

Invariante Objekterkennung mittels Ähnlichkeitsranglisten

Schriftliche Prüfungsarbeit
für die Bachelor-Prüfung des Studiengangs Angewandte Informatik
an der Ruhr-Universität Bochum

vorgelegt von
Raul Alexander Grieben

17.01.2013

Dr. habil. Rolf P. Würtz
M.Sc. Andreas Nilkens

Inhaltsverzeichnis

| | | |
|-------|---|----|
| | Tabellenverzeichnis | IV |
| | Abbildungsverzeichnis | V |
| 1 | Einleitung | 1 |
| 1.1 | Invariante Objekterkennung | 1 |
| 1.2 | Verfahren zur Objekterkennung | 2 |
| 1.3 | Ziel dieser Arbeit | 2 |
| 2 | Material und Methoden | 3 |
| 2.1 | Bild-Datenbanken | 3 |
| 2.1.1 | Columbia University Image Library (COIL-100) | 3 |
| 2.1.2 | Amsterdam Library of Object Images (ALOI) | 3 |
| 2.2 | Ähnlichkeitsbestimmung | 5 |
| 2.2.1 | <i>Graph matching</i> | 5 |
| 2.2.2 | <i>Color Histogram Intersection (CHI)</i> | 7 |
| 2.2.3 | Korrelationskoeffizient (PCC) | 7 |
| 2.3 | Bestimmung der Ähnlichkeit mittels Ranglisten | 8 |
| 2.3.1 | Ranglisten Erstellung | 8 |
| 2.3.2 | Ranglisten vergleichen | 10 |
| 2.3.2 | Erkennungsverfahren | 10 |
| 2.3.3 | Kombination von Ähnlichkeitsfunktionen | 11 |
| 2.4 | Schätzung des Betrachtungswinkel | 11 |
| 2.5 | Verwendete Hard- und Software | 12 |
| 2.6 | Versuchsaufbau | 12 |
| 2.6.1 | Leistungsvergleich | 12 |
| 2.6.2 | Aufbau | 16 |
| 3 | Ergebnisse | 18 |
| 3.1 | Farbtemperatur | 18 |
| 3.2 | Beleuchtung | 19 |
| 3.3 | Betrachtungswinkel | 21 |
| 3.3.1 | COIL-100 | 21 |
| 3.3.2 | ALOI | 23 |
| 3.3.3 | Bestimmung des Winkels | 26 |
| 3.3.4 | Laufzeiten | 26 |
| 4 | Diskussion | 28 |
| 4.1 | Farbtemperatur | 28 |
| 4.2 | Beleuchtung | 28 |
| 4.3 | Betrachtungswinkel | 29 |
| 4.4 | Vergleich mit anderen Verfahren | 30 |

| | | |
|-----|----------------------------|----|
| 4.5 | Fazit..... | 30 |
| 4.6 | Ausblick..... | 30 |
| | Zusammenfassung..... | 32 |
| | Literaturverzeichnis | 33 |

Tabellenverzeichnis

| | |
|--|----|
| Tabelle 1: Erkennungsrate in Prozent der einzelnen Ähnlichkeitsfunktionen und ihrer Kombinationen..... | 14 |
| Tabelle 2: Durchschnittliche Erkennungsrate in Prozent für die ausgewählten Kombinationen nach 100 zufällig gemischten Testläufen..... | 15 |
| Tabelle 3: Erkennungsrate in Prozent für verschiedene Graphenstrukturen..... | 15 |
| Tabelle 4: Erkennungsrate in Prozent für verschiedene Gittergraphen..... | 16 |
| Tabelle 5: Die zwei Versuchsaufbauten..... | 16 |
| Tabelle 6: Erkennungsrate auf der ALOI Datenbank bei variierender Farbtemperatur in Prozent..... | 18 |
| Tabelle 7: Erkennungsrate auf der ALOI Datenbank bei variierender Beleuchtung in Prozent..... | 19 |
| Tabelle 8: Erkennungsrate auf der COIL-100 Datenbank bei variierendem Betrachtungswinkel in Prozent..... | 21 |
| Tabelle 9: Erkennungsrate auf der ALOI Datenbank bei variierendem Betrachtungswinkel in Prozent..... | 24 |
| Tabelle 10: Erkennungsrate nach Schätzung des Betrachtungswinkels in Prozent..... | 26 |
| Tabelle 11: Laufzeiten in Sekunden für die Generierung der 900 Graphen..... | 26 |
| Tabelle 12: Laufzeiten in Sekunden für die Winkelschätzung und die Erkennungsroutine..... | 27 |

Abbildungsverzeichnis

| | |
|--|----|
| Abbildung 1: Die erste Bildreihe aus der COIL-100 Datenbank. | 3 |
| Abbildung 2: Alle 100 Objekte der COIL-100 Datenbank (Frontal-Ansicht). | 3 |
| Abbildung 3: Die erste Bildreihe aus der ALOI Datenbank (Betrachtungswinkel)..... | 4 |
| Abbildung 4: Die erste Bildreihe aus der ALOI Datenbank (Beleuchtung)..... | 4 |
| Abbildung 5: Die erste Bildreihe aus der ALOI Datenbank (Farbtemperatur). | 4 |
| Abbildung 6: Die ersten 100 Objekte aus der ALOI Datenbank (Frontal-Ansicht)..... | 4 |
| Abbildung 7: Verschiedene Graphenstrukturen. (<i>Gittergraph-16, NB-Gittergraph-16, Canny-FAST</i>) | 5 |
| Abbildung 8: <i>Feature detection</i> anhand einer Kantenextraktion mit nachfolgender Eckenextraktion. | 6 |
| Abbildung 9: Bestimmung der Ähnlichkeit mittels Ranglisten. (Die angegebenen Ranglisten sind fiktive Beispiele.) Grafik angelehnt an [MTBW12]. | 9 |
| Abbildung 10: Erkennungsrate ($S_{Abs} \diamond S_{Canb}$) in Abhängigkeit von Kamera- und Lichtposition. | 20 |
| Abbildung 11: Erkennungsrate ($S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$) in Abhängigkeit von Kamera- und Lichtposition. | 20 |
| Abbildung 12: Erkennungsrate auf COIL-100 ($S_{Abs} \diamond S_{Canb}$) in Abhängigkeit vom Betrachtungswinkel..... | 22 |
| Abbildung 13: Erkennungsrate auf COIL-100 ($S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$) in Abhängigkeit vom Betrachtungswinkel..... | 22 |
| Abbildung 14: Erkennungsrate auf COIL-100 in Abhängigkeit vom Betrachtungswinkel (alle 72 Blickwinkel). Rot ist die Kombination $S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$ und blau die Kombination $S_{Abs} \diamond S_{Canb}$ | 23 |
| Abbildung 15: Erkennungsrate auf ALOI ($S_{Abs} \diamond S_{Canb}$) in Abhängigkeit vom Betrachtungswinkel..... | 24 |
| Abbildung 16: Erkennungsrate auf ALOI ($S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$) in Abhängigkeit vom Betrachtungswinkel..... | 25 |
| Abbildung 17: Erkennungsrate auf ALOI in Abhängigkeit vom Betrachtungswinkel (alle 72 Blickwinkel). Rot ist die Kombination $S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$ und blau die Kombination $S_{Abs} \diamond S_{Canb}$ | 25 |

1 Einleitung

1.1 Invariante Objekterkennung

Eine der außergewöhnlichsten Leistungen unseres Gehirnes ist für die Menschen dermaßen alltäglich, dass die meisten sie gar nicht als solche wahrnehmen, sondern als etwas das jeder gesunde Mensch beherrscht.

Die Rede ist von der visuellen Objekterkennung, die es uns ermöglicht in weniger als einer halben Sekunde ein Objekt mit einer hohen Präzision zu erkennen (vgl. [TFM96]). Dabei scheint es einem auf dem ersten Blick eine wesentlich erstaunlichere Leistung zu sein, die 10.000ste Nachkommastelle von π im Kopf berechnen zu können. Aus der Sicht der Informatik ist die Berechnung der Nachkommastellen von π ein triviales Problem, während die visuelle Objekterkennungsfähigkeit des Gehirns nachzuahmen, eine hoch anspruchsvolle und bis dato unerreichte Herausforderung darstellt. Auch wenn über die Jahre hinweg große Fortschritte in dem Bereich erzielt worden sind, leidet der aktuelle Stand der Technik immer noch unter einigen starken Defiziten (vgl. [PBCD11]).

Die größte Stärke des Gehirns und gleichzeitig die größte Schwäche von computerbasierten Verfahren, ist die invariante Objekterkennung (vgl. [DZR12] und [PBCD11]). Diese ermöglicht dem Gehirn ein Objekt in hunderten von verschiedenen Situationen wiederzuerkennen und dies vollkommen unabhängig von Skalierung, Blickwinkel, Beleuchtung oder farblicher Veränderung (vgl. [DZR12]). Wie das Gehirn diese Aufgabe bewältigt, ist im Großen und Ganzen noch immer nicht geklärt, obwohl bereits verschiedene Theorien zu diesem Thema entwickelt wurden (vgl. [DZR12]).

Eine andere Eigenschaft, welche wir uns auch in dieser Arbeit zunutze machen wollen, ist die Fähigkeit Objekte einzig durch Ähnlichkeitsbeziehungen einer Objektklasse zuzuordnen bzw. wiederzuerkennen. Und auch wenn die Zuordnung eines neuen Objekts unter ungewohnten Umständen nicht immer perfekt verläuft, so verbessert sich die Effizienz des Erkennungsprozess trotzdem ungemein (vgl. [BN06]).

1.2 Verfahren zur Objekterkennung

In den letzten Jahren wurden einige Verfahren zur Objekterkennung vorgestellt. Die folgenden stellen eine kleine Auswahl verschiedener Ansätze dar und sollen gleichsam in der gebotenen Kürze den aktuellen Stand der Technik umreißen.

[WFKM97] entwickelte das *Elastic Graph Matching*, das sich sehr gut für die Gesichtserkennung eignet.

[OM02] verwendet *Local Affine Frames* und erreicht auf der COIL-100 Datenbank hohe Erkennungsraten.

