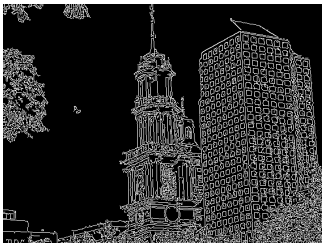
17. Image Features – Interest Points

Bildmerkmale

Bildmerkmale (engl. *image features*) sind mathematische Beschreibungen eines gesamten Bildes oder von Bildteilen, die eine leichter unterscheidbare Repräsentation als reine Pixelwerte liefern und somit helfen, relevante Information für gewisse Aufgaben zu extrahieren. In diesem Sinne können Bildmerkmale **global** oder **lokal** sein: globale Merkmale beschreiben das gesamte Bild (z.B. ein Grauwerthistogramm), sind aber im Allgemeinen nicht gut unterscheidbar und nicht robust genug, um sie in komplexen Anwendungen zu verwenden.
Lokale Merkmale beschreiben kleine Bildregionen und werden für Anwendungen wie Bildregistrierung, Panorama Stitching, 3D Modellierung oder Objekterkennung verwendet. Ihr Ziel ist es, eine mathematische Beschreibung „interessanter" Bildbereiche und deren lokaler Nachbarschaft zur Verfügung zu stellen. Dafür müssen die folgenden beiden grundsätzlichen Schritte durchgeführt werden:

1. **Merkmalsdetektion**: Zuerst werden Bildbereiche identifiziert, die einen hohen visuellen Informationsgrad aufweisen und deren räumliche Anordnung möglichst eindeutig bestimmt werden kann. Auch die räumliche Ausdehnung oder *Skalierung* (*Skala*, siehe auch Kap.16) des Merkmals wird in diesem Schritt bestimmt. Die erkannten Regionen werden oft als **Interest Points** oder **Keypoints** bezeichnet.
2. **Merkmalsbeschreibung:** Es wird ein Merkmalsvektor berechnet, der die lokale Bildstruktur beschreibt. Die Menge aller möglichen Merkmalsvektoren bildet einen Merkmalsraum (engl. *feature space*).

Kantenerkennung



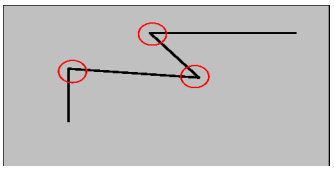
Wie schon in Kap. 13 ausgeführt, enthalten Kanten die nicht redundante Information in Bildern. Kantendetektion hat das Ziel, Punkte in digitalen Bildern zu erkennen, an denen starke Helligkeitsveränderungen auftreten oder an denen sich, formaler ausgedrückt, Unstetigkeiten befinden.
Nach der Anwendung eines Kantendetektors auf ein Bild erhält man als Ergebnis eine Menge von Pixeln an denen sich im Ausgangsbild Unstetigkeitsstellen befinden. Daher reduziert ein Kantendetektors die Menge an zu verarbeitenden Daten erheblich, da Informationen von niedrigerer Relevanz herausfiltert werden, während wichtige strukturelle Eigenschaften des Bildes erhalten bleiben – die hohen Bildfrequenzen. Allerdings sind Kanten zur Erfassung der relevanten Information nur beschränkt geeignet, da sie weder rotations- noch skalierungsinvariant sind, gleiche Merkmale in verschiedenen Bildern können somit nur schwer lokalisiert werden.

Interest Points

Eckpunkte sind nicht nur für uns Menschen auffällig, sondern sind auch aus technischer Sicht robuste Merkmale, die in 3D-Szenen nicht zufällig entstehen und von verschiedenen Blickwinkeln sowie unter verschiedenen Beleuchtungsbedingungen zuverlässig zu lokalisieren sind. Ein guter "Corner Detector" muss mehrere Kriterien erfüllen: Er soll Eckpunkte zuverlässig auch unter realistischem Bildrauschen finden, er soll die gefundenen Eckpunkte möglichst genau lokalisieren können und zudem effizient arbeiten. Wie immer gibt es nicht nur einen Ansatz für diese Aufgabe, aber im Prinzip basieren die meisten Verfahren zum Auffinden von Eckpunkten oder Interest Points im weiteren Sinn auf einer gemeinsamen Grundlage - während eine Kante in der Regel definiert wird als eine Stelle im Bild, an der der Gradient der Bildfunktion in einer bestimmten Richtung besonders hoch und orthogonal dazu besonders niedrig ist, weist ein Eckpunkt einen starken Helligkeitsunterschied in mehr als einer Richtung gleichzeitig auf. Allgemein ist ein Interest Point ein Punkt in einem Bild, der folgendermaßen charakterisiert werden kann:

1. Er hat eine mathematisch eindeutige Definition
2. Er hat eine klar definierte Position im Bildraum
3. Die ihn umgebende lokale Bildstruktur hat einen hohen Informationsgehalt
4. Er ist gegenüber lokalen und globalen Störungen des Bildes stabil - inklusive Deformationen, die durch perspektivische Transformationen oder Variationen in der Beleuchtung/Helligkeit entstehen.
5. Er sollte skalierungsinvariant sein.

Eckendetektion

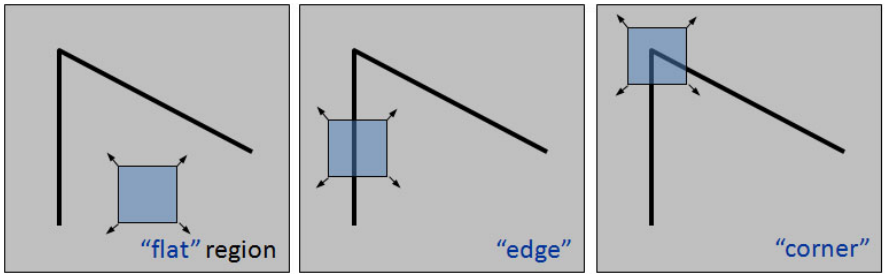


Die Eckendetektion ist eine Form der Interest Point Erkennung, bei der Ecken als jene Punkte im Bild bezeichnet werden, die für eine gegebene Aufgabenstellung "von Interesse" sind. In der Praxis sind die meisten Eckendetektoren nicht nur speziell empfindlich gegenüber Ecken, sondern auch gegenüber lokalen Bildregionen, die einen hohen Grad an Variation in alle Richtungen aufweisen. Ein Eckpunkt ist dort gegeben, wo der Gradient des Bildes in mehr als einer Richtung einen hohen Wert aufweist. Insbesondere sollen Stellen entlang von Kanten, wo der Gradient zwar hoch, aber nur in einer Richtung ausgeprägt ist, nicht als Eckpunkte gelten. Darüber hinaus sollen Eckpunkte natürlich unabhängig von ihrer Orientierung, d.h. in isotroper Weise, gefunden werden. Ein Eckpunkt kann somit als Schnittpunkt zweier Kanten definiert werden oder auch als Punkt, in dessen lokaler Nachbarschaft zwei verschiedene und dominante Kantenrichtungen existieren. Das bedeutet, dass ein Eckpunkt beispielsweise auch der Endpunkt einer Linie oder ein Punkt auf einer Kurve, an dem die Krümmung maximal ist, sein kann. Die Qualität eines Eckendetektors wird durch die Fähigkeit bestimmt, dieselbe Ecke in Bildern unter verschiedenen Lichtverhältnissen und geometrischen Transformationen wie Translation oder Rotation wieder zu erkennen.

Moravec-Eckendetektor

Dies ist einer der frühesten Eckendetekoren und definiert eine Ecke als einen Punkt mit geringer Eigenähnlichkeit. *Hans P. Moravec* führte das Konzept der "Points of Interest" ein: Sein Operator verwendet ein lokales Fenster in einem Bild und bestimmt die Veränderung der Intensitätsunterschiede, indem das Fenster in kleinem Ausmaß in alle vier Richtungen verschoben wird:

$$E(u,v) = \sum_{x,y} w(x,y)\left[I(x+u,y+v) - I(x,y)\right]^2$$



$w(x,y)$ ist ein Fensterfunktion, die das für den Vergleich verwendete quadratische Fenster definiert, $I(x+u,y+v)$ ist die Intensität an der verschobenen Stelle und $I(x,y)$ die ursprüngliche Intensität. Diese Operation wird für jede Pixelposition wiederholt. Jeder Position wird ein *Interest Value* zugewiesen, welcher der minimalsten Veränderung, die durch diese Verschiebungen produziert wird, entspricht. Sind alle Veränderungen gering, befindet sich das Fenster auf einer einfärbigen (*flat*) Region. Sind nur Veränderungen in einzelne Richtungen hoch, handelt es sich um einen Kantenbereich. Sind Veränderungen in alle Richtungen hoch, bleibt auch das Minimum aller Veränderungen hoch und es handelt sich einem Eckpunkt. Ein Nachteil des Moravec-Eckendetektors liegt jedoch in seiner anisotropischen Antwort: Wenn eine Kante vorhanden ist, die nicht in der Richtung der vier Nachbarn verläuft, dann wird diese auch einen vergleichsweise hohen Interest Value aufweisen. Weiters ist der Operator empfindlich gegenüber Bildrauschen, das entlang der Kanten auftritt, da er nur die minimalen Intensitätsveränderungen für jede Pixelposition in Betracht zieht (im Gegensatz zu der Variation zwischen diesen).

