

# Grundlagen der Ähnlichkeit

(Ähnlichkeitsmaße)

Georg Ruß  
Otto-von-Guericke-Universität Magdeburg  
georg.russ@student.uni-magdeburg.de

# Gliederung

- Einleitung / Motivation
- Ähnlichkeitstheorien
- Beispiele
- Zusammenfassung / Ausblick

# Motivation

- Traditionelles Matching in MM-DB kaum anwendbar
- Stattdessen Ähnlichkeitsanfragen
  - Query-by-Content
  - Query-by-Example
- Ähnlichkeitsmaße basierend auf menschlicher Wahrnehmung
- Metrische Maße nur bedingt geeignet

# Feature Contrast Model (1)

- Amos Tversky 1977, Psychological Review

- Annahmen:

- Matching  $s(a, b) = F(A \cap B, A - B, B - A)$

- Monotonie  $s(a, b) > s(a, c)$

- wenn  $A \cap C \subseteq A \cap B$

- $A - B \subseteq A - C$

- $B - A \subseteq C - A$

# Feature Contrast Model (2)

## – Unabhängigkeit von Merkmalen

- Wenn  $(a,b)$  und  $(c,d)$  sowie  $(a',b')$  und  $(c',d')$  in denselben beiden Komponenten übereinstimmen sowie  $(a,b)$  und  $(a',b')$  sowie  $(c,d)$  und  $(c',d')$  in der verbleibenden dritten Komponente übereinstimmen, so gilt:

$$s(a,b) \geq s(a',b') \Leftrightarrow s(c,d) \geq s(c',d')$$

Linearkombination einer Funktion der gemeinsamen Merkmale und der unterscheidenden Merkmale

$$S(a,b) \geq S(c,d) \Leftrightarrow s(a,b) \geq s(c,d)$$

$$S(a,b) = f(A \cap B) - \alpha \cdot f(A - B) - \beta \cdot f(B - A)$$

# Feature Contrast Model (3)



**a**



**b**



**a'**



**b'**



**c**



**d**



**c'**



**d'**

# Fuzzy Feature Contrast Model (1)

- Erweiterung des Feature Contrast Model um Fuzzy-Eigenschaften
- Ermittlung der Wahrheitswerte aus den Messungen auf  $\Omega$

$$T(P\omega) = \mu(\phi(\omega))$$

- Menge der wahren Prädikate über  $\phi$

$$\mu(\phi) = \{\mu_1(\phi), \dots, \mu_n(\phi)\}$$

- Auffälligkeit der Fuzzy-Menge

$$f(\mu) = \sum_{i=1}^p \mu_i$$

# Fuzzy Feature Contrast Model (2)

- Redefinition von Schnitt und Differenz:

$$\mu_{\cap}(\phi, \psi) = \{\min\{\mu_1(\phi), \mu_1(\psi)\}, \dots, \min\{\mu_p(\phi), \mu_p(\psi)\}\}$$

$$\mu_{-}(\phi, \psi) = \{\max\{\mu_1(\phi) - \mu_1(\psi), 0\}, \dots, \max\{\mu_p(\phi) - \mu_p(\psi), 0\}\}$$

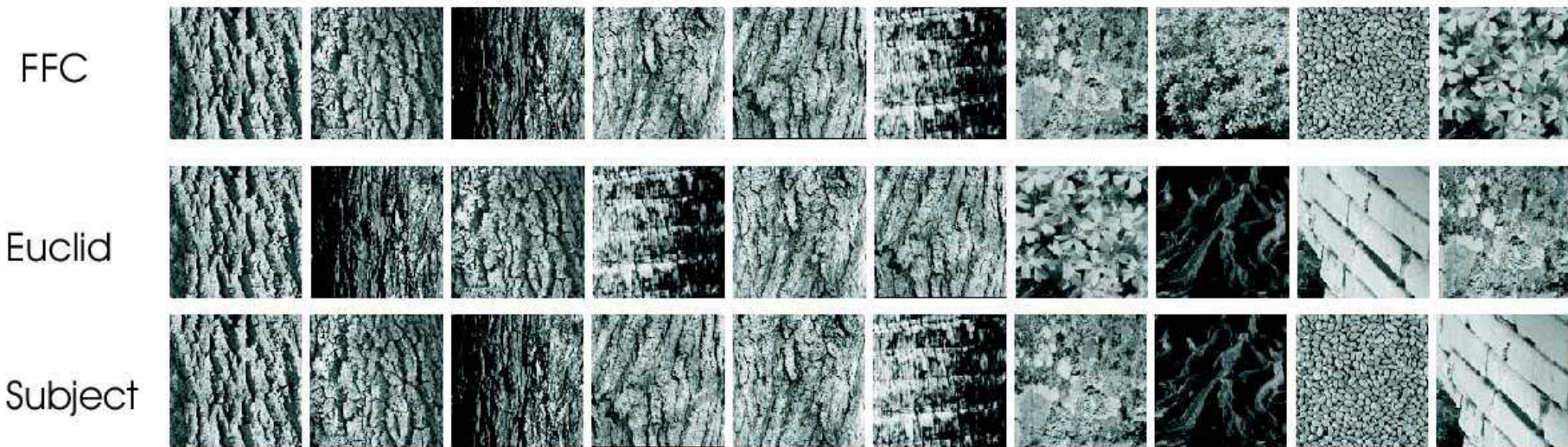
- Ähnlichkeitsfunktion:

$$S(\phi, \psi) = \sum_{i=1}^p \min\{\mu_i(\phi), \mu_i(\psi)\} - \alpha \cdot \sum_{i=1}^p \max\{\mu_i(\phi) - \mu_i(\psi), 0\} - \beta \cdot \sum_{i=1}^p \max\{\mu_i(\psi) - \mu_i(\phi), 0\}$$