[WW09] verbindet *feature* und *correspondence* basierte Methoden.

[LL04] erreicht eine hohe Erkennungsleistung durch hochdimensionale *Composed Receptive Field* Histogramme.

[LW12] stellt ein Verfahren auf Basis von hierarchischen Netzwerken vor.

1.3 Ziel dieser Arbeit

Ziel dieser Arbeit ist die invariante Objekterkennung anhand von Ähnlichkeitsranglisten. Das zugrunde liegende System wurde durch Müller et al. entwickelt und erfolgreich bei der invarianten Gesichtserkennung angewandt (vgl. [MTBW12] und [MTBW12E]) und soll für die Verwendung mit Objekten, welche im Gegensatz zu Gesichtern keine korrespondierenden lokalen Merkmale aufweisen angepasst werden.

Die Hauptfrage, auf welche diese Arbeit eine Antwort geben soll ist die, ob ein System allein durch Ähnlichkeitsbeziehungen zu bekannten Objekten in derselben Situation, ein unbekanntes Objekt, dem menschlichen Gehirn gleich (vgl. [BN06]), erkennen kann.

2 Material und Methoden

2.1 Bild-Datenbanken

Für die Evaluierung des hier vorgeschlagenen Verfahrens zur invarianten Objekterkennung wurden folgende zwei Bilddatenbanken verwendet.

2.1.1 Columbia University Image Library (COIL-100)

Die COIL-100 Datenbank [NNM96] enthält 100 Objekte, die in einem Abstand von jeweils 5° zueinander, aus 72 Betrachtungswinkeln aufgenommen wurden. Die einzelnen Objektaufnahmen wurden auf die Größe des Bildes skaliert. Dies kann je nach Objekt, zu einem größeren Skalenunterschied innerhalb einer Bildreihe führen.

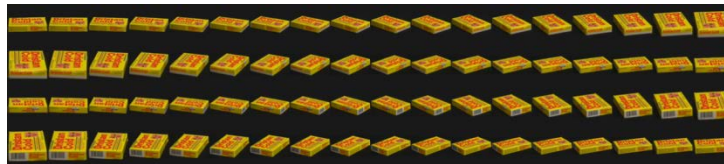


Abbildung 1: Die erste Bildreihe aus der COIL-100 Datenbank.

Die Datenbank besteht insgesamt aus 7.200 Farbbildern mit einer Auflösung von je 128x128 Pixeln.



Abbildung 2: Alle 100 Objekte der COIL-100 Datenbank (Frontal-Ansicht).

2.1.2 Amsterdam Library of Object Images (ALOI)

Die ALOI Datenbank [GBS05] besteht aus vier Bildgruppen, von denen drei in dieser Arbeit verwendet wurden. Alle Gruppen umfassen Aufnahmen der gleichen 1000 Objekte unter verschiedenen Situationen.

Beim ersten Set wurden die 1000 Objekte ebenso wie bei der COIL-100 aus 72 Betrachtungswinkeln im Abstand von 5° zueinander aufgenommen. Im Gegensatz zu COIL-100 wurden die Aufnahmen der ALOI allerdings nicht skaliert.



Abbildung 3: Die erste Bildreihe aus der ALOI Datenbank (Betrachtungswinkel).

Das zweite Set besteht aus jeweils 24 Aufnahmen der Objekte in Frontalansicht. Für diese Gruppe wurde nicht der Aufnahme- sondern der Beleuchtungswinkel variiert.



Abbildung 4: Die erste Bildreihe aus der ALOI Datenbank (Beleuchtung).

Auch in der dritten für diese Arbeit verwendeten Gruppe der ALOI Datenbank wurden alle Objekte in Frontalansicht dargestellt. Dieses Mal wurde aber die Farbtemperatur von 2175K bis 3075K in 12 Schritten verändert.

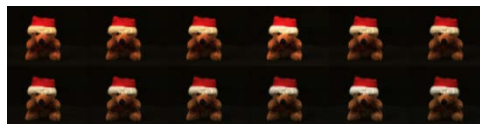


Abbildung 5: Die erste Bildreihe aus der ALOI Datenbank (Farbtemperatur).

Der verwendete Teil der ALOI Datenbank besteht aus insgesamt 108.000 Farbbildern mit einer Auflösung von je 192x144 Pixeln.

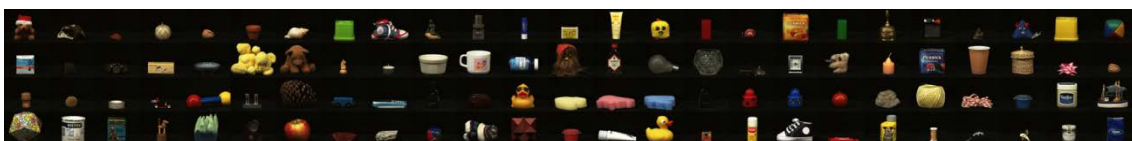


Abbildung 6: Die ersten 100 Objekte aus der ALOI Datenbank (Frontal-Ansicht).

2.2 Ähnlichkeitsbestimmung

Um die Ähnlichkeit zweier Objekte zu ermitteln, wurden im Rahmen dieser Arbeit die folgenden drei Ansätze verfolgt:

2.2.1 *Graph matching*

Die Erkennung mittels *graph matching* [LVBL93] vergleicht zwei Graphen mit N Knoten anhand einer Ähnlichkeitsfunktion.

Jeder Knoten enthält an seiner Position lokale Merkmale eines Bildes, in Form eines *Gabor jet* (vgl. [MTBW12E]).

Da die verwendeten Graphen in diesem Verfahren nicht elastisch sind, wurde auf Kanten verzichtet.

Graphenstruktur. Um die Auswirkungen der Graphenstruktur auf die Erkennungsleistung auszuloten, wurden drei verschiedene Graphenstrukturen verwendet.

Weil nicht angenommen werden konnte, dass Probe- und Galeriebilder korrespondierende lokale Merkmale besitzen, waren Gittergraphen eine naheliegende Graphenstruktur, da diese ein regelmäßiges Gitter mit Abstand n zwischen den einzelnen Knoten über das Bild spannen. Ihr Vorteil gegenüber anderen Graphenstrukturen, liegt darin begründet, dass Gittergraphen in ihrem Aufbau identisch sind. Sie wurden in der Arbeit mit *Gittergraph- n* betitelt.

In einer zweiten Variante wurden nur jene Knoten des *Gittergraph- n* verwendet, welche auf dem Objekt selbst liegen. Sie wurden mit *NB-Gittergraph- n* deklariert.

Der Vollständigkeit halber wurde in einer letzten Verfahrensvariante ein Graph anhand einer (einfachen) *feature detection* erstellt. Bei einer *feature detection* werden sogenannte *points of interest* im Bild gesucht und markiert.

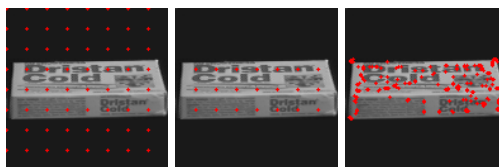


Abbildung 7: Verschiedene Graphenstrukturen. (*Gittergraph-16*, *NB-Gittergraph-16*, *Canny-FAST*)

Die *feature detection* beim letzten Graph bestand aus den Ecken, welche der FAST Algorithmus ([Ros12] und [RD05]) aus dem Ergebnis der Kantenextraktion durch den Canny Algorithmus [Can86] extrahiert hat und als *Canny-Fast* bezeichnet wurde.

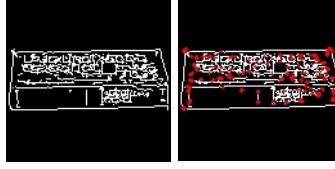


Abbildung 8: *Feature detection* anhand einer Kantenextraktion mit nachfolgender Eckenextraktion.

Die eben vorgestellten Varianten haben allerdings den Nachteil, dass sie für jedes Bild eine andere Größe und Form besitzen.

Gabor jets. *Gabor-Wavelets* sind in der Objekterkennung weit verbreitet (vgl. [LVBL93] und [MTBW12E]) und nehmen folgende Form an (vgl. [MTBW12E]):

$$\psi_{\vec{k}}(\vec{x}) = \frac{\vec{k}^2}{\sigma^2} e^{-\frac{\vec{k}^2 \vec{x}^2}{2\sigma^2}} \left(e^{i\vec{k}\vec{x}} - e^{-\frac{\sigma^2}{2}} \right) \quad (1)$$

Es wurden fünf Frequenzskalen und acht Richtungen verwendet, was 40 *Gabor-Wavelets* $\psi_{\vec{k}_j}$ ($j = 1, \dots, 40$) generiert. Die Faltung des Bildes an der Stelle \vec{x} mit einem *Gabor-Wavelet* $\psi_{\vec{k}_j}$ ergibt eine komplexwertige Antwort $a_j \cdot e^{i\phi_j}$ bestehend aus einer Amplitude a_j und einer Phase ϕ_j (vgl. [MTBW12E]).

Die Antworten aller 40 Faltungen an der Stelle \vec{x} werden *Gabor jet* \mathcal{J} genannt und repräsentieren die Texturinformation an der gegebenen Position (vgl. [MTBW12E]).

Ähnlichkeitsfunktionen. Um die *Gabor jets* mit Amplitude a_j zu vergleichen (vgl. [MTBW12E]) wurden folgende Ähnlichkeitsfunktionen nach [WFKM97], [GW09] und [GBTS07] verwendet:

$$S_{Abs}(\vec{J}_1, \vec{J}_2) = \frac{\sum_{j=1}^K a_{1,j} \cdot a_{2,j}}{\sqrt{(\sum_{j=1}^K a_{1,j}^2)(\sum_{j=1}^K a_{2,j}^2)}} \quad (2)$$

$$S_{Cantb}(\vec{J}_1, \vec{J}_2) = K - \sum_{j=1}^K \frac{|a_{1,j} - a_{2,j}|}{\max(|a_{1,j}| + |a_{2,j}|, 10^{-6})} \quad (3)$$

$$S_{Manh}(\vec{J}_1, \vec{J}_2) = -\frac{\sum_{j=1}^K |a_{1,j} - a_{2,j}|}{\sum_{j=1}^K |a_{1,j}| \sum_{j=1}^K |a_{2,j}|} \quad (4)$$

2.2.2 Color Histogram Intersection (CHI)

Die normalisierte *Color Histogram Intersection*, vorgeschlagen unter [SB91] ist eine Distanzfunktion zwischen zwei Farbhistogramme. Sie wird durch

$$H(I, M) = \frac{\sum_{j=1}^n \min(I_j, M_j)}{\sum_{j=1}^n M_j} \quad (5)$$

definiert (vgl. [SB91]).