Harris Eckendetektor

Harris und *Stephens* verbesserten den Moravec-Kantendetektor dahingehend, dass zur Berechnung nicht die absoluten Pixeldifferenzen von verschobenen Bildausschnitten zum Einsatz kommen, sondern die Variation der lokalen Bildstruktur betrachtet wird. Grundlage des Harris-Detektors sind die Gradienten des Bildes, an einem Eckpunkt sollten Gradienten sowohl in der Hauptrichtung als auch normal dazu vorhanden sein. Das Ergebnis ist ein weitaus verlässlicherer Detektor in Bezug auf Detektion und Wiedererkennung, jedoch zum Preis signifikant höherer Berechnungskosten. Trotz der hohen rechnerischen Anforderung wird dieser Algorithmus weitgehend in der Praxis

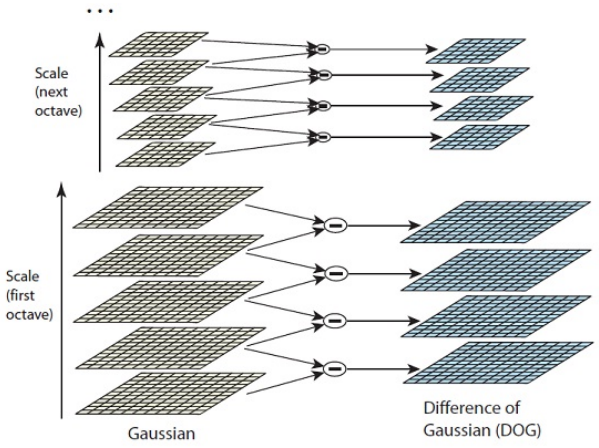
verwendet. Der Harris-Operator hat im Gegensatz zum Moravec-Operator eine isotrope Antwort und ist somit rotationsinvariant. Sowohl der Moravec- als auch der Harris-Detektor sind aber nicht skalierungsinvariant.

Scale-Invariant Feature Transform (SIFT)

Scale-Invariant Feature Transform (SIFT) ist ein Algorithmus, der in der Computer Vision zur Detektion und Beschreibung lokaler Merkmale in Bildern Verwendung findet. Dieser Algorithmus wurde von *David Lowe* im Jahr 1999 veröffentlicht und in den Vereinigten Staaten patentiert. Die Methode ist invariant gegenüber der Skalierung und Rotation des Bildes sowie bis zu einem gewissen Grad invariant gegenüber affinen Transformationen und Beleuchtungsveränderungen. Der SIFT-Algorithmus besteht aus vier Schritten:

1. Finden von Interest Points – Skalierung

SIFT verwendet eine durch Difference-of-Gaussians (*DoG*) approximierte Laplacepyramide (siehe Kap. 16). Der Unterschied zur ursprünglichen Laplacepyramide ist hier aber, das zunächst keine Größenänderungen von einer Ebene zur nächsten durchgeführt werden, das heißt die Auflösung bleibt in den Ebenen gleich. Somit werden zum Beispiel 5 Gaußfilter hintereinander angewendet und dadurch 4 bandpassgefilterte Ergebnisse erzeugt (siehe Abbildung rechts). Dies wird auch als Oktave bezeichnet. Anschließend wird die nächste Ebene der Gaußpyramide (2. Oktave) in der gleichen Weise behandelt. Die einzelnen Oktaven werden benötigt, um anschließend Extrema finden zu können. Die Ebenen der ersten Oktave haben die Auflösung des Originalbildes, die Ebenen der zweiten Oktave die halbe Auflösung des Originalbildes usw.

