

Auf dem Weg zum künstlichen aktiven Sehen: Modellfreie Bewegungsverfolgung durch Kameranachführung

K. Daniilidis, M. Hansen, Ch. Krauss, G. Sommer

Institut für Informatik und Prakt. Mathematik

Christian-Albrechts Universität Kiel

Preusserstr. 1-9, 24105 Kiel, email:{kd,mha,chk,gs}@informatik.uni-kiel.de

Zusammenfassung. Diese Arbeit beschreibt die aktive Verfolgung eines bewegten Objektes von beliebiger Form mittels Blicksteuerung. Die Verfolgung erfolgt modellfrei durch die Subtraktion des von der Kamerabewegung induzierten optischen Normalflusses an den Stellen eines hohen Grauwertgradienten. Der von der Verarbeitungsverzögerung geprägte Regelkreis basiert auf der Prädiktion der Objektbewegung mittels eines stationären Kalman-Filters und wird einem Achsenregler übergeordnet. Die Verarbeitungszeit von 30ms ermöglicht eine Regelrate von 25Hz.

1 Einführung

Das maschinelle Sehen war lange Zeit durch die Kamera als passiven Beobachter geprägt. Dabei wurde das Sehen als eine Rekonstruktionsaufgabe verstanden, d.h. man versuchte, die Abbildungstransformation Welt-Bild zu invertieren, um so möglichst viel Information über die Welt zu erhalten. Es hat sich gezeigt, daß die 3-D Rekonstruktion der Welt als Aufgabe instabil ist, enorme Datenmengen erzeugt, und nicht echtzeitfähig ist. Das Paradigma des aktiven Sehens als verhaltensbezogener Ansatz [11] der Informationsverarbeitung stellt einen hoffnungsvollen Weg zur Überwindung der Probleme des passiven maschinellen Sehens dar. Indem nur die für ein bestimmtes Systemverhalten notwendigen visuellen Informationen ermittelt werden, wird die Komplexität der visuellen Wahrnehmung reduziert und die Robustheit verbessert. Ein Kamerasystem, das im Sinne von Bajcsy [2] als ein aktiver Sensor in den Bildgewinnungsprozeß eingreift, erfordert geeignete Regelungsmechanismen, um bei der visuellen Wahrnehmung stabiles Verhalten zu zeigen. Der Leser sei bezüglich des oben beschriebenen paradigmatischen Wechsels auf die Beitragssammlungen [1, 4] verwiesen.

Gegenstand dieser Arbeit ist die monokulare Blicksteuerung als fundamentales okulomotorisches Verhalten in einer dynamischen Umgebung. Unter Blicksteuerung verstehen wir die Blickänderung und die Blickstabilisierung (Fixierung oder aktive Verfolgung). Während die Blickänderung ein komplexes Problem an sich darstellt, das in dieser Arbeit nur anhand einer vom Beobachter unabhängigen Bewegung betrachtet wird, ist die Blickstabilisierung ein wohl definiertes Problem. In biologischen Systemen erfolgt eine Blickstabilisierung einerseits im zentralen Bereich der Retina durch aufeinanderfolgende Augenfolgebewegungen

und Korrektursakkaden und andererseits im gesamten Bild durch den optokinetischen Nystagmus. Das Ziel der monokularen Nachführung kann sowohl die Verfolgung von bewegten Objekten für Überwachungsaufgaben sein, als auch –und das ist für das maschinelle Sehen von Bedeutung– die Möglichkeit zu robusten Lösungen bei Navigations- oder Hand-Auge-Koordinationsaufgaben bieten. Die monokulare Kameranachführung gilt als erster Test bei Arbeiten mit einer aktiven Kameraplattform. Im Gegensatz zu Verfolgungsmethoden, die auf zeitlicher Korrelation oder besonderen Merkmalen des verfolgten Objektes basieren, liegt dieser Arbeit nur die Annahme einer Bewegung des Objektes zugrunde. Diese unabhängige Bewegung kann detektiert werden, wenn die Kamera eine reine Rotationsbewegung durchführt. In Bezug auf die Annahmen und die Methodik ist unserer Ansatz mit den Arbeiten [9, 10, 12] zu vergleichen. Dort wird der Blick auf eine unabhängige Bewegung gerichtet und das bewegte Objekt wird nur anhand dieser Bewegung verfolgt. Während der Ansatz von Murray et al. [10] in der Echtzeitleistung vergleichbar ist, wurde der Ansatz von Murray und Basu [9] nicht on-line getestet. Tölg [12] arbeitet mit einer Verarbeitungsrate von 2Hz. Auf Hinweise zu weiteren Publikationen zur Kameranachführung sei hier aus Platzgründen verzichtet.