Üblicherweise und unter [SB91] vorgeschlagen, wird das Farbhistogramm quantisiert, so dass die Anzahl an Klassen meist deutlich kleiner ist, als die Zahl an möglichen Farben im RGB-Farbraum (2^{24}).

In dieser Arbeit wurde die *Color Histogram Intersection* als S_{CHI} abgekürzt. Des Weiteren wurde auf eine Quantisierung verzichtet und stattdessen für jede der drei Farbkomponenten im RGB-Farbraum ein eigenes Histogramm generiert und die Distanz zwischen ihnen nach Formel (5) berechnet.

Als Ähnlichkeitswert wurde die normalisierte Summe aller drei Histogrammvergleiche verwendet.

$$CHI(I_{RGB}, M_{RGB}) = \frac{H(I_R, M_R) + H(I_G, M_G) + H(I_B, M_B)}{3} \quad (6)$$

2.2.3 Korrelationskoeffizient (PCC)

Der Korrelationskoeffizient (hier nach Pearson) ist ein Maß für die Stärke und Richtung eines linearen Zusammenhangs zwischen zwei quantitativen Merkmalen. Das Maß ist dimensionslos und kann Werte zwischen 1 und -1 annehmen. Der Korrelationskoeffizient ist durch

$$r(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7)$$

definiert, wobei \bar{x} und \bar{y} das jeweilige arithmetische Mittel von x beziehungsweise y darstellen (vgl. [Kat12]).

Der Korrelationskoeffizient nach Pearson findet unter anderem Verwendung als Ähnlichkeitsmaß beim *template Matching* (vgl. [TLM06]).

In dieser Arbeit wurde der Korrelationskoeffizient als eine mögliche Ähnlichkeitsfunktion verwendet und als S_{PCC} abgekürzt. Als x bzw. y wurden die Grauwertmatrizen der zwei miteinander zu vergleichenden Bilder verwendet und der resultierende Korrelationskoeffizient als Ähnlichkeitswert gespeichert.

2.3 Bestimmung der Ähnlichkeit mittels Ranglisten

Das hier verwendete Verfahren zur Bestimmung der Ähnlichkeit mittels Ranglisten wurde von Müller *et al.* entwickelt und bereits erfolgreich bei der invarianten Gesichtserkennung angewandt (vgl. [MTBW12] und [MTBW12E]).

Bei der Ranglistenenerstellung wurde in der vorliegenden Arbeit auf zwei unterschiedliche Ansätze zurückgegriffen. Der erste Ansatz ist eine leichte Abwandlung des Verfahrens von Müller *et al.* (vgl. [MTBW12] und [MTBW12E]). Der zweite Ansatz ist das Originalverfahren (vgl. [MTBW12] und [MTBW12E]).

2.3.1 Ranglistenenerstellung

Um eine Rangliste erstellen zu können, gilt es zunächst einmal eine in jeder Situation als bekannt vorgegebene Modelldatenbank anzufertigen. Die Ansammlung an Graphen M_m^v mit der Situation v und der Position m wird verwendet um die Ähnlichkeitsrangliste des Probe- bzw. Galeriebildes zu den Modellen aus der Datenbank zu generieren (vgl. [MTBW12E]).

Eine Rangliste pro Graphen. Im ersten Ansatz wurde der Graph P vom Probebild in der Situation v mit allen Modellgraphen M_m^v verglichen und die durchschnittliche lokale Ähnlichkeit aller Knoten N berechnet. Anhand der durchschnittlichen Ähnlichkeit zu jedem Modellbild wurden diese absteigend von ähnlich zu nicht ähnlich sortiert und im Anschluss daran als Rangliste p^v gespeichert. Das Galeriebild wurde mit allen Modellgraphen M_m^0 in der Situation 0 verglichen und die Rangliste im gleichen Verfahren wie beschrieben erstellt (vgl. [MTBW12E]).

Die allgemeine Darstellung der Rangliste lautet:

$$p^v(P, m) \quad (8)$$

Wobei m die Position des Modellgraphen, v die Situation und P der zu vergleichende Graph ist (vgl. [MTBW12E]).

Für jeden Galeriegraph G_g wurde eine Rangliste γ_g erstellt (vgl. [MTBW12E]):

$$\gamma_g(m) = p^0(G_g, m), \quad m = 1 \dots M \quad (9)$$

Für das Probebild P^v in der bekannten Situation v definieren wir die Rangliste π (vgl. [MTBW12E]):

$$\pi^v(m) = p^v(P^v, m), \quad m = 1 \dots M \quad (10)$$

Eine Rangliste pro Knoten im Graphen. Im zweiten Ansatz, welcher dem Originalansatz entspricht, wurde eine Rangliste pro Knoten im Graph erstellt und als p_n^v gespeichert (vgl. [MTBW12E]).

Die allgemeine Darstellung der Rangliste lautet in diesem Fall (vgl. [MTBW12E]):

$$p_n^v(P, m) \quad (11)$$

Für jeden Galeriegraph $G_{g,n}$ wurde eine Rangliste $\gamma_{g,n}$ erstellt (vgl. [MTBW12E]):

$$\gamma_{g,n}(m) = p_n^0(G_{g,n}, m), \quad m = 1 \dots M \quad (12)$$

Und für das Probebild P^v in der bekannten Situation v definieren wir die Rangliste π (vgl. [MTBW12E]):

$$\pi_n^v(m) = p_n^v(P^v, m), \quad m = 1 \dots M \quad (13)$$

Die Abbildung 9 stellt das Verfahren der Ranglistenerstellung vereinfacht dar.

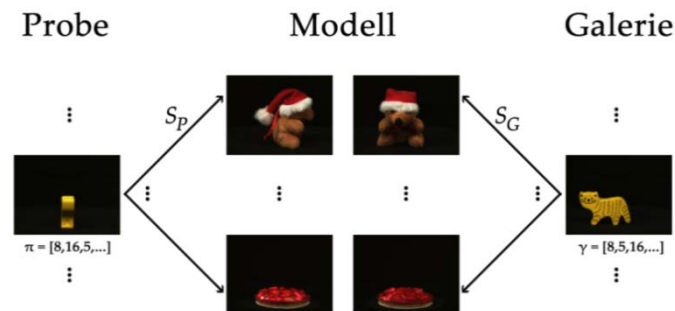


Abbildung 9: Bestimmung der Ähnlichkeit mittels Ranglisten. (Die angegebenen Ranglisten sind fiktive Beispiele.) Grafik angelehnt an [MTBW12].

2.3.2 Ranglisten vergleichen

Um die erstellten Ranglisten für Probe- und Galeriegraphen zu vergleichen, bedarf es einer Funktion $S_{rank}(\pi, \gamma)$, welche in der Lage ist, die Ähnlichkeit zweier Ranglisten zu bestimmen (vgl. [MTBW12E]).

Die Anforderungen an diese Funktion sind nach [MTBW12E] und [MTBW12]:

1. Der Ähnlichkeitswert sollte zwischen 0 und 1 liegen und maximal für zwei gleiche Ranglisten sein.
2. Wenn identische Modellindizes oft denselben Rang besitzen sollte auch die Ähnlichkeit hoch sein.
3. Je kleiner der Ranglisteneintrag (die Rangliste ist absteigend sortiert), desto stärker sollte dieser gewichtet werden.

Diese Anforderungen werden erfüllt, wenn $S_{rank}(p_1, p_2)$ folgende Form einnimmt (vgl. [MTBW12E]):

$$S_{rank}(p_1, p_2) = \frac{1}{F} \sum_{m=1}^M f(p_1(m) + p_2(m)) \quad (14)$$

$$F = \sum_{m=1}^M f(2m) \quad (15)$$

f sollte eine monoton fallend Funktion sein und F ein Normierungsfaktor, der die maximale Ähnlichkeit auf eins begrenzt (vgl. [MTBW12E]).

Als Funktion f wurde in dieser Arbeit anders als bei [MTBW12E] die Funktion $f(x) = (x + 1)^d$ mit $d \in [-2, 0)$ aus [MHTS07] verwendet, da diese bei einem Test auf der COIL-100 Datenbank zu besseren Erkennungsraten geführt hat.

Die daraus resultierende Formel lautet:

$$S_{rank}(\pi, \gamma_g) = \frac{1}{F} \sum_{m=1}^M (\pi(m) + \gamma_g(m) + 1)^d \quad (16)$$

2.3.2 Erkennungsverfahren

Im Gegensatz zu [MTBW12E] sind in dem hier verwendeten Verfahren immer alle Knoten vorhanden und die Anzahl N konstant. Dies hat zur Folge, dass die Erkennungsfunktion leicht abgewandelt wurde (vgl. [MTBW12E]).

Für den ersten Ansatz gilt:

$$S_{rec}(g) = S_{rank}(\pi^v, \gamma_g) \quad (17)$$

Und für den zweiten Ansatz:

$$S_{rec}(g) = \frac{1}{N} \sum_{n=1}^N S_{rank}(\pi_n^v, \gamma_{g,n}) \quad (18)$$

Der Index g , welcher die höchste Erkennungsrate aufweist, wird als gesuchtes Objekt erkannt (vgl. [MTBW12E]).

2.3.3 Kombination von Ähnlichkeitsfunktionen

Die eben vorgestellte Erkennungsfunktion kann auch für Kombinationen aus Ähnlichkeitsfunktionen verwendet werden (vgl. [MTBW12E]).

Für alle nicht leeren Teilmengen $A \subseteq \{Abs, Canb, Manh, CHI, PCC\}$ können die Ranglisten $\pi_n^{v,a}$ und $\gamma_{g,n}^a$ erstellt werden (vgl. [MTBW12E]).

