

# 14. Global Operations

## Bildtransformationen

Bisher haben wir uns mit Punkt- und lokalen Operationen befasst, Ausgangspunkt war entweder ein Pixel oder die Umgebung eines Pixels inklusive des betrachteten Pixels. Globale Operationen verwenden hingegen das gesamte Bild als Ausgangsbasis. Das gesamte Bild wird aus dem Ortsraum in einen anderen Raum (z.B. den Frequenzraum) transformiert, da dort manche Eigenschaften in Bildern besser sichtbar sind. So kann man dort z.B. Daten besser dekorrelieren oder sie dem menschlichen visuellen System besser anpassen. Es gibt eine Reihe von ähnlichen Methoden für eine Bild-Frequenzraum-Transformation, wobei die wichtigsten beiden (**Fourier-** und **Cosinustransformation**) hier behandelt werden.

Eine Transformation ist eine Matrixmultiplikation des Eingabesignals mit einer quadratischen  $N \times N$  Transformationsmatrix  $A_N$ . Die Transformation eines Spaltenvektors  $x$  mit  $N$  Komponenten entspricht  $y = A_N x$ . Für orthonormale Transformationen gilt:  $AA^T = E$ , wobei  $E$  die Einheitsmatrix ist. Damit gilt:  $y = Ax$  und  $x = A^T y$ . Daher ist auch

$$\begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \quad \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} a_{00} & a_{10} \\ a_{01} & a_{11} \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = y_0 \begin{bmatrix} a_{00} \\ a_{01} \end{bmatrix} + y_1 \begin{bmatrix} a_{10} \\ a_{11} \end{bmatrix} \quad \text{und}$$

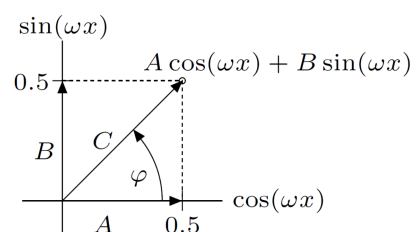
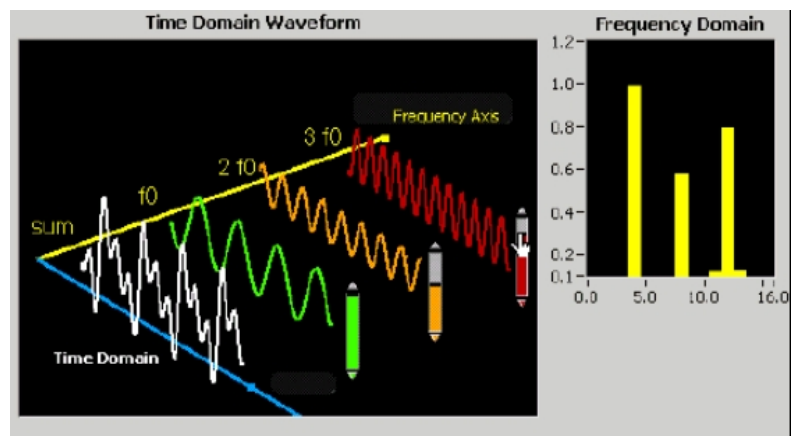
Wobei  $y_0$  und  $y_1$  die Transformationskoeffizienten und  $[a_{00} \ a_{01}]^T$  und  $[a_{10} \ a_{11}]^T$  die Basisvektoren sind.

