

3. Bildbearbeitung

3.1. Aufgaben der Bildbearbeitung

Wie schon in der Einführung erläutert, bedürfen die mit einem SPM-System gewonnenen Daten der Nachbearbeitung. Im folgenden werden zunächst die Aufgaben der Bildbearbeitung zusammengefaßt.

- Da das Einbringen der Probe in das System per Hand erfolgt, kann nicht sichergestellt werden, daß die vertikale Bewegung der Abtastspitze "orthogonal" zur Probe erfolgt, Orthogonal in dem Sinn, daß die bestapproximierende Ebene der Oberfläche senkrecht zur Spitze liegt. Die Situation ist in der Skizze dargestellt. Dieses Problem motiviert zur Ausgleichskorrektur.

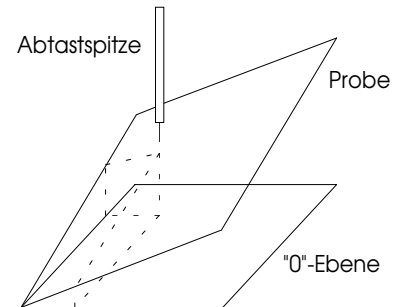


Abb. 11.: Nicht z-orthogonale Abtastung

- Während der Abtastung können immer wieder einzelne Werte extrem gestört werden.

Bemerkung: Eigentlich gibt es kein Datum, welches nicht gestört ist. Die Terminologie "gestörter Bildpunkt" meint hier wirklich nur extreme Störungen. Von einer solchen kann auch eine ganze Bildzeile betroffen sein. Eine weitere wichtige Aufgabe besteht also in der Erkennung und Elimination solcher Störungen.

- Die Anwendung von Filtern zur Beseitigung komplexerer Störungen, wie zum Beispiel des weißen Rauschens, ist der nächste logische Schritt.

- Um einen besseren visuellen Eindruck zu erzeugen, können/müssen diverse Grauwertoperationen auf die Daten angewandt werden. Dies betrifft insbesondere die Skalierung und Kontrastierung der Daten.

Der Begriff *Bildbearbeitung* könnte also sinnvoll als:

"Folge von Transformation, die dazu dienen, ein Bild in eine auswertbare Gestalt zu überführen"

erklärt werden.

3.2. Grauwertoperationen

Alle Bilder S , deren Wertebereich $\subset \mathbf{Z}$ und endlich ist, können mittels einer Abbildung (*lineare Skalierung*) in ein Grauwertbild \tilde{S} durch

$$(3.1) \quad \tilde{s}_{i,j} = \text{rnd} \left(g_{\max} \cdot \frac{s_{i,j} - s_{\min}}{s_{\max} - s_{\min}} \right)$$

überführt werden. $\text{rnd}(x)$ ist eine Rundungsfunktion, s_{\min} und s_{\max} sind das Minimum und Maximum von s , g_{\max} ist der größte Grauwert. Allgemein läßt sich eine lineare Skalierung als

$$\tilde{s}_{i,j} = (s_{i,j} + c_1) \cdot c_2$$

darstellen, wobei c_1 die Helligkeit und c_2 den Kontrast steuert. $c_1=0$ und $c_2=1$ ist die identische Abbildung, $c_1 = -s_{\min}$ und $c_2 = g_{\max} / (s_{\max} - s_{\min})$ ergibt genau den obigen Spezialfall.

3.3. Grundoperationen

3.3.1. Ausgleichskorrektur

Das in 3.1. dargelegte Problem der nicht z-orthogonalen Abtastung kann anhand der Meßdaten nicht als solches erkannt werden. Man stelle sich vor, daß die Probe ein Keil ist. In diesem Fall wäre eine Ausgleichskorrektur nicht zulässig, da sie Informationen echt verfälscht. An dieser Stelle ist der Physiker aufgefordert, über die Anwendbarkeit der Korrektur zu entscheiden.

Def.: Sei $s: \mathbf{G}^{n_y, n_x} \subset \mathbf{R}^{n_y, n_x}$ das Bild und $\varphi: \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ eine Abbildung aus dem Unterraum $U \subseteq C^k(\mathbf{R} \times \mathbf{R})$, die durch

$$(3.2) \quad \sum_{i=0}^{n_y-1} \sum_{j=0}^{n_x-1} (\varphi(j, i) - s_{i,j})^2 \rightarrow \min!$$

bestimmt wird. Dann heißt φ die Ausgleichsfunktion und

$$\tilde{s}_{i,j} = \text{rnd}(s_{i,j} - \varphi(j, i)) \quad \forall (i, j) \in \mathbf{I}_y \times \mathbf{I}_x$$

die Ausgleichskorrektur zu s .

Meist benutzt man $\varphi \in \Pi^{n,m}(\mathbf{R} \times \mathbf{R})$ (Polynome vom Grad (n,m) über $\mathbf{R} \times \mathbf{R}$). Für den Fall der nicht orthogonalen z-Abtastung wird man $n=1$ und $m=1$ wählen. Die Ausgleichskorrektur wird a priori nicht auf den bilinearen Fall beschränkt. Sinnvoll anzuwenden ist eine höhere Dimension, wenn bekannt ist, daß die Probe global gesehen eine bestimmte Form besitzt und man an einem Kartenwurf interessiert ist.

Sei $V = [\varphi_i(x, y)]^T$ die Zeilenmatrix der Basisfunktionen des Unterraumes U von C^k , und f sei aus C^k , mit $f(j, i) = s_{i,j}$, dann ist eine Funktion $\varphi \in U$; $\varphi = V \cdot \underline{c}$; $\underline{c} \in \mathbf{R}^{\dim(U)}$ der Art gesucht,

daß gilt $\left(\sum_{i \in I_y} \sum_{j \in I_x} (\varphi(j,i) - f(j,i))^2 \right)^{1/2} =: \|\varphi - f\|_{I_x, I_y} \rightarrow \min!$. Sei weiter $V(x,y)$ die Zeile, die

die Werte der Basisfunktionen an der Stelle (x,y) enthält.

Die eindeutige Abbildung $Goed: I_y \times I_x \rightarrow I_{Goed} \subseteq N$, die jedem $(i,j) \in I_y \times I_x$ genau eine Zahl aus I_{Goed} zuordnet und die sich für alle (i,j) in endlich vielen Schritten berechnen läßt (ebenso ihre Umkehrung), heißt Gödelabbildung. Im folgenden sei $I_{Goed} = N | card(I_y \times I_x)$.

1. Lösungsvariante:

Die Lösung ist die Lösung des Normalgleichungssystems $A^T A \cdot \underline{c} = A^T \cdot \underline{f}^{(*)}$,

$$A^T A = \left[\langle \varphi_i, \varphi_j \rangle_{I_x, I_y} \right] \text{ und } A^T \cdot \underline{f} = \left[\langle \varphi_i, \underline{f} \rangle_{I_x, I_y} \right] \text{ mit } A_{Goed(i,j)} = V(x_j, y_i) \text{ und } f_{Goed(i,j)} = f(x_j, y_i).$$

Im Spezialfall $U = \Pi^{n,n}$ sei

$$\varphi_{i + \frac{k(k+1)}{2}}(x, y) = x^{k-i} \cdot y^i; k=0(1)n, i=0(1)k.$$

Es ist dann

$$\langle \varphi_i, \varphi_j \rangle_{I_x, I_y} = \sum_{k=0}^{n_x-1} \sum_{l=0}^{n_y-1} x_k^{n_i+n_j} \cdot y_l^{m_i+m_j} \text{ und } \langle \varphi_i, \underline{f} \rangle_{I_x, I_y} = \sum_{k=0}^{n_x-1} \sum_{l=0}^{n_y-1} x_k^{n_i} \cdot y_l^{m_i} \cdot f(x_k, y_l).$$

Für die Ausgleichskorrektur eines Bildes wählt man $x_k=k, y_l=l$ und damit $f(x_k, y_l) = s_{l,k}$. $A^T A$ ist symmetrisch und positiv definit, deshalb kann die Lösung von $(*)$ mittels der Cholesky-Zerlegung von $A^T A$ in LL^T (L untere Dreiecksmatrix) gemäß $L \cdot \underline{z} = A^T \underline{f}$ und $L^T \cdot \underline{c} = \underline{z}$ erfolgen.

2. Lösungsvariante:

Seien A und \underline{f} wie oben erklärt. Das Problem lautet jetzt $\|A \underline{c} - \underline{f}\|_{I_x, I_y} \rightarrow \min!$. Man bestimme

Q und R so, daß $A=QR$, $Q^T Q=I$ und R obere Dreiecksmatrix wird. Dann sei $\underline{d} = Q^T \underline{f}$ und \underline{c} wird aus $R \cdot \underline{c} = \underline{d}$ bestimmt.

Bemerkung: Die beiden Lösungsvarianten leisten analytisch dasselbe ($R=L^T, \underline{z}=\underline{d}$), nur mit unterschiedlich hohem Aufwand. Für die erste Variante wird $(p+q/3)q^2/2 + (p+q) \cdot q + [p^2 \cdot q]$ angegeben, für die QR-Faktorisierung mit Lösung $p \cdot q^2 + 2p + q/2$. p ist die Anzahl der Basisfunktionen, q die Anzahl der Stützstellen, der Term $[]$ kommt von der Skalarproduktbildung zur Generierung der Matrix $A^T A$ ([Kiel88]).

Die Rechnung mittels der QR-Faktorisierung ist numerisch stabiler.

Folgendes Beispiel illustriert die Wirkung der Ausgleichskorrektur für den Fall $n=1, m=1$.

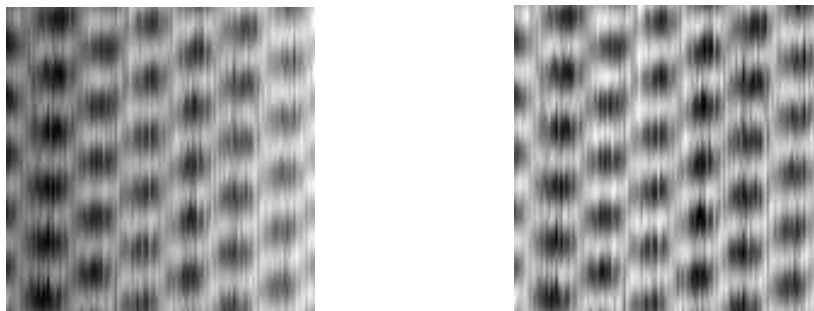


Abb. 12.: Originaldaten (links), gewonnen von einer Graphit-Probe, rechts nach Ausgleichskorrektur mit Polynom vom Grad (1,1). Das linke Bild wird im weiteren als "HOPG-Bild" bezeichnet.