Durch das Schließen des durch Wahrnehmung und Handlung gebildeten Regelkreises stellt die Echtzeitproblematik nicht mehr nur einen Effizienzaspekt bezüglich Anwendungen dar, sondern muß den Verarbeitungsmethoden inhärent sein. Unter diesem Aspekt zeichnet sich der hier präsentierte Ansatz dadurch aus, daß die Echtzeitbedingungen trotz Anwendung von formal begründbaren Algorithmen eingehalten werden können.

2 Geometrie und Kinematik des Kamerasystems

Die binokulare Plattform besitzt vier Freiheitsgrade beschrieben durch den Schwenkwinkel χ um den "Hals" des Systems, den Neigungswinkel ϕ und je einen Vergenzwinkel θ_l und θ_r (Abb. 1). Für die monokulare Kameranachführung

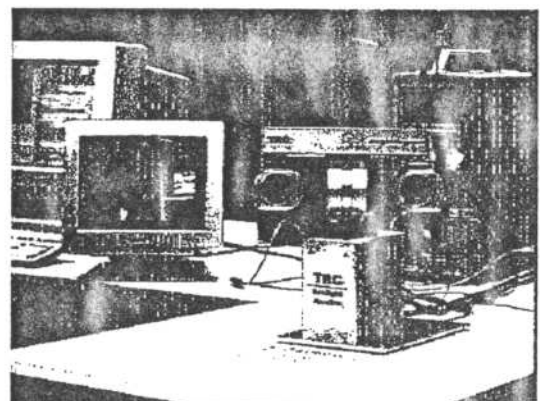
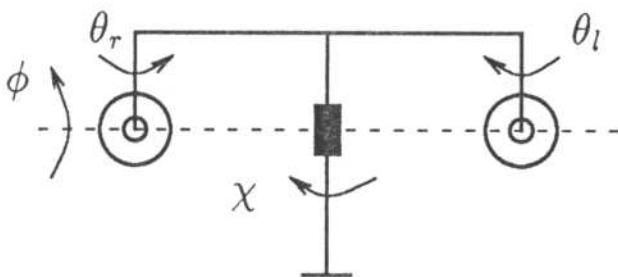


Abb. 1. Die vier mechanischen Freiheitsgrade und eine Abbildung der Kameraplattform.

benötigen wir nur die Neigung und einen der Vergenzwinkel, die wir desweiteren mit ϕ und θ bezeichnen. Der Positionsvektor eines Raumpunktes ist nicht ermittelbar. Deshalb führen wir den Begriff des Sehstrahles $(x, y, 1)$ ein, der einen Punkt (x, y) in einer Ebene $Z = 1$ beschreibt. Das Weltkoordinatensystem hat

seinen Ursprung im Schnittpunkt der Neigungs- und der Vergenzachse. Ein Sehstrahl $\mathbf{p}_w = (x_w, y_w, 1)^T$ im Weltkoordinatensystem läßt sich durch

$$\lambda \mathbf{p}_w = R_\phi R_\theta \mathbf{p}_c \quad (1)$$

auf einen Sehstrahl $\mathbf{p}_c$ im sich bewegenden Kamerakoordinatensystem abbilden, wobei R_ϕ und R_θ Rotationen um die x - und y -Achse beschreiben. Die Abbildung stellt eine projektive Kollineation zwischen Strahlenbündeln dar. Diese zeigt die bekannte Eigenschaft, daß sich jede Rotation des Kamerakoordinatensystems, im Gegensatz zu einer Translation, unter Kenntnis des Winkels und ohne Wissen über die Tiefen der abgebildeten Punkte kompensieren läßt. Aus den drei Komponenten der Vektorgleichung (1) läßt sich der Parameter λ eliminieren, woraus man