## Fourier-Transformation

Die Darstellung und Analyse von Bildern im Frequenzbereich basiert auf der Zerlegung von Bildsignalen in sogenannte harmonische Funktionen, also Sinus- und Kosinusfunktionen, mithilfe der Fouriertransformation. Das allgemeine Konzept von Frequenzen und der Zerlegung von Schwingungen in elementare, harmonische Funktionen entstand ursprünglich im Zusammenhang von Schall, Tönen und Musik, da das menschliche Ohr eine ähnliche Aufnahmeweise besitzt. Schall befindet sich in der Zeitdomäne, Töne haben eine Lautstärke (Amplitude) und eine Frequenz (Schwingungen/Sekunde). Wellen wie Schallwellen, Wasserwellen oder elektromagnetische Wellen mit beliebiger Frequenz, Amplitude und Phasenlage können als Summe entsprechend gewichteter Kosinus- und Sinusfunktionen dargestellt werden. Die Kosinusfunktion  $\cos(x)$  hat den Wert eins am Ursprung ( $\cos(0) = 1$ ) und durchläuft bis zum Punkt  $x = 2\pi$  eine volle Periode, das Gleiche gilt für die Sinusfunktion  $\sin(x)$  mit dem Unterschied, dass deren Wert am Ursprung null ist ( $\sin(0) = 0$ ). Die Anzahl der Perioden von  $\cos(x)$  innerhalb einer Strecke der Länge  $T = 2\pi$  ist eins:  $\omega = 2\pi/T = 1$ . Die Beziehung zwischen der Kreisfrequenz  $\omega$  und der Frequenz  $f$  ist  $\omega = 2\pi f$ , wobei  $f$  in Zyklen pro Raum- oder Zeiteinheit, z.B. 1000 Zyklen pro Sekunde (=Hertz) gemessen wird.

Wird eine Kosinusfunktion entlang der  $x$ -Achse um eine Distanz  $\phi$  verschoben, also  $\cos(x) \rightarrow \cos(x - \phi)$ , dann ändert sich die Phase der Kosinusschwingung und  $\phi$  bezeichnet den Phasenwinkel der resultierenden Funktion. Damit ist die Sinusfunktion nichts anderes als eine Kosinusfunktion, die um eine Viertelperiode ( $=\pi/2$ ) nach rechts verschoben ist. Kosinus- und Sinusfunktion sind also "orthogonal" und daher können neue sinusoidale Funktionen mit beliebiger Frequenz, Phase und Amplitude zu erzeugt werden. Insbesondere entsteht durch die Addition einer Kosinus- und Sinusfunktion mit identischer Frequenz  $\omega$  und Amplituden  $A$  bzw.  $B$  eine weitere sinusoidale Funktion mit derselben Frequenz  $\omega$ , wobei die resultierende Amplitude  $C$  und der Phasenwinkel  $\phi$  ausschließlich durch die beiden Amplituden  $A$  und  $B$  bestimmt sind:

$$C = \sqrt{A^2 + B^2} \quad \phi = \tan^{-1}\left(\frac{B}{A}\right)$$



Es war *Jean Baptiste Joseph de Fourier* (1768–1830), der die Addition von Sinus oder Kosinusfunktionen als Erster auf beliebige Funktionen erweiterte und zeigte, dass (beinahe) jede periodische Funktion  $g(x)$  mit einer Grundfrequenz  $\omega_0$  als (möglicherweise unendliche) Summe von "harmonischen" Sinusfunktionen dargestellt werden kann:

$$g(x) = \sum_{k=0}^{\infty} [A_k \cos(k\omega_0 x) + B_k \sin(k\omega_0 x)]$$

Dies bezeichnet man als Fourierreihe, die Gewichte  $A_k, B_k$  als Fourierkoeffizienten der Funktion  $g(x)$ . Die Frequenzen der in der Fourierreihe beteiligten Funktionen sind ausschließlich ganzzahlige Vielfache der Grundfrequenz  $\omega_0$ . Die Fourierkoeffizienten  $A_k, B_k$  können eindeutig aus der gegebenen Funktion  $g(x)$  berechnet werden, ein Vorgang, der als Fourieranalyse bezeichnet wird. Fourier wollte dieses Konzept nicht auf periodische Funktionen beschränken und postulierte, dass auch nicht periodische Funktionen in ähnlicher Weise als Summen von Sinus- und Kosinusfunktionen dargestellt werden können. Die resultierende Zerlegung nennt man ein Fourierintegral:

