

Echtzeit-Vergenzkontrolle eines binokularen Kamerasystems

Michael Hansen und Gerald Sommer

Institut für Informatik und Prakt. Mathematik

Christian-Albrechts-Universität zu Kiel *

Zusammenfassung

Dieser Beitrag beschreibt die Kontrolle der Vergenzbewegung eines aktiven Kamerasystems in Videoechtzeit. Es wird ein phasenbasierter Ansatz zur Berechnung von Disparitätskarten gewählt, der sich zur Implementierung auf der gegebenen Hardwarearchitektur besonders eignet und dem System diese hohe Regelrate ermöglicht. Neben der Filterwahl wird die Implementierung des Regelkreises beschrieben und auf dessen Eigenschaften im realen Experiment eingegangen. Als Anwendung wird eine Kombination dieser bewährten Vergenzregelung mit einer Blickkontrolle in einem aktiven System zur Tiefenkartengenerierung vorgestellt, das die verschiedenen Auflösungen der Disparitätskarten nutzt.

1 Einführung und Idee

Im Paradigma des Aktiven Sehens im Sinne von (Bajcsy, 1988) wird die Kamera in den Prozeß der Bildaufnahme und Bildverarbeitung integriert. In Abhängigkeit von der gestellten Wahrnehmungsaufgabe sind Regelungsstrategien zu entwickeln, die die verschiedenen Parameter eines künstlichen aktiven Sehsystems, wie Position, Blende und Fokus zielgerichtet optimieren. (Ballard & Brown, 1992) beschreiben diesen Ansatz als *animate* oder *purposive Vision* und verknüpfen dieses mit Befähigungen des menschlichen visuellen Systems wie Blickkontrolle und Lernen. Die Aktionen eines solchen aktiven Kamerasystems sind an seine Wahrnehmungsaufgabe gekoppelt. Durch einen Wahrnehmungs-Handlungs-Zyklus (Sommer, 1995) läßt sich z.B. ein Sehsystem zur Entfernungsbestimmung beschreiben. Die Vergenzbewegung (Handlung) des Sehsystems beeinflußt die Genauigkeit der Entfernungsbestimmung (Wahrnehmung) in einer Szene. Die Regelung der Vergenzbewegung ist damit in die Wahrnehmungsaufgabe des Systems integriert und ermöglicht bzw. optimiert diese.

Neben der Blickkontrolle stellt die Kontrolle der Vergenzbewegung für ein künstliches Sehsystem eine wichtige Grundfähigkeit dar. Dabei ist die Notwendigkeit für eine Vergenzbewegung beider Kameras in einem künstlichen System nicht so klar ersichtlich, wie es für das menschliche visuelle System ist, das durch die ortsvariante Auflösung der Retina die Vergenzbewegung der Augen benötigt, um eine optimierte Wahrnehmung zu ermöglichen. Die üblicherweise verwendeten Kameras besitzen mit wenigen Ausnahmen ortsinvariante Pixelauflösungen. In einem aktiven Kamerasystem verbessert die Vergenzbewegung und die damit verbundene Reduzierung der globalen Disparität in der Szene die Eigenschaft von Stereoalgorithmen zur Entfernungsschätzung. Durch die Reduzierung der Disparität können bei einem phasenbasierten Ansatz kleinere Filtermasken verwendet werden. In einem korrelationsbasierten Ansatz wird die Größe des Suchraumes reduziert. Erst die Konzeption der Vergenzkontrolle, verbunden mit spezieller Hardware ermöglicht es zur Zeit, Disparitätsbestimmungen mit aktiven Systemen in Echtzeit von 25 Hz zu realisieren.

Erste Arbeiten zur Vergenzregelung in einem aktiven System erfolgten von (Krotkov, 1989) und (Olsen & Coombs, 1991), die die Disparität mittels Matching- und Cepstrummethode bestimmten und zur Vergenzregelung nutzten. Korrelationsbasierte Ansätze für aktive Systeme wurden u.a. von (Ching, Toh, Chan, & Er, 1993) vorgestellt. In unserer Arbeit zeigen wir einen Zugang zur Bestimmung der horizontalen Stereodisparität, der hinsichtlich der Regelrate optimiert ist. Die Vergenzbewegung des Systems wird durch eine PD-Regelung realisiert. Dabei ist es das Ziel, die Disparität im Zentrum des Blickfeldes zu minimieren. Es wird ein phasenbasiertes Verfahren zur Disparitätsbestimmung genutzt, das hinsichtlich dieser Methodik zur Vergenzkontrolle mit den Arbeiten von (Theimer & Mallot, 1994), (Maki, 1996) und (Westelius, Knutsson, Wiklund, & Westin, 1994) zu vergleichen ist. Auf der Grundlage dieser Vergenzregelung wurde als Anwendung ein aktiver Zugang zur Tiefenschätzung in unbekanntem Szenen entwickelt (Hansen & Sommer, 1996), der in diesem Beitrag am Rande vorgestellt wird.

