

Rheinisch Westfälische Technische Hochschule Aachen
Lehrstuhl für Informatik VI
Prof. Dr.-Ing. Hermann Ney

Seminar Medizinische Bildverarbeitung im Sommersemester 2003

Elastic Bunch Graph Matching für die Objekterkennung

David Rybach

Betreuer: Daniel Keysers

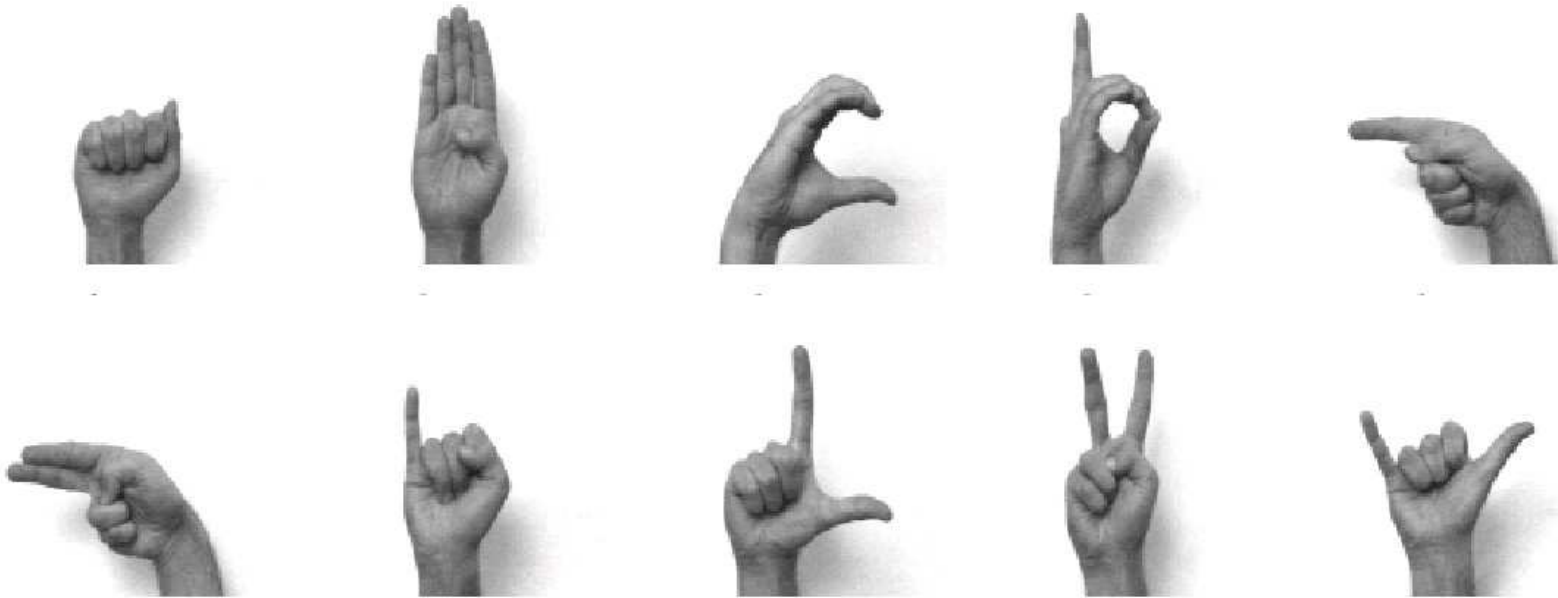
18. Juli 2003

Basis dieses Vortrags: J. Triesch and C. von der Malsburg, „A system for person-independent hand posture recognition against complex backgrounds“, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1449 – 1453, 2001

