

Eigenbewegungsschätzung bei aktivem Sehen: Ein neuronales Netz auf biologischer Grundlage

Markus Lappe* Antje Grigo*

Zusammenfassung

Es wird ein neuronales Netzwerk vorgestellt, das die von einem aktiven Beobachter ausgeführte Bewegung im Raum bestimmt. Dieses Netzwerk nutzt verschiedene, einem aktiven Sehsystem zur Verfügung stehende Informationen. Die Verarbeitung und Integration dieser Informationsquellen folgt neurobiologischen Prinzipien. Das Netzwerk ist in der Lage, eine Vielzahl von Eigenschaften der menschlichen Eigenbewegungsbestimmung zu reproduzieren.

1 Einleitung

Zielgerichtete Bewegungen von autonomen Systemen, egal ob Menschen, Tiere oder Roboter, in einer komplexen Umwelt erfordern eine ständige Eigenkontrolle dieser Bewegung im dreidimensionalen Raum. So ist zur Fortbewegung eines Individuums vom Ort A zum Ort B die stete Kontrolle der momentanen Eigenbewegungsrichtung notwendig. Einen wichtigen Beitrag zu dieser Eigenbewegungsschätzung leistet das Sehen. Die Bewegung durch eine strukturierte Umgebung erzeugt eine bewegte Abbildung der Umwelt auf der Netzhaut oder einem Kameratarget, ein optisches Bewegungs- bzw. Flussfeld. In einer statischen Umwelt erzeugte Flussfelder zeigen Eigenschaften, die charakteristisch für die Eigenbewegung des Individuums sind. Diese sind für einen Roboter im klassischen Sinn und für ein aktives Sehsystem verschieden.

2 Optische Bewegungsfelder aktiver biologischer Systeme

Für einen Roboter im klassischen Sinn ist die Kamerastellung zum Rumpf konstant. Die Aufgabe der Eigenbewegungsschätzung wird dann zur *passiven Navigation* (Bruss & Horn, 1983). Eine geradlinige Bewegung führt zu einem radialen optischen Bewegungsfeld, in dem der singuläre Punkt die Bewegungsrichtung anzeigt. Durch die einem aktiven Individuum gegebenen verschiedenen Möglichkeiten von Körper-, Kopf- und Augenstellungen und -bewegungen werden optische Bewegungsfelder dagegen sehr komplex, wobei der singuläre Punkt kein Indikator der Bewegungsrichtung mehr ist. Ein Beispiel ist in Abb. 1 dargestellt. Der Grund dafür ist, daß ein aktives, autonomes System im allgemeinen mehrere Verhaltensleistungen gleichzeitig vollbringen muß, die während einer Eigenbewegung beispielsweise das Fixieren und visuelle Verfolgen von interessierenden Objekten erfordern. Trotzdem kann das visuelle System des Menschen die Analyse des optischen Eingangs im Hinblick auf die korrekte Bewegungswahrnehmung sehr effektiv leisten. Allerdings erfordert gerade die Ausführung von Augen- und Kopfbewegungen während des Fortbewegens die zusätzliche Einbeziehung verschiedener Sinnesmodalitäten, wie z.B. okulomotorische und vestibuläre Informationen, um die Komplexität der Aufgabe zu verringern. Gleichfalls können hierzu aber auch Annahmen über die Art der von einem aktiven Sehsystem ausgeführten Augen- oder Kamerabewegungen genutzt werden (Fermüller & Aloimonos, 1993; Lappe & Rauschecker, 1993). Zudem muß man davon ausgehen, daß auch binokulare Disparitätsinformation zur Tiefenanalyse der Umgebung herangezogen wird. Alle diese Informationen werden im Gehirn in Form von neuronaler Aktivität kodiert. Untersuchungen an Primaten zeigen, daß die Neurone in zwei spezialisierten Hirnarealen, nämlich den Arealen MT und MST, spezifisch auf bestimmte großflächige Veränderungen der Netzhautabbildung reagieren. Dabei zeigt das Areal MST eine spezifische Bevorzugung von optischen Flussfeldreizen (Duffy & Wurtz, 1991). Zudem zeigen sie Aktivitätsänderungen in Abhängigkeit von Augenstellung und -bewegung sowie von der Tiefenstruktur der visuellen Szene. Wünschenswert wäre es, die Verarbeitung des visuellen Eingangs unter dem Aspekt der Eigenbewegungsbestimmung auf neuronaler Basis zu modellieren. Wir stellen hier ein neuronales Modell vor, das durch Integration von visuellen und nicht-visuellen Informationen die einem optischen Bewegungsfeld zugrundeliegende Eigenbewegungsrichtung bestimmt.