$$g(x) = \int_0^{\infty} A_{\omega} \cos(\omega x) + B_{\omega} \sin(\omega x) d\omega$$

wobei die Koeffizienten  $A_{\omega}$  und  $B_{\omega}$  wiederum die Gewichte für die zugehörigen Kosinus- bzw. Sinusfunktionen mit der Frequenz  $\omega$  sind. Jeder der Koeffizienten spezifiziert, mit welcher Amplitude die zugehörige Kosinus- bzw. Sinusfunktion der Frequenz  $\omega$  zur darzustellenden Signalfunktion  $g(x)$  beiträgt. Die Bestimmung der Koeffizienten erfolgt durch:

$$A_{\omega} = A(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} g(x) \cos(\omega x) dx \quad B_{\omega} = B(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} g(x) \sin(\omega x) dx$$

Die Koeffizientenfunktionen  $A(\omega)$  und  $B(\omega)$  sind kontinuierlich, sie enthalten das Spektrum der im ursprünglichen Signal enthaltenen Frequenzkomponenten. Das Fourierintegral beschreibt also die Funktion  $g(x)$  als Summe unendlich vieler Kosinus-/Sinusfunktionen mit kontinuierlichen (positiven) Frequenzwerten.

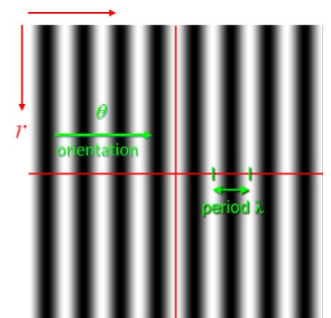
Die Fouriertransformation betrachtet im Unterschied zum Fourierintegral sowohl die Ausgangsfunktion wie auch das zugehörige Spektrum als komplexwertige Funktionen, wodurch sich die Darstellung vereinfacht. Das Fourierspektrum  $G(\omega)$  einer Funktion  $g(x)$  ist definiert als:

$$G(\omega) = \sqrt{\frac{\pi}{2}} [A(\omega) - iB(\omega)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [\cos(\omega x) - i \sin(\omega x)] dx$$

Unter Verwendung der Euler'schen Schreibweise für komplexe Zahlen ergibt sich die übliche Formulierung für das kontinuierliche Fourierspektrum:

$$G(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) \cdot e^{-i\omega x} dx$$

Der Übergang von der Funktion  $g(x)$  zu ihrem Fourierspektrum  $G(\omega)$  bezeichnet man als Fouriertransformation. Umgekehrt kann die ursprüngliche Funktion  $g(x)$  aus dem Fourierspektrum  $G(\omega)$  durch die inverse Fouriertransformation berechnet werden. Ortsraum und Frequenzraum sind zueinander "duale" Darstellungsformen, die sich nicht unterscheiden. Das Fourierspektrum erstreckt sich über positive und negative Frequenzen und ist, obwohl im Prinzip beliebige komplexe Funktionen auftreten können, um den Ursprung symmetrisch.



## Diskrete Fourier-Transformation (DFT)

Die Definition der kontinuierlichen Fouriertransformation ist für die numerische Berechnung am Computer nicht geeignet. Bei Bildern liegen immer diskrete Daten vor und daher wird eine Version der Fouriertransformation benötigt, in der sowohl das Signal wie auch das zugehörige Spektrum als endliche Vektoren dargestellt werden – die Diskrete Fourier-Transformation (DFT). Im Fall eines diskreten, periodischen Signals wird eine endliche Folge von  $M$  Abtastwerten benötigt, um sowohl das Signal  $g(u)$  als auch sein Fourierspektrum  $G(m)$  vollständig abzubilden. Durch die Darstellung als endliche Vektoren sind auch alle Voraussetzungen für die numerische Verarbeitung am Computer gegeben. Die diskrete Fouriertransformation ist, wie auch bereits die kontinuierliche FT, in beiden Richtungen identisch. Die Vorwärtstransformation (DFT) für ein diskretes Signal  $g(u)$  der Länge  $M$  ( $u = 0 \dots M-1$ ) ist für  $0 \leq m < M$  definiert als:

$$G(m) = \frac{1}{\sqrt{M}} \int_{u=0}^{M-1} g(u) \cdot e^{-i2\pi \frac{mu}{M}} \quad \text{und analog dazu die inverse Transformation (DFT}^{-1}\text{) für } 0 \leq u < M: \quad g(u) = \frac{1}{\sqrt{M}} \int_{m=0}^{M-1} G(m) \cdot e^{i2\pi \frac{mu}{M}}$$

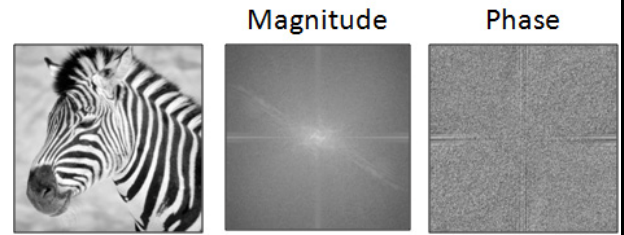
Sowohl das Signal  $g(u)$  wie auch das diskrete Spektrum  $G(m)$  sind komplexwertige Vektoren der Länge  $M$ .

Der Betrag (*magnitude*) des komplexwertigen Fourierspektrums

$$|G(m)| = \sqrt{G_{RE}^2(m) + G_{IM}^2(m)}$$

wird als Leistungsspektrum (*power spectrum*) eines Signals bezeichnet. Es beschreibt die Energie (Leistung), die die einzelnen Frequenzkomponenten des Spektrums zum Signal beitragen. Das Leistungsspektrum ist reellwertig und positiv und wird daher zur grafischen Darstellung der Fouriertransformierten verwendet. Da die Phaseninformation im Leistungsspektrum verloren geht, kann das ursprüngliche Signal aus dem Leistungsspektrum allein nicht rekonstruiert werden. Das Leistungsspektrum ist jedoch wegen der fehlenden Phaseninformation unbeeinflusst von Verschiebungen des zugehörigen Signals und eignet sich daher zum Vergleich von Signalen. Genauer gesagt ist das Leistungsspektrum eines zyklisch verschobenen Signals identisch zum Leistungsspektrum des ursprünglichen Signals.

Zur praktischen Berechnung der DFT existieren schnelle Algorithmen, in denen die Abfolge der Berechnungen so ausgelegt ist, dass gleichartige Zwischenergebnisse nur einmal berechnet und in optimaler Weise mehrfach wiederverwendet werden. Die Fast Fourier Transform (FFT), von der es mehrere Varianten gibt, reduziert i. Allg. die Zeitkomplexität der Berechnung von  $O(M^2)$  auf  $O(M \log_2 M)$ . Die Auswirkungen sind vor allem bei größeren Signallängen deutlich. Zum Beispiel bringt die FFT bei einer Signallänge  $M = 10^3$  bereits eine Beschleunigung um den Faktor 100 und bei  $M = 10^6$  um den Faktor 10.000.