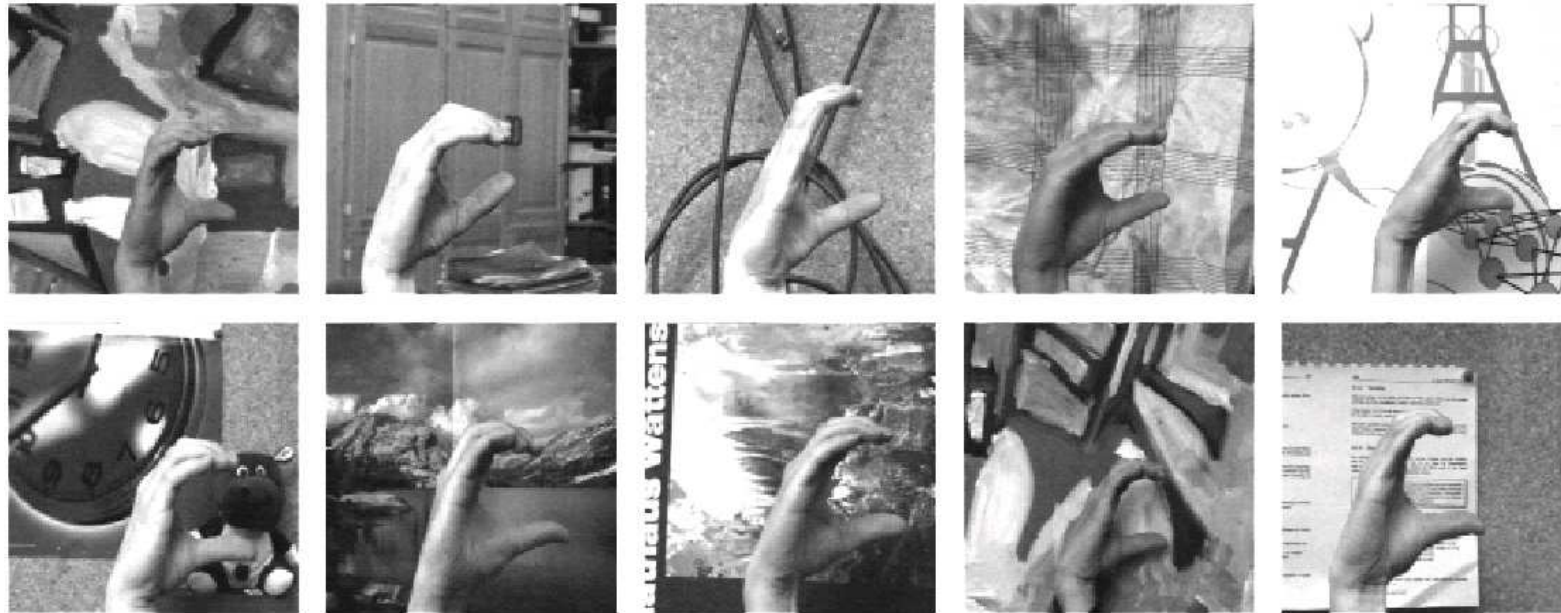
Elastic Bunch Graph Matching für die Objekterkennung

- Einführung
- Beispiel Handgesten-Erkennung
- Grundlagen
- Matching
- Training
- Erkennung
- Ergebnisse und Probleme
- Zusammenfassung