$$\begin{aligned} x_w &= \frac{x_c \cos \theta + \sin \theta}{-x_c \cos \phi \sin \theta + y_c \sin \phi + \cos \phi \cos \theta} \\ y_w &= \frac{x_c \sin \phi \sin \theta + y_c \cos \phi - \sin \phi \cos \theta}{-x_c \cos \phi \sin \theta + y_c \sin \phi + \cos \phi \cos \theta} \end{aligned} \quad (2)$$

erhält. Mittels (2) läßt sich das direkte kinematische Problem beschreiben. Wird ein Sehstrahl $\mathbf{p}_w = (x_w, y_w, 1)^T$ als Stellgröße für die optische Achse $(0, 0, 1)^T$ der Kamera angegeben, so erhält man die Lösung des inversen kinematischen Problems:

$$\tan \phi = -y_w \quad \tan \theta = \frac{x_w}{\sqrt{1 + y_w^2}}. \quad (3)$$

Aus einem gegebenen Sehstrahl $\mathbf{p}_c$ kann man durch die Transformation auf das Weltkoordinatensystem mittels (2) und (3) die Stellwinkel der optischen Achse in Richtung von $\mathbf{p}_c$ bestimmen (Sakkadenproblem). Abschließend ist die Winkelgeschwindigkeit des Kamerakoordinatensystems zu bestimmen, die zur Ermittlung des optischen Flusses benötigt wird. Sei $R(t)$ die von der Zeit abhängende Rotation eines Kamerakoordinatensystems und ω die Winkelgeschwindigkeit bezüglich des bewegten Koordinatensystems, so gilt $\dot{R}(t) = R(t)\Omega$, wobei Ω die aus der Winkelgeschwindigkeit ω resultierende antisymmetrische Matrix ist. Mit $R(t) = R_{\phi(t)}R_{\theta(t)}$ folgt

$$\omega = (\dot{\phi} \cos \theta \quad \dot{\theta} \quad \dot{\phi} \sin \theta). \quad (4)$$

Bei der weiteren Beschreibung der Kamerageometrie gehen wir von der Annahme aus, daß das Projektionszentrum der Kamera mit dem Schnittpunkt der Neigungs- und Vergenzachse zusammenfällt und daß die optische Achse mit der Z -Achse des oben definierten Kamerakoordinatensystems identisch ist.

Die projektive Transformation des Sehstrahls $\mathbf{p}_c$ in Kamerakoordinaten auf den Sehstrahl $\mathbf{p}_i = (x_i, y_i, 1)^T$ in Bildkoordinaten wird durch die affine Transformation $x_i = \alpha_x x_c + x_0$ und $y_i = \alpha_y y_c + y_0$ beschrieben, wobei α_x, α_y und (x_0, y_0) die internen Parameter der Kamera sind. Dabei hängen α_x und α_y von der Brennweite, der Pixelgröße des CCD-Sensors und der Abtastrate des A/D-Wandlers ab. Mit (x_0, y_0) wird der Schnittpunkt der optischen Achse mit der Bildebene bezeichnet. Die internen Parameter wurden sowohl durch ein klassisches Kalibrierungsverfahren [5] als auch mit einem speziell für aktive Systeme entwickelten Verfahren [8] bestimmt.

3 Verfolgung eines bewegten Objektes

Ein bewegtes Objekt wird durch die ständige Detektion der verursachten Bildbewegung verfolgt. Während man bei einer statischen Kamera ein bewegtes Objekt mit klassischen Verfahren der Änderungsdetektion ermitteln kann, ist das Problem bei einer Eigenbewegung der Kamera komplexer. In unserem Fall nutzen wir die Tatsache aus, daß sich die Winkelgeschwindigkeit der Kamera aus der Änderung der Winkelposition der Kameraplattform mit Hilfe von (4) ermitteln läßt. Daraus erhält man für den von der Kamerabewegung induzierten optischen Fluß

$$\mathbf{u}_c = \begin{pmatrix} x_c y_c & -(1 + x_c^2) & y_c \\ (1 + y_c^2) & -x_c y_c & -x_c \end{pmatrix} \boldsymbol{\omega}. \quad (5)$$