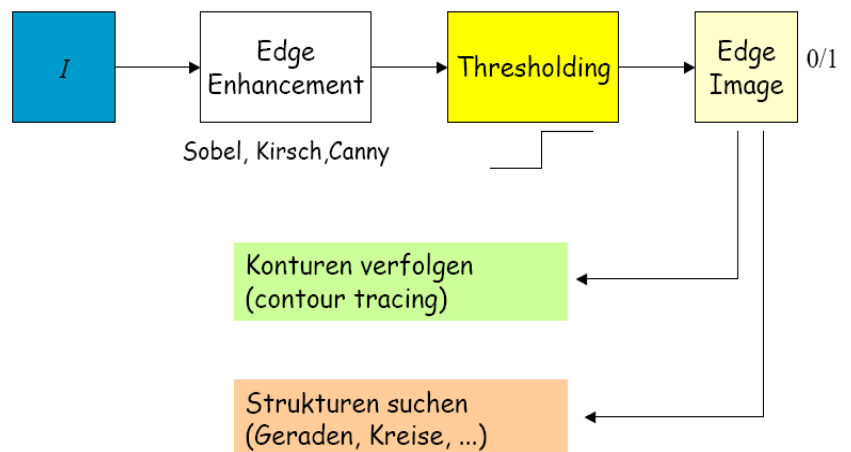
## Hough-Transformation

Als weiteres Beispiel für globale Operationen auf digitalen Bildern wird hier die Hough-Transformation vorgestellt, die als nachfolgender Schritt einer Kantendetektion (siehe Kap.11) zur Analyse der Bildinhalte verwendet wird. Die Kantendetektion ist jedoch eine "lokal beschränkte" Sichtweise, die generelle Struktur von zusammenhängenden Kanten fließt nicht in die Auswertung mit ein. Man muss daher davon ausgehen, dass in dem auf diese Weise produzierten Kantenbild viele vermeintliche Kantenpunkte markiert sind, die in Wirklichkeit zu keiner echten Kante gehören, und andererseits echte Kantenpunkte fehlen.

Ein intuitiver Ansatz zum Auffinden größerer Bildstrukturen besteht darin, beginnend bei einem beliebigen Kantenpunkt benachbarte Kantenpixel schrittweise aneinander zu fügen und damit die Konturen von Objekten zu bestimmen.

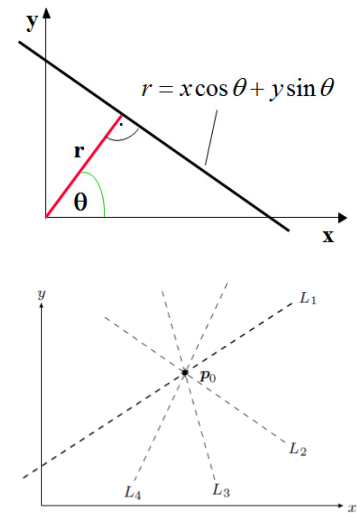
Das kann man sowohl im kontinuierlichen Kantenbild (mit Kantenstärke und Orientierung) als auch im binären Kantenbild durchführen. In beiden Fällen ist aufgrund von Unterbrechungen, Verzweigungen und ähnlichen Mehrdeutigkeiten mit Problemen zu rechnen und ohne zusätzliche Kriterien und Informationen über die Art der gesuchten Objekte bestehen nur geringe Aussichten auf Erfolg. Kantenbilder enthalten in der Regel zahlreiche irrelevante Strukturen und gleichzeitig sind wichtige Strukturen häufig unvollständig.

Daher besteht ein alternativer Ansatz darin, Strukturen im Kantenbild zu suchen, um die globale Information miteinfließen zu lassen. Die Methode von *Paul Hough*, ursprünglich als US-Patent publiziert und als "Hough-Transformation" (HT) bezeichnet, ist ein allgemeiner Ansatz, um beliebige, parametrisierbare Formen in Punktverteilungen zu lokalisieren. Zum Beispiel können geometrische Formen wie Geraden, Kreise und Ellipsen mit wenigen Parametern beschrieben werden. Da sich gerade diese Formen häufig im Zusammenhang mit von Menschenhand geschaffenen Objekten finden, sind sie für die Analyse von Bildern besonders interessant. Die Detektion von Geraden in binären Kantenbildern ist eine häufige Anwendung. Eine Gerade in 2D kann mit zwei reellwertigen Parametern beschrieben werden, z. B. in der klassischen Form  $y = kx + d$ , wobei  $k$  die Steigung und  $d$  die Höhe des Schnittpunkts mit der y-Achse bezeichnet. Ziel ist nun, jene Geradenparameter  $k$  und  $d$  zu finden, auf denen möglichst viele Kantenpunkte liegen. Eine Möglichkeit wäre, alle möglichen Geraden in das Bild zu "zeichnen" und die Bildpunkte zu zählen, die jeweils exakt auf einer bestimmten Geraden liegen, was aber ineffizient ist.