– Vgl. FCM:  $S(a, b) = f(A \cap B) - \alpha \cdot f(A - B) - \beta \cdot f(B - A)$

# Beispiel (1)

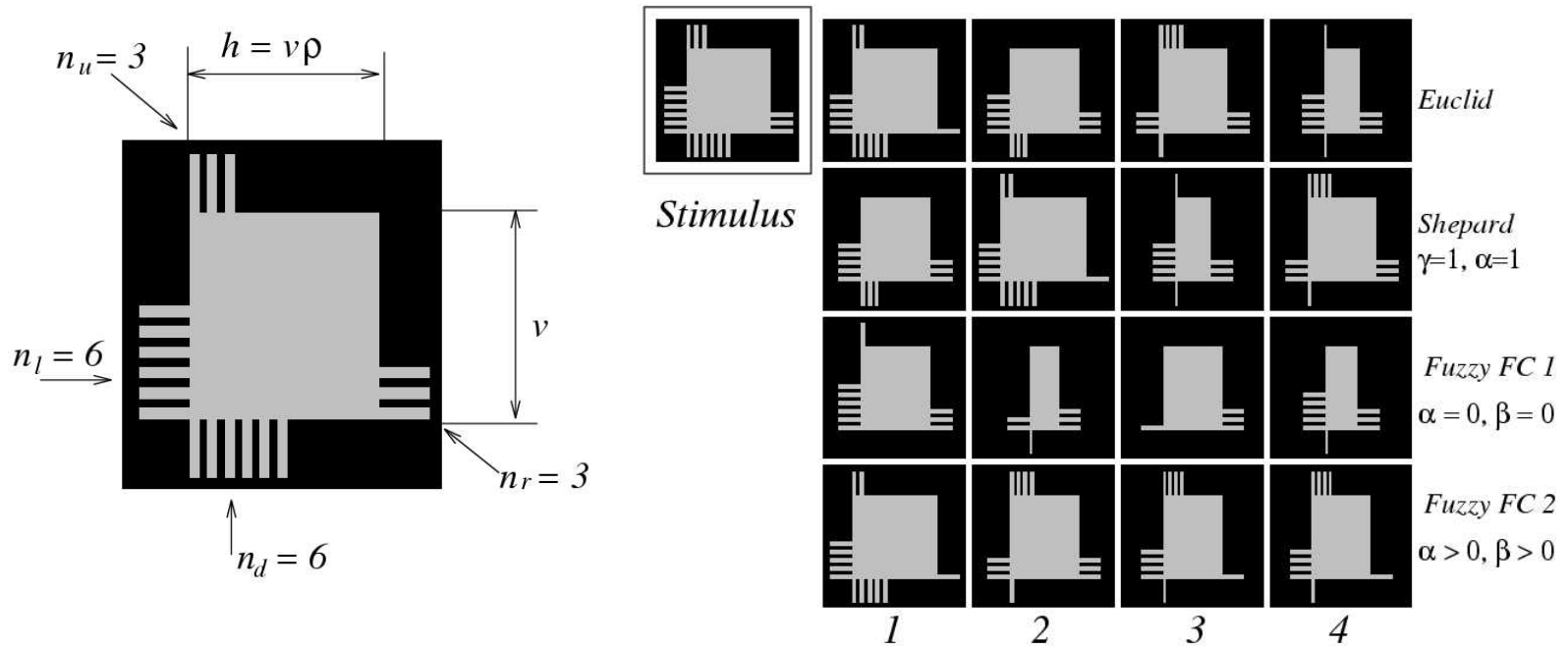
## Ähnlichkeitsbewertung von Texturen



Aus: Santini, Jain: Similarity Measures

# Beispiel (2)

## Ähnlichkeitsbewertung von Mustern



Aus: Santini, Jain: Similarity Queries in Image Databases

# Zusammenfassung / Ausblick

- Feature Contrast Model
- Fuzzy-Methoden verbessern FCM
- Weiter notwendig
  - Merkmals-Abhängigkeiten
  - Basisoperatoren für komplexe Datenbankabfragen
  - geeignete Merkmalsextraktion für verschiedene MM-Objekte

# Axiome einer Metrik

- Abstandsfunktion:  $d(S_A, S_B)$
- Selbstähnlichkeit:  $d(S, S) = 0 \forall S$
- Minimalität:  $d(S, S) \leq d(S, R)$
- Symmetrie:  $d(S, R) = d(R, S)$
- Dreiecksungleichung:  $d(S_A, S_C) \leq d(S_A, S_B) + d(S_B, S_C)$

# Thurstone-Shepard-Model

- Stimuli: normalverteilte hochdimensionale Vektoren
- Instanz: psychologische Momentaufnahme

- Distanzfunktion: Minkowski-Distanz 
$$d = \left[ \sum_{k=1}^n |x_k - y_k|^\gamma \right]^{\frac{1}{\gamma}}$$

- Ähnlichkeit:  $g(d) = \exp(-d^\alpha)$

- Spezialfälle: Euklidische Distanz  $\gamma = \alpha = 2$   
City-Block-Distanz  $\gamma = \alpha = 1$