Die Erkennungsfunktionen wurden gegenüber der Version von [MTBW12E] wie folgt angepasst.

Für den ersten Ansatz:

$$S_{rec}^A(g) = \frac{1}{|A|} \sum_{a \in A} S_{rank}^a(\pi^{v,a}, \gamma_g^a) \quad (19)$$

Und für den zweiten Ansatz:

$$S_{rec}^A(g) = \frac{1}{|A|} \frac{1}{N} \sum_{a \in A} \sum_{n=1}^N S_{rank}^a(\pi_n^{v,a}, \gamma_{g,n}^a) \quad (20)$$

2.4 Schätzung des Betrachtungswinkel

In den Versuchen wurde der Betrachtungswinkel als gegeben angesehen obwohl dieser eine Information darstellt, welche in einem realistischeren Versuchsaufbau nicht als bekannt vorausgesetzt werden dürfte (vgl. [MTBW12E]). Aus diesem Grund wurde

noch eine weitere Versuchsreihe durchgeführt, in welcher der Betrachtungswinkel zuerst abgeschätzt und dann der Erkennungsroutine übergeben wurde.

Zur Schätzung des Betrachtungswinkels wurde das Probebild mit allen Modellbildern verglichen und der Winkel ausgewählt, welcher die Ähnlichkeitsfunktion S_{ABS} maximiert (vgl. [MTBW12E]).

2.5 Verwendete Hard- und Software

Als Grundlage für die Versuche wurde die C++ Bibliothek *pragma* (Institut für Neuroinformatik, Ruhr-Universität Bochum) verwendet.

Zur Eckendetektion wurde eine vorkompilierte Version von FAST ([Ros12], [RD05] und [RD06]) (Win32) eingesetzt.

Die Versuchsprogramme wurden auf einem Desktop-Computer (Windows 7 Professional (64 Bit), Intel Core i7-2600 3,4 GHz und 12 GB RAM) unter Microsoft Visual Studio 2010 Ultimate kompiliert und *single threaded* ausgeführt. Alle Zeitangaben in dieser Arbeit beziehen sich auf dieses System.

2.6 Versuchsaufbau

2.6.1 Leistungsvergleich

Um einen möglichst geeigneten Versuchsaufbau zu ermitteln, wurde die Leistungsfähigkeit der einzelnen Kombinationen aus verschiedenen Ähnlichkeitsfunktionen und Graphenstrukturen getestet.

Für den Test wurden die 100 Objekte der COIL-100 Datenbank, soweit nicht anders angegeben, in der Originalreihenfolge in eine Modellmenge von 50 und eine Probenmenge von 50 aufgeteilt. Darüber hinaus wurden neun der 72 verfügbaren Aufnahmen pro Objekt (im Abstand von jeweils 40°) ausgewählt.

Ähnlichkeitsfunktionen. Es wurde ein Gittergraph mit einem Abstand von jeweils 16 Pixeln zwischen den einzelnen Knoten verwendet.

Bei der Erstellung der Ranglisten wurde zwischen einer Rangliste pro Graph und einer Rangliste pro Knoten im Graphen unterschieden.

| | Rangliste | Erkennungsrate |
|---|-----------|----------------|
| S_{Abs} | Bild | 44.25 |
| | Knoten | 57.75 |
| S_{Canb} | Bild | 42.5 |
| | Knoten | 60.25 |
| S_{Manh} | Bild | 44.75 |
| | Knoten | 61.25 |
| S_{CHI} | Bild | 29.75 |
| | Knoten | 29.5 |
| S_{PCC} | Bild | 46 |
| | Knoten | 46 |
| $S_{Abs} \diamond S_{Canb}$ | Bild | 51.5 |
| | Knoten | 62.75 |
| $S_{Abs} \diamond S_{Manh}$ | Bild | 47.75 |
| | Knoten | 60.5 |
| $S_{Abs} \diamond S_{CHI}$ | Bild | 61.5 |
| | Knoten | 60.5 |
| $S_{Abs} \diamond S_{PCC}$ | Bild | 59.25 |
| | Knoten | 61.75 |
| $S_{Abs} \diamond S_{Manh}$ | Bild | 45.75 |
| | Knoten | 62.5 |
| $S_{Canb} \diamond S_{CHI}$ | Bild | 56.75 |
| | Knoten | 58.25 |
| $S_{Canb} \diamond S_{PCC}$ | Bild | 54.75 |
| | Knoten | 63 |
| $S_{Manh} \diamond S_{CHI}$ | Bild | 59 |
| | Knoten | 57.75 |
| $S_{Manh} \diamond S_{PCC}$ | Bild | 58.5 |
| | Knoten | 62.25 |
| $S_{CHI} \diamond S_{PCC}$ | Bild | 61.25 |
| | Knoten | 61.25 |
| $S_{Abs} \diamond S_{Canb} \diamond S_{Manh}$ | Bild | 49.25 |
| | Knoten | 62 |
| $S_{Abs} \diamond S_{Canb} \diamond S_{CHI}$ | Bild | 61.25 |
| | Knoten | 65.75 |
| $S_{Abs} \diamond S_{Canb} \diamond S_{PCC}$ | Bild | 58.75 |
| | Knoten | 65.25 |
| $S_{Abs} \diamond S_{Manh} \diamond S_{CHI}$ | Bild | 61.25 |
| | Knoten | 66.5 |
| $S_{Abs} \diamond S_{Manh} \diamond S_{PCC}$ | Bild | 59 |
| | Knoten | 65 |
| $S_{Abs} \diamond S_{CHI} \diamond S_{PCC}$ | Bild | 67.25 |
| | Knoten | 68 |
| $S_{Canb} \diamond S_{Manh} \diamond S_{CHI}$ | Bild | 59.75 |
| | Knoten | 65.5 |

| | | |
|--|--------|--------------|
| $S_{\text{Canb}} \diamond S_{\text{Manh}} \diamond S_{\text{PCC}}$ | Bild | 58 |
| | Knoten | 66.5 |
| $S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ | Bild | 63.25 |
| | Knoten | 66.75 |
| $S_{\text{Canb}} \diamond S_{\text{Manh}} \diamond S_{\text{CHI}}$ | Bild | 59.75 |
| | Knoten | 65.5 |
| $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{Manh}} \diamond S_{\text{CHI}}$ | Bild | 61.25 |
| | Knoten | 68.75 |
| $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{Manh}} \diamond S_{\text{PCC}}$ | Bild | 57.25 |
| | Knoten | 66.25 |
| $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ | Bild | 66.25 |
| | Knoten | 70 |
| $S_{\text{Abs}} \diamond S_{\text{Manh}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ | Bild | 66.75 |
| | Knoten | 70.5 |
| $S_{\text{Canb}} \diamond S_{\text{Manh}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ | Bild | 64.25 |
| | Knoten | 70.75 |
| $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{Manh}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ | Bild | 65.5 |
| | Knoten | 71.25 |

Tabelle 1: Erkennungsrate in Prozent der einzelnen Ähnlichkeitsfunktionen und ihrer Kombinationen.

Anhand ihrer Erkennungsrate wurden zur Ermittlung einer durchschnittlichen Rate aus 100 zufälligen Testläufen folgende Kombinationen als geeignet befunden:

Die Kombinationen $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$, $S_{\text{Abs}} \diamond S_{\text{Manh}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$, $S_{\text{Canb}} \diamond S_{\text{Manh}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ und $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{Manh}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$, wegen ihrer besseren Erkennungsleistung.

Die Kombination $S_{\text{Abs}} \diamond S_{\text{Canb}}$ wurde ausgewählt, weil sie unter den Kombinationen ohne Farbinformation die beste Erkennungsleistung erzielte und somit zum direkten Vergleich beider Ansätze geeignet erschien.

Bei der Ranglistenerstellung wurde wegen ihrer besseren Leistung, die Knotenvariante angewandt.

Anhand von Tabelle 2 kann man sehen, dass die ersten zwei Kombinationen durchschnittlich die besseren Erkennungsraten vorweisen. Aus diesem Grund fiel die Entscheidung, für die späteren Versuche, auf diese beiden Kombinationen.

| | Erkennungsrate |
|---|---------------------|
| $S_{Abs} \diamond S_{Canb}$ | 63.93 ± 3.38 |
| $S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$ | 69.75 ± 3.26 |
| $S_{Abs} \diamond S_{Manh} \diamond S_{CHI} \diamond S_{PCC}$ | 68.58 ± 3.66 |
| $S_{Canb} \diamond S_{Manh} \diamond S_{CHI} \diamond S_{PCC}$ | 69.40 ± 2.84 |
| $S_{Abs} \diamond S_{Canb} \diamond S_{Manh} \diamond S_{CHI} \diamond S_{PCC}$ | 69.65 ± 2.76 |

Tabelle 2: Durchschnittliche Erkennungsrate in Prozent für die ausgewählten Kombinationen nach 100 zufällig gemischten Testläufen.

Graphenstruktur. Es wurden drei verschiedene Graphen verwendet. Ein Gittergraphen mit einem Abstand von 16 Pixeln zwischen den einzelnen Knoten (*Gitter-16*). Eine abgewandelte Form, welche nur die Knoten auf dem Objekt selbst enthält (*NB-Gitter-16*) und ein Graph bestehend aus den Ecken, welche der FAST Algorithmus [Ros12] [RD05] anhand des Ergebnisses der Kantenextraktion durch den Canny Algorithmus [Can86] extrahiert hat (*Canny-Fast*).