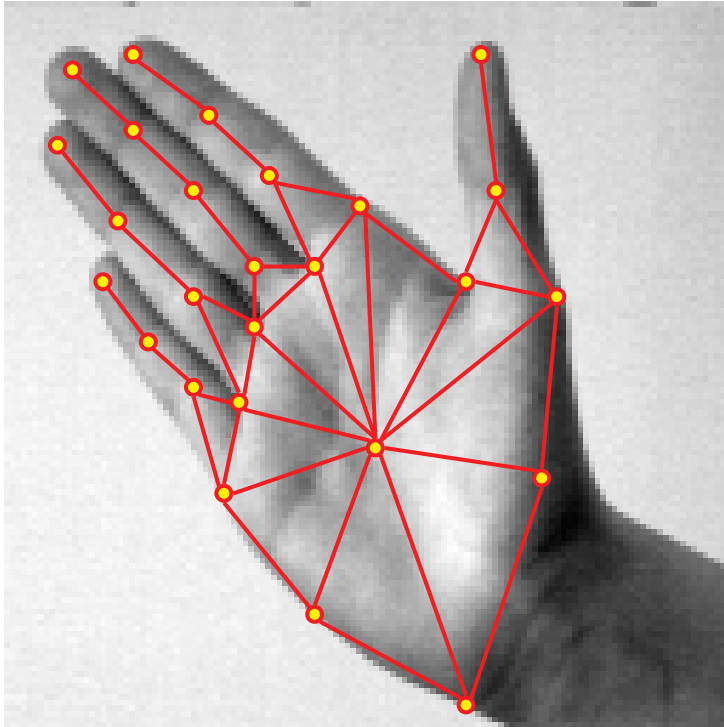
- Objekterkennung: Klassifizierung eines Bildes
- Elastic Bunch Graph Matching (EBGM): basiert auf lokalen und geometrischen Merkmalen
- Modellierung durch beschriftete Graphen
- Anwendungsbeispiele
 - Handgesten-Erkennung
 - Gesichtserkennung
 - Analyse von komplexen Szenen



- Forschungsprojekt „Hand Posture Classification“ am Institut für Neuroinformatik, Ruhr-Universität Bochum
- Integration in ein Gesture Interface für einen autonomen Roboter

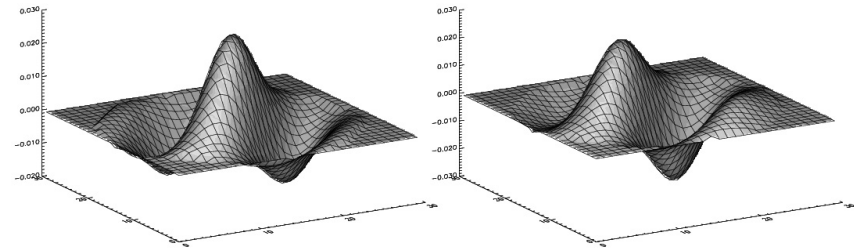


- Personenunabhängig
→ Unterschiede in der Handanatomie
- Komplexe Hintergründe
- Geringe Anforderungen an das Bild
→ Helligkeit, Größe und Position der Hand, Ringe, ...



- Modellierung der Hand: beschrifteter Graph
- Positionierung der Knoten: besondere Punkte der Hand (subjektiv)
- Objekt-adaptierter Graph
- Beschriftung der Knoten: lokale Merkmale
- Beschriftung der Kanten: geometrische Distanz

- Gabor-Wavelet: ebene Welle mit Gauß-Fensterfunktion
- Praktische Motivation: gute Ergebnisse in anderen Anwendungen
- Biologische Motivation: Modell für menschliches Sehen
- Wavelets für unterschiedliche Frequenzen und Orientierungen



Vorteile:

- Invariant zu Helligkeitsänderungen
- Robust gegen Translation und Verzerrung des Bildes

Nachteil:

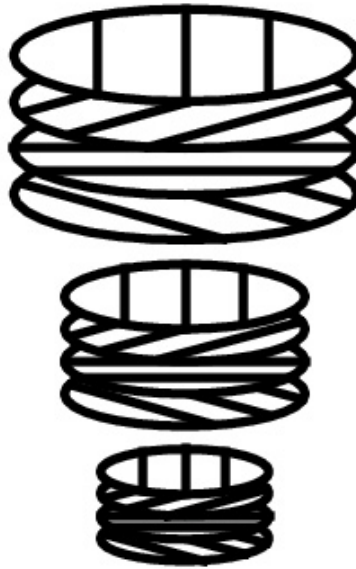
- Abhängig vom Hintergrund des Objekts

- Wavelets für unterschiedliche Frequenzen und Orientierungen
- Zusammenfassen der Filterantworten für ein Pixel