*Allgemeine Zoologie und Neurobiologie, Ruhr-Universität, 44780 Bochum.
Die Arbeit wurde unterstützt von der deutschen Forschungsgemeinschaft, Sonderforschungsbereich 509 'Neurobiologie des Sehens'.

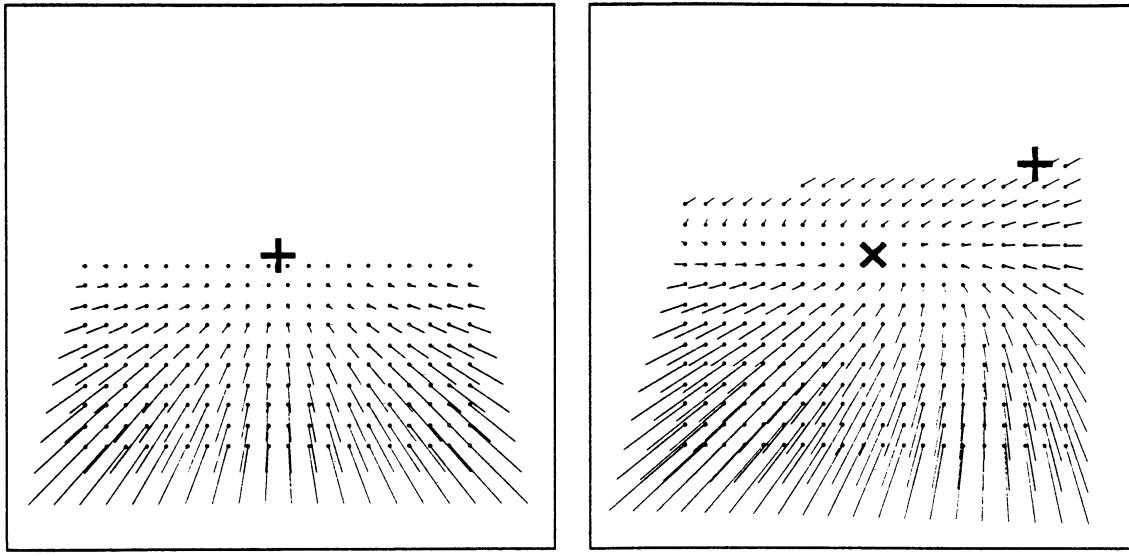


Abbildung 1: Einfluß von aktiven visuellen Folgebewegungen auf das Flußfeld. Die linke Abbildung zeigt das augenzentrierte optische Bewegungsfeld bei geradliniger Translation über eine Horizontalebene. In der rechten Abbildung ist dieser Eigenbewegung eine Rotation des visuellen Sensors überlagert. Diese Rotation ist so beschaffen, daß ein spezifischer Punkt der Horizontalebene konstant fixiert wird. Dies kann in einem aktiven Sehsystem notwendig sein, um z.B. eine Objektanalyse dieses Raumbereichs durchzuführen. Dadurch befindet sich der singuläre Punkt des Bewegungsfeldes an der Position des fixierten Objektes (x) und nicht in Richtung der Eigentranslation (+). Eine detaillierte mathematische Analyse findet man in Lappe und Rauschecker (1995b)

