



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Studienarbeit

Carsten Fries

Objekterkennung mit SIFT-Merkmalen

Carsten Fries

Carsten.Fries@haw-hamburg.de

Objekterkennung mit SIFT-Merkmalen

Studienarbeit eingereicht im Rahmen der Vorlesung Anwendungen 1

im Studiengang Informatik (Master)
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuer:

Prof. Dr.-Ing. Andreas Meisel

Prüfer:

Prof. Dr. rer. nat. Kai von Luck

Prof. Dr. rer. nat. Gunter Klemke

Abgegeben am 06. Januar 2010

Carsten Fries

Thema der Studienarbeit

Objekterkennung mit SIFT-Merkmalen

Stichworte

Objekterkennung, SIFT, Skalierungsinvariante Merkmale, Merkmalsextraktion, Greifarm, Knickarm-Roboter

Kurzzusammenfassung

Diese Arbeit befasst sich mit der Wiedererkennung von Gegenständen anhand der skaleninvarianten Merkmalsextraktion (engl. Scale-invariant feature transform, kurz SIFT). SIFT-Merkmale weisen eine Invarianz gegenüber Rotation, Translation, Skalierung, Änderung der Lichtverhältnisse und teilweise affiner Verzerrung auf. Das SIFT-Verfahren wurde für die Erfassung robuster Merkmale herangezogen und untersucht. Für das Verfahren dienen Kamerabilder als Informationsgrundlage.

Carsten Fries

Title of the paper

Recognition of objects using SIFT-features

Keywords

object detection, SIFT, scale-invariant features, feature extraction, manipulator-arm

Abstract

This thesis focuses on the recognition of objects using "Scale-invariant feature transform (SIFT)"-features. SIFT-features are invariant to rotation, translation, scale, illumination changes and partially invariant to affine distortion. The SIFT-method is used for robust features detection. The analyzed information comes from camera images.

Inhaltsverzeichnis

1	Problemstellung	1
2	Detektion robuster SIFT-Merkmale	2
2.1	Ermittlung potentieller Merkmale	2
2.2	Filterung und Lokalisierung potentieller Merkmalspunkte	4
2.3	Bestimmung der Hauptorientierungen	5
2.4	Erzeugung der Merkmalsvektoren	6
3	Merkmalsvergleich zur Identifikation von Gegenständen	11
3.1	Merkmalsvergleich auf Basis des euklidischen Abstandes	11
3.2	Verbesserte Korrespondenzsuche mit einem Suchbaum	11
3.2.1	Aufbau eines k-d Baumes	12
3.2.2	Suche im k-d Baum	13
4	Zusammenfassung und Ausblick	16
	Abbildungsverzeichnis	17
	Quellenverzeichnis	18

1 Problemstellung

Die bis heute rasante technologische Entwicklung führte zu einer zunehmenden Anzahl an existierenden Assistenzrobotern. Sie kommen unter anderem für unterstützende Tätigkeiten wie Manipulations-, Prüf- und Nachbearbeitungsaufgaben zum Einsatz. Dabei werden die Tätigkeiten ortsfest oder mobil und gegebenenfalls autonom durchgeführt.

Diese Arbeit befasst sich mit der Erkennung von Gegenständen basierend auf Kamerabil- dern. In weiterführenden Arbeiten kann aufbauend auf diesen Erkenntnissen ein autonomer Knickarmroboter entwickelt werden. Dieser soll zur autonomen Identifizierung, Lokalisierung und Manipulation von Gegenständen eingesetzt werden.

Bei dem einzusetzenden Knickarmroboter handelt es sich um einen 4kg leichten Katana-Greifarm der Firma NEURONICS¹. Dieser besitzt sechs Freiheitsgrade und kann ortgebunden Tätigkeiten in einem Radius von 60cm bewältigen. Einen sicheren Umgang zwischen Mensch und Maschine wird durch die maximale Nutzlast von 500g gewährleistet. Die Bewegungsgeschwindigkeit beträgt 90° pro Sekunde. Dieser Wert entspricht einer Bewegung von einem Meter pro Sekunde. Die Gelenke können dabei gleichzeitig bewegt werden. Ein montierter Winkelgreifer ermöglicht das Greifen von Gegenständen. Dieser besteht aus zwei Fingern, welche mit Infrarot-, Kraft-, Leitfähigkeitssensoren und Encoder für die Greiferöffnung bestückt werden können. Der in dieser Arbeit verwendete „Katana-Greifarm 6M S“ mit montiertem Winkelgreifer ist nachfolgend dargestellt.



Abbildung 1.1: Katana-Greifarm mit sechs Freiheitsgraden.²

Für die Identifikation von Gegenständen ist eine Kamera nötig. Diese ist am Winkelgreifer zu befestigen und soll Grauwertbilder zu einem nah positionierten Computer übertragen. Aufgrund aufgenommener Bilder ist ein im System bekannter Gegenstand vorerst zu identifizieren. Das anschließende Lokalisieren, Anfahren, Greifen und Bewegen von Gegenständen ist in nachfolgenden Arbeiten zu behandeln. Die Softwarelösung soll unter dem Aspekt der Genauigkeit und Geschwindigkeit realisiert werden.

¹Zu finden auf AG (2009)

²Bildquelle: AG (2009)

2 Detektion robuster SIFT-Merkmale

Inhalt

2.1	Ermittlung potentieller Merkmale	2
2.2	Filterung und Lokalisierung potentieller Merkmalspunkte	4
2.3	Bestimmung der Hauptorientierungen	5
2.4	Erzeugung der Merkmalsvektoren	6

Das Verfahren zur skaleninvarianten Merkmalsextraktion (engl. Scale-invariant feature transform, kurz SIFT) wurde 1999 von LOWE³ entwickelt und veröffentlicht. Dieses Verfahren ermöglicht das Auffinden von invarianten Merkmalen bezüglich Rotation, Translation, Skalierung, Änderung der Lichtverhältnisse und teilweise affiner Verzerrung in Bildern. Anhand dieser Merkmale kann eine zuverlässige Identifizierung von Gegenständen erfolgen.

Die Arbeitsweise vom SIFT-Algorithmus gliedert sich in vier Schritte, welche nachfolgend erläutert werden.

2.1 Ermittlung potentieller Merkmale

Die Grafik 2.1 veranschaulicht die Suche nach robusten Merkmalskandidaten. Auf der oberen Seite sind von links nach rechts vier Gauß-Pyramiden dargestellt. Eine Pyramide besteht aus fortlaufend stärker geglätteten Abbildungen des zu analysierenden Bildes $I(x,y)$. Dabei kommt ein Gauß-Filter $G_\sigma(x,y)$ aufgrund seiner Invarianz gegenüber Skalierung zum Einsatz. Unter Verwendung von unterschiedlichen Varianzen⁴ σ ergibt sich eine in der Gleichung 2.1 angegebene Faltungsgleichung, welche auf das zu untersuchende Bild $I(x,y)$ angewendet wird.

$$I(x,y) * G_\sigma(x,y) = I(x,y) * \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \right) \quad (2.1)$$

Jeweils zwei benachbarte Abbildungen der Gauß-Pyramide werden im Anschluss der Glättung subtrahiert. Diese Subtraktion wird als „Difference of Gaussian“ (DOG) bezeichnet. Die untere Seite stellt die Differenzbilder dar.