Bei der Erstellung der Ranglisten von *Gitter-16* wurde unterschieden zwischen einer Rangliste pro Graph und einer Rangliste pro Knoten im Graphen. Bei den anderen Graphen ist dies nicht möglich, da bei ihnen die Anzahl an Knoten variieren kann und sie folglich nicht mehr auf Knotenbasis vergleichbar sind.

| | Graph | Rangliste | Erkennungsrate |
|---|---------------------|-----------|----------------|
| $S_{Abs} \diamond S_{Canb}$ | <i>Gitter-16</i> | Bild | 51.5 |
| | | Knoten | 62.75 |
| | <i>NB-Gitter-16</i> | Bild | 50.5 |
| | | | |
| | <i>Canny-FAST</i> | Bild | 48.25 |
| | | | |
| $S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$ | <i>Gitter-16</i> | Bild | 66.25 |
| | | Knoten | 70 |
| | <i>NB-Gitter-16</i> | Bild | 66.75 |
| | | | |
| | <i>Canny-FAST</i> | Bild | 70.75 |
| | | | |

Tabelle 3: Erkennungsrate in Prozent für verschiedene Graphenstrukturen.

In der zweiten Kombination erreichte der *Canny-Fast-Graph* zwar eine höhere Erkennungsleistung, als *Gitter-16*, dafür aber die schlechteste in der ersten Kombination.

Aus Performancegründen und um beide Kombinationen besser vergleichen zu können, fiel die Entscheidung auf die Gittergraphen-Variante.

Zuletzt wurde überprüft, wie sich der Abstand zwischen den einzelnen Knoten im Gittergraphen auf die Erkennungsleistung auswirkt.

| Graph | ER $S_{\text{Abs}} \diamond S_{\text{Canb}}$ | ER $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ | Knoten |
|------------------|--|--|--------|
| <i>Gitter-4</i> | 64.0 | 70.75 | 1024 |
| <i>Gitter-8</i> | 64.25 | 71.25 | 256 |
| <i>Gitter-12</i> | 63.0 | 70.75 | 100 |
| <i>Gitter-16</i> | 62.75 | 70.0 | 64 |
| <i>Gitter-20</i> | 60.75 | 69.75 | 36 |
| <i>Gitter-24</i> | 60.0 | 68.75 | 25 |
| <i>Gitter-32</i> | 57.0 | 68.25 | 16 |
| <i>Gitter-64</i> | 45.5 | 65.25 | 4 |

Tabelle 4: Erkennungsrate in Prozent für verschiedene Gittergraphen.

Die besten Raten lagen hier bei den Abständen von 4 bis 16 Pixeln.

2.6.2 Aufbau

Aus den Ergebnissen des Leistungsvergleichs wurden die zwei folgenden Versuchsaufbauten konzipiert.

| Versuch | Ähnlichkeitsfunktionen | Rangliste | Graphen |
|---------|---|-----------|------------------|
| V_1 | $S_{\text{Abs}} \diamond S_{\text{Canb}}$ | Knoten | <i>Gitter-16</i> |
| V_2 | $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ | Knoten | <i>Gitter-16</i> |

Tabelle 5: Die zwei Versuchsaufbauten.

Der Graph *Gitter-16* wurde aus Performancegründen ausgewählt, auch wenn die Erkennungsrate etwas geringer als bei *Gitter-8* war. Dafür benötigt er aber nur ein Viertel der Knoten. Als Ranglistenmethode wurde eine Rangliste pro Knoten im Graphen generiert, da hier die Erkennungsrate in allen vorangehenden Vergleichen immer höher war.

Bei den Ähnlichkeitsfunktionen wurde in V_1 die Kombination mit der höchsten Erkennungsrate mit reinen *graph matching* ausgewählt und in V_2 die Kombination, welche insgesamt die beste Erkennungsleistung erreichte.

Bei jedem Test wurden immer 100 Objekte verwendet. Bei der ALOI Datenbank die ersten 100 der zufälligen Verteilung aller 1000 Objekte. Von der COIL-100 Datenbank wurden alle 100 Objekte verwendet.

Alle angegebenen Erkennungsraten in den Ergebnissen sind arithmetische Mittelwerte und die dazugehörige Standardabweichung um den Mittelwert, gewonnen aus einer Zufalls-kreuzvalidierung von 100 Testläufen (*repeated random sub-sampling validation*).

3 Ergebnisse

In allen Versuchen mit der ALOI Datenbank wurde die Situation 0 aus der ALOI Blickwinkelgruppe als Galeriebild bzw. als Situation 0 in der Modelldatenbank verwendet. Die restlichen Modell- und Probedbilder wurden, je nach Versuchsreihe, der entsprechenden ALOI Gruppe entnommen.

Die COIL100 Datenbank enthält nur eine Gruppe, weshalb hier nur eine Versuchsreihe durchgeführt wurde, für welche die Situation 0 als Galeriebild Verwendung fand.

Der genaue Versuchsaufbau kann unter Abschnitt 2.6.2 nachgelesen werden.

3.1 Farbtemperatur

Es wurden alle 12 Farbtemperaturen verwendet.

| Anteil an Modellbilder | ER $S_{\text{Abs}} \diamond S_{\text{Canb}}$ | ER $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ |
|------------------------|--|--|
| 3% | 99.99 ± 0.01 | 89.9 ± 6.7 |
| 5% | 99.99 ± 0.01 | 97.5 ± 2.1 |
| 10% | 100.0 ± 0 | 99.8 ± 0.2 |
| 20% | 100.0 ± 0 | 99.99 ± 0.01 |
| 30% | 100.0 ± 0 | 99.99 ± 0.01 |
| 40% | 100.0 ± 0 | 99.99 ± 0.01 |
| 50% | 100.0 ± 0 | 99.99 ± 0.01 |
| 60% | 100.0 ± 0 | 99.99 ± 0.01 |
| 70% | 100.0 ± 0 | 100.0 ± 0 |
| 80% | 100.0 ± 0 | 100.0 ± 0 |
| 90% | 100.0 ± 0 | 100.0 ± 0 |

Tabelle 6: Erkennungsrate auf der ALOI Datenbank bei variierender Farbtemperatur in Prozent.

Die Kombination $S_{\text{Abs}} \diamond S_{\text{Canb}}$ erreicht bereits ab eine Modellmenge von 10% eine Erkennungsrate von 100%, während die Kombination $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ diese erst ab 70% erreicht.

3.2 Beleuchtung

Es wurden alle 24 Beleuchtungswinkel verwendet.

| Anteil an Modellbilder | ER $S_{\text{Abs}} \diamond S_{\text{Canb}}$ | ER $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ |
|------------------------|--|--|
| 3% | 89.3 ± 3.1 | 53.6 ± 7.8 |
| 5% | 96.0 ± 1.1 | 61.8 ± 6.6 |
| 10% | 98.7 ± 0.4 | 79.5 ± 3.9 |
| 20% | 99.4 ± 0.3 | 91.6 ± 1.5 |
| 30% | 99.6 ± 0.2 | 94.9 ± 1.3 |
| 40% | 99.7 ± 0.2 | 96.6 ± 1.0 |
| 50% | 99.8 ± 0.2 | 97.6 ± 0.9 |
| 60% | 99.8 ± 0.2 | 98.1 ± 0.7 |
| 70% | 99.9 ± 0.1 | 98.5 ± 1.0 |
| 80% | 99.9 ± 0.1 | 99.3 ± 0.8 |
| 90% | 99.9 ± 0.1 | 99.5 ± 1.0 |

Tabelle 7: Erkennungsrate auf der ALOI Datenbank bei variierender Beleuchtung in Prozent.

Die Kombination $S_{\text{Abs}} \diamond S_{\text{Canb}}$ erreicht bereits ab einer Modellmenge von 20% eine hohe Erkennungsrate von 99.4 ± 0.3 , während die Kombination $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ eine vergleichbare Rate erst ab 80% erreicht.

Anhand der folgenden Liniendiagramme lässt sich erkennen, dass vor allem die Aufnahmen von Kamera 3 das größere Fehleinschätzungspotenzial aufweisen.

Zur besseren Übersicht wurde der Wertebereich der Y-Achse an die Erkennungsrate angepasst und nur die Modellmengen ab 50% eingezeichnet.

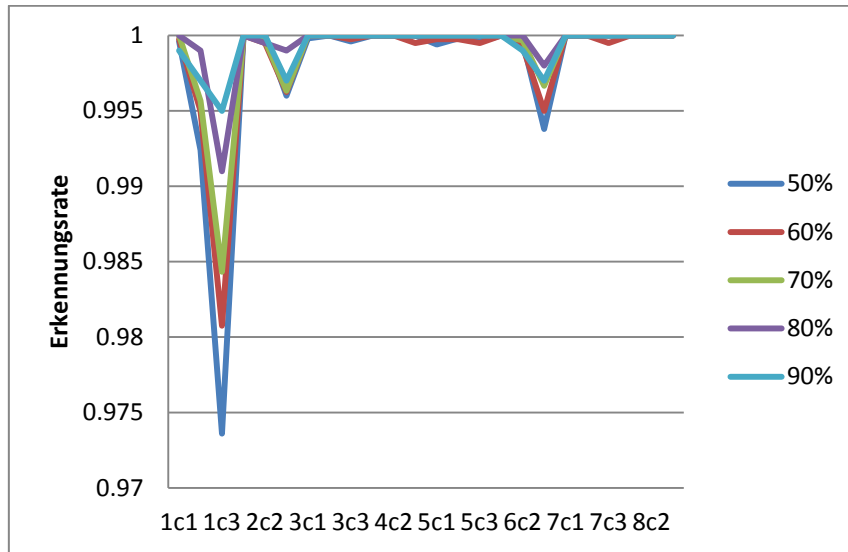


Abbildung 10: Erkennungsrate ($S_{\text{Abs}} \diamond S_{\text{Canb}}$) in Abhängigkeit von Kamera- und Lichtposition.