3 Minimierungsverfahren zur Eigenbewegungsanalyse

Ziel ist es, aus den visuellen Bewegungsfeldern die sie erzeugende Bewegung des Beobachters abzuschätzen. Mathematisch läßt sich die durch Bewegung des Beobachters im augenzentrierten Koordinatensystem verursachte Bewegung eines Punktes $\mathbf{R} = (X, Y, Z)^t$ im dreidimensionalen Raum ausdrücken als

$$(1) \quad \mathbf{V} = -\mathbf{T} - \boldsymbol{\Omega} \times \mathbf{R}.$$

Der optische Fluß ergibt sich aus der Projektion der Punktbewegung in der dreidimensionalen Welt auf eine zweidimensionale Bildfläche. Die Bewegung des an den Ort (x, y) auf die Bildfläche projizierten Punktes \mathbf{R} liefert den retinalen Fluß $\boldsymbol{\theta}(x, y)$. Dieser läßt sich in Matrixform schreiben (Heeger & Jepson, 1992):

$$(2) \quad \boldsymbol{\theta}(x, y) = \frac{1}{Z(x, y)} \mathbf{A}(x, y) \mathbf{T} + \mathbf{B}(x, y) \boldsymbol{\Omega}$$

Die Matrizen \mathbf{A} und \mathbf{B} enthalten nur den bekannten Projektionsort (x, y) sowie die festzulegende Entfernung der Bildfläche. Bekannt ist außerdem der gemessene optische Fluß $\boldsymbol{\theta}(x, y)$. \mathbf{T} ist der in die Richtung der Eigenbewegung zeigende Einheitsvektor. Seine zwei Komponenten sind unbekannt, ebenso die drei Rotationskomponenten von $\boldsymbol{\Omega}$ und die räumliche Entfernung $Z(x, y)$ des Punktes \mathbf{R} vom augenzentrierten Koordinatenursprung. Da \mathbf{T} und $\boldsymbol{\Omega}$ für jeden Ort des Flußfeldes identisch sind, lassen sich die Flußvektoren von m verschiedenen Punkten zu einer Matrixgleichung kombinieren:

$$(3) \quad \boldsymbol{\Theta} = \mathbf{C}(\mathbf{T}) \mathbf{q}.$$

Dabei ist $\boldsymbol{\Theta}$ ein aus den Komponenten der m einzelnen Bildgeschwindigkeiten bestehender $2m$ -dimensionaler Vektor und \mathbf{q} ein $(m+3)$ -dimensionaler Vektor, der die unbekannten Entfernungen $Z(x, y)$ der m Bildpunkte und die unbekannte Rotation $\boldsymbol{\Omega}$ enthält. Die $2m \times (m+3)$ Matrix $\mathbf{C}(\mathbf{T})$ ergibt sich aus Kombination der Matrizen $\mathbf{A}(x_i, y_i) \mathbf{T}$ und $\mathbf{B}(x_i, y_i)$ der einzelnen Bildpunkte und enthält daher als einzige unbekannte Größe die Translationsrichtung \mathbf{T} . Diese kann gemäß Gl. (3) bestimmt werden durch Minimierung des Fehlerquadrats

$$(4) \quad R(\mathbf{T}) = \|\boldsymbol{\Theta} - \mathbf{C}(\mathbf{T}) \mathbf{q}\|^2.$$

Dies entspricht der Lösung der Kostenfunktion

$$(5) \quad R(\mathbf{T}) = \|\Theta^t \mathbf{C}^\perp(\mathbf{T})\|^2,$$

wobei $\mathbf{C}^\perp(\mathbf{T})$ das orthogonale Komplement zu \mathbf{C}^\perp ist (Heeger & Jepson, 1992). Dieser Algorithmus gestattet es, die Translationsrichtung unabhängig von Rotation und Tiefenverteilung abzuschätzen, und lässt sich zudem als Grundlage für ein biologisch plausibles neuronales Modell verwenden.

4 Neuronale Implementation

Das neuronale Modell basiert auf einer neuronalen Implementation des Minimierungsverfahrens (Lappe & Rauschecker, 1995b). Es besteht aus zwei Schichten von Neuronen. Die erste Schicht repräsentiert das retinale Bewegungsfeld als Aktivitäten s_m richtungsselktiver Neuronen. Diese Neurone sind so modelliert, daß ihre Eigenschaften denen von Zellen im visuellen kortikalen Areal MT entsprechen. Da eine einzelne Zelle aber nicht alleine die Bewegungsrichtung kodieren kann, wird diese Aufgabe auf eine kleine Population von Neuronen verteilt, die dasselbe rezeptive Feld, aber verschiedene Bewegungspräferenzen \mathbf{e}_m aufweisen. Der einzelne Flußvektor läßt sich dann schreiben als