3.3.2. Elimination gestörter Bildpunkte

Def.: (intuitiv, für Grauwertbilder):

Ein Dropout (isolierte Punktstörung) ist ein Punkt, dessen Grauwert stark von seinem Erwartungswert abweicht.

Diese Definition induziert, daß ein Punkt als Zufallsgröße betrachtet wird. Die damit einhergehende Beschreibung von Bildern als *Zufallsfelder* wird sehr detailliert in [Rose82] erörtert.

a) Detektion:

An die Erkennung von Dropouts kann man auf zwei Arten herangehen:

- I) Man betrachte die Umgebung eines Punktes und entscheide dann, ob der Punkt zur Menge der Dropouts gehört (lokaler Ansatz, [Haber89]).
- II) Man betrachte das Histogramm h des gesamten Bildes und bestimme anhand einer vorgegebenen Fehlerwahrscheinlichkeit die Werte der Dropout-Menge und aus dieser dann alle Dropouts.

Def.: Sei $X = \{X_i : i \in I\}$ mit den Werten $x_p, x_{\max} = \max_{i \in I} \{x_i\}$ und $x_{\min} = \min_{i \in I} \{x_i\}$, dann heißt die Abbildung $h_m : N \rightarrow N$ mit

$$h_m(k) = |\{X_i : X_i \in X \wedge (k-1) \cdot \Delta x \leq x_i - x_{\min} < k \cdot \Delta x\}|; k=1..m \text{ und für}$$

$$k=m \text{ sei } x_i - x_{\min} \leq m \cdot \Delta x \text{ mit } \Delta x = \frac{x_{\max} - x_{\min}}{m} \text{ Histogramm der Menge } X \text{ mit}$$

m Klassen.

zu II) Die Grauwertmenge des Bildes S definiert den Wertebereich einer diskreten Zufallsgröße G . Für diese liefert das Histogramm eine Schätzung der Verteilung, so daß die Wahrscheinlichkeit $p_k = P(G < x)$ aus

$$(3.3) \quad P(G < x) = \frac{1}{|h_m|} \cdot \sum_{i: x - x_{\min} < i \cdot \Delta x} h(i); |h_m| := \sum_{i=1}^m h_m(i)$$

bestimmt werden kann.

Die Dropout-Wahrscheinlichkeit P_{DO} wird jetzt definiert als

$$(3.4) \quad P_{DO}(D_1, D_2) = P(G - h_{\min} < D_1) + P(h_{\max} - G < D_2).$$

Damit werden sowohl Ausreißer am unteren als auch am oberen Rand als Dropouts erfaßt. Die Dropout-Menge $\Delta(\delta_u, \delta_o)$ definiert man mit den Fehlerwahrscheinlichkeiten δ_u und δ_o als:

$$(3.5) \quad \Delta(\delta_u, \delta_o) = \{(i, j) : s_{i,j} < D_1 \vee s_{i,j} > D_2\}$$

mit

$$D_1 = \Delta s \cdot \max\{k : P(S - s_{\min} < k \cdot \Delta s) \leq \delta_u\} + s_{\min}$$

und

$$D_2 = \Delta s \cdot \max\{k : 1 - P(S - s_{\min} < k \cdot \Delta s) \leq \delta_o\} + s_{\min}.$$

b) Elimination

Für alle Punkte der Dropout-Menge bestimme man

$$\tilde{s}_{i_d, j_d} = \frac{1}{k} \sum_{i=-1}^1 \sum_{j=-1}^1 s_{i_d+i, j_d+j} : k \neq 0 \wedge s_{i_d+i, j_d+j} \notin \Delta(\delta_u, \delta_o) \wedge s_{i_d+i, j_d+j} \in S$$

k = "Anzahl der s_{i_d+i, j_d+j} ".

Ist $k=0$, so wird $\tilde{s}_{i_d, j_d} = s_{i_d, j_d}$ und die Störung kann nur in einem zweiten Durchlauf eliminiert werden.

Abbildung 13 zeigt eine Testfläche mit eingebauten Dropouts. Hinten befindet sich eine isolierte Punktstörung, rechts ein Kreuz und links eine Pyramide. Wie zu erkennen ist, konnten nur die hintere und rechte Störung im ersten Anlauf beseitigt werden ($\delta_u = 0\%$, $\delta_o = 10\%$). Erst im zweiten Durchlauf mit $\delta_u = 0\%$ und $\delta_o = 5\%$ wurde auch die Pyramide entfernt.

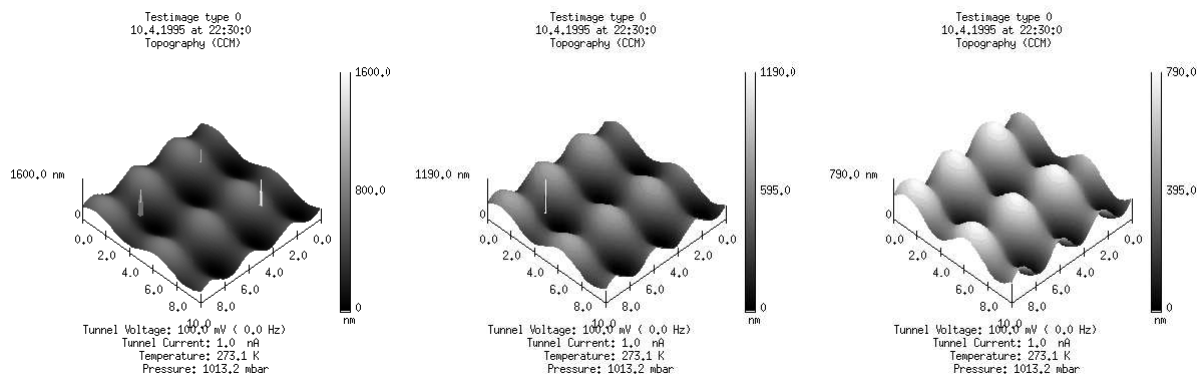


Abb. 13.: Links: Testfläche mit Dropouts, Mitte: erster Durchlauf, Rechts: zweiter Durchlauf

3.3.3. Elimination gestörter Bildzeilen

a) Detektion:

Zwei aufeinanderfolgende ungestörte Zeilen eines Bildes sind stark korreliert, wenn man voraussetzt, daß der wirkliche Übergang hinreichend glatt ist, was bei natürlichen Daten der Fall ist. Falls nicht, so kann keine Unterscheidung zwischen gestörten und nicht gestörten Zeilen getroffen werden.

Man bestimme zunächst alle Korrelationskoeffizienten

$$(3.6) \quad r_{y, y+1} = \frac{v_{y, y+1}}{\sqrt{v_{y, y} \cdot v_{y+1, y+1}}} \quad \forall 0 \leq y \leq n_y - 2$$

mit

$$(3.7) \quad v_{i, j} = \frac{1}{n_x} \sum_{x=0}^{n_x-1} (s_{i, x} - m_i) \cdot (s_{j, x} - m_j)$$

und

$$(3.8) \quad m_y = \frac{1}{n_x} \sum_{x=0}^{n_x-1} s_{y, x}.$$

Nach einem, in [Haber89] beschriebenen Verfahren, suche man jetzt nach $|r_{y, y+1}|$ die

unter einer gewissen Schranke liegen. Werden solche $r_{y,y+1}$ gefunden, so wird Zeile $y+1$ als gestört angesehen. Das Verfahren hat den Nachteil, daß die Startzeile korrekt sein muß. Das zweite Problem ist die Wahl der Schranke.

Folgendes Verfahren hat bei Versuchen gute Ergebnisse geliefert. Es verlangt keine Startzeile und keine Schranke für die Korrelationskoeffizienten.

1. Bestimme die Beträge aller Korrelationskoeffizienten $r_y := |r_{y,y+1}|$
2. Berechne

$$(3.9) \quad c_i := \begin{cases} 1 - \frac{r_i}{r_{i-1}}; r_i < r_{i-1} \\ 1 - \frac{r_{i-1}}{r_i}; r_i \geq r_{i-1} \end{cases} \quad \forall i = 1..(n_y - 1).$$

3. Annahme: Es existieren keine gestörten Bildzeilen, falls

$$(3.10) \quad s^2 = \frac{1}{n_y - 2} \sum_{i=1}^{n_y-1} (c_i - \bar{c})^2 < \varepsilon$$

ist (Tests haben gezeigt, daß s^2 im ungestörten Fall in der Größenordnung von 10^{-7} oder darunter, im gestörten Fall dagegen deutlich über 10^{-2} lag).

4. Für alle $i: |c_i - \bar{c}| > k \cdot s$ ($k > 0, k \in \mathbb{N}$) besteht der Verdacht auf Störung. k ist eine Konstante, meist gleich 1. Man fasse jeweils aufeinanderfolgende i zu einem Cluster $[i_1, i_2]$ zusammen.
5. Für diese Cluster bestimme man das j , für welches $r_j = \min\{r_i : i \in [i_1, i_2]\}$ gilt. Ist $r_{j+1} > r_{j-1}$, so ist Zeile $j+1$ gestört, im anderen Fall Zeile j .

Der Nachteil dieses Verfahrens besteht darin, daß nur isolierte Zeilenstörungen erkannt werden. Beispiel:



Abb. 14.: Bild mit erkannten (weiß markierten) gestörten Zeilen und korrigiertes Bild

Das Verfahren hat auch bei einem künstlich verrauschten Bild gute Ergebnisse geliefert.

b) Elimination:

Oft wird eine gestörte Zeile durch Interpolation zwischen den (beiden) benachbarten Zeilen überschrieben.

3.4. Filter

Ausgangspunkt für die Charakterisierung von Filtern ist das lineare zeitinvariante System, welches, wie unter 2.4.2. erläutert, durch seine Impulsantwort oder Übertragungsfunktion beschrieben wird. Nachfolgend wird nur erläutert, was Filter sind. Eine Diskussion von Eigenschaften und Verhalten wird im nächsten Kapitel geführt.