³Veröffentlichung von SIFT: LOWE (1999)

⁴Die Varianz ist ein Maß für die Streuung bezüglich einer Messgröße.

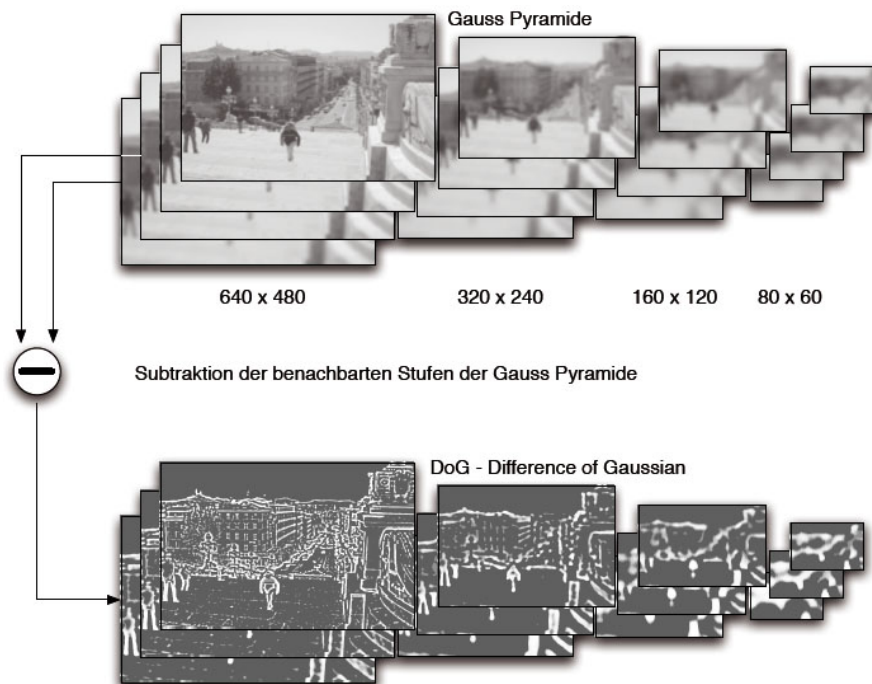


Abbildung 2.1: Ermittlung der Extrema im Skalenraum.⁵

Dieser beschriebene Vorgang wird n -mal mit jeweils um 50 % verkleinerten Abbildungen durchgeführt. Dabei ist als Ausgangsbild jeweils die letzte Abbildung aus der vorherigen Pyramide zu verwenden. Die somit entstehende Folge von unterschiedlichen Auflösungen eines Bildes bauen einen sogenannten „Skalenraum“ auf. Die Bildverkleinerung erfolgt, indem jeder zweite Pixel in x - und y -Richtung verwendet wird. Aufgrund der vorherigen Gaußglättung beinhaltet jeder Pixel Details von seinen Nachbarpixeln. Dadurch muss beim Verkleinern der Bilder keine zusätzliche Pixelinterpolation durchgeführt werden.

In der letzten Phase werden die DOG-Pyramiden auf minimale und maximale Pixelwerte untersucht. Ein Maximum ist gefunden, wenn alle der 26 benachbarten Pixel kleiner als er selbst sind. Die 26 Nachbarn ergeben sich aus den acht Nachbarn derselben Ebene und den jeweils neun Nachbarn der niedrigeren und höheren Ebene. Ein Minimum ist im umgekehrten Fall gefunden. Die Grafik 2.2 veranschaulicht die Nachbarschaft.

⁵Bildquelle: HEYMANN (2005)

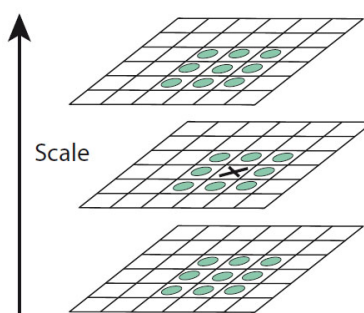


Abbildung 2.2: Potentieller Merkmalspixel wird mit seinen 26 Nachbarn verglichen.⁶

2.2 Filterung und Lokalisierung potentieller Merkmalspunkte

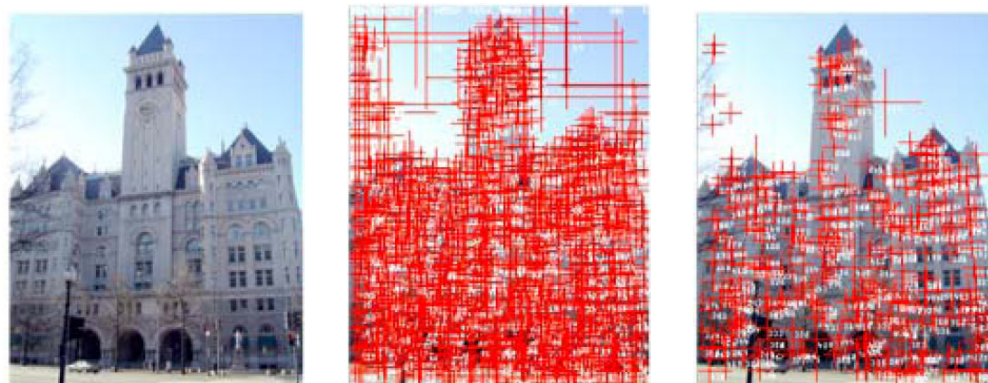
Alle im ersten Schritt ermittelten Extremwerte $L(x,y)$ sind Merkmalskandidaten. Im zweiten Schritt sind diese in einer Stabilitätsanalyse genauer zu untersuchen. Erfüllt ein Merkmalskandidat die festgelegten Stabilitätsbedingungen, ist seine exakte Position und Skalierung zu bestimmen. Andererseits ist der Merkmalskandidat nicht weiter von Bedeutung und wird verworfen. Stabilitätsmerkmale sind unter anderem Kontrast (Rauschunterdrückung), Klassifikation in Ecken, homogene Regionen und Kanten.

Ein Kantenpunkt kann anhand seiner zwei Hauptkrümmungen erkannt werden. Jeder Kantenpunkt besitzt eine große Hauptkrümmung quer über die Kante. Die zweite Hauptkrümmung verläuft entlang der Kante und ist deswegen sehr klein. Mit dem Verhältnis zwischen den beiden Hauptkrümmungen wird anschließend ermittelt, ob sich ein Punkt auf einer Kante befindet. Mit Hilfe einer 2×2 Hesse-Matrix lässt sich das Verhältnis der Hauptkrümmungen berechnen.