$$(6) \quad \theta = \sum_m s_m \mathbf{e}_m.$$

In der zweiten Schicht des Netzwerks werden die möglichen Translationsrichtungen der Eigenbewegung repräsentiert. Diese Schicht entspricht funktionell dem kortikalen Areal MST. Die Aktivität von Neuronenpopulationen signalisiert die Wahrscheinlichkeit, daß eine von ihnen kodierte Richtung die dem Flußfeld zugrundeliegende ist. Je größer die Wahrscheinlichkeit, desto höher ihre Aktivität. Bezugnehmend auf Gl. (5) heißt dies, je kleiner der Wert von $\|\Theta^t \mathbf{C}^\perp(\mathbf{T}_j)\|^2$, desto wahrscheinlicher das spezielle \mathbf{T}_j , und desto höher muß die Aktivität der diese Richtung repräsentierenden Neuronenpopulation sein. Daher ist es naheliegend, die synaptischen Gewichte zwischen den Neuronenpopulationen der zwei Schichten über den beschriebenen Minimierungsalgorithmus einzustellen. Da das Argument der Kostenfunktion in Gl. (5) ein einfaches Matrixprodukt ist, läßt es sich in einfacher Weise als synaptische Summation ausdrücken (Lappe & Rauschecker, 1993). Dazu müssen die synaptischen Gewichte J so beschaffen sein, daß

$$(7) \quad \sum_m J_{mj} s_m = \Theta^t \mathbf{C}^\perp(\mathbf{T}_j).$$

Dieses rein visuelle neuronale Netz ist in der Lage, verschiedene visuelle Eigenschaften kortikaler Neurone (Lappe, Bremmer, Pekel, Thiele, & Hoffmann, 1996) und der menschlichen Wahrnehmung (Lappe & Rauschecker, 1993, 1995a) zu reproduzieren. Allerdings versagt es analog zum visuellen System des Menschen in bestimmten Situationen, die nicht-visuelle Informationen benötigen.

5 Multimodale sensorische Integration

Für ein aktives, autonomes System ist die Integration multipler, auch redundanter, Signalquellen wesentlich (Clark & Yuille, 1990). Für die konkrete Aufgabe der Eigenbewegungsschätzung sind aus Untersuchungen an Menschen im wesentlichen drei die Flußfeldanalyse beeinflussende Faktoren bekannt: Informationen über Augen- und Kopfdrehungen, die als extraretinale Signale bezeichnet werden, stereoskopische Tiefenwahrnehmung, die zur Rauschunterdrückung im Flußfeld genutzt wird, und ethologische Randbedingungen, d.h. vernünftige Annahmen über die Struktur der Umgebung und die Arten von möglichen Eigenbewegungen. Diese zusätzlichen Informationen lassen sich in den Minimierungsalgorithmus integrieren, wodurch die Anforderungen an die Minimierung verringert werden. Extraretinale Signale liefern eine Schätzung der Rotation Ω , stereoskopische Tiefe eine Schätzung der Struktur der Umwelt. Ethologische Randbedingungen lassen sich ebenso einfügen. Ein Beispiel hierfür ist die Beobachtung, daß aktive Augenbewegung i.a. der Fixation oder Verfolgung eines Objektes dient. Ist dieses Objekt statisch im Raum, so ist die Augenbewegung mit der Eigenbewegung gekoppelt. Dadurch wird die Zahl der Freiheitsgrade eingeschränkt und so die Minimierung vereinfacht. Die Integration dieser Signale und Randbedingungen mit dem Flußfeld wird in Anlehnung an neurobiologische Erkenntnisse bei Primaten auf verschiedenen Ebenen des Netzwerkmodells vorgenommen (Abb. 2). Dabei wird davon ausgegangen, daß immer nur ein Teil der Neurone ein jeweiliges zusätzliches Signal empfängt. Die sensorische Integration wird dann durch die Neuronenpopulation geleistet.