3.4.1. Filter im Frequenzbereich

*Def.: Die Faltung $\tilde{s}(x, y) = s(x, y) * g(x, y)$ heißt Filterung von s . $g(x, y)$ heißt Impulsantwort des Filters. In der, durch den Übergang zum Frequenzbereich erhaltenen, Darstellung $\tilde{S}(u, v) = S(u, v) \cdot G(u, v)$ heißt $G(u, v)$ die Übertragungsfunktion des Filters.*

Beispiel.: Tiefpaß-Filter

$$G(u, v) = \begin{cases} 1; & \forall u, v: |u| \leq u_g \wedge |v| \leq v_g \\ 0; & \text{sonst} \end{cases}$$

$$\tilde{s} = \text{IFFT}(\text{FFT}(s) \cdot G), \quad (u_g = 0.64 \text{ nm}^{-1}, v_g = 0.64 \text{ nm}^{-1})$$



Abb. 15.: Verrauschtes HOPG-Bild, gefiltertes HOPG-Bild

Wie in Abbildung 15 zu sehen ist, hat sich der visuelle Eindruck im rechten Bild aufgrund der Anwendung eines Tiefpaß-Filters erheblich verbessert, das Rauschen ist entfernt worden. Natürlich stellt man jetzt sofort die Frage nach den Parametern u_g und v_g des angewendeten Filters oder anders formuliert:

Welche Gestalt muß die Übertragungsfunktion eines Filters haben, um eine optimale Trennung von Signal und Störung zu erreichen?

Die Beantwortung dieser Frage nach dem Filterdesign kann manuell erfolgen, indem man zunächst die Fouriertransformierte des Signals (=des Bildes) berechnet läßt und dann anhand dieser die Parameter für einen vorgegebenen Filtertyp bzw. den zu applizierenden Filtertyp selbst bestimmt. Diese Methode wird bisher am häufigsten benutzt ([Joerg93]), hat aber die entscheidenden Nachteile, langsam und sehr von der subjektiven Bewertung durch den Operator abhängig zu sein (Abb. 16). Ziel muß es also sein, automatisch einen Filter zu einem gegebenen Problem zu generieren, der Signal und Störung möglichst gut voneinander trennt und dabei die innere Struktur des Bildes erhält.

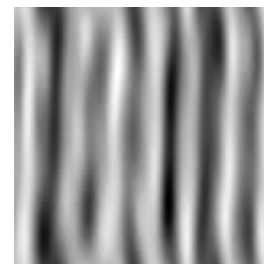


Abb. 16.: manuell ausgeführter Tiefpaß-Filter über dem HOPG-Bild

3.4.2. Digitale Filter im Ortsbereich

Die Filterung im Frequenzbereich ist im allgemeinen sehr langsam (Fouriertransformation → Filterung → Rücktransformation), deshalb wird häufig zu wesentlich schnelleren digitalen Filter im Ortsbereich gegriffen.

Def.: Sei $F \in \mathbf{R}^{2n+1, 2m+1}$, dann bezeichnet F den Filter im Ortsbereich, der s in \otimes vermöge

$$(3.11) \quad \tilde{s}_{i,j} := \sum_{k=-n}^n \sum_{l=-m}^m F_{k,l} \cdot s_{i-k, j-l}; \quad \forall (i, j) \in \mathbf{I}_y \times \mathbf{I}_x$$

überführt.

Dies ist die diskrete Darstellung der Faltung einer Funktion f , für die $f(x_b, y_k) = F_{k,l}$ ist und die für alle (x, y) außerhalb $[x_{-m}, x_m] \times [y_{-n}, y_n]$ verschwindet, mit s (siehe auch [Haber89]).

3.5. Diskussion von Filtern und Filterdesign

3.5.1. Motivation

Zunächst soll motiviert werden, daß es hinreichend ist, eine Diskussion 1-dimensionaler Filter durchzuführen und dann in Analogie auf das Verhalten 2-dimensionaler Filter zu schließen. Dazu betrachte man die Eigenschaften der Fouriertransformierten (2.3.2.). Weiterhin überzeuge man sich leicht davon, daß $\tilde{F}(u, v_0) = G(u, v_0) \cdot F(u, v_0)$, $v_0 \in \mathbf{R} \text{ const.}$ einen 1-dimensionalen Filter beschreibt. Wendet man diesen Filter für alle $v_0 \in [-v_{\max}, v_{\max}] \subset \mathbf{R}$ an und wiederholt danach die Prozedur für konstantes u , so hat man einen 2-dimensionalen Filter in zwei 1-dimensionale Filter zerlegt.

3.5.2. Tiefpaß-Filter

Def.: Ein Filter, dessen Übertragungsfunktion die Eigenschaften

$$(3.12) \quad G(0) = 1 \text{ und } \lim_{|u| \rightarrow \infty} G(u) = 0$$

hat, heißt Tiefpaß-Filter ([Schue91]).

Im allgemeinen wird eine Grenzfrequenz u_g angegeben, für die

$$G(u) = \begin{cases} 1; & |u| \leq u_g \\ 0; & \text{sonst} \end{cases}$$

und im diskreten Fall

$$G_k^{u_g} = \begin{cases} 1; & |k| \cdot u_0 \leq u_g \\ 0; & \text{sonst} \end{cases} \text{ gilt.}$$

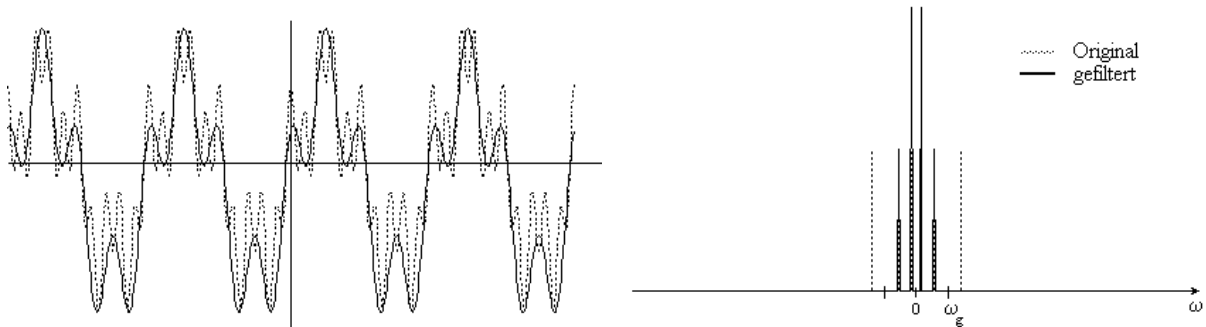


Abb. 17.: Ausschnitt aus der Funktion $2\sin(\pi x) + \cos(4\pi x) + \cos(10\pi x)$ (links) mit zugehöriger Fouriertransformierten vor und nach Filterung mit einem Tiefpaß-Filter mit der Grenzfrequenz $u_g = 2.5$

Geht man davon aus, daß ein Signal durch die Frequenz u_s bandbegrenzt ist und additiv mit weißem Rauschen überlagert wurde, so kann mittels des Tiefpaß-Filters mit der Grenzfrequenz $u_g > u_s$ ein großer Rauschanteil eliminiert werden. Nicht beseitigt werden können Rauschanteile, deren Spektrum im Bereich $\leq u_g$ liegt. Abbildung 17 zeigt ein Signal mit einer Störung mit der Frequenz 5. Wendet man auf dieses Signal einen Tiefpaß mit $u_g = 2.5$ an, so wird das Störsignal vollständig entfernt. Allerdings kann ein unerwünschter

Randeffekt auftreten, der sich, wie in Abb. 18 zu sehen, äußert. Zur Begründung betrachte man die l -periodische Funktion $f(x)$ mit $f(x) = f(x+l)$ und $f(l-0) \neq f(l)$. f kann auch als Summe einer l -periodischen Funktion \tilde{f} mit $\tilde{f}(l-0) = \tilde{f}(l)$,

$\tilde{f}(0) = \tilde{f}(l-0) = 0$ und der Sägezahnfunktion $z(x)$,

$z(x) = \alpha(x - [x/l] * l) + \beta$ dargestellt werden. Für die

Sägezahnfunktion tritt der unter 2.3. beschriebene Fehler bei Nichtverwendung des AVED auf. Der gleiche Effekt kann auch im 2-dimensionalen Fall demonstriert werden. Um ihn zu vermeiden, muß also vor der Filterung Lemma 2 angewendet werden. Dazu könnte man folgendermaßen vorgehen:

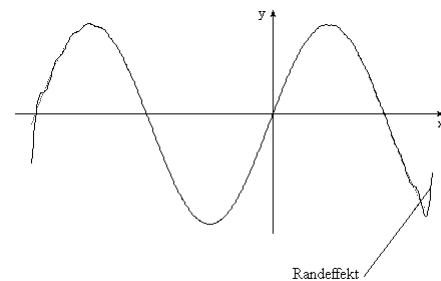


Abb. 18.: Randeffekt bei Tiefpaßfilterung

Gegeben seien die Werte f_k ; $k=0(1)N-1$. Man setze $f_{N-1} := \frac{1}{2}(f_0 + f_{N-1}^*)$, * kennzeichnet den ursprünglichen Wert mit dem Index $N-1$. Dann berechne man die diskrete Fouriertransformation, führe die Filterung und anschließend die Rücktransformation aus. Das Ergebnis \tilde{f}_{N-1} genügt sicherlich $\tilde{f}_{N-1} = \frac{1}{2}(\tilde{f}_0 + \tilde{f}_{N-1}^*)$ und damit kann man \tilde{f}_{N-1}^* rekonstruieren.

Beispiel: Die Funktion $f(x)=x$ wird gestört und auf $[-0.5, 0.5]$ mit 100 Stützstellen abgetastet. Dann wird ein Tiefpaß mit $u_g=30$ angewendet und zurücktransformiert. Wie in Abbildung 19 zu erkennen ist, tritt der Randeffekt bei Verwendung des vorgeschlagenen Verfahrens noch auf, ist aber geringer als ohne dessen Anwendung. Um ganz exakt zu arbeiten, müßte man versuchen, den Wert f_{-1} zu bestimmen und dann $f_{N-1} = (f_{-1} + f_{N-1}^*) / 2$ setzen. Weiterhin muß man an G dieselben Forderungen wie an s bezüglich des Verhaltens an Unstetigkeitsstellen stellen. Das heißt, beim idealen Tiefpaß müssen die Flanken stetig sein. Im diskreten Fall bedeutet das $G_{-u_g} = G_{u_g} = 0.5$.

