

# Übung Algorithm Engineering

Toni Barth und Max Haarbach

27. Juni 2018

## 1 Heuristiken

### 1.1 Heuristik 1: Zufallsdrehungen

Die erste Heuristik führt eine bestimmte Anzahl an Drehungen, die von der Größe der Instanz abhängt, an zufällig ausgewählte Knoten aus. Dieser Vorgang wird wiederum je nach Größe der Instanz mehrfach durchgeführt und am Ende die Sortierung mit dem geringsten Abstand als Ergebnis ausgegeben.

### 1.2 Heuristik 2: Optimal Leaf Ordering nach [BGJ01]

Die zweite Heuristik nutzt das Verfahren, das Bar-Joseph und weitere für eine möglichst schnelle und optimale Sortierung von hierarchisch geclusterten Datensätzen entwickelt haben. Dabei wird folgender rekursiver Ansatz verfolgt: Sollen Kosten für einen bestimmten Knoten berechnet werden, setzen sich diese aus den Kosten der beiden Kindknoten und dem Abstand der beiden inneren Blätter dieser beiden Teilbäume. Sofern der Knoten, für den Kosten berechnet werden sollen, ein Blatt bzw. einen Datensatz darstellt, betragen dessen Kosten 0. Dies ist daher das Rekursionsende. Begonnen wird üblicherweise mit dem Wurzelknoten, da man dadurch am Ende auch die gesamten Abstandskosten berechnet hat.

## 2 Ziele

Durch die Experimente sollen sowohl

- die Laufzeiten der Heuristiken bei unterschiedlichen Größenordnungen bezüglich der Anzahl der Testobjekte als auch
- die Güte aufgrund der Ähnlichkeiten zu den jeweiligen Originalbildern

ermittelt und verglichen werden.

## 3 Faktoren

Beim „Leaf-ordering“ sind lediglich 2 Faktoren von Bedeutung:

Zum Einen bestimmt die Größe der Bilder, die im Endeffekt die Anzahl der Testobjekte widerspiegelt, die Laufzeit der Heuristiken. Zum Anderen spielt auch deren Struktur oder Art eine Rolle, die sich allerdings schwer in konkrete Messgrößen oder Werte fassen lassen.

## 4 Testinstanzen

Gemäß der Faktoren werden auch die Testinstanzen, die durch Grauwert-Bilder realisiert sind, in die entsprechenden Kategorien unterteilt:

- Größen:
  - 10
  - 50
  - 100
  - 500
  - 1000
  - 2500
  - 5000

- Arten:
  - (symmetrische) Testbilder
  - Fotos der realen Welt
  - Farb- bzw. Grau-Übergänge

## 5 Gemessene Laufzeiten

An dieser Stelle werden die Laufzeiten der Heuristiken für die jeweiligen Testinstanzen dokumentiert. Die Einheiten der Messungen sind jeweils Sekunden (s).

### 5.1 Gradienten

### 5.2 Fotos

### 5.3 Testbilder

## Literatur

- [BGJ01] Ziv Bar-Joseph, David K Gifford und Tommi S Jaakkola.  
 „Fast optimal leaf ordering for hierarchical clustering“.  
 In: *Bioinformatics* 17.suppl\_1 (2001), S22–S29 (siehe S. 1).