5.1 Disparität

Disparität δ wird im Modell auf zwei unterschiedlichen Ebenen benutzt, die sich funktionell voneinander unterscheiden. Zunächst wird auf Ebene der Repräsentation des Flußfeldes eine disparitätsbasierte Rauschunterdrückung durchgeführt. Dabei wird angenommen, daß die Neuronen der ersten Schicht des Netzwerkes bereits eine räumliche Glättung

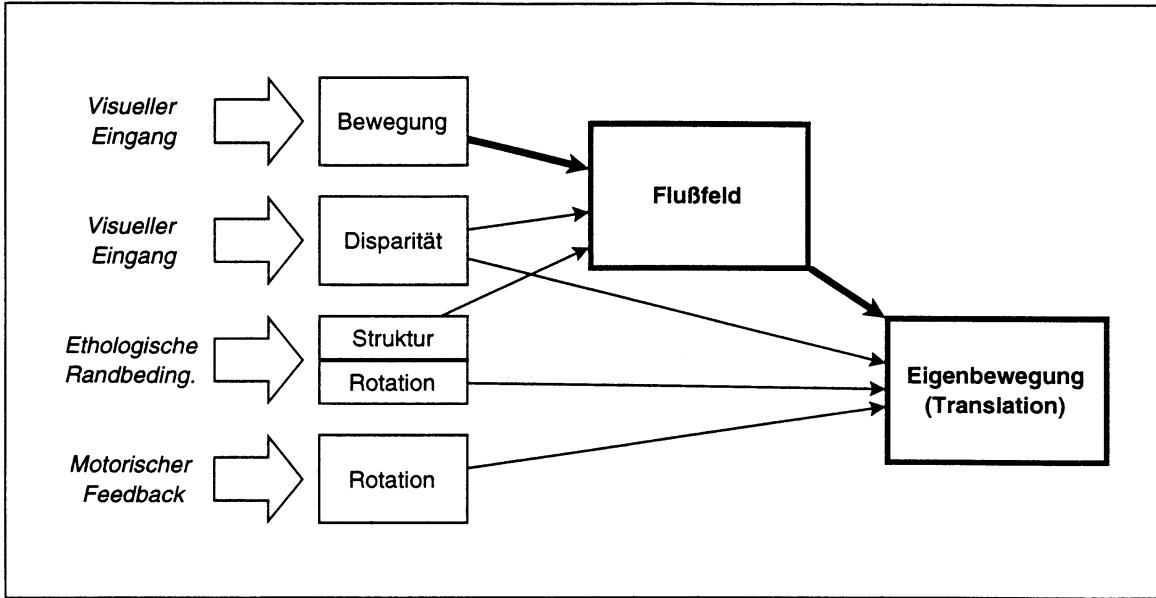


Abbildung 2: Schematische Darstellung der Integration verschiedener Informationsquellen zur Bestimmung der Eigenbewegung. Disparität, motorischer Feedback und ethologische Randbedingungen greifen auf verschiedenen Ebenen in die Flußfeldanalyse ein.

des Flußfeldes durchführen, indem sie Bewegungssignale innerhalb räumlich ausgedehnter rezeptiver Felder mitteln.

$$(8) \quad s_m = \sum_{i \in R} s_i / \sum_{i \in R} 1.$$

Entsprechend den neurobiologischen Befunden aus Area MT ist die Größenverteilung der rezeptiven Felder inhomogen und nimmt zur Peripherie hin zu. Die empirische Formel ist:

$$(9) \quad \text{RFSize} = 1.04\text{deg.} + 0.61\varepsilon.$$

Ebenfalls entsprechend neurobiologischer Daten aus MT ist die räumliche Mittelung disparitiesabhängig, d.h. sie wird nur innerhalb eines bestimmten Disparitätsintervalls $\pm D$ durchgeführt:

$$(10) \quad s_m = \sum_{i \in R, \delta < D} s_i / \sum_{i \in R, \delta < D} 1.$$