Ein Beispiel einer mit einem Kantenfilter angewandten Merkmalsfilterung, ist in der nachfolgenden Abbildung 2.3 dargestellt. Ausgehend eines Eingangsbildes (a) wurde dieses auf potentielle robuste Merkmale analysiert. Auf Basis der extrahierten Merkmalsmenge (b) konnten mit Hilfe eines Kantenfilters, Merkmale die sich auf Kantenregionen befanden, entfernt werden. Das rechte Bild (c) der Abbildung veranschaulicht die reduzierte Merkmalsmenge.

Dieser Fall ist ein gutes Beispiel für die gebräuchliche Anwendung eines Kantenfilters. Käme keine Kantenfilterung zum Einsatz, wäre die Wiedererkennung eines Kantenpunktes in einer anderen Aufnahme wahrscheinlich nicht möglich. Denn Punkte, welche sich auf derselben Kante befinden, haben oftmals die gleichen Merkmalseigenschaften und sind damit nicht zu unterscheiden.

⁶Bildquelle: LOWE (2004)



(a) Auf Merkmale zu untersuchende Aufnahme der Größe 320x240

(b) 1450 ermittelte Merkmale

(c) Kantengefilterte Menge mit 463 Merkmalen

Abbildung 2.3: Filterung von Merkmalen, welche sich auf Kanten befinden.⁷

2.3 Bestimmung der Hauptorientierungen

Die Hauptorientierung eines Merkmalspunktes kann mit den umliegenden Bildgradienten seines gaußgefiltertem Bildes bestimmt werden. Dabei ist das gaußgefilterte Bild, welches der Skalierung des Merkmalspunktes am nächsten kommt, zu verwenden. Ein Gradienten kennzeichnet die Richtung des größten Helligkeitsanstiegs und wird als Vektor mit seiner Länge $m(x,y)$ und Orientierung $\theta(x,y)$ angegeben. Die Berechnung eines Gradienten bezüglich eines Punktes $L(x,y)$, ist nachfolgend in den Gleichungen 2.2 und 2.3 beschrieben.

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (2.2)$$

$$\theta(x,y) = \tan^{-1} \frac{L(x+1,y) - L(x-1,y)}{L(x,y+1) - L(x,y-1)} \quad (2.3)$$

Für die Ermittlung der Merkmalsorientierung sind alle Gradienten in seiner näheren Umgebung zu berechnen. Anschließend sind die ermittelten Gradientenorientierungen θ_i zu gewichten. Die erste Gewichtung erfolgt mit der jeweiligen Gradientenlänge m_i . Das hat zur Folge, dass Gradienten mit einer größeren Länge, stärker bewertet werden. Die zweite Gewichtung erfolgt in Abhängigkeit eines Gewichtungsfaktors k . Dieser bewirkt, dass äußere Bereiche der 360°-Umgebung gedämpft einbezogen werden.

Nachfolgend ist ein Histogramm anhand der gewichteten Orientierungen θ_{gi} und den jeweiligen Gradientenlängen aufzubauen. Das entstehende Orientierungshistogramm ist in 36 Winkelbereiche unterteilt und hat somit eine Klassenbreite von 10°. Alle Gradientenlängen sind anhand ihrer gewichteten Orientierung an der passenden Stelle im Histogramm aufzuaddieren.

⁷Bildquelle: HEYMANN (2005)

Ist das Histogramm vollständig aufgebaut, kann die Hauptorientierung des Merkmales bestimmt werden. Dafür wird die größte Gradientenlänge m_{max} aus dem Histogramm ermittelt. Die Hauptorientierung des Merkmales wird anschließend mit der Länge m_{max} und dem dazugehörigen Orientierungswinkel θ_i zusammengesetzt.

In dem gesonderten Fall, dass die Länge eines weiteren Gradienten mindestens achtzig Prozent des größten Betrags aufweist, ist ein weiterer Merkmalspunkt zu erzeugen. Dies führt aufgrund der Möglichkeit den Merkmalspunkt für mehrere Richtungen zu betrachten, zu einer stabilitätssteigernden Eigenschaft. Beispielhaft sind derartige Merkmale (gleiche Position + mehrere Orientierungen) an Ecken zu finden.

Die nachfolgend dargestellte Abbildung 2.4 veranschaulicht ein Orientierungshistogramm. Die x-Achse kennzeichnet die Richtung in Grad und besitzt eine Klassenbreite von 10° . Die y-Achse stellt die akkumulierten Gradientenlängen dar. Die größte Gradientenlänge bestimmt die Richtung des Merkmales. Des Weiteren wurden zwei zusätzliche Richtungen ermittelt und dem Merkmal zugeordnet.

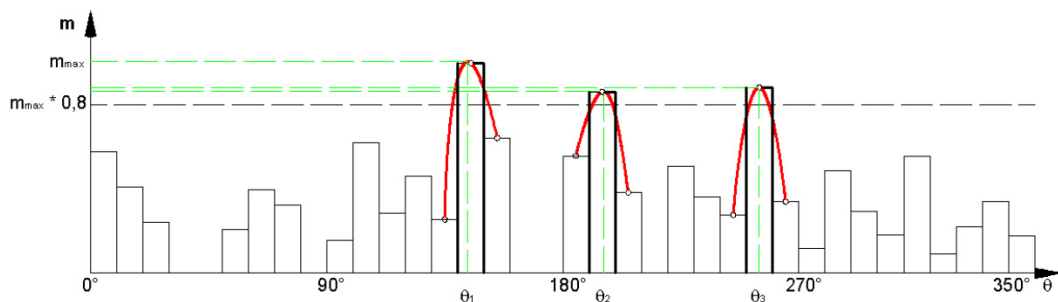


Abbildung 2.4: Richtungsbestimmung eines Merkmales mit einem Orientierungshistogramm.⁸

Diese eindeutige Hauptorientierung eines Merkmales ist für die Rotationsinvarianz von Nöten.

2.4 Erzeugung der Merkmalsvektoren

Ein Merkmal wird durch seine Bildposition (x, y) , seiner Hauptorientierung und einem Merkmalsvektor gekennzeichnet.

Ein 128-dimensionaler Merkmalsvektor ist die dritte und letzte fehlende Information, zur vollständigen Beschreibung eines robusten Merkmales. Anhand der Hauptorientierung, welche die Merkmalsregion definiert, beschreibt der Merkmalsvektor diese Umgebung. Er beinhaltet somit Erkenntnisse über sein Umfeld. Berechnet wird dieser relativ zur Hauptorientierung. Wird ein Bild beispielhaft rotiert, haben übereinstimmende Merkmale den gleichen Merkmalsvektor aber eine unterschiedliche Hauptorientierung. Nachfolgend ist in Abbildung 2.5 die Hauptorientierung eines Merkmales mit seinen 16×16 Umgebungspixeln dargestellt. Aufgrund der Rotation müssen gegebenenfalls Pixel, welche sich nach der Drehung zwischen zwei oder mehreren Pixelfeldern befinden, interpoliert werden.