Sei $\mathbf{u} = (u, v)$ der gemessene optische Fluß, so ist $\mathbf{u} - \mathbf{u}_c$ der aus der Objektbewegung resultierende Fluß. Damit kann man direkt, d.h. ohne zusätzliche Annahmen über den Flußverlauf, den der Objektbewegung entsprechenden Bildbereich detektieren. Dieses Vorgehen ist wesentlich für das Einhalten der Echtzeitforderung. Wir gehen von der Annahme der Grauwertverhaltung

$$g_x u + g_y v + g_t = 0$$

aus, wobei g_x , g_y und g_t die örtlich-zeitlichen Ableitungen der Grauwertfunktion sind. Die Projektion des optischen Flusses in Richtung des Gradienten wird als Normalfluß u_n bezeichnet. Die Differenz zwischen dem von der Eigenrotation induzierten Normalfluß u_{c_n} und dem gemessenen Normalfluß u_n

$$u_{c_n} - u_n = \frac{g_x u_c + g_y v_c}{\sqrt{g_x^2 + g_y^2}} + \frac{g_t}{\sqrt{g_x^2 + g_y^2}}$$

ergibt den Normalfluß aus der Objektbewegung. Weil die Meßunsicherheit vom Betrag des örtlichen Gradienten abhängt, werden nur die Punkte einbezogen, deren Normalflußdifferenz und Gradientenbetrag über bestimmten Schwellen liegen (siehe auch [9]). Überschreitet die Menge der so detektierten Bildpunkte eine bestimmte Schwelle, so wird der Schwerpunkt als gemessene Bildposition des Objekts angenommen und dem Schätzer als Objektposition übergeben. Die drei verwendeten Schwellwerte stellen ein absolutes Minimum an Vorwissen für flußbasierte Detektionsaufgaben dar.

Besondere Sorgfalt wurde auf die örtlich-zeitliche Filterung gelegt, so daß die Frequenzcharakteristiken der Ableitungen erhalten bleiben, ohne die Echtzeitbedingungen zu verletzen. Für die örtlichen Ableitungen kamen binomiale Faltungsmasken zur Anwendung, die eine Approximation der ersten Ableitung der Gaußfunktion sind. Jede Ableitung wird in der zu ihr orthogonalen Richtung geglättet. Für die zeitliche Filterung kamen aus Gründen der Speichereffizienz und der Zeitverzögerung rekursive IIR-Filter zur Anwendung. Das benutzte zeitliche Glättungsfilter ist die diskrete Version einer abgeschnittenen Exponentialfunktion [6] mit der z-Transformation

$$H(z) = q \frac{1 + z^{-1}}{1 + rz^{-1}}, \quad q = \frac{\tau}{\tau + 2} \quad r = \frac{\tau - 2}{\tau + 2}.$$

Die Kaskadierung von mehreren Tiefpaßfiltern $H(z)$ betont den Tiefpaßeffect und hat eine elegante Eigenschaft: Die Ableitung der Impulsantwort ist gleich der Differenz von zwei Filtern aufeinanderfolgender Ordnung. Damit sowohl örtliche als auch zeitliche Filter mit gleicher Verzögerung wirken, wird das zeitliche Tiefpaßfilter zweiter Ordnung auch auf die Antworten der örtlichen Ableitungen angewandt. Die zeitliche Verzögerung ist abhängig vom Modus ($1/\tau$) und Mittelwert ($2/\tau$) der Impulsantwort und wird entsprechend in der Prädiktionsstufe berücksichtigt.

4 Schätzung und Regelung

Es ist das Ziel der Kameranachführung, die Blickrichtung möglichst nahe an der Projektion eines bewegten Objektes zu halten. Dabei sind Stellgrößen der monokularen Kameranachführung der Neigungswinkel ϕ und der Vergenzwinkel θ . Die Meßgrößen sind die Bildposition des Objektes und die Winkelangaben der Achsenenkoder. Weil sowohl die Abbildung zwischen den Winkeln (ϕ, θ) und den Weltkoordinaten (x_w, y_w) (siehe (3)) als auch die Abbildung zwischen Kamerakoordinaten und Weltkoordinaten der Projektion des Objekts (Gl. 2) eineindeutig sind, werden wir von nun an in Weltkoordinaten rechnen. Diese entsprechen den Koordinaten in der Ebene $Z = 1$ bei Ruhelage der Plattform. Wir bezeichnen mit c den Schnittpunkt der optischen Achse der Kamera mit der Ebene $Z = 1$ und mit o die Projektion des Positionsvektors zum Objekt auf diese Ebene. Es seien v und a die Geschwindigkeit und Beschleunigung des zu verfolgenden Punktes. Wir fassen diese vier zweidimensionalen Vektoren in einem Zustandsvektor