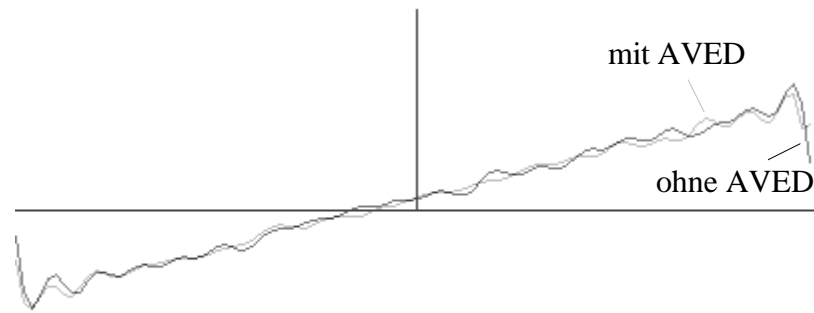


Abb. 19.: Tiefpaß-Filter und AVED

3.5.3. Hochpaß- und Bandpaß-Filter

Def.: Ein Filter, heißt:

- *Hochpaß-Filter, falls*

$$(3.14) \quad \lim_{|u| \rightarrow 0} G(u) = 0 \wedge \lim_{|u| \rightarrow \infty} G(u) = 1$$

- *Bandpaß-Filter, falls*

$$(3.15) \quad \lim_{|u| \rightarrow 0} G(u) = \lim_{|u| \rightarrow \infty} G(u) = 0 \wedge \exists u_0 \in \mathbb{R}: G(u) = 1$$

Ein Bandpaß kann durch die Hintereinanderausführung eines Tiefpaß- und eines Hochpaß-Filters erzeugt werden. Da sich die zu betrachtenden Signalkomponenten normalerweise im niederfrequenten Bereich befinden, kommt dem Hochpaß keine ausgezeichnete Rolle zu. Wesentlich interessanter ist der Bandpaß, da er die Abspaltung von bandbegrenzten Signalanteilen erlaubt (siehe dazu auch nachfolgende Abschnitte).

In 3.5.2. und 3.5.3. werden Standardfilter erläutert, die aus der Literatur hinlänglich bekannt sind und in jedem guten Buch der Systemtheorie oder der Nachrichtentechnik zu finden sind (zum Beispiel [Schue91], [Kress89] und [Lange75]).

3.5.4. Energie-Filter

Man betrachte das Signal $\tilde{s}(t) = s(t) + n(t)$. $n(t)$ sei weißes Rauschen. Nach 2.4.3. folgt für $n(t)$ eine hohe Wahrscheinlichkeit für kleine Amplituden und entsprechend eine geringe Wahrscheinlichkeit großer Amplituden (immer im Vergleich zum Nutzsignal $s(t)$). Analog kann man formulieren: Der Erwartungswert der Energie des Rauschsignals in einem festen Frequenzbereich ist klein. Daraus folgt, daß die Abspaltung niederenergetischer Anteile einer Unterdrückung von Rauschanteilen im Signal entspricht. Den Energie-Filter definiert man folgendermaßen:

Def.: Es sei $s(t)$ ein Signal und $S(u)$ seine Fouriertransformierte, dann bezeichne $E(f_1, f_2)$ die Energie des Signals im Frequenzband $[f_1, f_2]$. Der Grenzwert

$$(3.16) \quad L(u) = \lim_{\Delta u \rightarrow 0} \frac{E(u, u + \Delta u)}{\Delta u}$$

heißt Energiedichte des Signals $s(t)$.

$$(3.17) \quad G_{L_0}(u) = \begin{cases} 1; & L(u) \geq L_0 \\ 0; & \text{sonst} \end{cases}$$

heißt *Energiefilter*.

Zur Bestimmung der Energiedichte $L(u)$: Die Energie eines Signals $s(t)$ bestimmt man aus

$$E = \int_{-\infty}^{\infty} s^2(t) dt.$$

Die Parsevalsche Beziehung liefert eine entsprechende Darstellung im Frequenzbereich:

$$E = \int_{-\infty}^{\infty} s^2(t) dt = \int_{-\infty}^{\infty} |S(u)|^2 du.$$

Die Energie im Frequenzband $[f, f + \Delta f]$ ergibt sich also zu

$$E(f, f + \Delta f) = \int_f^{f+\Delta f} |S(u)|^2 du$$

und damit folgt die Energiedichte

$$(3.18) \quad L(f) = \lim_{\Delta f \rightarrow 0} \frac{E(f, f + \Delta f)}{\Delta f} = \lim_{\Delta f \rightarrow 0} \frac{1}{\Delta f} \int_f^{f+\Delta f} |S(u)|^2 du$$

$$\Rightarrow L(f) = |S(f)|^2$$

Im 2-dimensionalen Fall berechnet man zunächst $S(u, v) = \mathcal{F}(s(x, y))$ und bestimmt dann $L(u, v) = |S(u, v)|^2$ und daraus die Übertragungsfunktion $G(u, v)$. In Abbildung 20 wird die Arbeitsweise des einfachen Energiefilters über dem HOPG-Bild demonstriert.

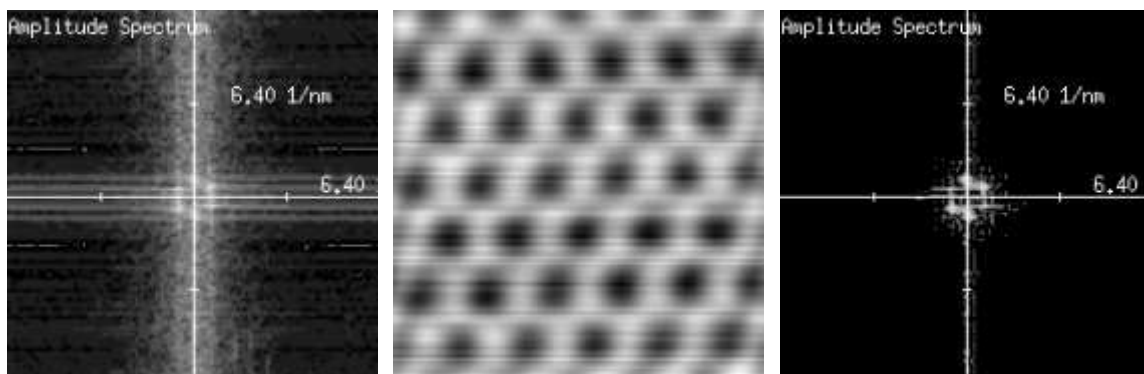


Abb. 20.: Fouriertransformierte des Originalbildes, gefiltertes Bild und Fourierbereich

Es stellt sich jetzt die Frage, wie gut man Signal und Rauschen voneinander trennen kann. Dazu betrachte man das folgende Beispiel. Sei $s(t) = 2 \sin(\pi x) + \cos(4\pi x) + 2.5 \cdot \psi(1)$, die Funktion $\psi(x)$ liefert gleichverteilte Zufallszahlen im Intervall $[-x, x]$ (Abb. 21).

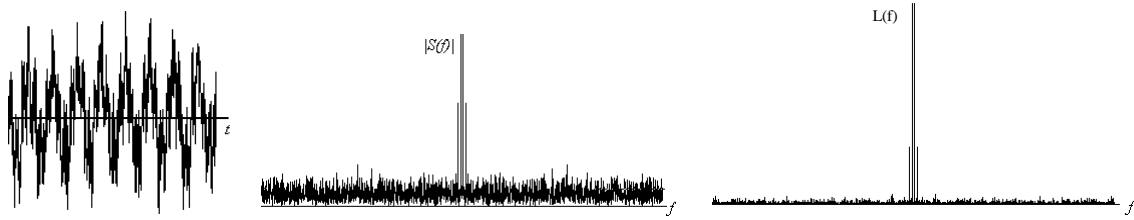


Abb. 21.: Signal, zugehöriges Amplitudenspektrum und Energiedichtespektrum

Man kann vermuten, daß die verschiedenen Spektren eine unterschiedlich gute Trennung zulassen. Um eine Abschätzung der Trenngüte geben zu können, wird jetzt das mittlere Trennverhältnis zwischen zwei Spektren A und B eingeführt:

Def.: Das mittlere Trennverhältnis $\Theta(A, B, f_1, f_2)$ im Frequenzband $[f_1, f_2]$ wird erklärt als:

$$(3.19) \quad \Theta(A, B, f_1, f_2) = \frac{\Sigma(A, f_1, f_2)}{\Sigma(B, f_1, f_2)}$$

mit dem Spektralwert (siehe auch [Kress89])

$$(3.20) \quad \Sigma(X, f_1, f_2) = \frac{1}{(f_2 - f_1)} \int_{f_1}^{f_2} |X(u)| du$$

für ein kontinuierliches Spektrum und

$$(3.21) \quad \Sigma(X, f_1, f_2) = \frac{1}{n_2 - n_1} \sum_{i=n_1}^{n_2} |X(u_i)|; \quad n_i : u_{n_i} \leq f_i < u_{n_i+1}$$

für den diskreten Fall.

Je größer $|\log(\Theta(A, B, f_1, f_2))|$ ist, desto besser sind die Signale A und B im Frequenzband $[f_1, f_2]$ voneinander unterscheidbar.

Bemerkung: Für ein stochastisches Signal $n(t)$ wird $\lim_{\substack{f_1 \rightarrow -\infty \\ f_2 \rightarrow \infty}} \Sigma(N, f_1, f_2) = \mu_N$; μ_N ist der

Erwartungswert von N . Zur Abschätzung von $\Sigma(N, f) = \Sigma(N, -f, f)$ aus $x(t) = s(t) + n(t)$ mit $n(t) = A \cdot \psi(1)$ kann man zunächst so argumentieren: Das (bandbegrenzte) Nutzsinal S nimmt nur einen sehr geringen Teil des Frequenzbereiches ein, das Rauschen N ist hingegen über das gesamte Spektrum verteilt. Es erscheint deshalb sinnvoll, zur Schätzung für $\Sigma(N, f)$ den Modalwert von X zu verwenden.

Def.: Sei Z eine (stetige) Zufallsgröße, dann liegt ein Modalwert $\text{Modal}(Z)$ an der Stelle, an der die Dichtefunktion ein relatives Maximum hat.

Die Dichtefunktion kann durch eine Klasseneinteilung geschätzt werden.

Der Modalwert $\text{Modal}(Z, k)$ ist dann die Klassenmitte, für die bei einer Einteilung in k Klassen die zugehörige Klasse die meisten Elemente enthält.