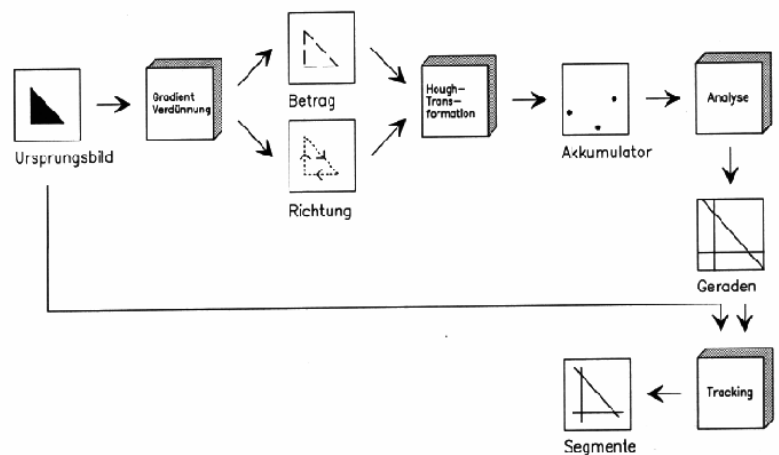
Die Hough-Transformation geht an dieses Problem auf dem umgekehrten Weg heran, indem sie alle möglichen Geraden ermittelt, die durch einen einzelnen, gegebenen Bildpunkt laufen. Dazu verwendet man eine andere Geradendarstellung nämlich die Gerade in Parameterform (Hessesche Normalform):  $r = x \cos(\theta) + y \sin(\theta)$ , wobei  $r$  den Normalabstand der Geraden zum Ursprung und  $\theta$  den Winkel des Normalabstandes zur x-Achse bezeichnet. Jede Gerade  $L_j$ , die durch einen Punkt  $p = (x, y)$  läuft, muss die Gleichung  $r = x \cos \theta + y \sin \theta$

erfüllen, somit entspricht das einem Bündel von unendlich vielen Geraden, die alle durch den gegebenen Punkt  $p$  laufen, wobei die Parameter  $r$  und  $\theta$  variieren. Die Lösungsmenge  $\{(r_j, \theta_j)\}$  beschreibt die Parameter aller möglichen Geraden  $L_j$ , die durch einen Bildpunkt  $p_0 = (x_0, y_0)$  führen. Jede einzelne Gerade dieser Menge wird im so genannten Parameterraum (auch "Hough-Raum" genannt), der durch die Koordinaten  $r, \theta$  aufgespannt wird, durch einen einzigen Punkt repräsentiert, der genau den Parametern der betrachteten Geraden entspricht.