$$s = (c^T \ o^T \ v^T \ a^T)^T$$

zusammen. Sei $\Delta u(k)$ die von einem Bewegungsbefehl resultierende Änderung der Kameraposition zum Zeitpunkt k . Unter der Annahme, daß sich das Objekt mit gleichmäßiger Beschleunigung bewegt, führen wir die Strecke unseres diskreten Systems ein:

$$s(k+1) = \Phi s(k) + \Gamma \Delta u(k) \quad \text{mit} \quad \Phi = \begin{pmatrix} I_2 & O_2 & O_2 & O_2 \\ O_2 & I_2 & \Delta t I_2 & \Delta t^2/2 I_2 \\ O_2 & O_2 & I_2 & \Delta t I_2 \\ O_2 & O_2 & O_2 & I_2 \end{pmatrix},$$

$\Gamma = (1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)^T$, I_2 eine 2×2 Einheitsmatrix und O_2 eine 2×2 Nullmatrix. Nach dem Separationsprinzip [7] kann man eine optimale Regelung durch einen getrennten Entwurf von Regler und Schätzer aufbauen. Wir machen uns eine lineare Regelfunktion $\Delta u(k) = -K\hat{s}(k)$ zu nutze, wobei $\hat{s}$ die Zustandsschätzung ist. Regelungsziel ist die Minimierung der Abweichung $\|o - c\|$. Die Minimierung dieser Abweichung kann durch einen Linearen Quadratischen Regler (LQR) ohne Bedingung bezüglich des Stellvektors Δu durch Minimierung der Kostenfunktion $\sum_{k=0}^N s^T(k) Q s(k)$ erfolgen. Dabei ist Q eine symmetrische Matrix mit den einzigen nicht verschwindenden Elementen $Q_{11} = 1$, $Q_{12} = Q_{21} = -1$ und $Q_{22} = 1$. Aus Gründen der Zeiteffizienz

wird ein konstanter Verstärkungsvektor K angewendet. Dies entspricht der Annahme eines stationären Zustandes. Der Vektor K läßt sich durch Lösung der in diesem Fall einfachen algebraischen Riccati Gleichung ermitteln und lautet $K = (1 - 1 - \Delta t - \Delta t^2/2)$. Dies bedeutet, daß die Stellposition der Kamera der Prädiktion der Objektposition entspricht.

Im stationären Zustand werden wir auch die Schätzung behandeln. Eine spezielle Form des stationären optimalen Kalman-Filters ist das α - β - γ -Filter [3]:

$$\hat{s}^+(k+1) = \hat{s}^+(k) + (\alpha \beta / \Delta t \gamma / \Delta t^2)^T (\mathbf{m}(k+1) - \mathbf{m}^-(k+1)),$$

wobei $\mathbf{s}^+$ den Zustand nach der Aktualisierung und $\mathbf{m}^-$ die prädizierte Messung darstellen. Die Koeffizienten α, β und γ lassen sich aus dem sogenannten Manöverindex berechnen, der das Verhältnis zwischen Prozeßrauschen und Meßrauschen darstellt. Je niedriger der Manöverindex ist, desto glatter verläuft die Zustandsschätzung. Ist der Manöverindex hoch, so haben wir großes Vertrauen in die Messung und es erfolgt keine Prädiktion während der Filterung. Ein solches Schätzungsfilter gibt uns auch die Möglichkeit, das Totzeitproblem anzugehen. Obwohl unsere Messungen und Bewegungsbefehle mit gleicher Abtastrate erfolgen, fallen die Zeitpunkte des Anfangs jeder Bewegung und der Messung nicht zusammen. Dieses Problem läßt sich dadurch lösen, daß mit Hilfe der α - β - γ -Schätzung ein beliebiges Zeitintervall im voraus prädiziert wird.