*Lehrstuhl für Kognitive Systeme, Preußerstr. 1-9, 24105 Kiel, {mha,gs}@informatik.uni-kiel.de

2 Disparitätsbestimmung aus lokalen Phasendifferenzen

Die Idee der Disparitätsbestimmung aus lokalen Phasendifferenzen basiert auf dem Fourier-Verschiebungstheorem

$$(1) \quad f(x) \text{ --- } \bullet F(\omega) \quad f(x + D) \text{ --- } \bullet F(\omega)e^{i\omega D}.$$

Eine Verschiebung D eines Signals $f(x)$ führt im Frequenzraum zu einer Phasenverschiebung ωD und kann dort entsprechend detektiert werden. Um die lokale Disparität bestimmen zu können, ist zunächst die Berechnung der lokalen Phase $\phi(x)$ in beiden Bildern (links/rechts) erforderlich. Aus dieser lokalen Phasendifferenz $\Delta\Phi(x) = \phi_l(x) - \phi_r(x)$ wird direkt die lokale Disparität $D(x) = \frac{\Delta\Phi(x)}{\omega}$ berechnet. Beim *konstanten* Frequenzmodell wird als Frequenz ω die Filtermittenfrequenz gewählt. Die Berechnung einer *lokalen* Frequenz als Ableitung der lokalen Phase kann die Disparitätsschätzung verbessern, erhöht jedoch die Laufzeit des Algorithmus.

Verschiedene Filter sind geeignet, um die lokale Phase zu extrahieren. Von (Westelius, 1995) werden u.a. Gabor-, Difference of Gaussians und Lognormfilter hinsichtlich ihrer Eignung diskutiert. Am meisten verbreitet ist die Anwendung von Gaborfiltern. (Sanger, 1988) und (Fleet, Jepson, & Jenkin, 1991) haben diesen Zugang in einer Dimension entwickelt. Die Gaborfunktion ist definiert durch

$$(2) \quad G(x; \sigma, \omega) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} e^{-i\omega x}.$$

Eine Erweiterung auf zwei Dimensionen findet sich bei (Langley, Atherton, Wilson, & Larcombe., 1990) und (Theimer & Mallot, 1994). Erfolgt die Filterung im Skalenraum als Gabor-Wavelet-Transformation, sind die Parameter über einen Bandbreitenfaktor $t = \frac{1}{\sigma\omega} \in (0, 1)$ gekoppelt.

2.1 Filterwahl für eine echtzeitfähige Implementierung

Zur Disparitätsschätzung werden für unserm Algorithmus approximierte Gaborfilter entwickelt. Dabei hängt die Wahl der Filterparameter (σ, ω) von den Anforderungen der gestellten Aufgabe und der zur Realisierung verwendeten Hardwarekomponenten ab. Ein Filter für eine Echtzeitanwendung ist hinsichtlich seiner Größe σ und nicht unbedingt hinsichtlich seiner Genauigkeit zu optimieren. Das Ergebnis der Disparitätsbestimmung kann dann durch eine Regelung der Vergenzbewegung aktiv verbessert werden. Dagegen ist in einem statischen System mehr Sorgfalt auf die Wahl oder Variation der Filterfrequenz ω zu legen, um die Disparitätsschätzung zu optimieren.

Unsere approximierten Gaborfilter haben eine Größe von 7x7 Pixel und eine horizontale Wellenlänge $\lambda = 2\pi/\omega_x = 6$ Pixel. Sie sind auf die Hardware * optimal abgestimmt, so daßFaltungsoperationen auf Bildern der Größe 512x512 Pixel in 14 ms realisiert werden kann und damit eine Taktrate von 25 Hz ermöglicht.