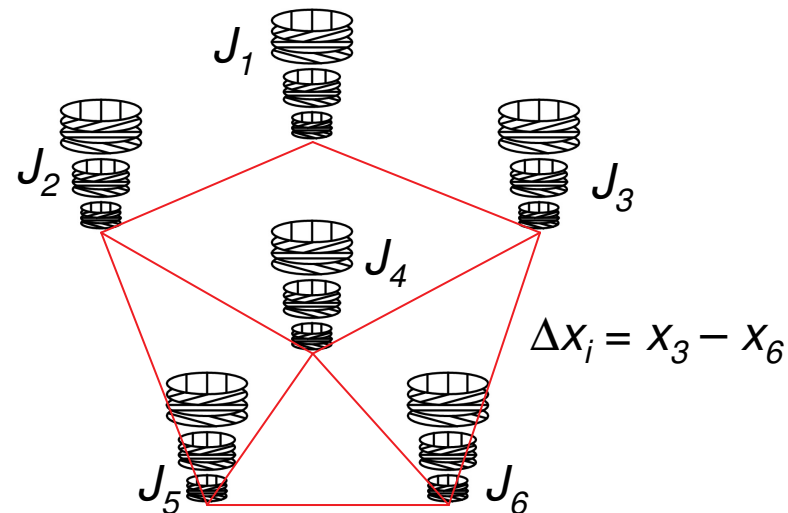
$$\text{Jet } \mathcal{J} = \{J_j\}$$

komplexwertige Filterantworten in Polar-Darstellung $J_j = a_j \exp(i\phi_j)$

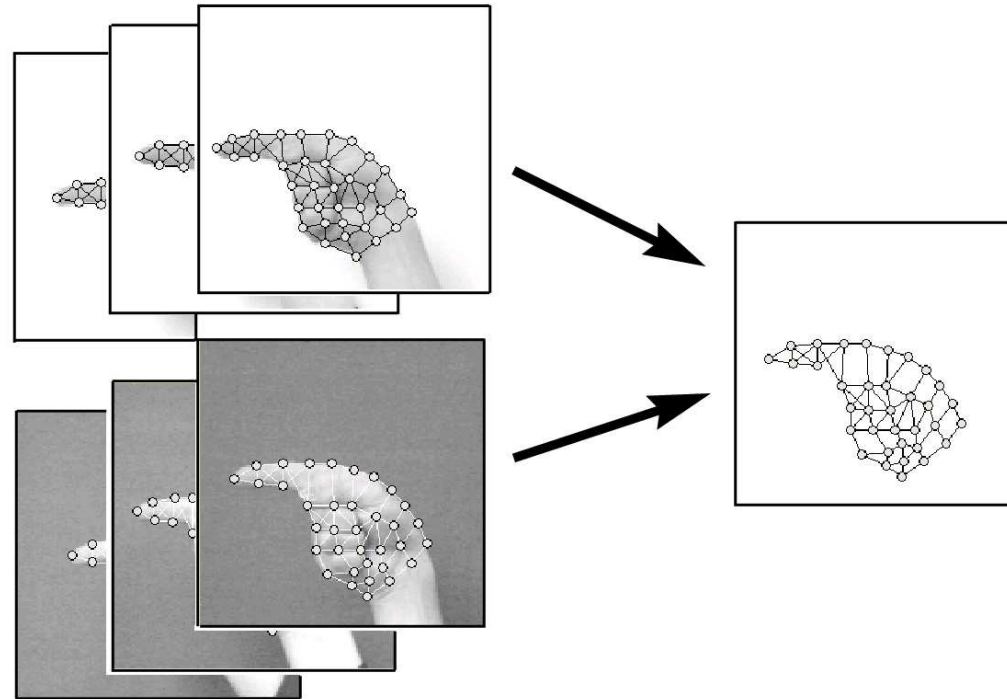
- Amplitude: langsame Änderung mit der Position
- Phase: teilweise abrupte Änderung