Hier soll der Modalwert betrachtet werden, für den die Dichte ein globales Maximum annimmt. Damit kann man jetzt einen signalgesteuerten Filter (SDF) konstruieren. Hierfür sei $\Sigma(N, f_{\max}) \approx \mu_N \approx \hat{\mu}_N = \text{Modal}(X, k)$ ($\text{Modal}(X, k)$ der Modalwert mit k Klassen der Menge

$\mathbf{X} = \{|X_i| : X_i \text{ ist Wert der DFT von } x(t)\}$ die Schätzung für den Spektralwert des Störsignals n auf dem Signal $x(t) = s(t) + n(t)$. Man filtert dann derart, daß alle Komponenten, für die $|X_i| \geq \hat{\mu}_N$ gilt, als Signalkomponenten akzeptiert werden. Die nicht als Signal akzeptierten Komponenten werden auf 0 gesetzt. Das mittlere Trennverhältnis ergibt sich jetzt zu

$$\Theta(S, N, -f_{\max}, f_{\max}) = \frac{\Sigma(S, f_{\max})}{\Sigma(N, f_{\max})}$$

$$= \frac{|\mathbf{X}| - |\mathbf{X}_S|}{|\mathbf{X}_S|} \cdot \frac{\sum_{X_i \in \mathbf{X}_S} |X_i|}{\sum_{X_i \in \mathbf{X} \setminus \mathbf{X}_S} |X_i|}; \mathbf{X}_S = \{X_i : |X_i| \geq \hat{\mu}_N\}.$$

Abbildung 22 zeigt die Wirkung des soeben beschriebenen Filterverfahrens.

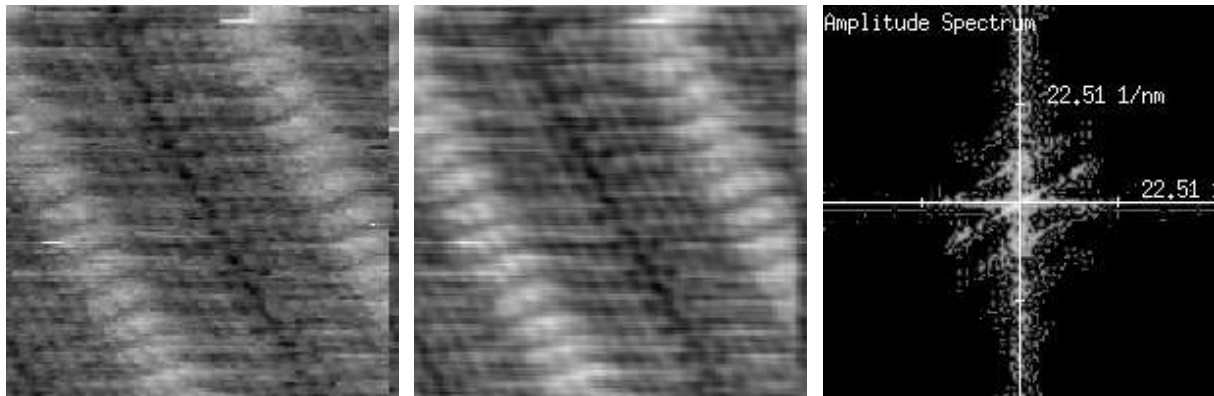


Abb. 22.: Didodecyl-Benzen (DDB) Moleküle, links: Originaldaten, Mitte: μ_N Abtrennung des Signals im Amplitudenspektrum, rechts: DFT des gefilterten Bildes

Wieder fragt man sofort, ob sich das Trennverhältnis unter Benutzung von Spektren höherer Ordnung, wie zum Beispiel dem Energiedichtespektrum, noch verbessern läßt. Sei die Anzahl c der Klassen zur Bestimmung des Modalwertes so gewählt, daß die Breite $\Delta_1 X$ der Klassen für $|X|^l$ größer gleich 1 ist, dies gilt dann auch für die Einteilungen von $|X|^p$. Sei weiterhin der Modalwert von $|X|^p$ Element der k -ten Klasse, also $\hat{\mu}_{N^p} \geq (k-1) \cdot \Delta_p X$ und $\hat{\mu}_{N^p} < k \cdot \Delta_p X$; $\Delta_p X = (\max\{|X_i|^p\} - \min\{|X_i|^p\}) / c$. Dann folgt $\hat{\mu}_{N^{p+1}} < k \cdot \Delta_{p+1} X$ (siehe dazu Abb. 23), d.h. der Modalwert kann bei wachsendem p nicht in eine höhere Klasse springen, woraus folgt, daß sich das Trennverhältnis nicht verschlechtern kann. Seien jetzt die Rauschanteile $|N_i|$ alle kleiner als eine gewisse Schranke ϵ , dann rücken bei wachsendem p der Modalwert $\hat{\mu}_{N^p}$ und ϵ^p immer mehr zusammen. Die Trennung zwischen Nutzsignal und Rauschen wird damit besser. Ideal wären p und k derart, daß die Klasse, durch die $\hat{\mu}_{N^p}$ bestimmt wird, nach oben durch ϵ^p begrenzt wird.

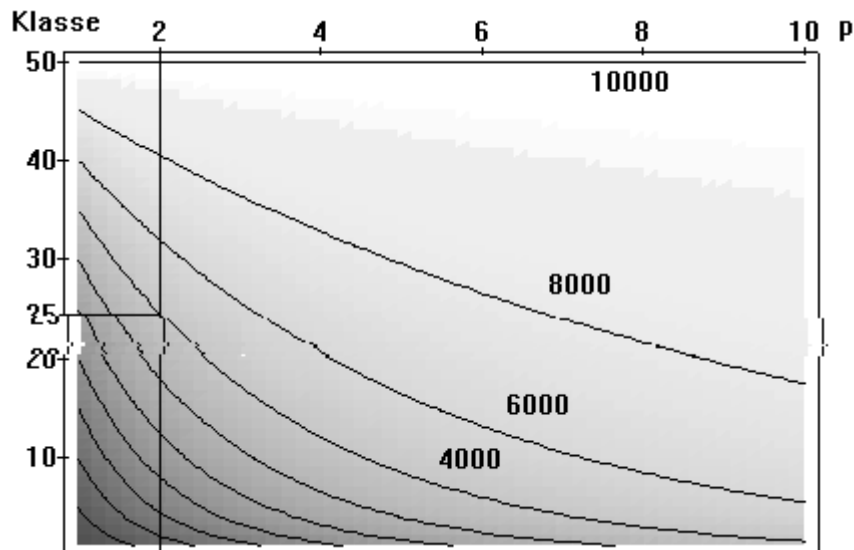


Abb. 23.: Das Diagramm zeigt an, welche Amplitudenwerte ($|X^l|$) bei der Verwendung von 50 Klassen bei der Klasseneinteilung im Spektrum $|X^p|$ in der Klasse k zu finden sind.
Bsp.: Sei $p=2$, befindet sich der Amplitudenwert 4000 in der Klasse 25.

3.5.5. Inverser- und Wiener-Filter

Ein, durch seine Übertragungsfunktion H beschriebenes, abbildendes System, verändert ein Eingangssignal S so, daß das Ausgangssignal \tilde{S} die Gestalt $\tilde{S} = S \cdot H + N$ annimmt (N charakterisiert additives Rauschen). Um das Signal s zu rekonstruieren, wendet man den Filter $1/H$ an und erhält $\tilde{S} / H = S + N / H$. Dabei kommt es allerdings zu Problemen, falls H Werte gleich oder nahe 0 hat. In diesen Fällen ist die Rekonstruktion des Signals s nicht möglich. Sinnvoll erscheint es hier, den Filter nur dort anzuwenden, wo H Werte ungleich 0 hat. Der Filter wird in [Rose82] genauer erläutert.

Das nächste, wesentlich größere Problem stellt jedoch die Bestimmung der Übertragungsfunktion H dar. Es seien s_1 und s_2 zwei Realisierungen des Signals s mit $s_1 = s * h + n_1$ und $s_2 = s * h + n_2$, die durch unkorrelierte Störsignale n_1 bzw. n_2 überlagert wurden. Dann gilt für die Kreuzkorrelation $C_{s_1, s_2} = C_{s * h, s * h} + C_{s * h, n_1} + C_{s * h, n_2} + C_{n_1, n_2}$. Da n_1 und n_2 jeweils unkorreliert mit s und h und untereinander unabhängig sind, gilt also (wenigstens in sehr guter Näherung) $C_{s_1, s_2} = C_{s * h, s * h}$. Aus dem Wiener-Chinitchin Theorem und (3.18) folgt sofort

$$(3.22) \quad \mathcal{F}(C_{s_1, s_2}) = L_{s * h} = |S|^2 \cdot |H|^2$$

und damit zumindest eine Aussage über den Betrag der Übertragungsfunktion H . Führt man dem System ein bekanntes Signal $S_0 \neq 0$ zu und bildet dann (eventuell nach Filterung von \tilde{S} mit G zur Beseitigung von N)

$$(3.23) \quad \tilde{H} = \frac{G \cdot \tilde{S}}{S_0},$$

so erhält man eine Abbildung, die sicherlich in der Nähe von H anzutreffen ist. Leider ist die einzig mögliche Abschätzung $\|H - \tilde{H}\| \leq \max_{x \in C} \{|H(x)| + |\tilde{H}(x)|\}$ nicht sehr aussagekräftig.

Damit ist die Wiener-Filterung gemäß

$$G = \frac{1}{H} \cdot \frac{|H|^2}{|H|^2 + S_n / S_s} \approx \frac{1}{H} \cdot \frac{|H|^2}{|H|^2 + SNR}$$

(S_n - Fouriertransformierte der Autokorrelation des Rauschens und S_s die der AKF des Signals) zwar optimal bezüglich $\|s - g * \tilde{s}\| \rightarrow \min!$, jedoch meist nicht verwendbar, da sich H nicht oder nicht gut genug bestimmen läßt. Die Konstante SNR heißt *signal to noise ratio* und gibt das Verhältnis von Nutzsignalenergie zu Rauschenergie an.

3.5.6. Signalangepaßter Tiefpaß-Filter

Der Tiefpaß-Filter wurde bereits unter 3.5.2. erläutert. Zur Bestimmung der Übertragungsfunktion eines Tiefpaßfilters wird jetzt seine Einwirkung auf ein Signal simuliert. Dazu betrachte man das Verhältnis der Signalenergie am Ausgang des Filters zur Gesamtenergie des Eingangssignals als Funktion der Grenzfrequenz.