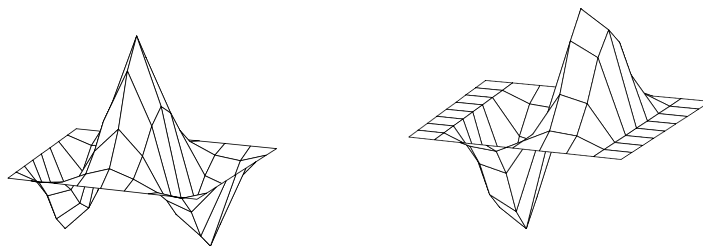


Abbildung 1: Gerader und ungerader Gaborfilter 7x7, Aspektverhältnis $\sigma_x/\sigma_y = 2$.

Die Gaborfilter genügen trotz ihrer geringen Größe den Kriterien, die für Filter zur lokalen Phasenberechnung gefordert sind (Westelius et al., 1994). Entscheidend ist, daß die Phase monoton und ohne ein Phasenumklappen von $-\pi$ zu $+\pi$ verläuft. Um einen optimalen (linearen) Verlauf der lokalen Phase zu erhalten, ist der Gleichanteil des Gaborfilters zu minimieren. Abbildung (2) zeigt (1.) das Fourier Spektrum und den linearen Phasenverlauf des nach (Michaelis & Sommer, 1994) korrigierten Gaborfilters. Das Spektrum ist aufgrund von Randeffekten in der linken (negativen) Halbebene nicht überall gleich null.

*Datacube, MaxVideo 200

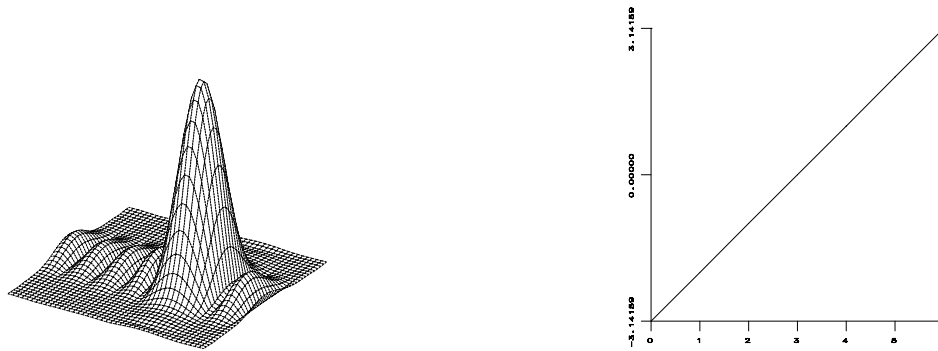


Abbildung 2: Zentriertes Fourier Spektrum und linearer Phasenverlauf von $-\pi$ bis $+\pi$ des bzgl. Gleichanteil korrigierten Gaborfilter.

2.2 Stabilitätsbetrachtungen

Neben der direkten Berechnung muß eine Betrachtung der Stabilität der lokalen Phase $\phi(x)$ einhergehen. Um einen stabilen Phasenwert zu erhalten, muß der Betrag der Filterantwort signifikant sein, dh. es müssen ausreichend Frequenzanteile des Bildes innerhalb der Filterbandbreite liegen. Dieses ist in realen Szenen vorwiegend an Ecken und Kanten der Fall. Wie bei anderen Autoren wird die Stabilität der Phase und damit der Disparität über den Betrag der Gaborfilterantwort (20 % des Maximums) beurteilt. Eine resultierende Konfidenzkarte beschreibt die Orte x , wo die Disparitätsbestimmung in beiden Bildern stabil ist, dh. wo die Summe der Beträge beider Filterantworten einen Schwellwert (40 % des Maximums) übersteigt.

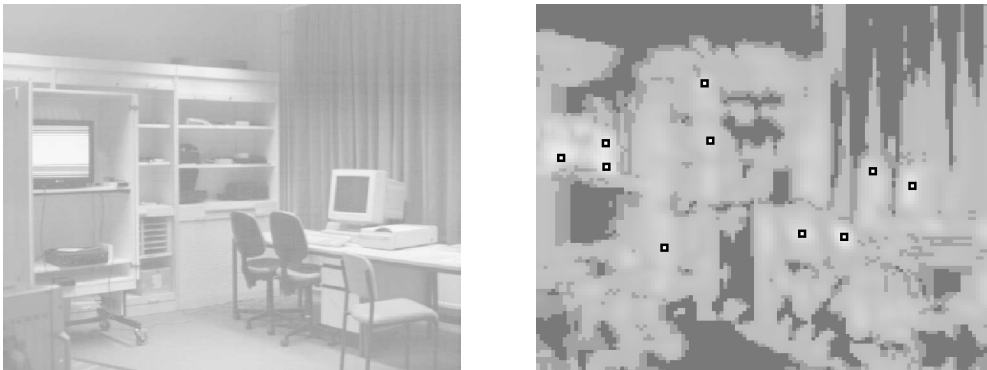


Abbildung 3: Eine Laborszene und die zugehörige Konfidenzkarte stabiler Phaseninformation mit lokalen Maxima.