Die zweite Form der Disparitätsnutzung basiert auf Ergebnissen psychophysischer Experimente bei Menschen (Grigo & Lappe, 1996). Dort wurde gefunden, daß Bewegungssignale, die von weit entfernten Objekten stammen, bei der Bestimmung der Eigenbewegung stärker berücksichtigt werden als Bewegungssignale von nahen Objekten. Im Modell läßt sich dies durch eine zusätzliche disparitiesbezogene Gewichtung in der Kostenfunktion implementieren. Dazu muß Gl. (4) modifiziert werden:

$$(11) \quad R(\mathbf{T}) = \|\mathbf{G}(\delta)(\Theta - \mathbf{C}(\mathbf{T})\mathbf{q})\|^2.$$

Diese für die Eigenbewegungsbestimmung spezifische Integration der Disparität wird damit erst in der zweiten Schicht des Netzwerkes realisiert.

5.2 Extraretinale Signale (Motorischer Feedback)

Da die das Flußfeld erzeugende Bewegung im allgemeinen eine Überlagerung aus Vorwärtsbewegung und Augen- oder Kamerarotation ist, sind Rückmeldungen über motorische Kommandos, die Richtung und Geschwindigkeit der Rotation anzeigen, geeignet, den rotationsbedingten Anteils des Flußfeldes zu kompensieren. Da bei einer Rotation die induzierte Bewegung der visuellen Objekte unabhängig von ihrer Entfernung ist, genügt allein die Kenntnis der Rotationsparameter, um die rotationsbedingte Komponente jedes einzelnen Flußvektors zu kennen. Allerdings muß man davon ausgehen, daß die Kenntnis der Rotationsparameter meist nur unvollständig ist, so daß auch nach Ausnutzen

des motorischen Feedbacks das Flußfeld immer noch Rotationsanteile enthält. Analog zur Kodierung der Flußvektoren durch Neuronenpopulationen wird auch die Rotation durch eine Population von, in diesem Fall rotationsselektiven. Neuronen p_k dargestellt:

$$(12) \quad \Omega = \gamma \sum_1 p_k \mathbf{e}_k.$$

wobei γ ein Faktor ist, der die Ungenauigkeit der Kenntnis der Rotation berücksichtigt. Der rotationsspezifische Anteil des Flußfeldes ist damit:

$$(13) \quad \Theta^P = \gamma \sum_{k=1}^n p_k \mathbf{B}(x, y) \mathbf{e}_k.$$

Die synaptischen Gewichte müssen dann

$$(14) \quad \sum_m J_{mj} s_m + \sum_k J_{kj}^P p_k = (\Theta - \Theta^P)^t \mathbf{C}^\perp(\mathbf{T}_j)$$

erfüllen, wobei J^P die Gewichte zwischen den rotationsselektiven Neuronen und den Neuronen der zweiten Schicht sind.

5.3 Ethologische Randbedingungen

Unter dem Begriff 'ethologische Randbedingungen' lassen sich Umwelt- und Verhaltensbedingungen verstehen, die konstant sind oder sich zumindest für die zu lösende Aufgabe in immer gleicher Weise auswirken. Durch Ausnutzen derartiger immer gleicher Zusammenhänge lässt sich z.B. die Zahl der abzuschätzenden Parameter oder die Anzahl notwendiger Messungen reduzieren. Auf einer stärker kognitiven Stufe, etwa bei der Objekterkennung, würde man dies als 'Vorwissen' bezeichnen und gegebenenfalls nutzen. Ein Beispiel aus dem Bereich der Robotik ist etwa die Ausnutzung der Struktur der Umwelt zur Vereinfachung der Bestimmung des optischen Flußfeldes (Mallot, Bülthoff, Little, & Bohrer, 1991). Hier soll der Einfluß und Nutzen ethologischer Randbedingungen in zwei Bereichen berücksichtigt werden.