Abbildung 24 zeigt eine solche Simulation. Dreht man die Energiecharakteristik mittels $1 - E_p(f)/E$ ($E_p(f) := E(-f, f)$), so erhält man eine Kurvenform, die der Gestalt der Übertragungsfunktion eines Tiefpasses sehr ähnlich sieht. Auffällig sind die Knickpunkte bei 20 und 70. Geht man wie üblich davon aus, daß dem Nutzsignal additiv ein Rauschen überlagert ist, so kann man anhand der Kurve argumentieren:

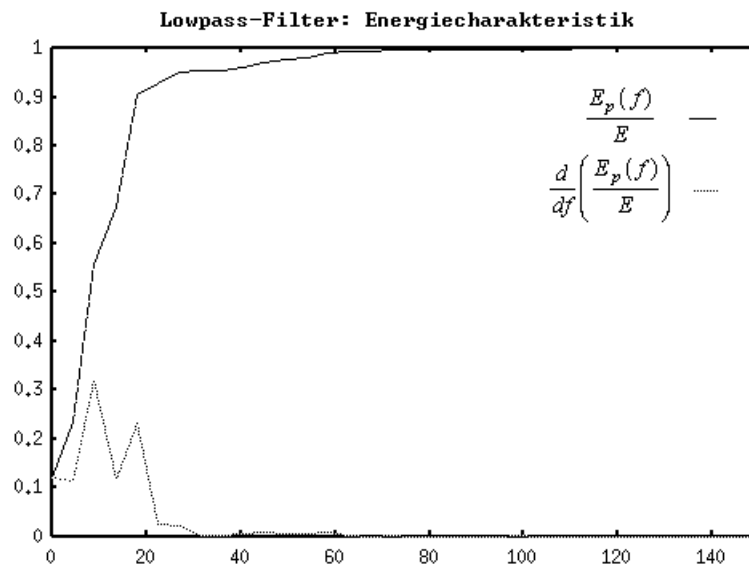


Abb. 24.: Simulation der Tiefpaßfilterung von Zeile 80 aus dem linken Bild von Abb. 13

Der Energieanteil oberhalb $u=70$ liefert keinen wesentlichen Beitrag zur Gesamtenergie und kann deshalb vernachlässigt werden. Es erscheint also sinnvoll, das Signal durch einen Tiefpaß mit der Grenzfrequenz $u_g=70$ zu filtern. Dieser Wert entspricht auch in etwa der per Hand ermittelten, scheinbar optimalen Grenzfrequenz (3.4.1., Abb. 15). Durch Wegschneiden aller Frequenzanteile über $u=70$ kann somit ein großer Rauschanteil eliminiert werden und die Eigenschaften des Nutzsignals bleiben fast unbeeinflusst.

Die Festlegung der Grenzfrequenz für einen idealen Tiefpaß erscheint interaktiv relativ einfach möglich zu sein. Gesucht ist aber ein Kriterium, welches diesen Vorgang automatisiert, um die Subjektivität des Menschen auszuschalten. Man verwendet daher keinen idealen Tiefpaß, sondern wie schon vermutet, $G(u) = 1 - E_p(|u|)/E$ als Übertragungsfunktion. $G(u)$ hat Tiefpaßcharakteristik ($G(0)=1$ und $\lim_{|u| \rightarrow \infty} G(u) = 0$). Die Abbildungsfolgen 26 und 27 zeigen die Simulation eines MELP (Matched Energy Lowpass)-Filters bei der Anwendung auf das Signal

$\sin(\pi \cdot x) + a \cdot \psi(1)$, Folge 28 für die Funktion $2\sin(\pi \cdot x) + 2\cos(4\pi \cdot x) + a \cdot \psi(1)$. Die MELP-Filterung scheint für diese Signale sehr gut zu arbeiten. Allerdings betrachte man Abbildung 28.b, dort werden die zu 4π korrespondierenden Anteile so gedämpft, daß das gefilterte Signal extrem verfälscht wird. Noch deutlicher kann man dies anhand der Funktion $2\sin(\pi \cdot x) + \cos(4\pi \cdot x) + a \cdot \psi(1)$ demonstrieren. Folgerung: Der MELP-Filter arbeitet gut bei einfachen Signalen, bei denen die Energien der einzelnen Frequenzanteile etwa gleich groß sind. Andernfalls kann es vorkommen, daß man ein im Vergleich zum Nutzsignal verfälschtes Ausgangssignal erhält. Der Übergang zum 2-dimensionalen Fall ist wieder trivial. Man berechne dazu

$$(3.25) \quad E(u, v) = \int_{-|u|}^{|u|} \int_{-|v|}^{|v|} |S(\tilde{u}, \tilde{v})|^2 d\tilde{v} d\tilde{u}$$

sowie

$$(3.26) \quad G(u, v) = 1 - \frac{E(u, v)}{E(u_{\max}, v_{\max})}.$$

Für (diskrete) Bilder muß man die Integration natürlich wieder durch eine Summenbildung ausdrücken. Es ist dann

$$(3.27) \quad G_{i,j} = 1 - \frac{\sum_{k=-|i|}^{|i|} \sum_{l=-|j|}^{|j|} |S_{k,l}|^2}{\sum_{k=-|n_x/2|}^{|n_x/2|} \sum_{l=-|n_y/2|}^{|n_y/2|} |S_{k,l}|^2}.$$

Abbildung 25 zeigt die Anwendung auf das HOPG-Bild.

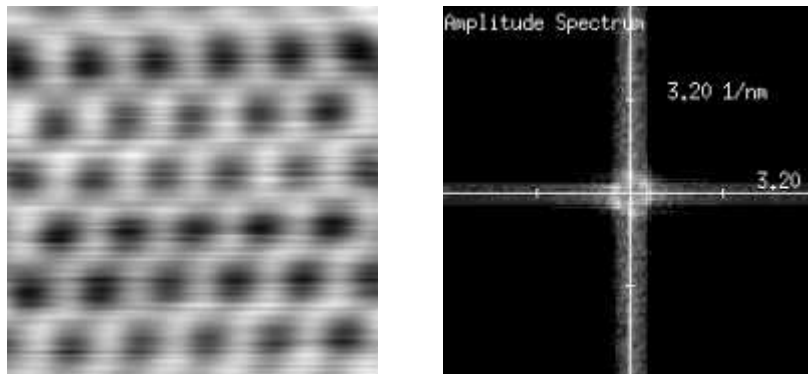


Abb. 25.: 2-fach gefiltertes HOPG-Bild und Fouriertransformierte

Bunte Bildchen MELP Filterung

Abb. 26.a: $\sin(\pi x) + 2\psi(1)$, 1-facher MELP

Abb. 26.b: $\sin(\pi x) + 2\psi(1)$, 2-facher MELP-Filter

Abb. 27.a: $2\sin(\pi x) + 2\psi(1)$, 1-facher MELP

Abb. 27.b: $2\sin(\pi x) + 2\psi(1)$, 2-facher MELP

Abb. 28.a: $2\sin(\pi x) + 2\cos(4\pi x) + 2\psi(1)$, 1-facher MELP

Abb. 28.b: $2\sin(\pi x) + 2\cos(4\pi x) + 2\psi(1)$, 2-facher MELP

Abb. 1.

Abb. 2.

Abb. 3.

3.5.7. Digitale Filter im Ortsbereich

In diesem Abschnitt findet nicht wie bisher erst eine Diskussion im 1-Dimensionalen statt, sondern gleich im Bildbereich. Wie bereits erläutert, handelt es sich bei digitalen Filtern im Ortsbereich um gewichtete Summen. Um die Arbeitsweise des Filters erklären zu können, müssen zuerst einige Begriffe eingeführt werden. Es sei zunächst $F \in \mathbf{R}^{2n+1, 2m+1}$; n, m fest $\in \mathbf{N}$, dann heißt $P(i, j) \in S$ ein:

- innerer Punkt, falls $n \leq i < n_y - n \wedge m \leq j < n_x - m$
- Randpunkt, falls $(\neg(n \leq i < n_y - n) \wedge m \leq j < n_x - m)$
oder $(n \leq i < n_y - n \wedge \neg(m \leq j < n_x - m))$
- Eckpunkt, falls nicht innerer Punkt und nicht Randpunkt
- äußerer Punkt, falls $P(i, j) \notin S$.

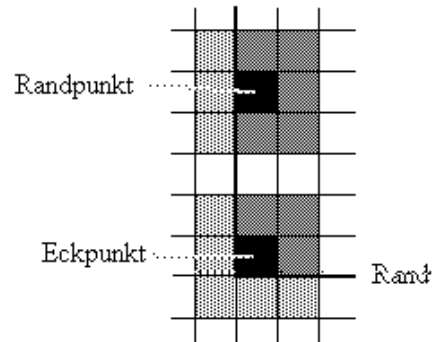


Abb. 29.: Spezielle Punkte

Man erkennt sofort, daß der Filter F nicht auf allen Bildpunkten erklärt ist. Für alle Punkte, die nicht innere Punkte sind, muß er eine andere Gewichtsmatrix haben. Meist definiert man einfach

$$(3.28) \quad \tilde{s}_{i,j} = \sum_{k=-n}^n \sum_{l=-m}^m in(i-k, j-l) \cdot F_{k,l} \cdot s_{i-k, j-l}; \quad in(\mu, \nu) = \begin{cases} 1; & s_{\mu, \nu} \in S \\ 0; & \text{sonst} \end{cases}$$

Damit haben die äußeren Punkte keinen Einfluß auf die Filterung. Sinnvoll kann aber auch die periodische Fortsetzung des Bildes über den Rand hinaus sein (siehe Fouriertransformation). Im weiteren werden einige wichtige digitale Filter erläutert:

a) convolution lowpass Filter oder bewegter Mittelwert

$$(3.29) \quad F = \frac{1}{(2n+1)(2m+1)} \cdot \underline{1}; \quad \underline{1} \in N^{2n+1, 2m+1} \wedge \underline{1}_{i,j} = 1.$$

Sei $\mathfrak{R} = s + r$ ein, durch unkorreliertes Rauschen additiv überlagertes Signal/Bild und hat r den Erwartungswert 0 und die Streuung σ^2 , dann gilt nach [Haber89]

$$E(r'_{i,j}) = E\left(\frac{1}{(2n+1)(2m+1)} \sum_{k=-1}^1 \sum_{l=-1}^1 r_{i-k, j-l}\right) = \frac{1}{(2n+1)(2m+1)} \sum_{k=-1}^1 \sum_{l=-1}^1 E(r_{i-k, j-l}) = 0$$

$$E(r'_{i,j}{}^2) = E\left(\left(\frac{1}{(2n+1)(2m+1)} \sum_{k=-1}^1 \sum_{l=-1}^1 r_{i-k, j-l}\right)^2\right)$$

und

$$= \left(\frac{1}{(2n+1)(2m+1)}\right)^2 E\left(\left(\sum_{k=-1}^1 \sum_{l=-1}^1 r_{i-k, j-l}\right)^2\right)$$

$$= \dots = \left(\frac{1}{(2n+1)(2m+1)}\right)^2 \cdot \sigma^2$$

Das heißt, das Rauschen im Bild wird für innere Punkte um den Faktor $((2n+1) \cdot (2m+1))^2$ gedämpft, analog für Randpunkte um einen geringeren Faktor.