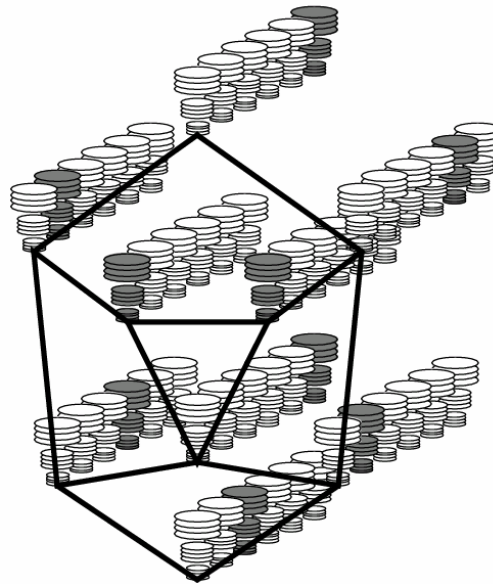
$$\begin{aligned} \mathcal{J} &= \{ J_1, J_2, \dots, J_{12} \} \\ &= \{ a_1 \exp(i \phi_1), \dots, a_{12} \exp(i \phi_{12}) \} \end{aligned}$$



- Image Graph G : N Knoten, E Kanten
- Beschriftung der Knoten:
Jets J_n an Positionen x_n , $n = 1, \dots, N$
- Beschriftung der Kanten:
Distanzvektoren $\Delta x_e = x_n - x_{n'}$
zwischen Knoten n und n'
- Graph ist nicht vollständig: nur Kanten für „sinnvolle“
Distanzmessungen, Redundanz vermeiden



- Repräsentant einer Klasse: Bunch Graph
- Zusammenfassung mehrerer Image Graphs
→ Image Graphs gleicher Struktur
- Überdeckung vieler Variationen in den Trainingsbildern
Annahme: begrenzte Variationen



Konstruktion eines Bunch Graphs B aus M Image Graphs G^{Bm} :

- Zusammenfassen der Jets eines Knotens
→ Menge von Jets → „Bunch“ (Bündel)
- Beschriftung der Knoten mit Bunches
- Beschriftung der Kanten mit durchschnittlicher Distanz:

$$\Delta \mathbf{x}_e^B = \frac{1}{M} \sum_m \Delta \mathbf{x}_e^{Bm}$$

1. Vergleich von Jets

- Ohne Berücksichtigung der Phase:

$$S_a(\mathcal{J}, \mathcal{J}') = \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j^2 \sum_j a'_j{}^2}}$$

- normiertes Skalarprodukt, normalisierte Kreuz-Korrelation
- Robust gegen unterschiedliche Beleuchtungsverhältnisse

- Phasensensitiv

$$S_\phi(\mathcal{J}, \mathcal{J}') = \frac{1}{2} \left(1 + \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_j)}{\sqrt{\sum_j a_j^2 \sum_j a'_j{}^2}} \right)$$

- Gut geeignet für die Positionierung von Knoten
- Unterscheidung von Mustern mit gleichem absoluten Anteil