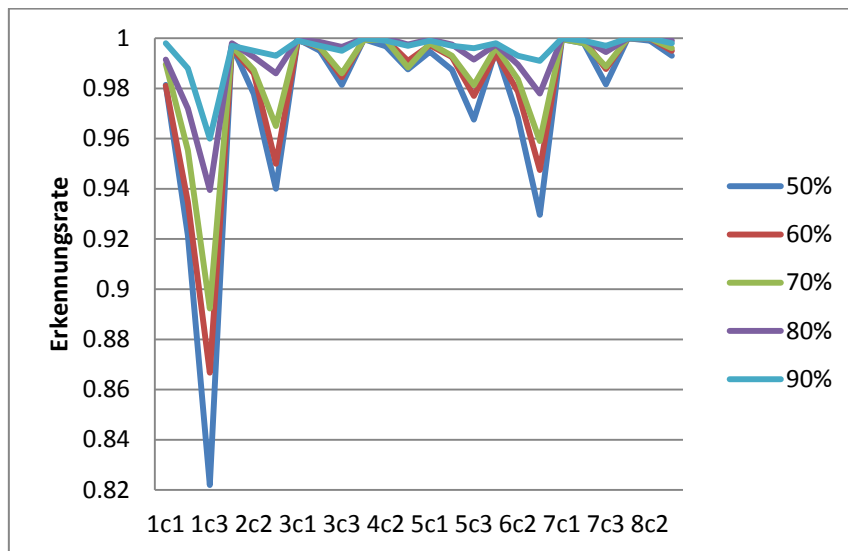


Abbildung 11: Erkennungsrate ($S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$) in Abhängigkeit von Kamera- und Lichtposition.

3.3 Betrachtungswinkel

3.3.1 COIL-100

Für den ersten Versuch wurden nur neun der 72 Betrachtungswinkel im Abstand von 40° verwendet.

| Anteil an Modellbilder | ER $S_{\text{Abs}} \diamond S_{\text{Canb}}$ | ER $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ |
|------------------------|--|--|
| 3% | 20.4 ± 6.5 | 17.64 ± 5.88 |
| 5% | 29.9 ± 5.6 | 21.82 ± 4.36 |
| 10% | 40.58 ± 3.16 | 35.22 ± 4.21 |
| 20% | 50.24 ± 2.46 | 50.3 ± 3.1 |
| 30% | 55.73 ± 2.6 | 59.78 ± 2.48 |
| 40% | 59.34 ± 3.06 | 65.28 ± 3.27 |
| 50% | 63.59 ± 3.59 | 69.96 ± 3.53 |
| 60% | 67.31 ± 4.17 | 74.16 ± 4.22 |
| 70% | 71.49 ± 4.31 | 77.75 ± 4.28 |
| 80% | 76.43 ± 5.36 | 82.04 ± 5.68 |
| 90% | 83.16 ± 7.42 | 88.41 ± 5.51 |

Tabelle 8: Erkennungsrate auf der COIL-100 Datenbank bei variierendem Betrachtungswinkel in Prozent.

Die Kombination $S_{\text{Abs}} \diamond S_{\text{Canb}}$ zeigt bis zu einer Modellmenge von 20% die besseren Erkennungsraten, während die Kombination $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ im Gegenzug ab einer Modellmenge von 40% eine etwas höhere Erkennungsleistung erreicht.

Aus dem folgenden Liniendiagramm lässt sich der Schluss ziehen, dass die seitlichen Aufnahmen am fehleranfälligstem sind.

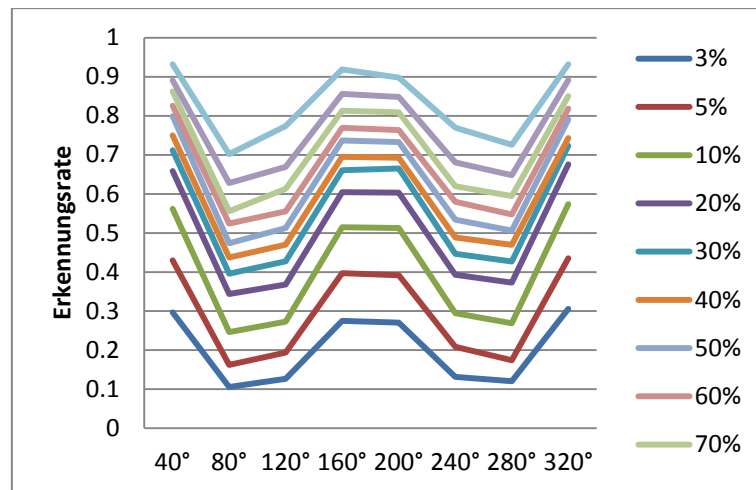


Abbildung 12: Erkennungsrate auf COIL-100 ($S_{\text{Abs}} \diamond S_{\text{Canb}}$) in Abhängigkeit vom Betrachtungswinkel.

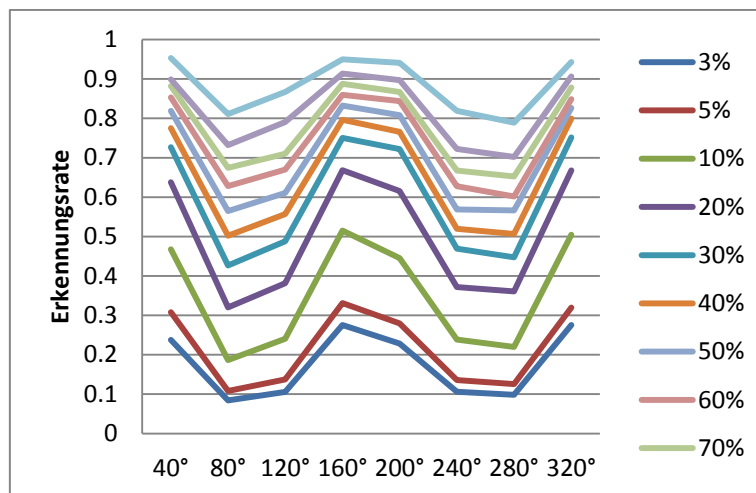


Abbildung 13: Erkennungsrate auf COIL-100 ($S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$) in Abhängigkeit vom Betrachtungswinkel.

Die zweite Kombination zeigt zwar in den seitlichen Aufnahmen eine Verbesserung, hat aber dennoch auch dort ihre Schwächen.

Zuletzt wurde noch ein Versuch mit allen 72 Blickwinkeln durchgeführt. Die Modellmenge betrug hier 90%, da diese die beste Erkennungsrate aufgewiesen hat.

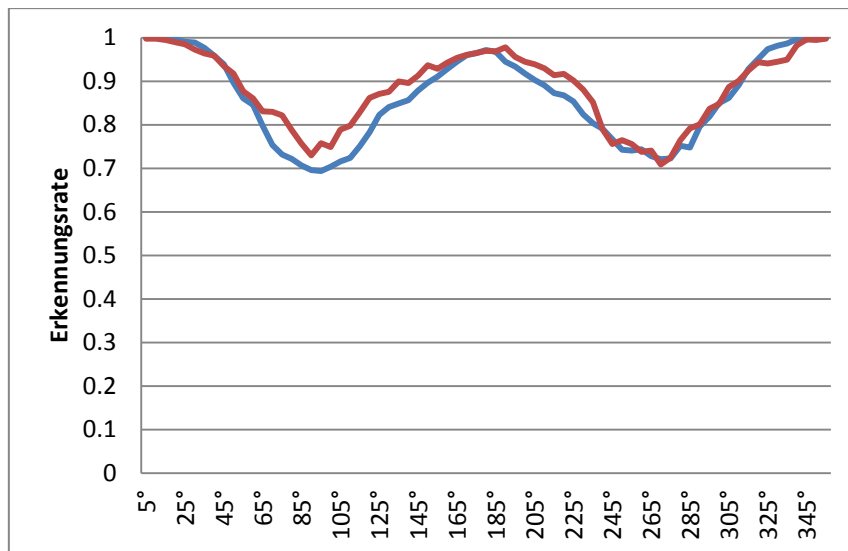


Abbildung 14: Erkennungsrate auf COIL-100 in Abhängigkeit vom Betrachtungswinkel (alle 72 Blickwinkel). Rot ist die Kombination $S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$ und blau die Kombination $S_{Abs} \diamond S_{Canb}$

Hier lässt sich gerade bei geringen Abweichungen im Blickwinkel ($\pm 30^\circ$) eine sehr hohe Erkennungsrate von bis zu 100% ausmachen. Ansonsten entspricht der Verlauf dem des vorherigen Versuchs mit neun Blickwinkeln.

3.3.2 ALOI

Für den ersten Versuch wurden ebenfalls nur neun Betrachtungswinkel im Abstand von 40° verwendet.

| Anteil an Modellbilder | ER $S_{Abs} \diamond S_{Canb}$ | ER $S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$ |
|------------------------|--------------------------------|--|
| 3% | 33.8 ± 9.19 | 28.42 ± 8.64 |
| 5% | 47.18 ± 6.35 | 34.72 ± 6.07 |
| 10% | 59.34 ± 3.66 | 48.37 ± 5.19 |
| 20% | 65.83 ± 3.18 | 64.54 ± 4.14 |
| 30% | 68.29 ± 3.68 | 71.86 ± 3.49 |
| 40% | 70.8 ± 3.34 | 76.9 ± 3.25 |
| 50% | 72.9 ± 3.91 | 79.97 ± 3.05 |

| | | |
|-----|------------------|------------------|
| 60% | 75.68 ± 4.11 | 82.46 ± 3.31 |
| 70% | 78.02 ± 3.91 | 86.14 ± 3.87 |
| 80% | 81.29 ± 5.18 | 88.87 ± 4.25 |
| 90% | 88.24 ± 6.17 | 93.86 ± 4.05 |

Tabelle 9: Erkennungsrate auf der ALOI Datenbank bei variierendem Betrachtungswinkel in Prozent.

Auch hier ergibt sich das gleiche Gesamtbild, wie beim Versuch mit der COIL-100 Datenbank. Der einzige Unterschied liegt darin, dass die Erkennungsrate hier generell etwas höher liegt.

Aus dem Liniendiagramm lässt sich als Grund für die höhere Erkennungsrate die bessere Leistung bei seitlichen Aufnahmen herauslesen.

