichtserkennungen haben.

Wichtig für die korrekte Gesichtserkennung sind unter anderem:

- gleiche Bildwinkel und gleiche Gesichtsausschnitte (nicht teilweise von Stirn bis Kinn und manchmal von Augen bis Oberkörper)
- gleiche Belichtung und Kamera
- Brillen, Mützen, Ohrringe...
- Bildfehler, Rauschen, Kontrast