⁸Bildquelle: RONCAT (2005)

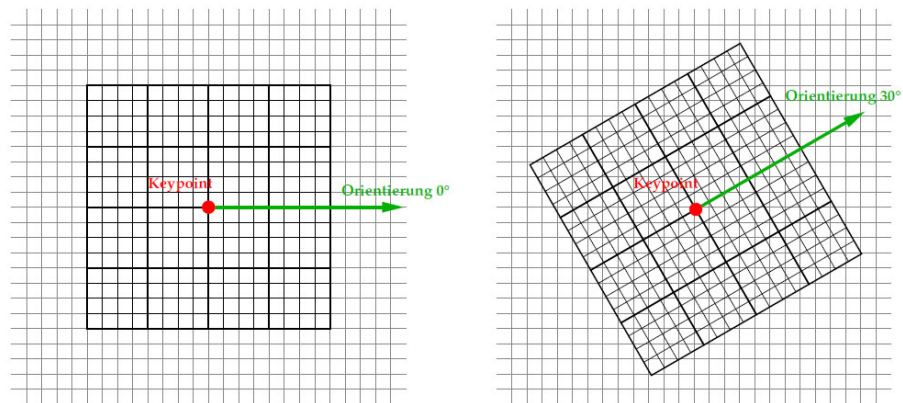


Abbildung 2.5: Rotation einer 16x16 Pixel großen Merkmalsumgebung anhand seiner Hauptorientierung.⁹

Der Merkmalsvektor dient somit zum Vergleich von Merkmalen und bringt die Invarianz bezüglich Helligkeits- und Kontrastveränderungen mit sich. Er wird auch als sogenannter „Deskriptor“ bezeichnet.

LOWE¹⁰ verwendet einen $4 \times 4 \times 8 = 128$ -dimensionalen Deskriptor. Der Deskriptor setzt sich aus $4 \times 4 = 16$ Orientierungshistogrammen, welche jeweils eine gleichgroße Fläche abdecken, zusammen. Jedes Histogramm wird ähnlich wie in Schritt 3 „Bestimmung der Hauptorientierungen“ berechnet. Die Klassenbreite beträgt in diesem Fall 45° . Damit hat ein Histogramm $360^\circ / 45^\circ = 8$ Winkelbereiche, welche jeweils 45° seiner Subregion abdecken.

Die Abbildung 2.6 veranschaulicht einen $2 \times 2 \times 8$ -dimensionalen Deskriptor. Im linken Bild der Abbildung sind die ermittelten Gradienten dargestellt. Diese beschreiben das Umfeld des Merkmals. Der blaue Kreis kennzeichnet dabei die Gewichtung der Gradienten. Die rechte Seite zeigt einen aus $2 \times 2 = 4$ Histogrammen bestehenden Deskriptor, wobei jedes Histogramm acht Richtungsbereiche beinhaltet. Die acht Richtungspfeile stellen anhand ihrer Länge und Orientierung die Stärke und den Richtungsverlauf in den jeweiligen 45° -Bereich dar.

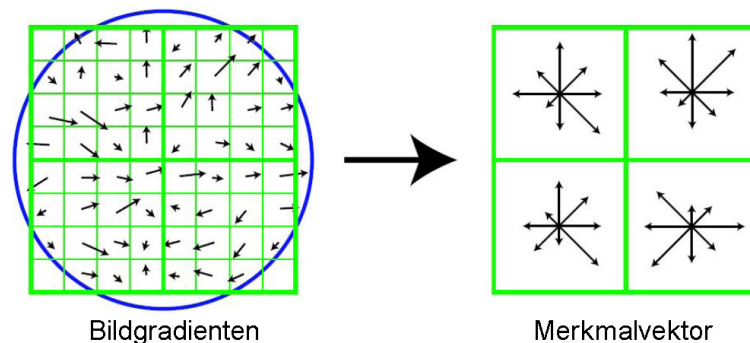


Abbildung 2.6: $2 \times 2 \times 8$ -dimensionaler Deskriptor eines Merkmals. Dieser beschreibt seine 8×8 Pixelumgebung, welche in vier Subregionen mit jeweils 4×4 Pixeln aufgeteilt wurde.¹¹

⁹Bildquelle: ZIMMERMANN (2007)

¹⁰In LOWE (2004) und LOWE (1999) beschrieben.

¹¹Bildquelle: BARSKI (2007)

In der nachfolgenden Abbildung 2.7 ist ein weiterer 2x2x8-dimensionaler Deskriptor abgebildet. Dieser Deskriptor dient lediglich zur Veranschaulichung und ist aufgrund dessen eher unwahrscheinlich in der Praxis anzutreffen. Der abstrakte Deskriptor stellt anschaulich den Übergang zwischen den Gradienten der Merkmalsumgebung (links) und den vier Orientierungshistogrammen (rechts) dar. Die Umgebung des Merkmals ist dabei in vier Subregionen eingeteilt. Aus jeder Subregion ist ein Orientierungshistogramm mit jeweils acht Richtungseinträgen zu bestimmen. In der sich oben links befindenden ersten Subregion, zeigen viele große Gradienten nach oben rechts. Das aus diesen Gradienten berechnete Orientierungshistogramm, stellt anhand der stark nach oben rechts zeigenden Richtungspfeile, den Zusammenhang dar.

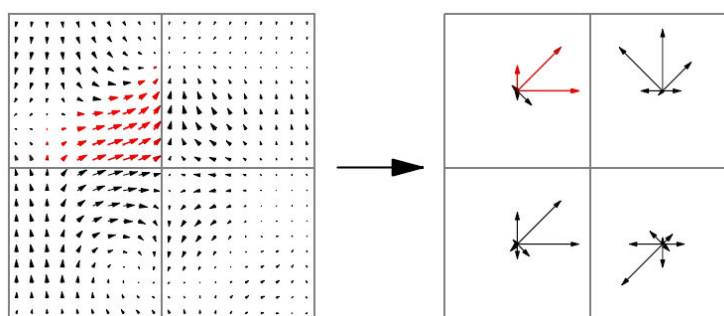


Abbildung 2.7: Abstrakter 2x2x8-dimensionaler Deskriptor¹²

Um eine Invarianz bezüglich Beleuchtungsveränderungen zu ermöglichen, wird der Gradientenbetrag des Deskriptors, indem er durch seine Länge dividiert wird, auf Einheitslänge normiert. Dies erfolgt, da eine Kontrastveränderung eine gleichmäßige Erhöhung aller Gradientenbeträge zur Folge hat, welche durch die Normalisierung aufgehoben wird.