2. Vergleich von Graphen

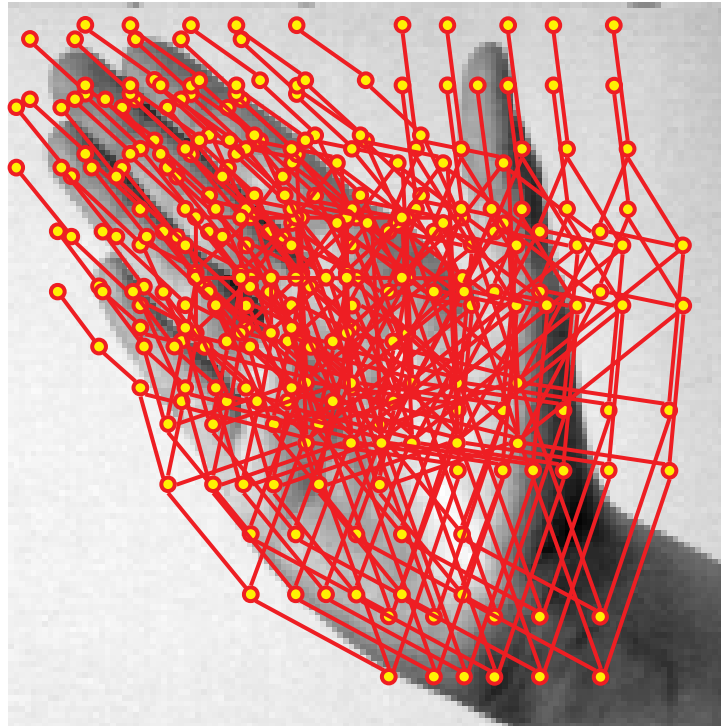
- Berücksichtigt Ähnlichkeit von Jets und geometrische Ähnlichkeit
- Für jeden Knoten: bester Jet im Bunch Graph
→ „Local Expert“
- Für Bunch Graph B mit M Jets pro Bunch und Image Graph G^I mit $N = 1, \dots, n$ Knoten, $E = 1, \dots, e$ Kanten:

$$S_B = \frac{1}{N} \sum_n \max_m \left\{ S_\phi \left(j_n^I, j_n^{Bm} \right) \right\} - \frac{\lambda}{E} \sum_e \frac{(\Delta \mathbf{x}_e^I - \Delta \mathbf{x}_e^B)^2}{(\Delta \mathbf{x}_e^B)^2}$$

- λ Gewichtung der geometrischen Ähnlichkeit
- Benutzung von S_a oder S_ϕ je nach Anwendung

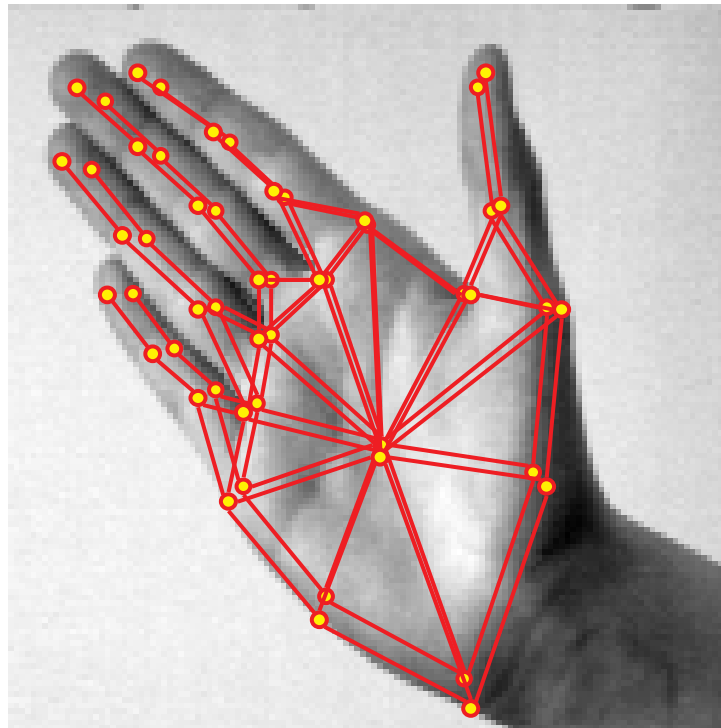
- Ziel: Berechnung eines Image Graphs für ein Bild und einen Bunch Graph
- Genaue Positionierung der Knoten
- Möglichst geringe Deformation des Graphen
- Matching in mehreren Schritten
- Pro Schritt bestes Ergebnis auswählen
- Ergebnis des vorherigen Schrittes als Basis für den nächsten Schritt
- Initialer Graph: Struktur des Bunch Graphs