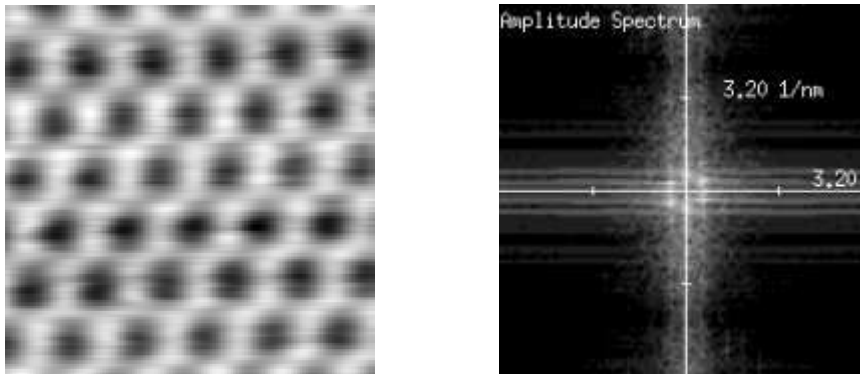


Abb. 30.: convolution lowpass des HOPG-Bildes und Fouriertransformierte

b) Differenz-, Sobel- und Laplace-Operator

Die im folgenden zu behandelnden Operatoren stellen keine Filter im Sinne der Definition dar. Sie werden aber hier beschrieben, da sie meist im Zusammenhang mit digitalen Filtern im Ortsbereich erscheinen und durch die gleiche Vorschrift berechnet werden.

Man betrachte die Matrix $F = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \in \mathbf{Z}^{1,2} \Rightarrow n=0, m=1$. Wurde das Bild durch äquidistante Abtastung gewonnen, ist also der Abstand in X-Richtung $=\Delta x = \text{const.}$, so ergibt $s_{i,j} = s_{i,j+1} - s_{i,j} = \Delta s$ mit $\Delta s / \Delta x$ den Differenzenquotienten in X-Richtung. Analog erhält man für F^T den Differenzenquotienten in Y-Richtung. Die einfache Differenzenbildung ist in der Praxis zu anfällig gegenüber Störungen, etwa Rauschen ([Haber89]). Man verwendet deshalb oft den Sobel-Operator, um Grauwertübergänge zu extrahieren. Hierfür benutzt man die Approximation $s_x(x, y) \approx \frac{s(x+h, y) - s(x-h, y)}{2h}$ und bildet dann das gewichtete Mittel $(s_x(x, y-h) + 2s_x(x, y) + s_x(x, y+h)) / 4$. Die Skalierung $1/8h$ kann im Prinzip weggelassen werden, wenn sie bei der Interpretation des Ergebnisses berücksichtigt wird. Der Sobel-Operator wird durch

$$(3.30) \quad H_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{und} \quad H_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix},$$

mit H_x für die Ableitung nach x und H_y für die Ableitung nach y , beschrieben. Auch hier muß der Rand gesondert behandelt werden, beispielsweise mit einer analogen Mittelung über dem einfachen Differenzenoperator. Die Matrizen sind so gestaltet, daß sie mit der Berechnungsvorschrift aus 3.4.2. konform sind.

Um den Gradienten eines Bildes zu bestimmen, muß man gemäß

$$(3.31) \quad \nabla(s)(x, y) = \left[\frac{\partial}{\partial x} s(x, y) \quad \frac{\partial}{\partial y} s(x, y) \right]^T$$

nur noch $s_x(x, y)$ und $s_y(x, y)$ berechnen. Damit ergeben sich dann sofort Richtung und Betrag (Abbildung 32).

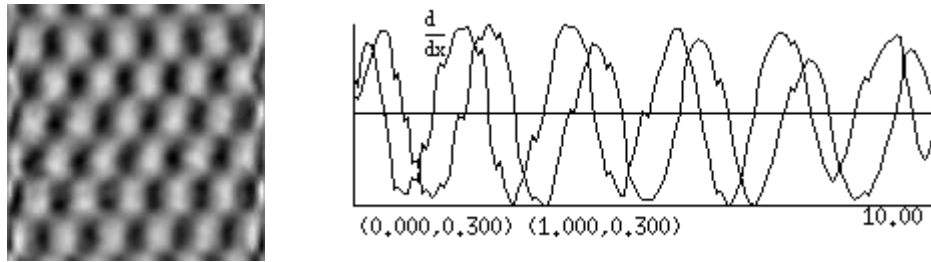


Abb. 31.: Sobel-Operator H_x auf Tiefpaß gefiltertes HOPG-Bild angewandt, die rechte Abbildung zeigt ein Zeilenprofil mit Ableitung (Profil aus H_x -Bild)

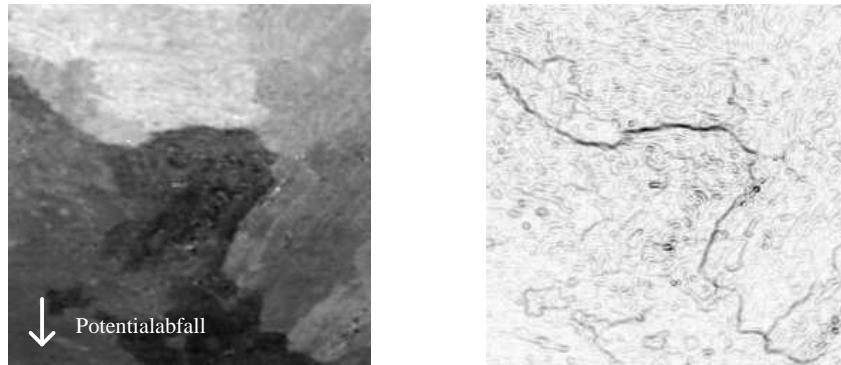


Abb. 32.: Das linke Bild zeigt den Potentialabfall über einer, mit Gold bedampften Kohlenstoffprobe rechts ist der Betrag des Gradienten zu sehen. Sehr deutlich tritt die Stufe hervor

Ausgehend von der Diskretisierung des Laplace-Operators $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ folgt aus

$$(3.32) \quad \Delta s_{i,j} = \frac{s_{i,j+1} - 2s_{i,j} + s_{i,j-1}}{(\Delta x)^2} + \frac{s_{i+1,j} - 2s_{i,j} + s_{i-1,j}}{(\Delta y)^2}$$

bei äquidistanter Abtastung mit $\Delta x / \Delta y = c$ die Filtermatrix

$$(3.33) \quad F = \begin{bmatrix} 0 & 1/c^2 & 0 \\ 1 & -2(1+1/c^2) & 1 \\ 0 & 1/c^2 & 0 \end{bmatrix} \text{ bzw. } F = \begin{bmatrix} 0 & 1 & 0 \\ c^2 & -2(1+c^2) & c^2 \\ 0 & 1 & 0 \end{bmatrix}.$$

Die in diesem Abschnitt behandelten Operatoren sind weniger geeignet, die Qualität eines Bildes zu verbessern, sie liefern vielmehr Informationen über die innere Struktur der Daten und spielen demzufolge in der Bildauswertung eine große Rolle (siehe 4. Datenanalyse).

c) Rangordnungsoperatoren (nach [Haber89])

Es sei $F_{k,l} \in \{0,1\}$ für alle Elemente von F , dann beschreibt F die Nachbarschaftsbeziehungen in der n,m -Umgebung eines Punktes $S_{i,j}$ ($S_{k,l}$ heißt Nachbar von $S_{i,j}$, falls $F_{k-i,l-j}=1$). Es sei weiterhin $(p)_n$ die geordnete Folge der Werte $s_{k,l}$ aller Nachbarn von $S_{i,j}$. Dann heißt (F, med) Median-Filter, wenn $\tilde{s}_{i,j} = med((p)_n)$. Der Median-Filter ist für die Dropout-Elimination geeignet, verändert aber bei der Anwendung auf alle Punkte des Bildes auch die nicht gestörten. Besser wäre es, (F, med) nur zur Detektion zu verwenden und dann die gestörten Punkte durch Interpolation zu den Nachbarn zu beseitigen (siehe 3.3.2.).

3.5.9. Schlußfolgerungen

Den idealen Filter gibt es nicht. Jedoch erweist es sich als günstig, den interaktiv gesteuerten Tiefpaß, den MELP, den ebenfalls interaktiv gesteuerten Energiefilter sowie den signalgesteuerten Filter zu verwenden. Es wurde gezeigt, daß ein Tiefpaßfilter durch den convolution lowpass ersetzt werden kann.

Für den MELP und den signalgesteuerten Filter konnte gezeigt werden, daß die Struktur des Bildinhaltes unverändert bleibt und diese Filter deshalb bevorzugt einzusetzen sind, wenn nachfolgend Auswertungen im Bild stattfinden sollen. Der Erhalt der Signalstruktur ist beim MELP eingeschränkt gemäß 3.5.6.

"Schönere", für Repräsentationen geeignete Bilder liefern hingegen der Tiefpaß und der convolution lowpass. Allerdings sind die Ausgaben dieser Filter in bestimmten Fällen Strukturveränderungen unterworfen (siehe zum Beispiel Abbildung 15). Es läßt sich nur durch den Vergleich mit den Originaldaten, mit MELP oder mit SDF gefilterten Daten erkennen, ob die Bildstruktur erhalten geblieben ist.