¹²Bildquelle: SCHATZ (2006)

Der gesamte Ablauf der Merkmalsextraktion ist in Abbildung 2.8 zusammenfassend dargestellt. Ein Eingangsbild ist auf Merkmale zu analysieren. (1a) Dazu wird das Eingangsbild fortlaufend geglättet. Es entsteht eine Gauß-Pyramide. Das letzte gaußgeglättete Bild wird anschließend um 50 % runterskaliert und bildet das Eingangsbild für die nächste Gauß-Pyramide. Dieser Vorgang wird n -mal fortgeführt. (1b) Jeweils zwei benachbarte Bilder aus der Gauß-Pyramide werden miteinander subtrahiert. Die subtrahierten Bilder bauen DoG-Pyramiden auf. (1c) Jeder Punkt in der DoG-Pyramide wird auf einen Extrempunkt geprüft. (2) Anschließend wird jedes gefundene Maxima einer Stabilitätsanalyse unterzogen. (3) Anhand der Umgebungsgradienten eines Merkmalspunktes wird die Hauptorientierung des Merkmales bestimmt. (4) Deskriptoren beschreiben die Merkmalsumgebung und sind im letzten Schritt zu ermitteln.

Ein Objekt wird somit durch eine Menge von SIFT-Merkmalen beschrieben. Anhand dieser robusten Merkmale kann ein Gegenstand in einem rotierten, verschobenen, skalierten, teilweise affine verzerrten und unter einem anderen Lichtverhältnis aufgenommen Bild, wiedererkannt werden. Dies bildet die Grundlage zur 3D-Objekt- und Lageerkennung von Gegenständen.

Die Identifikation von Objekten auf Grundlage extrahierter Merkmale, wird im nachfolgenden Kapitel 3 beschrieben.

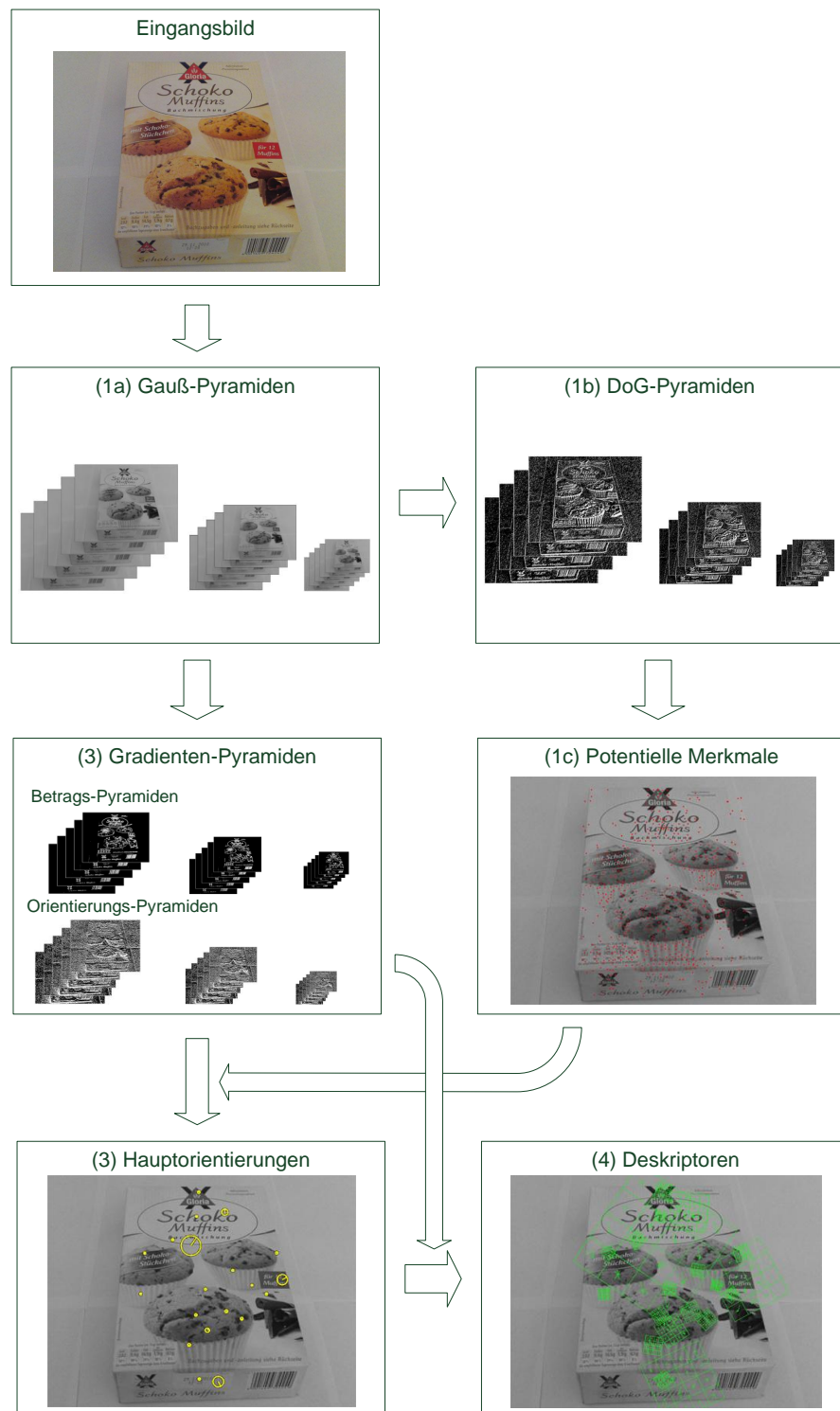


Abbildung 2.8: Ablauf einer Merkmalsextraktion.

3 Merkmalsvergleich zur Identifikation von Gegenständen

Inhalt

3.1	Merkmalsvergleich auf Basis des euklidischen Abstandes	11
3.2	Verbesserte Korrespondenzsuche mit einem Suchbaum	11
3.2.1	Aufbau eines k-d Baumes	12
3.2.2	Suche im k-d Baum	13

Die Merkmale der Gegenstände müssen bekannt sein, um Gegenstände in Bildern wiedererkennen zu können. Dazu sind robuste Merkmale der Gegenstände mit dem in Kapitel 2 beschriebenen Verfahren zu ermitteln. Anschließend kann der gleiche Gegenstand in einer beispielsweise rotierten Ansicht auf Übereinstimmung untersucht werden. Dafür sind die Merkmalsvektoren aus der einen Ansicht, mit den Merkmalsvektoren aus der anderen Ansicht, zu vergleichen.

3.1 Merkmalsvergleich auf Basis des euklidischen Abstandes

Beim Vergleich kann der nächste Nachbar zu einem Merkmalspunkt mit Hilfe des euklidischen Abstandes gesucht werden. Der euklidische Abstand zweier Vektoren V_1 und V_2 wird mit der Formel 3.1 berechnet. N kennzeichnet die Länge der Vektoren.

$$\sqrt{\sum_{i=1}^N (V_{1_i} - V_{2_i})^2} \tag{3.1}$$

Beispielsweise existieren Deskriptoren D_{1_i} aus der initialen Gegenstandsaufnahme und Deskriptoren D_{2_i} aus der aktuellen Aufnahme. Somit ist der korrespondierende Deskriptor aus der Menge D_2 derjenige, welcher die geringste euklidische Distanz zum Deskriptor D_{1_i} aufweist. Dieses Vorgehen wird als „Brute Force“-Ansatz bezeichnet.