Grobe Positionierung



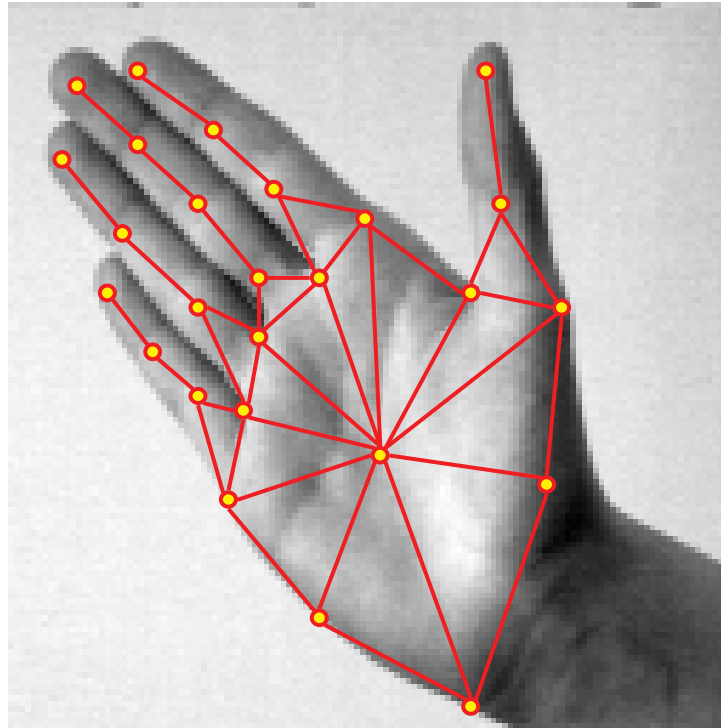
- Bewegung des Graphen auf einem groben Raster
- Keine Verformung: $\lambda = \infty$
- Verwendung von S_a

Anpassung der Größe und genauere Positionierung



- Verkleinern und Vergrößern des gesamten Graphen
- Rotation um den Schwerpunkt
- Bewegung um einige Pixel in x- und y-Richtung
- Keine Verformung: $\lambda \rightarrow \infty$