Die Simulation (Abb. 2) zeigt das Verhalten unseres Reglers bei der Verfolgung eines Objektes, deren Projektion in der Ebene $Z = 1$ sinusförmig ist. Im ersten Experiment wurde ein niedriger Manöverindex (0.1), im zweiten Experiment ein höherer Manöverindex (1.0) angenommen. Man beachte die glatte Trajektorie bei Vertrauen in die Prädiktion (links) und die gestörte Trajektorie bei Vertrauen in die Messung (rechts).

5 Systemarchitektur und Experimente

Die in den vorherigen Abschnitten beschriebene Bildverarbeitung und Regelung wurden auf einem System realisiert, das sich aus kommerziellen Hardware-Komponenten zusammensetzt (Abb. 1 rechts). Die Kameraplattform BiSight der Fa. TRC hat zusätzlich zu den im zweiten Abschnitt beschriebenen vier mechanischen auch je zwei regelbare optische Freiheitsgrade (Zoom und Fokus) für jede Kamera. Die untergelagerte Servoregelung der Achsen erfolgt auf der Steuereinheit PMAC der Fa. DeltaTau, welche auch die Winkelpositionen ausliest. Das Videosignal der zwei CCD-Kameras wird in je einem MaxVideo-200 Pipeline Rechner der Fa. Datacube digitalisiert. Im gleichen Rechner finden alle im dritten Abschnitt beschriebenen Filter- und Schwellwertoperationen statt.

Das detektierte Bewegungsgebiet wird zu einer Sun-Sparcstation 10/40 über VME-SBus geleitet, wo der Schwerpunkt des Objektes berechnet wird und die im vierten Abschnitt beschriebene Regelung erfolgt. Wir fassen zur Übersicht die Verarbeitungsschritte der Bewegungsdetektion und Regelung zusammen:

1. Die Winkel der Achsen werden vor jeder Bildaufnahme abgelesen.
2. Das eingelesene Videosignal wird digitalisiert und unterabgetastet (128x128).

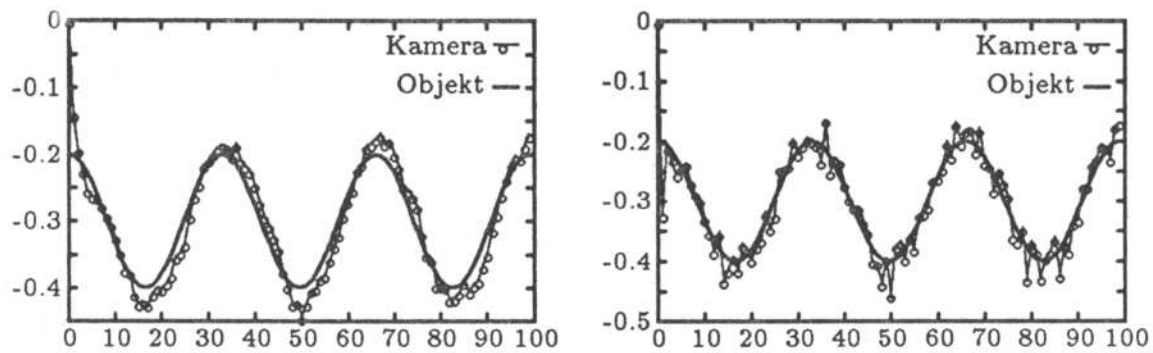


Abb. 2. Verlauf der Kamera- und der Objektposition in Weltkoordinaten für einen niedrigen (links) und einen hohen Manöverindex (rechts).

3. Die Berechnung der örtlichen Gradienten erfolgt durch Faltung mit zwei 7x7 Binomialmasken.
4. Die örtlichen Gradienten werden rekursiv zeitlich geglättet. Ebenfalls rekursiv erfolgt die Berechnung des zeitlichen Gradienten.
5. Über LUT wird der Betrag des örtlichen Gradienten berechnet und die Normalflußdifferenz gebildet. Zwei Schwellwertoperationen werden auf diese Differenz und den Gradientenbetrag angewendet. Die Ergebnisse werden mit einem logischen AND verknüpft.
6. Das resultierende Binärbild wird auf die Workstation übertragen, wo die Schwerpunktsberechnung des Objektes erfolgt.
7. Mittels der Winkelmessung wird der Schwerpunkt in das Weltkoordinatensystem transformiert.
8. Der Systemzustand wird mit dem α - β - γ -Filter aktualisiert und prädiziert.
9. Die Prädiktion wird auf Stellgrößen für die Winkel abgebildet.
10. Die Winkelstellgrößen werden zur Steuereinheit übertragen.