3.2 Verbesserte Korrespondenzsuche mit einem Suchbaum

Ein Vergleich mit allen Merkmalsvektoren ist aufgrund der hohen Vektordimension sehr zeitintensiv. Eine bessere Lösung ist der Aufbau eines sogenannten „k-d Baumes“. Dieser, zur Verwaltung von Merkmalspunkten k-dimensionale Suchbaum, erleichtert das Finden vom nächsten Nachbarn.

3.2.1 Aufbau eines k-d Baumes

Der k-d Baum ist ein binärer Suchbaum der Tiefe t , welcher einen Raum in 2^t Unterräume aufteilt. Jeder Knoten gliedert jeweils zwei Unterräume auf seine zwei Nachfolgerknoten auf. Die Blätter des Baumes repräsentieren die Merkmalspunkte mit ihren Deskriptoren. Mit n Merkmalsvektoren ergibt sich daraus eine Maximaltiefe von $\log_2 n$.

Für den Aufbau eines k-d Baumes sind die höheren Baumebenen priorisiert zu gewichten. Dafür sind die Dimensionen der Merkmalsvektoren anhand ihrer Varianz zu sortieren. In Gleichung 3.2 ist die Berechnung der Varianz nachfolgend dargestellt. \bar{v} kennzeichnet den Deskriptormittelwert.

$$\frac{1}{N} \sum_{i=1}^N (v_i - \bar{v})^2 \quad (3.2)$$

Nachdem die Vektordimensionen anhand ihrer Varianz sortiert wurden, müssen die Zweigbedingungen für den Wurzelknoten bestimmt werden. Dazu wird der Median über die höchstpriorisierteste Dimension aller Deskriptoren gebildet. Die priorisierteste Dimension ist das erste Element aus der sortierten Menge. Der Median einer geordneten Menge \tilde{x} ist folgendermaßen zu berechnen:

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & n \text{ ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n+1}{2}}) & n \text{ gerade} \end{cases} \quad (3.3)$$

Es liegen beispielhaft drei zweidimensionale Punkte $P_1(0,2)$, $P_2(1,4)$ und $P_3(2,8)$ vor. Dann ist der Median der zweiten Dimension 4. Die Deskriptoren sind anschließend, anhand der in der zweiten Dimension befindenden Werte, in den Baum einzuordnen. In dem Fall, dass der Wert echt kleiner als der Median ist, wird der Deskriptor in den linken Ast des Wurzelknotens ein-geordnet. Ansonsten ist der Deskriptor in den rechten Ast einzusortieren. Anschließend ist der Median über alle Deskriptoren eines Unterraumes in Bezug auf die zweithöchst-priorisierteste Dimension zu bestimmen. Das Verfahren läuft iterativ wie bisher beschrieben fort. Der Median bewirkt einen ausgeglichen verteilten Suchbaum.

Die Abbildung 3.1 veranschaulicht einen einsortierten k-d Baum im zweidimensionalen Fall. Der k-d Baum wurde mit den Vektoren $v_1(0,1)$, $v_2(-1,3)$, $v_3(4,0)$, $v_4(2,2)$ und $v_5(3,1)$ aufgebaut. Die erste Dimension weißt die größte Varianz auf. Aufgrund dessen wurde der Median über die erste Dimension aller fünf Vektoren gebildet. Aus den Vektorwerten der ersten Dimension $\{0, -1, 2, 3, 4\}$ wurde der Median 2 ermittelt. Dieser Wert ist die Entscheidungsbedingung für den Wurzelzweig. Bei den Vektoren $v_1(0,1)$ und $v_2(-1,3)$ ist der erste Dimensionswert echt kleiner als 2. Das hat zur Folge, dass diese beiden Vektoren in den linken Ast einzusortieren sind. Die Vektoren $v_3(4,0)$, $v_4(2,2)$ und $v_5(3,1)$ müssen dementsprechend in den rechten Ast einsortiert werden. Anschließend wird der Baum nach dem gleichen Verfahren weiter aufgeteilt. Der Median für die zwei neuen Knoten ist anhand der zweiten Dimension zu berechnen. Für den linken Knoten ergibt sich somit ein Median von $(1+3)/2 = 2$. Der rechte Knoten besitzt einen Median von 1. Die Vektoren sind dementsprechend weiter nach unten in den Baum einzusortieren.

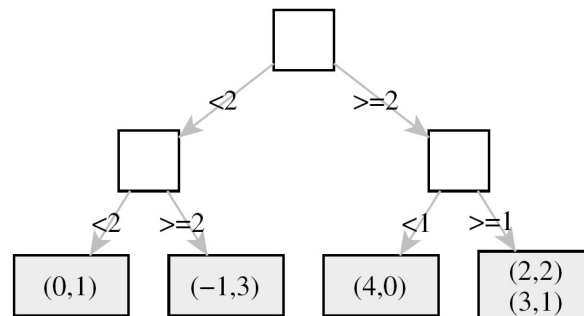


Abbildung 3.1: Aufbau eines k-d Baumes mit einer zweidimensionalen Punktmenge.¹³

3.2.2 Suche im k-d Baum

Beispielhaft soll der nächste Nachbar zum Vektor $v_x(-1,2)$, in dem im Unterkapitel 3.2.1 aufgebauten k-d Baum gefunden werden. Die Zweigbedingungen sind jeweils an den Ästen vermerkt. Die Blätter sind Deskriptoren und somit die potentiellen nächsten Nachbarn des Vektors v_x .

Nachfolgend ist die Suche des nächsten Nachbarn zum Vektor v_x anhand der Abbildung 3.2 beschrieben.

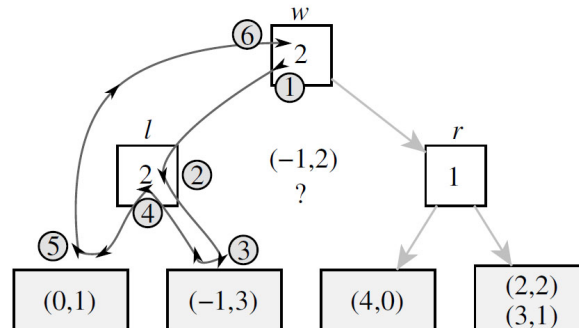


Abbildung 3.2: Suchablauf in einem k-d Baum.¹⁴

1. Auf der Ebene der Wurzel w ist aufgrund der Priorisierung die erste Dimension des Vektors v_x zu betrachten. Anhand des Wertes der ersten Dimension wird entschieden, ob nach links oder rechts im Baum traversiert wird. -1 ist echt kleiner als der Zweigbedingungswert 2 . Aufgrund dessen ist nach links im Baum zu traversieren. In einer sogenannten „Backtracking-Warteschlange Q “ ist der quadratische Abstand $(2 - (-1))^2 = 9$ zwischen der ersten Dimension zur Wurzel zu vermerken. Dieser kennzeichnet den Mindestabstand, den alle Deskriptoren des rechten Unterbaumes zum Vektor v_x besitzen. Die Warteschlange enthält somit einen Eintrag: $Q[(w, 9)]$
2. Am Nachfolgerknoten l ist im Anschluss die zweite Dimension des Vektors v_x zu betrachten. Die zweite Dimension des Vektors v_x enthält den Wert 2 . Da 2 größer gleich