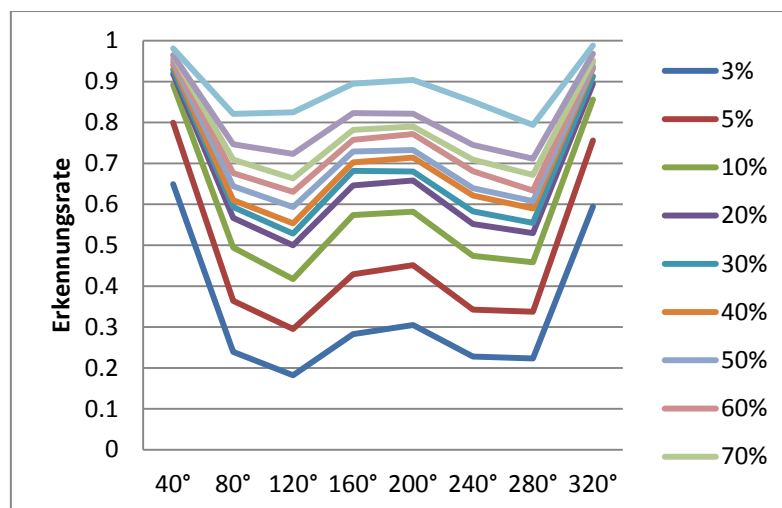


Abbildung 15: Erkennungsrate auf ALOI ($S_{Abs} \diamond S_{Canb}$) in Abhängigkeit vom Betrachtungswinkel.

Wie man dem folgenden Liniendiagramm entnehmen kann, zeigt auch hier die zweite Kombination eine Verbesserung in den lateralen Aufnahmen.

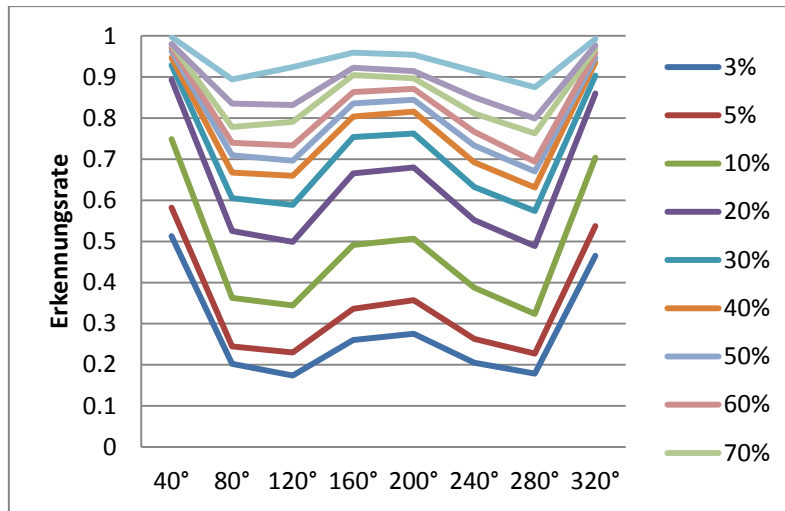


Abbildung 16: Erkennungsrate auf ALOI ($S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$) in Abhängigkeit vom Betrachtungswinkel.

Auch hier wurde ein Versuch mit allen 72 Blickwinkeln durchgeführt. Die Modellmenge betrug ebenfalls 90%.

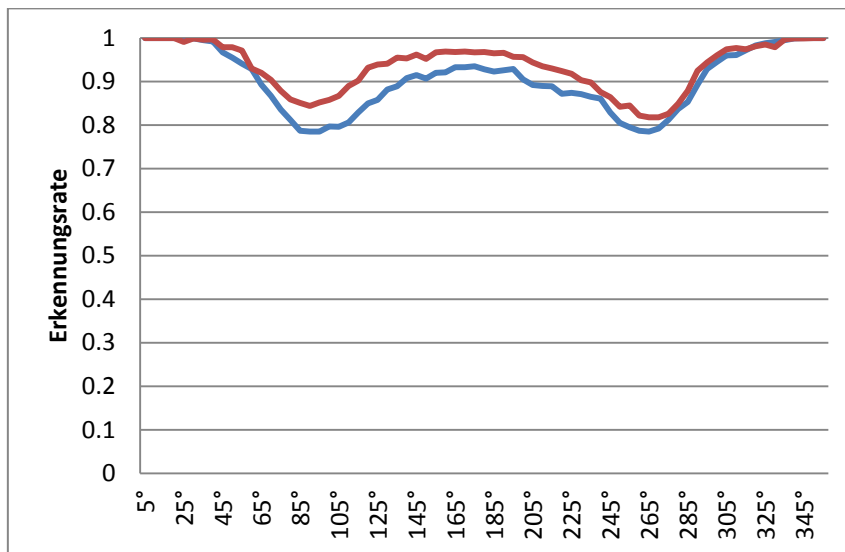


Abbildung 17: Erkennungsrate auf ALOI in Abhängigkeit vom Betrachtungswinkel (alle 72 Blickwinkel).

Rot ist die Kombination $S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$ und blau die Kombination $S_{Abs} \diamond S_{Canb}$.

Hier ergeben sich vor allem bei geringen Abweichungen von $\pm 45^\circ$ im Blickwinkel sehr hohe Erkennungsraten von bis zu 100%. Ansonsten entspricht der Verlauf dem der vorherigen Versuche dieser Versuchsreihe.

3.3.3 Bestimmung des Winkels

Wie unter Abschnitt 2.4 beschrieben, wurde in einer letzten Versuchsreihe der Winkel durch das System geschätzt und der Erkennungsroutine übergeben.

Als Modellmenge wurde hier, um einen besseren Vergleich zu ermöglichen, 50% gewählt und neun Betrachtungswinkel verwendet.

Die Schätzungsleistung vom System war vor allem auf der ALOI Datenbank sehr gering. Dennoch hat sich die Erkennungsleistung aber nur geringfügiger verschlechtert.

Bei dem Versuch auf der COIL-100 Datenbank war die Schätzungsleistung zwar besser als bei ALOI, im Gegenzug verschlechterte sich die Erkennungsrate aber deutlicher als bei ALOI.

| Datenbank | ER Winkel | ER $S_{Abs} \diamond S_{Canb}$ | ER $S_{Abs} \diamond S_{Canb} \diamond S_{CHI} \diamond S_{PCC}$ |
|-----------|----------------|--------------------------------|--|
| COIL-100 | 36.9 ± 1.9 | 60.3 ± 3.2 | 62.8 ± 4.2 |
| COIL-100 | gegeben | 63.59 ± 3.59 | 69.96 ± 3.53 |
| ALOI | 18.8 ± 2.0 | 71.7 ± 4.3 | 76.1 ± 3.4 |
| ALOI | gegeben | 72.9 ± 3.91 | 79.97 ± 3.05 |

Tabelle 10: Erkennungsrate nach Schätzung des Betrachtungswinkels in Prozent.

3.3.4 Laufzeiten

Informationen zur verwendeten Hard- und Software finden sich unter 2.5. Es wurden 100 Objekte und neun Betrachtungswinkel verwendet. 50% der Objekte als Modellbilder.

| Datenbank | Alle Graphen | Ein Objekt | Ein Graph |
|-----------|--------------|------------|-----------|
| COIL-100 | 180.9 | 1.8 | 0.2 |
| ALOI | 323.1 | 3.2 | 0.36 |

Tabelle 11: Laufzeiten in Sekunden für die Generierung der 900 Graphen.

Die zweite Kombination hat zwar bei den Blickwinkelversuchen die bessere Erkennungsleistung, im Gegenzug aber eine deutlich schlechtere Laufzeit.

| Datenbank | Winkelschätzung | Winkel $S_{\text{Abs}} \diamond S_{\text{Canb}}$ | Winkel $S_{\text{Abs}} \diamond S_{\text{Canb}} \diamond S_{\text{CHI}} \diamond S_{\text{PCC}}$ |
|-----------|-----------------|--|--|
| COIL-100 | 4.3 | 10.1 | 57.2 |
| ALOI | 7.1 | 16.7 | 106.9 |

Tabelle 12: Laufzeiten in Sekunden für die Winkelschätzung und die Erkennungsroutine.

4 Diskussion

Das in dieser Arbeit vorgeschlagene Verfahren zur invarianten Objekterkennung mittels Ähnlichkeitsranglisten basiert auf Ähnlichkeitsbeziehungen eines Objekts in einer bestimmten Situation, zu einer Menge an Modellobjekten in derselben Situation. Der Lerneffekt hängt, ebenso wie die Erkennungsrate, bis zu einem bestimmten Grad direkt von den gewählten Modellobjekten ab. Eine reine Angabe der Erkennungsleistung auf eine bestimmte vorgegebene Modellmenge wäre somit keineswegs repräsentativ für die realistische Lernleistung des Verfahrens. Aus besagtem Grund wurden alle Ergebnisse durch einer Zufallskreuzvalidierung von 100 Testläufen (*repeated random subsampling validation*) ermittelt. Es sollte also davon ausgegangen werden können, dass diese unabhängig von den gewählten Modellobjekten betrachtet werden dürfen und somit auch das tatsächliche Lernergebnis des Verfahrens bestmöglich darstellen.

4.1 Farbtemperatur

Die Erkennungsleistung lässt sich bei beiden Kombinationen kaum bis gar nicht von der wechselnden Farbtemperatur beeinflussen. Das etwas schwächere Ergebnis von Kombination 2 lässt sich von der Verwendung von S_{CHI} ableiten. In dieser Versuchsreihe, in der es speziell um eine wechselnde Farbinformation geht, stellt die Verwendung der selbigen einen Nachteil dar. Trotzdem ist auch die Leistung von Kombination 2 als hoch zu werten. Kombination 1 ist hier nur deutlich robuster bei einer Rate von 100% ab 10% Modellanteil.

4.2 Beleuchtung

In der Versuchsreihe mit variierendem Beleuchtungswinkel zeigt Kombination 1 ebenfalls die bessere Leistung. Auch hier gereicht der zweiten Kombination die Verwendung von S_{CHI} zum Nachteil, aber auch S_{PCC} verschlechtert das Ergebnis um einige Prozent.

Ogleich die Erkennungsraten als hoch angesehen werden können, so ist Kameraposition 3 mit Lichtposition 1 (die genauen Spezifikationen finden sich unter [GBS05]) die Situation, welche die höchste Fehlerrate in dieser Versuchsreihe aufweist.

Die Versuche haben deutlich gemacht, dass die unterschiedliche Beleuchtung die Erkennungsleistung nur geringwertig verschlechtert.