Die Schritte 1-10 benötigen insgesamt 30ms. Damit bleibt ein Zeitintervall von 10ms für weitere Bildverarbeitungsschritte übrig. Die Bewegung erfolgt nach einer Totzeit, die sich aus der Zeitverzögerung des rekursiven Filters (40ms) und der Verarbeitungszeit (30ms) zusammensetzt.

Ein Beispiel aus einer Verfolgung eines beliebigen Objektes wird in Abb. 3 dargestellt. Es ist zu beachten, daß bei einer ausgedehnten Projektion des unabhängig bewegten Objektes (hier die ganze Person mit ihrer Hand) sich eine sehr unregelmäßige Trajektorie des ermittelten Schwerpunktes ergibt. Dieses muß jedoch nicht als Objektbewegung, sondern eher als Meßfehler gewertet werden. Denn es ist das Ziel, das Objekt und nicht genau seinen Schwerpunkt im zentralen Gesichtsfeld zu fixieren.

In dieser Arbeit haben wir die ersten Schritte auf dem Weg zu einem künstlichen aktiven visuellen System präsentiert. Der Beitrag dieser Arbeit liegt in einer modellfreien Verfolgungsleistung, die in Echtzeit (25Hz) erbracht wird, ohne daß ad-hoc Techniken oder heuristische Methoden angewandt werden, was beginnend vom dem Sobeloperator für Gradientenberechnung über die einfache Bilddifferenz bis zur Dreipunktregelung oft bei Echtzeitanwendungen bisher notwendig war.

Danksagung: Ohne die Beiträge von Henrik Schmidt zur Programmierung



Abb. 3. Drei Aufnahmen "on the fly" während der Verfolgung einer Person mit winkender Hand. Markiert sind die Punkte, die die Schwellen bei der Bewegungsdetektion überschritten haben, sowie der Schwerpunkt der detektierten Fläche.

der Kameraplattform, von Jörg Ernst zur Kalibrierung, sowie von Gerd Diesner zur Programmierung des MaxVideo200 wäre das System nicht realisierbar gewesen. Wir bedanken uns bei Ulf Cahn von Seelen (GRASP Lab, Philadelphia) für sehr hilfreiche Diskussionen.

Literatur

1. Y. Aloimonos, editor. *Active Perception*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.
2. R. Bajcsy. Active Perception. *Proceedings of the IEEE*, 76:996–1005, 1988.
3. Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*. Academic Press, New York, NY, 1988.
4. A. Blake and A. Yuille, editors. *Active Vision*. MIT Press, Cambridge, MA, 1992.
5. O. Faugeras. *Three-dimensional Computer Vision*. MIT-Press, Cambridge, MA, 1993.
6. D.J. Fleet and K. Langley. Recursive Filters for Optical Flow. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17:61–67, 1995.
7. G.F. Franklin, J.D. Powell, and M.L. Workman. *Digital Control of Dynamic Systems*. Addison-Wesley, 1992.
8. M. Li. Camera Calibration of a Head-Eye System for Active Vision. In *Proc. Third European Conference on Computer Vision*, pp. 543–554, Stockholm, Sweden, May 2–6, 1994.
9. D. Murray and A. Basu. Motion Tracking with an Active Camera. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16:449–459, 1994.
10. D.W. Murray, P.L. McLauchlan, I.D. Reid, and P.M. Sharkey. Reactions to peripheral image motion using a head/eye platform. In *Proc. Int. Conf. on Computer Vision*, pp. 403–411, Berlin, Germany, May 11–14, 1993.
11. G. Sommer. Verhaltensbasierter Entwurf technischer visueller Systeme. *Künstliche Intelligenz*, 5(3), 1995.
12. S. Tölg. *Strukturuntersuchungen zur Informationsverarbeitung in neuronaler Architektur am Beispiel der Modellierung von Augenbewegungen für aktives Sehen*. VDI-Verlag, Düsseldorf, 1992.