2.3 Schema der Auflösungs-hierarchie

Die maximal meßbare Disparität ist durch die Filtergröße auf $\pm\lambda/2 = \pm 3$ Pixel begrenzt. In einer realen Szene treten jedoch wesentlich größere Disparitäten auf. Bei paralleler Ausgangsposition beider Kameras ist je nach Distanz zum betrachteten Objekt ein Meßbereich bis zur halben Bildgröße notwendig, um die Vergenzbewegung zu regeln. Dazu ist die Berechnung einer Auflösungs-hierarchie notwendig, die durch Filterung mit Binomialfilter (7x7) und anschließender Unterabtastung realisiert ist. Die Gaborantwort $f_g(x)_i$ berechnet sich wie folgt:

$$(3) \quad f_g(x)_i = G(x, \sigma, \omega) * f_i(x) \quad \text{mit} \quad f_i(x) = \text{Sub}(\text{Low} * f_{i-1}(x)).$$

Die Auflösung variiert dabei in 6 Stufen von 512x512 bis 16x16 Pixel. Die maximal messbare Disparität beträgt damit ± 96 Pixel auf der untersten Stufe der Hierarchie. Die Auflösungsstufe wird während der Vergenzbewegung des Systems je nach berechneter Disparität im Bildzentrum variiert. Treten größere Disparitäten auf, u.a. bei Divergenzstellung der Kameras, sind andere Strategien wie z.B. die Verwendung von *Depth from Focus* zum Erzwingen der Vergenzbewegung erforderlich.

3 Vergenzkontrolle

Das Kieler Visionsystem verfügt über die vier mechanischen Freiheitsgrade Schwenkachse χ , Neigungsachse ϕ und zwei Vergenzachsen θ_l/θ_r . Die rechte Kamera unseres System wird durch eine übergeordnete Blicksteuerung mit Regelung der Schwenk-, Neigungs- und rechten Vergenzachse geführt, die abhängig von der gestellten Aufgabe des Systems ist. Bei dem von der Arbeitsgruppe (Hansen & Sommer, 1996) vorgestellten Modell zur Erkundung der Entfernungen in einer zunächst unbekanntem Umgebung wird die rechte Kamera sequentiell auf interessante Blickpunkte in der Szene fixiert. Lokale Maxima der Konfidenzkarte können solche interessanten Blickpunkte darstellen. Die Vergenzregelung richtet die linke Kamera auf den gleichen Blickpunkt, indem die im Bildzentrum gemessene Disparität D_c in einem geschlossenen Regelkreis reduziert wird. Nach erfolgter Vergenzbewegung läßt sich die absolute Entfernung des fixierten Blickpunktes zum System mittels Triangulation bestimmen.

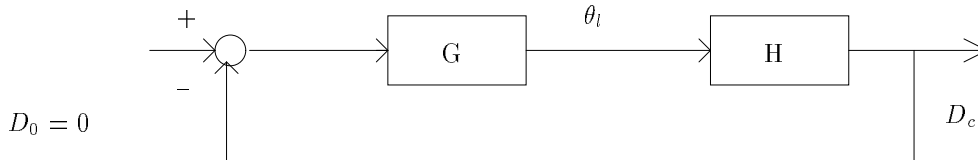


Abbildung 4: Diagramm der Vergenzregelung.