Der erste betrifft die schon angesprochene Überlagerung von Vorwärtsbewegung und Augen/Kamerarotationen. Das menschliche Auge besitzt drei Rotationsfreiheitsgrade. Jedoch ist das okulomotorische System des Menschen so organisiert, daß sich in der Praxis die Anzahl aller durchführbaren Augenstellungen durch zwei Freiheitsgrade beschreiben lässt. Der dritte, nämlich die Augentorsion, ist mit Elevation und Azimuth gekoppelt (Listing'sche Gesetze). Mit dieser Randbedingung lässt sich die Zahl der Parameter im Minimierungsverfahren verringern:

$$(15) \quad \Omega = f(\Omega_X, \Omega_Y)$$

Eine weitere Verringerung lässt sich erreichen, wenn man davon ausgeht, daß die Augenbewegungen mit der Eigenbewegung gekoppelt sind. Dies ist auch im allgemeinen der Fall, da Augenfolgebewegungen dazu dienen, einzelne Objekte während der Eigenbewegung zu verfolgen. Sind diese Objekte statisch im Raum, so wird ihre visuelle Bewegung allein durch die Eigenbewegung und die Entfernung vom Beobachter bestimmt. Unter dieser Annahme reduziert sich die Zahl der zusätzlichen Freiheitsgrade der Augenbewegung auf einen, nämlich die Entfernung Z_F des Folgeziels vom Beobachter.

$$(16) \quad \Omega = g(\mathbf{T}, Z_F)$$

Eine weitere ethologische Randbedingung, die auch in bewegungsverarbeitenden Großhirnarealen höherer Säugetiere wie Katzen und Affen ausgenutzt wird, ist die Annahme, daß Eigenbewegungsinformation im wesentlichen nur bei Vorwärtsbewegung benötigt wird. Das bedeutet, daß relevante Flußfelder immer eine stark radiale Struktur aufweisen. Wie man aus Abb. 1 ersehen kann, ist dies nicht nur für geradlinige Bewegung der Fall, sondern auch, wenn ethologisch relevante (siehe oben) Augenfolgebewegungen ausgeführt werden. Diese Struktur des Flußfeldes gestattet es, die Zahl der zur Repräsentation des Flußfeldes notwendigen Neurone zu reduzieren bzw. ihre Verteilung zu optimieren. So ist ein Verzicht auf zentripetale Bewegungssignale, d.h. Bewegungssignale, die auf das Zentrum der Retina hinführen, ohne größere Einbußen an Genauigkeit der Eigenbewegungsbestimmung möglich.

6 Vergleichende Simulationen

Ziel des Modells ist neben der Entwicklung von Algorithmen auch zum wesentlichen Teil die Modellierung des Sehsystems. Die Simulationen mit Hilfe des Modells erfolgen auf zwei verschiedenen Ebenen. Zum einen ist die Systemebene wichtig, d.h. die Leistung des Gesamtsystems wird geprüft und mit psychophysischen Untersuchungen an Menschen verglichen. Zum anderen werden aber auch die Eigenschaften einzelner Komponenten, also einzelner Modellneurone