4.3 Betrachtungswinkel

Die Unterschiede zwischen den Erkennungsraten bei COIL-100 und ALOI könnten unter anderem mit den unterschiedlichen Skalierungsverfahren der beiden Datenbanken zusammenhängen. Diese Hypothese wurde in der vorliegenden Arbeit allerdings nicht überprüft. Dennoch legt der Umstand, dass in diesem Fall auch die Größe des Objekts als Ähnlichkeitsmaß mit einbezogen werden kann, die Vermutung nahe, dass die Messung der Ähnlichkeitsbeziehungen einfacher ist, wenn die Objekte ohne besondere Skalierung verglichen werden.

Beim Betrachtungswinkel hat das hier vorgeschlagene Verfahren zwei Schwächen. Zum einen benötigt das Verfahren einen hohen Anteil an Modellbildern. Zum anderen nimmt die Erkennungsrate beim seitlichen Blickwinkel (45° bis 125° Unterschied zur Frontalpose) deutlich ab.

Besonders positiv fallen dafür die Identifizierungsraten in den Bereichen von 5° bis 45° und von 125° bis 180° Unterschied zur Frontalpose aus.

Dieses Phänomen rührt vermutlich daher, dass zwei Objekte sich frontal (und um 180° gedreht) zwar stark ähneln können, dafür lateral aber eine geringere Ähnlichkeit aufweisen.

Beim Betrachtungswinkel profitiert Kombination 2 zum ersten Mal von S_{CHI} und S_{PCC} , vor allem in der lateralen Erkennung. Durch die Verwendung von weiteren geeigneten Ähnlichkeitsmaßen sollte es möglich sein, die angesprochenen Schwächen zu minimieren und die Erkennungsleistung so zu maximieren.

Das hier verwendete Verfahren zur Schätzung des Winkels erwies sich als unbrauchbar, aber trotz des geringen Schätzungserfolges verschlechterte sich die Erkennungsleistung nur um wenige Prozent.

4.4 Vergleich mit anderen Verfahren

Ein direkter Vergleich mit anderen Verfahren ist schwer, da das hier vorgeschlagene Verfahren ungelernte Objekte in bekannten Situationen erkennt, während auf dem aktuellen Stand der Technik versucht wird, ein gelerntes Objekt in einer unbekanntem Situation wieder zu erkennen.

Ein Vergleich einzig und allein anhand der tatsächlichen Erkennungsleistung, lässt sich mit dem *Local Affine Frames* Verfahren ziehen, das eine sehr hohe Erkennungsrate von 99.9% auf der COIL-100 Datenbank, bei Verwendung von 18 der 72 Blickwinkel als Trainingsdaten, erreichte (vgl. [OM02]) und dem hier vorgestellten Verfahren somit überlegen.

4.5 Fazit

Die Fragestellung dieser Arbeit, ob ein System in der Lage ist, allein durch Ähnlichkeitsbeziehungen zu bekannten Objekten, ein neues Objekt in einer bekannten Situation seinem Galeriebild zuzuordnen, kann unter Vorbehalt mit Ja beantwortet werden. Denn auch wenn das hier vorgeschlagene Verfahren noch einer Optimierung bedarf, so konnte es bereits seine Stärken zeigen und aus seinen Schwächen lässt sich ermitteln, wie man das Verfahren noch verbessern kann. Trotzdem, kann die erreichte Leistung nicht mit dem aktuellen Stand der Technik mithalten.

4.6 Ausblick

Das hier vorgestellte Verfahren zur invarianten Objekterkennung mittels Ähnlichkeitsranglisten könnte auf mehreren Ebenen verbessert werden. Zum einen durch Hinzufügen von neuen geeigneten Ähnlichkeitsfunktionen bzw. -maßen, die eine genauere Eingrenzung der Ähnlichkeit, vor allem im lateralen Bereich, ermöglichen könnten. Zum anderen könnte auch die Ähnlichkeitsfunktion zum Vergleich der Ranglisten optimiert werden. Erste kleinere Experimente in diese Richtung brachten bereits eine Verbesserung von einigen Prozent, wurden aber nicht weiter verfolgt, da dies ansonsten den Rahmen dieser Arbeit gesprengt hätte.

Wie in der vorliegenden Arbeit gezeigt, kann auch die Wahl der Graphenstruktur die Leistung beeinflussen. Auch hier ließen sich somit einige Optimierungen vornehmen. Inwiefern sich der Einsatz eines elastischen Graphen gegenüber der hier verwendeten lohnen würde, wäre ebenfalls eine Untersuchung wert.

Auch würde es sich anbieten, ein verbessertes Verfahren zur Schätzung der Situation zu entwickeln, um die Erfolgsrate zu steigern.

Zusammenfassung

In dieser Arbeit wird ein Verfahren zur invarianten Objekterkennung mittels Ähnlichkeitsranglisten vorgestellt, welches in der Lage ist, ein nicht vorher gelerntes Objekt in einer bekannten Situation seinem Galeriebild zuzuordnen. Es basiert auf der Annahme, dass Objekte, die sich in einer Situation ähneln, dies auch in einer veränderten Situation tun werden. Dazu leitet das System Ähnlichkeitsbeziehungen von einer Gruppe Modellobjekte in der gleichen Situation ab, welche in Form einer Rangliste als Schlüssel zur Ermittlung des korrespondierenden Galeriebildes dienen.

Literaturverzeichnis

- [BN06] I. Bülthoff and F. N. Newell, "The role of familiarity in the recognition of static and dynamic objects.," *Progress in Brain Research*, vol. 154, pp. 315-325, 2006.
- [Can86] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, Nov. 1986.
- [DZR12] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, "How Does the Brain Solve Visual Object Recognition?," *Neuron*, vol. 73, no. 3, pp. 415-134, Feb. 2012.
- [GBS05] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam library of object images," *Int. J. Comput. Vision*, vol. 61, no. 1, pp. 103-112, Jan. 2005.
- [GBTS07] D. Gonzáles, et al., "Distance Measures for Gabor Jets-Based Face Authentication: A Comparative Evaluation," in *Advances in Biometrics, ICB 2007. LNCS 4642*, S. .-W. Lee and S. Z. Li, Eds. Springer-Verlag Berlin Heidelberg, 2007, pp. 474-483.
- [GW09] M. Günther and R. P. Würtz, "Face detection and recognition using maximum likelihood classifiers on Gabor graphs.," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 3, pp. 433-461, 2009.
- [Kat12] K. Schmidt. Korrelationskoeffizient. [Online].
<http://wirtschaftslexikon.gabler.de/Definition/korrelationskoeffizient.html>
- [LL04] O. Linde and T. Lindeberg, "Object Recognition Using Composed Receptive Field Histograms of Higher Dimensionality," *Proceedings of ICPR. ICPR '04, Washington, DC, USA, IEEE Computer Society*, pp. 1-6, 2004.
- [LVBL93] M. Lades, et al., "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *IEEE Transactions on Computers*, vol. 42, no. 3, pp. 300-311, 1993.
- [LW12] M. Lessmann and R. P. Würtz, "Learning of Invariant Object Recognition in a Hierarchical Network," *Proceedings of New Challenges in Neural Computation, Graz*, vol. 3, pp. 104-112, 2012.
- [MHTS07] M. K. Müller, A. Heinrichs, A. H. Tewes, A. Schäfer, and R. P. Würtz, "Similarity rank correlation for face recognition under unenrolled pose.," in *Advances in Biometrics. LNCS*, S. .-W. Lee and S. Z. Li, Eds. Springer, 2007, pp. 67-76.
- [MTBW12] M. K. Müller, M. Tremer, C. Bodenstein, and R. P. Würtz, "Lernen situationsunabhängiger Personenerkennung," *Informatikspektrum*, vol. 32, no. 2, pp. 112-118, 2012.
- [MTBW12E] M. K. Müller, M. Tremer, C. Bodenstein, and R. P. Würtz, "Learning invariant face recognition from examples," *Neural Networks*, 2012.
- [NNM96] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia Object Image Library

- (COIL-100)," *Technical Report CUCS-006-96*, Feb. 1996.
- [OM02] Š. Obdržálek and J. Matas, "Object Recognition using Local Affine Frames on Distinguished Regions," *British Machine Vision Association*, 2002.
- [PBCD11] N. Pinto, Y. Barhomi, D. D. Cox, and J. J. DiCarlo, "Comparing State-of-the-art Visual Features on Invariant Object Recognition Tasks," *Proceedings of the 2011 IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 463-470, Jan. 2011.
- [RD05] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," *IEEE International Conference on Computer Vision*, pp. 1508-1511, Oct. 2005.
- [RD06] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," *European Conference on Computer Vision*, pp. 430-443, May 2006.
- [Ros12] E. Rosten. (2012, Sep.) FAST Corner Detection -- Edward Rosten. [Online]. <http://www.edwardrosten.com/work/fast.html>
- [SB91] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.
- [TFM96] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, pp. 520-522, Jun. 1996.
- [TLM06] M. Tatsuno, P. Lipa, and B. L. McNaughton, "Methodological Considerations on the Use of Template Matching to Study Long-Lasting Memory Trace Replay," *The Journal of Neuroscience*, vol. 26, no. 42, p. 10727-10742, 2006.
- [WFKM97] L. Wiskott, J. -M. Fellous, N. Krüger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching.," *IEEE Transactions on Pattern Analysis and Machine*, vol. 19, no. 7, pp. 775-779, 1997.
- [WW09] G. Westphal and R. P. Würtz, "Combining Feature- and Correspondence-Based Methods for Visual Object Recognition," *Neural Computation*, vol. 21, pp. 1952-1989, 2009.

Erklärung

Ich erkläre, dass das Thema dieser Arbeit nicht identisch ist mit dem Thema einer von mir bereits für ein anderes Examen eingereichten Arbeit.

Ich erkläre weiterhin, dass ich die Arbeit nicht bereits an einer anderen Hochschule zur Erlangung eines akademischen Grades eingereicht habe.

Ich versichere, dass ich die Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, habe ich unter Angabe der Quellen der Entlehnung kenntlich gemacht. Dies gilt sinngemäß auch für gelieferte Zeichnungen, Skizzen und bildliche Darstellungen und dergleichen.

Datum

Unterschrift