3.6. Elimination systematischer Störungen

Systematische Störungen werden durch das abbildende System hervorgerufen. Kann man die Übertragungsfunktion des Systems bestimmen, so kann man mit dem Inversen- oder dem Wiener-Filter eine Korrektur durchführen. Andernfalls muß man sich mit der Elimination von wenigen, beschreibbaren Komponenten (im Beispiel der Rastersondenmikroskopie: Kriechen, Drift und Hysteresis) der Übertragungsfunktion begnügen.

a) Wird die an einem piezokeramischen Element angelegte Spannung geändert, so ändert dieses seine Länge, wobei allerdings der durch ein Modell von VIEIRA (in [Vieira86]) beschriebene Kriecheffekt auftritt

$$\Delta l = \Delta U \cdot (\alpha + \beta \cdot \ln t), t \geq 1.$$

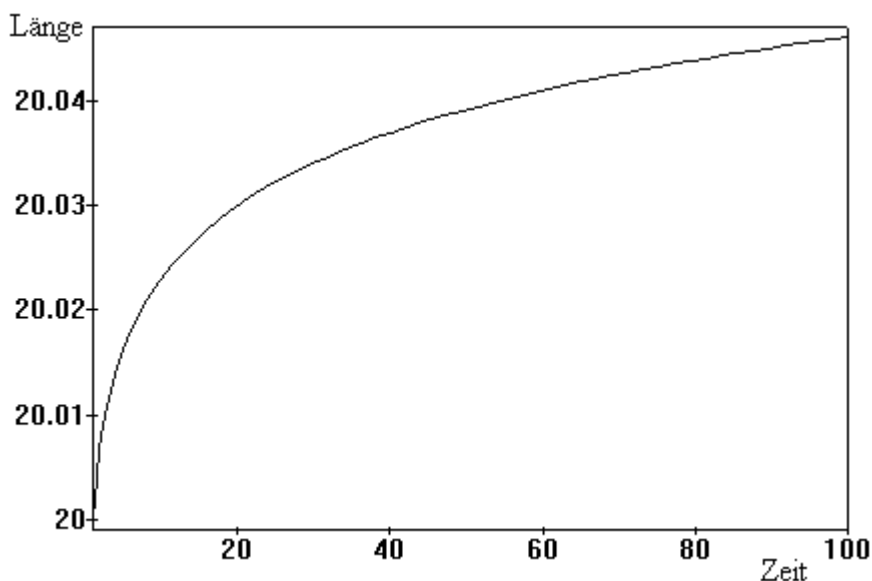


Abb. 33.: Kriecheffekt bei konstantem ΔU , α und β sind hier willkürlich gewählt

Bezugnehmend auf [Joerg93] scheint eine automatische Ermittlung der Parameter des Modells aus einer Messung so gut wie unmöglich zu sein. Außerdem kann man den Effekt zumindest in x-Richtung durch eine geschickte Wahl der Steuerprozedur und hinreichend schnelles Messen

schon im Experiment verringern. Im IFW wird mit einem Zeilen- und Spaltenvorlauf gearbeitet. Dabei wird nicht der gesamte abgefahrene Bereich gemessen, sondern nur ein Teil desselben (vgl. Abbildung 34).

b) Ein piezokeramisches Element hat, wie jeder andere Werkstoff auch, die Eigenschaft, sich bei Temperaturänderungen auszudehnen bzw. zusammenzuziehen. Setzt man eine lineare Temperaturabhängigkeit voraus, so ist

$$(3.34) \quad dl = l \cdot c_l \cdot dT_p.$$

c_l ist der Temperaturexpansionskoeffizient der Piezokeramik. c_l ist meist temperaturabhängig. Das nachfolgend hergeleitete Modell ist deshalb nur für einen (kleinen) Temperaturbereich, in welchem c_l als konstant angesehen werden kann, gültig. Um eine Temperaturänderung dT_p des Piezos zu erzielen, ist die Zufuhr einer Energie

$$(3.35) \quad dQ = c_p \cdot m_p \cdot dT_p$$

(c_p - spezifische Wärmekapazität der verwendeten Piezokeramik, m_p - Masse des Piezos) nötig. Unter Berücksichtigung des Wärmeüberganges zwischen dem Piezo und seiner Umgebung

$$(3.36) \quad dQ = \alpha \cdot A \cdot dt \cdot (T - T_p)$$

erhält man das Differentialgleichungssystem

$$(3.37) \quad \begin{aligned} \alpha \cdot A \cdot (T(t) - T_p(t)) &= c_p \cdot m_p \cdot \frac{d}{dt} T_p(t) \\ \frac{d}{dt} l(t) &= l(t) \cdot c_l \cdot \frac{d}{dt} T_p(t). \end{aligned}$$

α ist der Wärmeübergangskoeffizient, A die Oberfläche des Piezos und damit

$$(3.38) \quad T_p(t) = \frac{\alpha \cdot A}{c_p \cdot m_p} \cdot \int_{t_0}^t T(t) \cdot e^{-\frac{\alpha \cdot A}{c_p \cdot m_p} (t-\tau)} d\tau + T_p(t_0) \cdot e^{-\frac{\alpha \cdot A}{c_p \cdot m_p} (t-t_0)}$$

sowie

$$(3.39) \quad l(t) = l(t_0) \cdot e^{c_l \cdot (T_p(t) - T_p(t_0))}.$$

In diesem Modell wird darauf verzichtet, die Veränderung der Fläche zu berücksichtigen. Man kann argumentieren, daß die Änderung der Oberfläche anteilmäßig sehr klein ist. Läßt man außerdem die während des Meßvorganges stattfindende Längenänderung durch Anlegen einer Spannung außer acht, so kann man für die x -Position des Meßkopfes in der Zeile i

$$\tilde{x}_j = l \left(t_i + j \cdot \frac{t_{zeile}}{n-1} \right) - l(t_0) + x_j \text{ beim Hinlauf}$$

und

$$\tilde{x}_j = l \left(t_i + 2 \cdot t_{zeile} - j \cdot \frac{t_{zeile}}{n-1} \right) - l(t_0) + x_j \text{ beim Rücklauf}$$

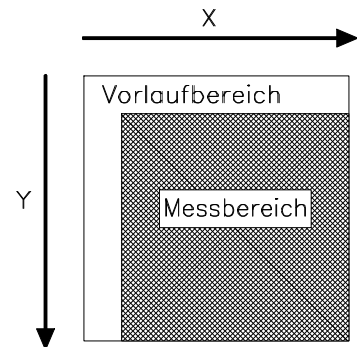


Abb. 34. Steuerung des Meßkopfes zur Vermeidung des Kriecheffektes

angeben. Dieses Modell sollte zumindest eine Abschätzung ermöglichen, wie lange man vor dem Beginn einer Messung warten muß (Temperaturausgleich Meßkopf - Mikroskop), bis die Temperaturdrift vernachlässigt werden kann. Einige wenige Anmerkungen zu den verwendeten Meßpiezos sind im Anhang C.2. zu finden.

c) Sei jetzt $h(x, \underline{c})$ eine Abbildung, die in Abhängigkeit von Parametern \underline{c} , den wahren Ort des Piezos angibt, dann wird nicht $s(x)$ gemessen, sondern $\tilde{s}(x) = s(h(x, \underline{c}))$. h wird Hysterisis genannt, da in Experimenten bisher immer $h_{hin}(0) = h_{rück}(0)$ gemessen wurde. Um diesen Effekt korrigieren zu können, ist es also notwendig, die Parameter \underline{c} zu bestimmen. Setzt man voraus, daß die Parameter \underline{c} bei gleichen Bedingungen (Meßbereich, Temperatur, ...) gleich bleiben, so kann man versuchen, ein vorgegebenes Signal $s(x)$ zu messen und mittels des Verfahrens

$$(3.40) \quad \hat{c}_{n+1} = \hat{c}_n - \varepsilon \cdot \nabla \left(\|s(h(x, \hat{c}_n)) - \tilde{s}(x)\|^2 \right)$$

die c_i zu bestimmen. Ist $\tilde{s}_i = \tilde{s}(x_i)$ und $h(x, \underline{c}) = c_0 \cdot x + \varphi(x, c_1, \dots, c_m)$ sei bis auf die Parameter \underline{c} bekannt, dann kann man in Näherung

$$(3.41) \quad \frac{\partial}{\partial c_i} \|s(h(x, \underline{c})) - \tilde{s}(x)\|^2 \approx \sum_{i=0}^{n-1} 2 \cdot (s(h(x_i, \underline{c})) - \tilde{s}_i) \cdot \frac{s(h(x_i, \underline{c} + \tau \cdot e_i)) - s(h(x_i, \underline{c}))}{\tau}$$

zur Berechnung von $\nabla \|\cdot\|^2$ verwenden. Als Startwert benutze man $c_i = \delta_{i,0}$. Es wird also erwartet, daß kein Hysterisis-Effekt in x auftritt. Ein Signal $s(x)$ wird dann aus $s_{rek}(x) = \tilde{s}(h^{-1}(x, \underline{c}))$ (h sollte eineindeutig sein) rekonstruiert. Bei [Joerg93] wird eine quadratische Ansatzfunktion h vorgeschlagen.

Beispiel: $x=0..100$, $c_0=1$, $c_1=1$, $h(x)=c_0*x+c_1*(1-(x-50)^2/2500)$

Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$s(x)$	0.0	5.3	10.5	15.8	21.1	26.3	31.6	36.8	42.1	47.4	52.6	57.9	63.2	68.4	73.7	78.9	84.2	89.5	94.7	100.0
$s(h(x))$	0.0	6.3	12.4	18.4	24.4	30.2	35.9	41.5	47.0	52.4	57.6	62.8	67.8	72.7	77.6	82.3	86.9	91.4	95.7	100.0
Anzahl Iterationen: 3263 $c_1=1.0000$ $c_2=1.0000$ Abstand=0.000000																				

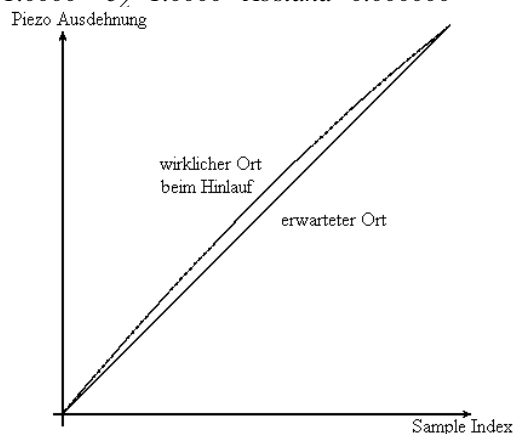


Abb. 35.: Simulierte Hysterisis