## Hough-Transformation für Linien

Ausgehend von einem Eingangsbild  $I(u, v)$  wird zunächst ein Kantenbild erstellt, wobei für jedes Kantenpixel der Betrag und die Richtung des Gradienten berechnet wird (z.B. durch den Sobel-Filter). Für die Hough-Transformation wird ein zweidimensionales Akkumulator-Array erzeugt. Dazu wird der  $(r, \theta)$ -Raum diskretisiert (z.B.  $r$  in Pixel,  $\theta$  in  $1^\circ$  oder  $5^\circ$  Stufen), sodass  $\theta$  in der x-Richtung und  $r$  in y-Richtung eingetragen werden kann. Das daraus entstehende endliche, zweidimensionale Feld  $(r, \theta)$  wird als Akkumulator bezeichnet, jedes Element im Akkumulator entspricht einer Geraden im  $(x, y)$ -Koordinatensystem. Der Akkumulator wird zunächst mit Null initialisiert, anschließend werden alle Kantenpunkte mit der Hough-Transformation in den  $(r, \theta)$ -Raum überführt und in den Akkumulator eingetragen (die betroffenen Zellen werden inkrementiert). Jede Stelle im Akkumulator, deren Wert einen gegebenen Schwellwert überschreitet, repräsentiert genau eine Gerade im Bild (dazu muss man in der Praxis Punkthäufungen (Cluster) auf einen Punkt reduzieren). Das Ergebnis der Hough-Raum-Analyse ist ein Bild mit Geraden ohne Anfangs- und Endpunkte. Um deren Anfangs- und Endpunkte zu finden, wird ein sogenanntes Tracking durchgeführt. Dabei dienen die durch die Hough-Transformation gefundenen Geraden als "Wegweiser" durch das Originalbild. Man fährt auf den Geraden das Originalbild mit einem Fenster ab und registriert wann die Grauwertdifferenz (des Bereichs rechts und links von der Geraden) im Fenster einen Schwellwert überschreitet. An diesen Stellen befindet sich das Fenster zu einer hohen Wahrscheinlichkeit auf einer Objektkontur. Der Vorteil bei der Verwendung von Betrag und Richtung von Kantenpixeln ist, dass man die Richtung der Kante im Originalbild schon kennt. Man kann natürlich auch Filter 2. Ordnung verwenden, dann muss man jedoch an jeder Stelle der Kantenpixel alle möglichen Geraden durch diesen Punkt eintragen, was zu einer sinusförmigen Linie im Akkumulator führt.



Das Ergebnis der Hough-Raum-Analyse ist ein Bild mit Geraden ohne Anfangs- und Endpunkte. Um deren Anfangs- und Endpunkte zu finden, wird ein sogenanntes Tracking durchgeführt. Dabei dienen die durch die Hough-Transformation gefundenen Geraden als "Wegweiser" durch das Originalbild. Man fährt auf den Geraden das Originalbild mit einem Fenster ab und registriert wann die Grauwertdifferenz (des Bereichs rechts und links von der Geraden) im Fenster einen Schwellwert überschreitet. An diesen Stellen befindet sich das Fenster zu einer hohen Wahrscheinlichkeit auf einer Objektkontur. Der Vorteil bei der Verwendung von Betrag und Richtung von Kantenpixeln ist, dass man die Richtung der Kante im Originalbild schon kennt. Man kann natürlich auch Filter 2. Ordnung verwenden, dann muss man jedoch an jeder Stelle der Kantenpixel alle möglichen Geraden durch diesen Punkt eintragen, was zu einer sinusförmigen Linie im Akkumulator führt.

## Weitere Hough-Transformationen

Prinzipiell kann man jede parametrisierbare Struktur mit Hilfe der Hough Transformation finden, die häufigsten sind Kreise und Ellipsen. Die Kreiserkennung funktioniert ähnlich der Geradenerkennung. Durch jeden Punkt im Originalbild wird ein Kreis mit vorher festgelegtem Radius gelegt. Punkte im Hough-Transformationsbereich, in denen sich solche Kreise schneiden, sind Kandidaten für Mittelpunkte von Kreisen (mit dem gewählten Radius) im Originalbild. Die Punkte, in denen sich die meisten Kreise schneiden, werden als Kreismittelpunkte in den Originalbereich rücktransformiert. Die Hough-Transformation auch zur Ellipsenerkennung verwendet werden. Im Gegensatz zum Kreis wird eine Ellipse neben dem Mittelpunkt  $(x, y)$  auch noch durch die 2 Halbachsen  $a$  und  $b$  parametrisiert, sodass der Houghraum 4 Dimensionen hat, die bei der Suche berücksichtigt werden müssen.