¹³Bildquelle: HÄRTEL (2006)

¹⁴Bildquelle: HÄRTEL (2006)

der Zweigbedingung des Knotens l ist, wird nach rechts traversiert. Der Mindestabstand des Knotens l ist in die Warteschlange mit dem Wert $(2 - 2)^2 = 0$ einzutragen. 0 kennzeichnet den Mindestabstand aller Vektoren im linken Blatt des Knotens l . Dieser wird in der Warteschlange vermerkt: $Q[(l, 0), (w, 9)]$

3. Im Anschluss ist der Vektor v_x mit allen an dem Blatt befindenden Vektoren zu vergleichen. $(-1, 3)$ ist der einzige Vektor an diesem Ort. Damit ist dieser der beste nächste Nachbar des Vektors v_x . Der quadratische Abstand zwischen diesen beiden Vektoren beträgt 1. Der quadratische Abstand zwischen zwei Vektoren wird ermittelt, indem die beiden quadratischen Abstände der zwei Dimensionen summiert werden.
4. Der sogenannte „Backtracking“-Schritt kommt an dieser Stelle zum Einsatz. Dazu ist der Knoten mit dem kleinsten Mindestabstand aus der Backtracking-Warteschlange zu entnehmen. In dem Fall hat l den geringsten Mindestabstand. In diesem Teilbaum wird der alternative Ast betrachtet.
5. An diesem alternativen Ast werden erneut alle Vektoren des Blattes mit dem Vektor v_x verglichen. Der Vektor $(0, 1)$ hat dabei einen quadratischen Abstand von 2. Da der Abstand nicht besser geworden ist, bleibt der nächste Nachbar mit dem Abstand von 1, der Vektor $(-1, 3)$.
6. Im letzten Schritt wird noch geprüft, ob in der Warteschlange ein weiterer Knoten existiert, welcher einen kleineren Abstand als der bisher ermittelte kleinste Abstand von 1 besitzt. Der Abstand 9 von w ist größer als der Abstand 1 von $(-1, 3)$. Damit bleibt der Vektor $(-1, 3)$ der nächste Nachbar und somit der am besten übereinstimmende Vektor zu v_x .

LOWE¹⁵ verwendet zur effizienteren Suche im k-d Baum diesen verbesserten Suchalgorithmus mit „Backtracking“. Diese Methode wird auch als „Best-Bin-First (BFF)“ bezeichnet. Dabei ist in diesem modifizierten k-d Baum eine 95% Suchübereinstimmung bei verbesserter Geschwindigkeit erzielbar.

Es besteht die Möglichkeit, die Dimension eines Merkmalsvektors zu reduzieren. Die sogenannte „Hauptkomponentenanalyse“ (engl. Principal Component Analysis, kurz PCA)¹⁶ und Diskriminanzanalyse (engl. Linear Discriminant Analysis, kurz LDA)¹⁷ sind Verfahren, welche eine Dimensionsreduzierung realisieren können. In KRZANOWSKI (2000) und FAHRMEIR U. A. (1996) sind die Hauptkomponenten- und Diskriminanzanalyse ausführlich beschrieben und gegebenenfalls zu entnehmen.

Für weitere Informationen bezüglich der Erzeugung und Suche in k-d Bäumen ist auf die Literatur HEYMANN (2005) und BEIS UND LOWE (1997) verwiesen.

¹⁵Zu finden unter BEIS UND LOWE (1997) und LOWE (2004)

¹⁶Die Hauptkomponentenanalyse wurde in KRZANOWSKI (2000) detailliert beschrieben.

¹⁷Die Diskriminanzanalyse ist in FAHRMEIR U. A. (1996) zu finden.

Die nachfolgende Abbildung 3.3 veranschaulicht das Prinzip der Objektidentifikation auf Basis robuster Merkmale. In einem initialen Prozess sind robuste Merkmale eines Gegenstandes einmalig zu erfassen und für einen späteren Vergleich zwischenspeichern. Eine andere Aufnahme des gleichen Gegenstandes wird in einem fortlaufenden Prozess auf robuste Merkmale untersucht. Befindet sich beispielhaft der Gegenstand in der zweiten Aufnahme an einer anderen Position, kann dieser aufgrund seiner Merkmalseigenschaften durch einen Merkmalsvergleich auf Übereinstimmung geprüft werden. Eine erfolgreiche Objektidentifikation ist in der Abbildung 3.4 dargestellt. In dieser sind zur Veranschaulichung acht von insgesamt 851 richtigen Merkmalsübereinstimmungen abgebildet.

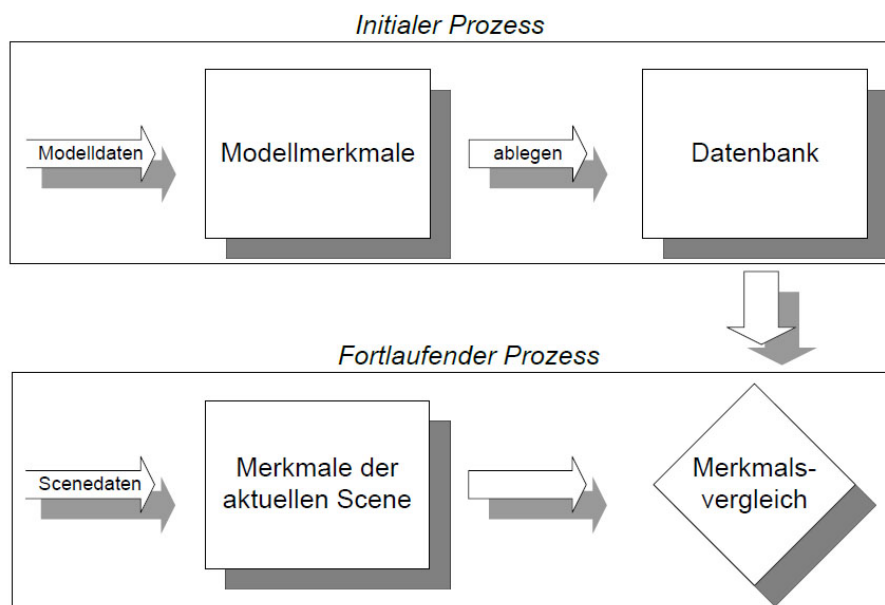


Abbildung 3.3: Ablauf einer Objektidentifikation auf Basis von SIFT-Merkmalen.



Abbildung 3.4: Erfolgreiche Merkmalsübereinstimmung.

4 Zusammenfassung und Ausblick

In dieser Arbeit wurde die Wiedererkennung von Gegenständen auf Basis von Kamerabildern untersucht. Eine am Katana-Roboterarm befestigte Kamera liefert unterschiedliche Aufnahmen von Gegenständen. Die Kamera befindet sich am Winkelgreifer und überträgt Grauwertbilder zu einem nah positionierten Computer. Jede Aufnahme enthält robuste Merkmale, welche einen Gegenstand eindeutig identifizieren können.