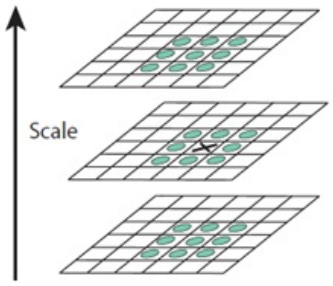


In den DoG Bildern – oder auch Laplacebildern – werden die Frequenzen separiert. Dieser Vorgang lokalisiert somit Kanten und Ecken im Bild.

2. Finden von Interest Points – Position

Das Finden von Interest Points ist ein dreiteiliger Prozess:

1. Lokalisierung der Extrema (Maxima/Minima) in den DoG-Bildern



Der erste Schritt besteht darin, die Maxima und Minima grob zu lokalisieren. Dies wird dadurch erzielt, das iterativ alle Nachbarn im DoG-Skalenraum sowohl in der gleichen Ebene als auch in den Ebenen darüber und darunter überprüft werden. Auf diese Weise werden für jedes Pixel insgesamt 26 Abfragen durchgeführt (9 Punkte darüber, 9 Punkte darunter und die 8 Nachbarn). Interest Points sind dann all jene Punkte, die größer oder kleiner als alle 26 benachbarten Punkte sind.

2. Bestimmung der Position der Extrema mit Subpixel-Genauigkeit

Anschließend werden die approximierten Maxima und Minima der Interest Punkte näherungsweise bestimmt, da die Maxima/Minima nicht notwendigerweise genau auf einem Pixel, sondern auch zwischen den Pixeln liegen können. Dafür werden anhand der zur Verfügung stehenden Pixeldaten mittels der Taylorreihenentwicklung um den approximierten Interest Point herum Subpixelwerte generiert.

2. Eliminierung ungeeigneter Interest Points

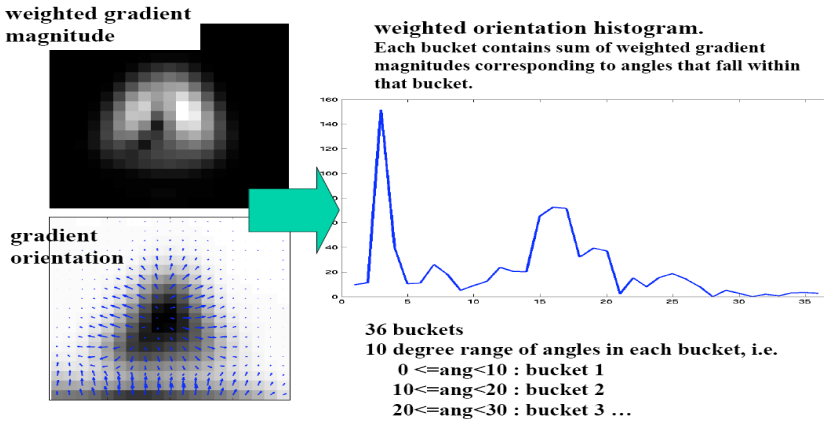
Manche der Interest Points liegen entlang einer Kante oder besitzen nicht genügend Kontrast. In beiden Fällen sind diese als Merkmal nicht brauchbar und werden daher eliminiert. Zur Eliminierung von Merkmalen mit geringem Kontrast werden die Intensitätswerte im DoG-Bild kontrolliert. Die Taylorreihenentwicklung wird verwendet, um die Intensitätswerte am Ort des Subpixels zu bestimmen. Ist dieser Wert geringer als ein gewisser Schwellwert, wird der Interest Point verworfen. Die Eliminierung von Interest Points an einer Kante erfolgt mittels eines Ansatzes, der dem Harris-Ecken-Detektor ähnlich ist. Um Kanten zurückzuweisen, werden zwei Gradienten, die normal zueinander

stehen, am Interest Point berechnet. Basierend auf der Umgebung des Interest Points, gibt es drei Möglichkeiten. Das Bild kann:

1. Eine flache Region sein – Ist dies der Fall sind beide Gradienten klein.
2. Eine Kante sein – hier wird ein Gradient groß (senkrecht zur Kante) und der andere wird klein (entlang der Kante).
3. Eine „Ecke“ sein – hier werden beide Gradienten groß.

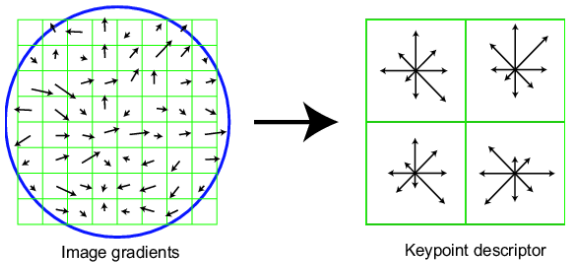
Da Ecken gute Interest Points darstellen, werden nur diese (beide Gradienten sind groß) in Betracht gezogen und der Rest verworfen.