Abbildung 4 zeigt ein Diagramm des Regelkreises. Die gemessene Disparität D_c ist das Referenzsignal des Regelkreises. Die Berechnung der Stellgröße (Vergenzwinkel θ_l) kann durch Gleichungen der inversen Kinematik des Sehsystems beschrieben werden. Der neue Winkel θ_l^* ist eine Funktion der aktuellen Vergenz-, Schwenk- und Neigungswinkel, sowie der Disparität D_c . Im Prinzip ist eine sakkadenhafte Vergenzbewegung innerhalb eines Zeittaktes möglich. Durch Trägheitseffekte der Motorachsen und nicht exakt definierte Zeitintervalle zwischen Bildaufnahme und Achsenbewegungen kommt es in der Anwendung zu einem starken Überschwingen der linken Kamera. Ein eingeführter Dämpfungsfaktor $\alpha \in (0, 1)$ mit $\theta_l^* = \alpha f(\theta_l, \phi, \chi; D_c)$ verbessert das Verhalten. Durch den Übergang zur PD-Reglung, die auch die Komplexität der Berechnung reduziert, konnte eine glatte Vergenzbewegung erzielt werden:

$$(4) \quad \Delta\theta_l = K_p D_c + K_d \dot{D}_c$$

Die Vergenzbewegung ist innerhalb von 10-12 Takten ($t \leq 0.5$ s) abgeschlossen. Es wird dabei keine symmetrische Vergenz erzwungen. Die Koeffizienten K_p und K_d des PD Reglers wurden in einer Simulation mit der Methode nach der Ziegler-Nichols (Franklin, Powell, & Workman, 1992) optimiert.

3.1 Die Vergenzregelung

Die Vergenzregelung arbeitet mit einer Taktrate $t < 40$ ms. Geringe Zeitschwankungen hängen von der aktuellen Auflösungsstufe ab. Der Zyklus startet mit einer Auflösung von 16x16 Pixel und besteht aus folgenden Schritten:

1. Digitalisierung der analogen Videosignale (rechts/links) in einer Auflösung von 512x512 Pixel.
2. Tiefpaßfilterung mit einem Binomialfilter und Unterabtastung bis zur festgelegten Auflösungsstufe i .
3. Faltung der geraden und ungeraden Gabormaske mit beiden unterabgetasteten Bildern.
4. Berechnung der Phase und des Betrages der Gaborantwort in einer Lookup-Table. Dabei wird implizit die Schwellwertoperation des Betrages durchgeführt.
5. Berechnung der Summe der Beträge mit einer zweiten Schwellwertoperation. Das Ergebnis ist eine Konfidenzkarte mit Gebieten stabiler Phasenschätzung.
6. Berechnung der Phasendifferenz. Aufgrund des Modulos bei der 8-Bit-Arithmetik des Bildverarbeitungssystems liegt $\Delta\Phi(x) = \phi_l(x) - \phi_r(x)$ zwischen $-\pi$ and $+\pi$.
7. Die Disparitätskarte \dagger wird mit einer binären Konfidenzkarte maskiert.

\dagger Beim bisher verwendeten konstanten Frequenzmodell unterscheiden sich Phasendifferenzen und Disparitäten nur um den Faktor $1/\omega$.

8. Der Median im Zentrum der Disparitätskarte wird als Fehlersignal für die PD-Regelung berechnet.
9. Der neue Vergenzwinkel θ_i wird errechnet und auf die untergeordnete Motorenregelung übertragen.
10. Falls der gemessene Wert von D_c kleiner ist als die auf einer feineren Skala messbare Disparität, wird die Auflösungsstufe i erhöht.

Die Schritte 1-4 bis werden parallel auf zwei MaxVideo 200 Systemen berechnet. Für die folgenden Schritte 5-7 werden die Daten (Phase und Betrag) auf beiden Systemen derart verteilt, daß die Berechnung der Disparitäts- und Konfidenzkarten parallel erfolgt. Die Zeit für die Bildverarbeitung (1-7) liegt zwischen 33 und 36 ms. Die Schritte 8-10 werden auf einer Workstation berechnet ($t \leq 3ms$).