Die Wiedererkennung von Gegenständen in Kamerabildern soll schnell und exakt erfolgen. Aufgrund dessen kommt zur Wiedererkennung das SIFT-Verfahren zum Einsatz. Die Arbeitsweise vom SIFT-Algorithmus gliedert sich in vier Schritte. Im ersten Schritt sind anhand Extrema im Skalenraum potentielle skalierungsinvariante Merkmale zu ermitteln. Anschließend erfolgt durch eine Stabilitätsuntersuchung eine Filterung der Merkmalskandidaten. Für alle verbliebenen Merkmalskandidaten wird dabei die exakte Position bestimmt. Im dritten Schritt ist zu jedem potentiellen Merkmalspunkt eine Hauptorientierung zu berechnen. Eine Hauptorientierung ist für die Rotationsinvarianz von Merkmalen von Nöten. Zur vollständigen Beschreibung eines Merkmales wird im letzten Schritt ein 128-dimensionaler Merkmalsvektor ermittelt. Dieser dient zum Vergleich von Merkmalen.

Für den Merkmalsvergleich wurden zwei Verfahren vorgestellt. Zum einen kann ein Vergleich auf Basis des euklidischen Abstandes erfolgen. Zum anderen erzielt ein Suchbaum eine verbesserte Korrespondenzsuche.

Das SIFT-Verfahren ermöglicht somit die Suche von Merkmalen, welche eine Invarianz gegenüber Rotation, Translation, Skalierung, Änderung der Lichtverhältnisse und teilweise affiner Verzerrung aufweisen. Diese robusten Merkmale ermöglichen die Wiedererkennung von Gegenständen in unterschiedlichen Kameraaufnahmen.

Auf der Basis dieser Ausarbeitung kann ein autonomer Roboterarm zur Identifizierung, Lokalisierung und Manipulation von Gegenständen realisiert werden.

Abbildungsverzeichnis

1.1	Katana-Greifarm mit sechs Freiheitsgraden	1
2.1	Ermittlung der Extrema im Skalenraum	3
2.2	Potentieller Merkmalspixel wird mit seinen 26 Nachbarn verglichen	4
2.3	Filterung von Merkmalen, welche sich auf Kanten befinden	5
2.4	Richtungsbestimmung eines Merkmales mit einem Orientierungshistogramm	6
2.5	Rotation einer 16x16 Pixel großen Merkmalsumgebung anhand seiner Hauptorientierung	7
2.6	2x2x8-dimensionaler Deskriptor eines Merkmales	7
2.7	Abstrakter 2x2x8-dimensionaler Deskriptor	8
2.8	Ablauf einer Merkmalsextraktion	10
3.1	Aufbau eines k-d Baumes mit einer zweidimensionalen Punktmenge	13
3.2	Suchablauf in einem k-d Baum	13
3.3	Ablauf einer Objektidentifikation auf Basis von SIFT-Merkmalen	15
3.4	Erfolgreiche Merkmalsübereinstimmung	15

Quellenverzeichnis

Literaturquellen

- [Barski 2007] BARSKI, Anatoli: *Objekterkennung in 2D-Bildern und Tiefenbildern*. Universität Karlsruhe, 2007
- [Beis und Lowe 1997] BEIS, Jeffrey S. ; LOWE, David G.: *Shape indexing using approximate nearest-neighbour search in high-dimensional spaces*. Conference on Computer Vision and Pattern Recognition, 1997
- [Brown und Lowe 2002] BROWN, Matthew ; LOWE, David G.: *Invariant features from interest point groups*. British Machine Vision Conference, 2002
- [Fahrmeir u. a. 1996] FAHRMEIR, Ludwig ; HAMERLE, Alfred ; TUTZ, Gerhard: *Multivariate statistische Verfahren*. 1996. – ISBN 3-11-013806-9
- [Helms und Meyer 2005] HELMS, E. ; MEYER, C.: *Assistor - Mensch und Roboter rücken zusammen: Einsatz von sicheren und einfach bedienbaren Assistenzrobotern im Produktionsalltag*. Fraunhofer-Institut für Produktionstechnik und Automatisierung : Werkstattstechnik online, 2005 ISSN 1436-4980
- [Heymann 2005] HEYMAN, Sebastian: *Implementierung und Evaluierung von Video Feature Tracking auf moderner Grafikhardware*. Diplomarbeit - Bauhaus Universität Weimar, 2005
- [Härtel 2006] HÄRTEL, Volker: *Effiziente Repräsentation von hochdimensionalen Merkmalsvektoren für die Korrespondenzfindung*. Diplomarbeit - Christian Albrechts Universität zu Kiel, 2006
- [Ichbiah 2005] ICHBIAH, Daniel: *Roboter: Geschichte - Technik - Entwicklung*. Knesebeck, 2005. – ISBN 3-896-60276-4
- [Krzanowski 2000] KRZANOWSKI, Wojtek J.: *Principles of Multivariate Analysis: A User's Perspective*. Oxford University Press, 2000. – ISBN 978-0198507086
- [Lindeberg 1994] LINDBERG, Tony: *Scale-space theory: A basic tool for analysing structures at different scales*. Journal of Applied Statistics, 1994
- [Lowe 1999] LOWE, David G.: *Object Recognition from Local Scale-Invariant Features*. International Conference on Computer Vision, 1999
- [Lowe 2001] LOWE, David G.: *Local Feature View Clustering for 3D Object Recognition*. International Conference on Computer Vision, 2001
- [Lowe 2004] LOWE, David G.: *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision, 2004

- [Mikolajczyk 2002] MIKOLAJCZYK, Krystian: *Detection of local features invariant to affine transformations*. Institut National Polytechnique de Grenoble, 2002
- [Roncat 2005] RONCAT, Andreas: *Nahbereichsphotogrammetrie*. Diplomarbeit - Bauhaus Universität Weimar, 2005
- [Schatz 2006] SCHATZ, Alexej: *Visuelle Navigation mit Scale Invariant Feature Transform*. Diplomarbeit - Universität Bielefeld, 2006
- [Zimmermann 2007] ZIMMERMANN, Fabian: *Keypoint-Detektion und Deskriptoren-Berechnung auf der Grafikkarte*. Diplomarbeit - Technische Universität Kaiserslautern, 2007

Internetquellen

- [AG 2009] AG, Neuronics: *Neuronics AG - intelligent & personal robotics*. Stand: 27.10.2009. 2009. –
URL: [HTTP://WWW.NEURONICS.CH/](http://www.neuronics.ch/)
- [IPA 2009] IPA, Fraunhofer: *Care-O-bot 3 Serviceroboter*. Fraunhofer-Institut für Produktionstechnik und Automatisierung - Stand: 29.10.2009. 2009. –
URL: [HTTP://WWW.CARE-O-BOT.DE/](http://www.care-o-bot.de/)