3. Finden von Interest Points – Orientierung



Im nächsten Schritt wird jedem Interest Point eine Orientierung zugeordnet, um den SIFT-Merkmalsvektor rotationsinvariant zu machen - das bedeutet, dass der gleiche Interest Point in einem rotierten Bild den gleichen Merkmalsvektor haben soll. Dazu werden rund um jeden Interest Point die Längen (Gradientenbeträge) und Richtungen der Gradienten gesammelt und anschließend die Orientierung der stärksten Gradienten dieser Region bestimmt (die "dominante" Gradientenrichtung dieser Region). Zu diesem Zweck wird ein Histogramm erstellt, in dem die 360 Orientierungsgrade in 36 Bins aufgeteilt werden (alle 10 Grad). Nehmen wir an, die Gradientenrichtung hat an einem bestimmten Punkt 18,7 Grad, dann würde diese in den Histogramm-Bin von 10-19 Grad fallen. Der Wert, der diesem Bin hinzugefügt wird, ist gleich der Länge des Gradienten an diesem Punkt. Hat man dies für alle Pixel um den Interest Point getan, wird das Bin mit dem höchsten Wert im Orientierungshistogramm als dominante Gradientenrichtung definiert.

4. Erstellen einer Beschreibung der Merkmale



Jeder Interest Point benötigt eine einzigartige Signatur (Beschreibung), die leicht zu berechnen ist. Außerdem sollte die Signatur relativ robust gegenüber kleinen Störungen sein, da lokale Strukturen in zwei verschiedenen Bildern in der Regel nicht hundertprozentig identisch sind. Um diese Signatur zu generieren, wird ein 16x16 Fenster um den Interest Point aufgrund dessen Skalierung definiert und in 16 4x4 Fenster unterteilt. Innerhalb jedes 4x4 Fensters werden die Richtungen und Längen der Gradienten berechnet. Diese Richtungen werden in einem Histogramm mit 8 Bins zusammengefasst. Alle Gradientenrichtungen im Bereich von 0 - 44 Grad werden dem 1. Bin zugeordnet, 45 - 89 Grad zum nächsten usw. Der Wert, der zu einem Bin hinzugefügt wird, hängt von der Gradientenlänge und der Entfernung vom Interest Point ab. Längere Gradienten (d.h. ein höherer Kontrast) werden stärker gewichtet, weiter entfernte Gradienten werden mithilfe einer Gaußschen Gewichtungsfunktion weniger stark gewichtet als näher liegende Gradienten. Macht man dies für alle 16 Pixel, werden 16 Orientierungen in 8 vordefinierte Gruppen unterteilt. Dies wird für alle 16 4x4 Regionen gemacht und man erhält 4x4x8 = 128 Werte (bitte beachten, dass in der obigen Abbildung zu besseren Veranschaulichung nur ein 8x8 Fenster mit 4x4 Subfenstern abgebildet ist). Nachdem alle 128 Werte berechnet wurden, werden diese normalisiert (durch die Wurzel der summierten Quadrate dividiert). Diese 128 Werte bilden den "Merkmalsvektor". Der Interest Point ist somit durch diesen Merkmalsvektor identifizierbar. Jedoch hat dieser zwei Einschränkungen:

1. Rotationsabhängigkeit: Da der Merkmalsvektor Gradientenrichtungen verwendet, verändern sich diese bei einer Rotation des Bildes. Um Rotationsinvarianz zu erreichen, wird die dem Interest Point zugewiesene Orientierung (siehe 3. Schritt) von jeder Gradientenrichtung subtrahiert. Daher ist jede Gradientenrichtung relativ zu der Orientierung des Interest Points gegeben.
2. Beleuchtungsabhängigkeit: Lichteffekte können einen starken Anstieg der Länge einzelner Gradienten bewirken. Um dem entgegenzuwirken, werden alle Werte im Merkmalsvektor, die größer als 0.2 sind, auf 0.2 gesetzt. Der resultierende Merkmalsvektor wird normalisiert und ist schlussendlich weniger sensibel gegenüber einzelner Ausreißer, die aufgrund unterschiedlicher Beleuchtungsbedingungen auftreten können.