4 Beispiele und Ergebnisse

Mit einer Taktrate von 25 Hz werden Disparitätskarten unterschiedlicher Auflösungen berechnet, die neben der Vergenzkontrolle auch zur lokalen Entfernungsbestimmung dienen. Die Abbildung 5 zeigt eine Laborszene mit einigen Gegenständen. Die Disparitätskarten zeigen Eigenschaften des phasenbasierten Zuganges. Das weiße Quadrat im Bildzentrum kennzeichnet den Fixationspunkt auf dem linken Flaschenhals. In den schwarzen Gebieten ist die Disparitätsschätzung unzuverlässig. Dieses sind im wesentlichen homogene Regionen, deren Bildstruktur zu gering ist oder nicht im Frequenzbereich des Gaborfilters liegt. Kanten, Ecken und Texturen, die z.B. besonders gut im Muster des Kalibrierungswürfels auftreten, ergeben stabile Disparitätswerte. Helle Bereiche sind näher, dunkle entfernter. Zur Verdeutlichung ist in grauen Gebieten mit schwarzen Pixel der Disparitätswert null kodiert. Die Vergrößerung der Auflösung erweitert den Messbereich und reduziert gleichzeitig jedoch die Filterfrequenz ω . Die einzelnen Objekte bilden nur noch eine zusammenhängende Region. Die Kanten im Hintergrund ergeben eine zunehmend flächenhafte Region stabiler Disparitätsschätzung.

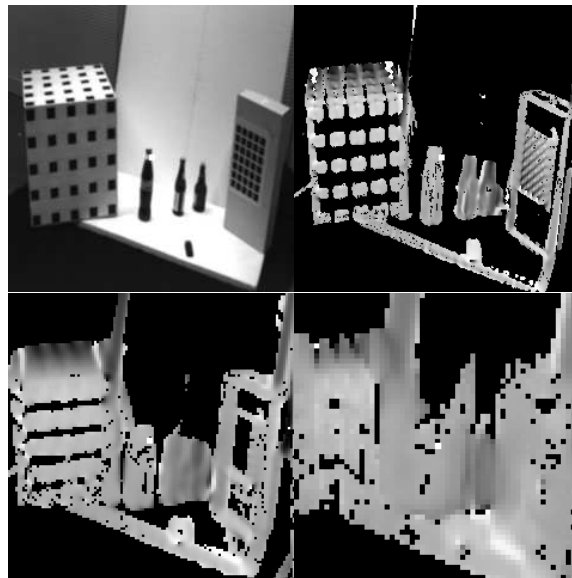


Abbildung 5: (o.l.) Ein Stillleben aus unserem Labor mit den zugehörigen Disparitätskarten in verschiedenen Auflösungsstufen von (o.r.) 256x256 bis (u.r.) 64x64.

Die Disparität im Zentrum dieser Karten werden zur Vergenzkontrolle genutzt. Zur Erhöhung der Stabilität wird der 3x3 Median und nicht der Mittelwert berechnet. Die Vergenzkontrolle des Systems arbeitet stabil mit einer Taktrate von 25 Hz. Probleme treten methodikbedingt bei sehr großen Disparitäten (ab ca. 100 Pixel) auf und in nicht statischen Szenen, wenn z.B. der fixierte Punkt das Blickfeld der Kamera verläßt.

Abbildung 6 zeigt die Abfolge einer 10 Takte dauernden Vergenzbewegung des Systems. Die rechte Kamera fixiert einen beliebigen Blickpunkt einer vorstehenden Ecke an einer gut strukturierten Wand. Im ersten Takt wird in der größten Auflösungsstufe ein Disparitätswert ermittelt und die Vergenzbewegung beginnt (o.r.). Zum 4., 7. und 9. Takt wird jeweils eine feinere Auflösungsstufe gewählt. Die feinste Stufe 512x512 liefert in diesem Beispiel keine stabile Phaseninformation am fixierten Blickpunkt. Es wird die nächst gröbere und stabile Auflösungsstufe gewählt.