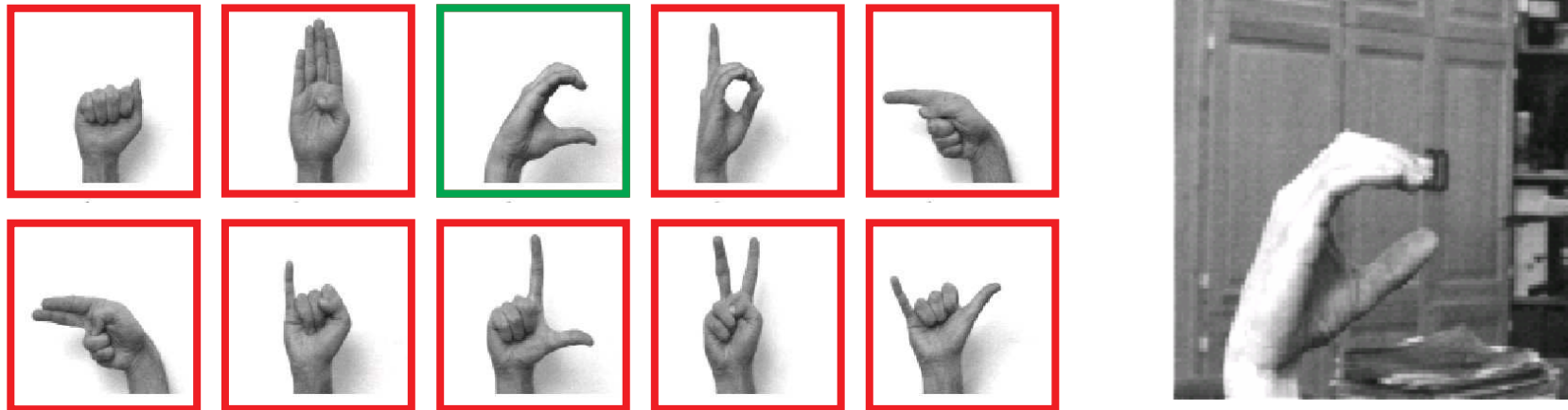
Lokale Verformung



- Knoten werden einzeln um wenige Pixel verschoben
- Verwendung eines geeigneten λ
- → Ähnlichkeitswert für Image Graph

- Training: Erstellen von Bunch Graphs für jede Klasse
- Erstellen von Image Graphs für jedes Trainingsbild
- Zusammenfassen zu Bunch Graphs
- Methode 1: manuelle Positionierung der Knoten
→ Aufwändig → Wenig Trainingsbilder
- Methode 2: halbautomatische Positionierung
 - Erstellen der ersten Image Graphs manuell
 - Erzeugen weiterer Image Graphs mit EBGM

- Verbesserung: EM-Iteration



- Matching Algorithmus für alle Bunch Graphs
- Bunch Graph mit höchstem Ähnlichkeitswert ist der Repräsentant der Klasse
- Zu geringer Ähnlichkeitswert → „nicht erkannt“



- Erkennungsraten Handgesten-Erkennung:
(Mit Compound Bunch Graphs)
Einheitlicher Hintergrund: 92,9%
Komplexer Hintergrund: 85,8%
- Fehler bei starken geometrischen Verzerrungen durch
 - unterschiedliche Anatomie der Hand
 - unterschiedliche Ausführung der Geste
 - Tiefenrotation
- Verbesserung: Active-Shape-Models
- Ergebnisse bei der Gesichtserkennung:
 - 98% bei „guten“ Voraussetzungen
 - weniger als 20% bei „schlechten“ Voraussetzungen

Ergebnisse bei der Gesichtserkennung auf der FERET-Datenbank:

- Bilder in unterschiedlichen Posen: Frontal, Halbprofil, Profil
- Drehwinkel für Halbprofil nicht exakt

Methode	Model Galery	Testbilder	Erkennungsrate
Normalisierte Kreuz-Korrelation	194 fa	194 fb	62 %
Radial basis function network	100 fa	100 fb	83 %
Trained matching pursuit filters	311 fa	311 fb	95 %
Principal Component Analysis	150 fa	150 fb	99 %
	150 hr	150 hl	38 %
	150 pr	150 pl	32 %
Elastic Bunch Graph Matching	250 fa	250 fb	97 %
	250 hr	181 hl	57 %
	250 pr	250 pl	84 %

- Tiefenrotation kann nicht gut modelliert werden
→ Lineare Jet Transformation
- Geometrische Verformungen folgen oft keiner erkennbaren statistischen Verteilung
- Graph hat wenig Freiheitsgrade:
 - gute Positionierung
 - Verformung und Drehung werden nur grob approximiert
- Manuelle Positionierung
→ stark eingeschränkte Trainingsdatenbank

- Gute Ergebnisse, solange die Objekte nicht stark verzerrt oder gedreht sind
- Gute Ergebnisse auch in schwierigen Situationen (komplexe Hintergründe)
- Für neue Objekte kein intensives Training
- Geringe Menge von typischen Bildern erzeugt robustes Modell
- Bei der Gesichtserkennung: 1 Bild pro Person
- Einsatzmöglichkeiten in der Medizin:
 - Objekte, die sich durch besondere Regionen auszeichnen
 - Klassenweise ähnliche geometrische Form
 - z.B. Lokalisierung und Segmentierung von Knochen auf Röntgenbildern