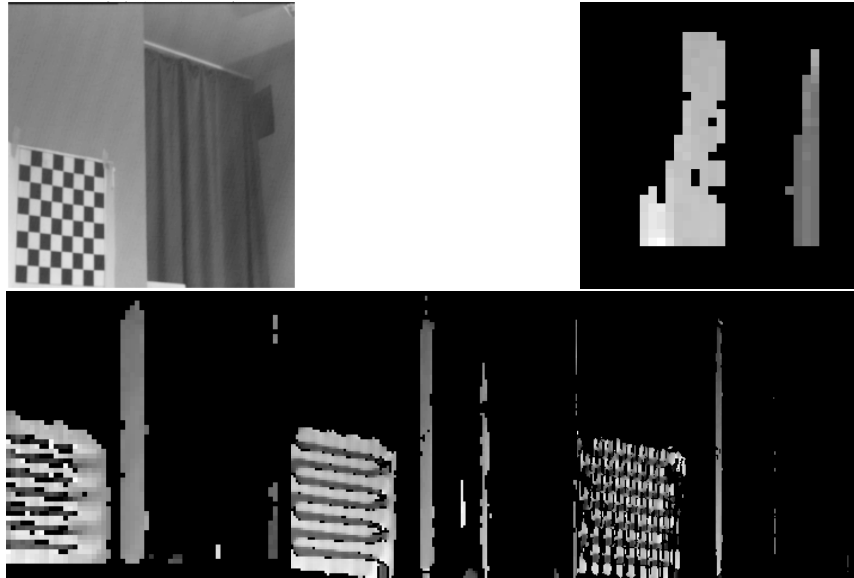


Abbildung 6: (o.l.) Folge der Vergenzbewegung (10 Takte) an einer strukturierten Wand mit den zugehörigen Disparitätskarten in verschiedenen Auflösungsstufen vom 1. Takt (o.r.) 32x32, über den 4. (u.l.) und 7. (u.m.) bis zum 9. Takt (u.r.) 256x256.

5 Zusammenfassung und Ausblick

In diesem Beitrag wurde der Aufbau eines Verfahrens zur Vergenzregelung vorgestellt. Basis ist ein echtzeitfähiger Algorithmus zur Disparitätsschätzung. Die Anwendung dieser erprobten Regelung zur Vergenzbewegung erfolgte in einem System zur aktiven Tiefenschätzung einer unbekanntenen Umgebung (Hansen & Sommer, 1996). Dort sind die Blick- und die Vergenzkontrolle basierend auf Gaborfilterantworten kombiniert, um absolute Entfernungswerte der fixierten Blickpunkte zu ermitteln. In einem weiteren Schritt werden die Gaborfilter für die Disparitätsschätzung dahingehend optimiert, daß sie nicht nur zur Vergenzbewegung, sondern auch zu einer exakten Tiefenrekonstruktion der lokalen Umgebung fixierter Blickpunkte beitragen. In einer mobilen Version sollen die aktiv generierten Tiefenkarten dem System zu einem autonomen Verhalten in seiner Umwelt verhelfen.

Literatur

- Bajcsy, R. (1988). Active perception. *Proceedings of the IEEE*, 76, 996–1005.
- Ballard, D., & Brown, C. (1992). Principles of animate vision. *CVGIP: Image Understanding*, 56, 3–21.
- Ching, W., Toh, P., Chan, K., & Er, M. (1993). Robust vergence with concurrent detection of occlusion and specular highlights. In *Proceeding of ICCV'93*, 384–393.
- Fleet, D. J., Jepson, A. D., & Jenkin, M. R. M. (1991). Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2).
- Franklin, G., Powell, J., & Workman, M. (1992). *Digital Control of Dynamic Systems*. Addison-Wesley.
- Hansen, M., & Sommer, G. (1996). Active depth estimation with gaze and vergence control using gabor filters. In *Proc. 13th ICPR, Vienna*, Vol. A, 344–348.
- Krotkov, E. (1989). *Active Computer Vision by Cooperative Focus and Stereo*. Springer Verlag, Berlin et al.
- Langley, K., Atherton, T., Wilson, R., & Larcombe, M. (1990). Vertical and horizontal disparities from phase. In *Proc. 1st ECCV*, 315–325.
- Maki, A. (1996). *Stereo Vision in attentive Scene Analysis*. Ph.D. thesis, CVAP Stockholm.
- Michaelis, M., & Sommer, G. (1994). Basis functions for early vision. Tech. rep., Christian-Albrechts-Universität zu Kiel.
- Olsen, T., & Coombs, D. (1991). Real-time vergence control for binocular robots. *IJCV*, 1, 76–89.
- Sanger, T. (1988). Stereo disparity computation using gabor filters. *Biol. Cybernetics*, 59, 405–418.
- Sommer, G. (1995). Verhaltensbasierter Entwurf technischer visueller Systeme. *Künstliche Intelligenz*, 5(3), 42–45.
- Theimer, W. M., & Mallot, H. P. (1994). Phase-based binocular vergence control and depth reconstruction using active vision. *CVGIP: Image Understanding*, 60(2), 343–358.
- Westelius, C. (1995). *Focus of Attention and Gaze Control for Robot Vision*. Ph.D. thesis, Linköping University.
- Westelius, C., Knutsson, H., Wiklund, J., & Westin, C. (1994). Phased-based disparity estimation. In Crowley, J., & H.I.Christensen (Eds.), *Vision as Process*, 157–178. Springer Verlag, Berlin et al.