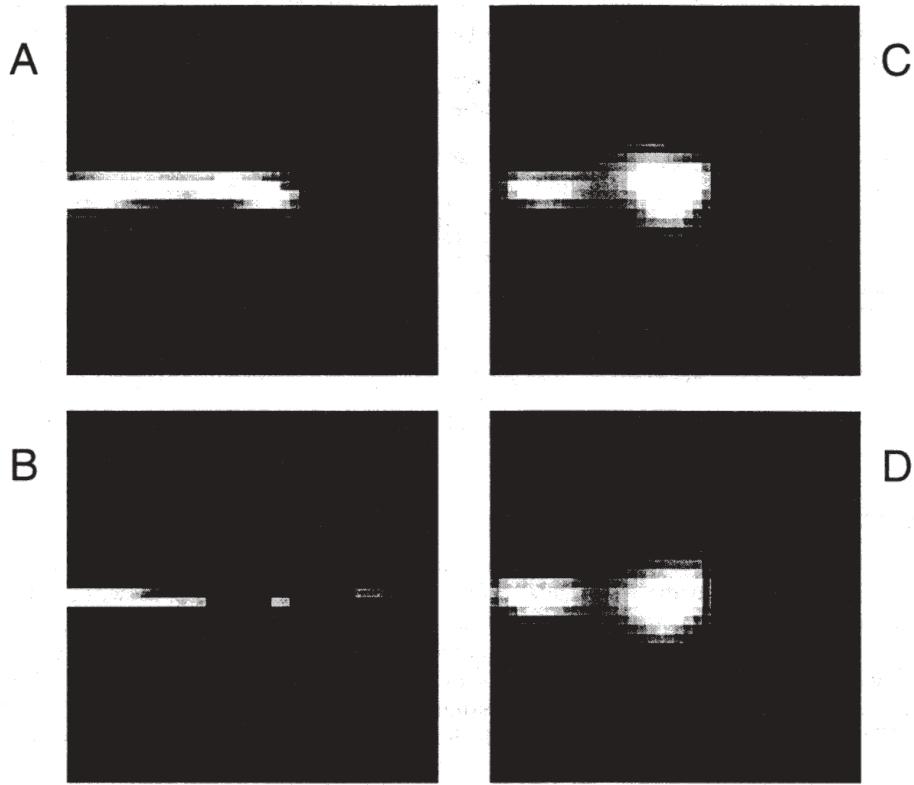


Abbildung 3: Nutzung motorischen Feedbacks und ethologischer Randbedingungen bei der Analyse uneindeutiger Flußfelder. Die einzelnen Abbildungen zeigen Graustufenkarten der vom Netzwerk errechneten Werte der Kostenfunktion $R(T_X, T_Y)$ wenn ein mehrdeutiges Flußfeld gegeben ist. Die Abbildungen A, C und D zeigen jeweils zwei mögliche Lösungen an. Nur das linke Maximum entspricht der korrekten Lösung. In Abb. B ist die korrekte Lösung gefunden. In diesem Fall erhält des Netzwerk zusätzlich zum Flußfeld auch motorischen Feedback und stellt außerdem ethologisch sinnvolle Randbedingungen an die Augenrotation.

der zweiten Schicht. in Simulationen bestimmt und mit den Eigenschaften der Neurone in bewegungsverarbeitenden Arealen des Primatenkortex verglichen. Dabei zeigen sich sowohl auf Neuronen- wie auch auf Systemebene zahlreiche Übereinstimmungen (Lappe et al., 1996; Lappe & Rauschecker, 1993, 1995b). Im folgenden soll dies an zwei Beispielen dargestellt werden.

6.1 Mehrdeutige Flußfelder erfordern multimodale Integration

Bewegung relativ zu einer ebenen Fläche erzeugt ein uneindeutiges optisches Bewegungsfeld. Zu jedem gegebenen Zeitpunkt lassen sich zwei mögliche Sätze von Bewegungsparametern angeben, um dieses Flußfeld zu erzeugen (Longuet-Higgins, 1984). Menschliche Versuchspersonen unterliegen in diesem Fall einer charakteristischen Fehleinschätzung ihrer Eigenbewegung, wenn sie ihre Entscheidung allein auf das optische Flußfeld stützen müssen. Wenn die Versuchspersonen dagegen motorischen Feedback zu Hilfe nehmen können, wird diese Fehleinschätzung aufgehoben (Warren & Hannon, 1990). Abb. 3 zeigt den Ausgang des Netzwerkmodells in vier verschiedenen Bedingungen, die diesen Experimenten entsprechen. Dargestellt sind die errechneten Werte der Kostenfunktion $R(T_X, T_Y)$ als Graustufenabbildung. Als Eingang wurde ein synthetisches Bewegungsfeld verwendet, das Bewegung auf eine vertikale Wand zu bei gleichzeitiger Augenrotation simuliert. Abb. 3A zeigt eine Simulation, in der nur dieser visuelle Eingang vorlag. Man erkennt deutlich zwei Helligkeitsmaxima, die den beiden möglichen Lösungen entsprechen. Dabei entspricht das linke Maximum der korrekten Lösung, das Maximum in der Mitte einer inkorrekt Lösung, die eine geradlinige Translation ohne Augenrotation beschreibt. In Abb. 3B dagegen liegt zusätzlich zum Flußfeld noch ein motorisches Feedback-Signal am Eingang an. Man erkennt, daß dadurch die Lösung in der Mitte unterdrückt und nur noch die korrekte Lösung gefunden wird. Motorischer Feedback kann also vom Modell sinnvoll genutzt werden. Dies ist allerdings nur in Verbindung mit der Berücksichtigung ethologischer Randbedingungen der Fall, wie die Abbildungen C

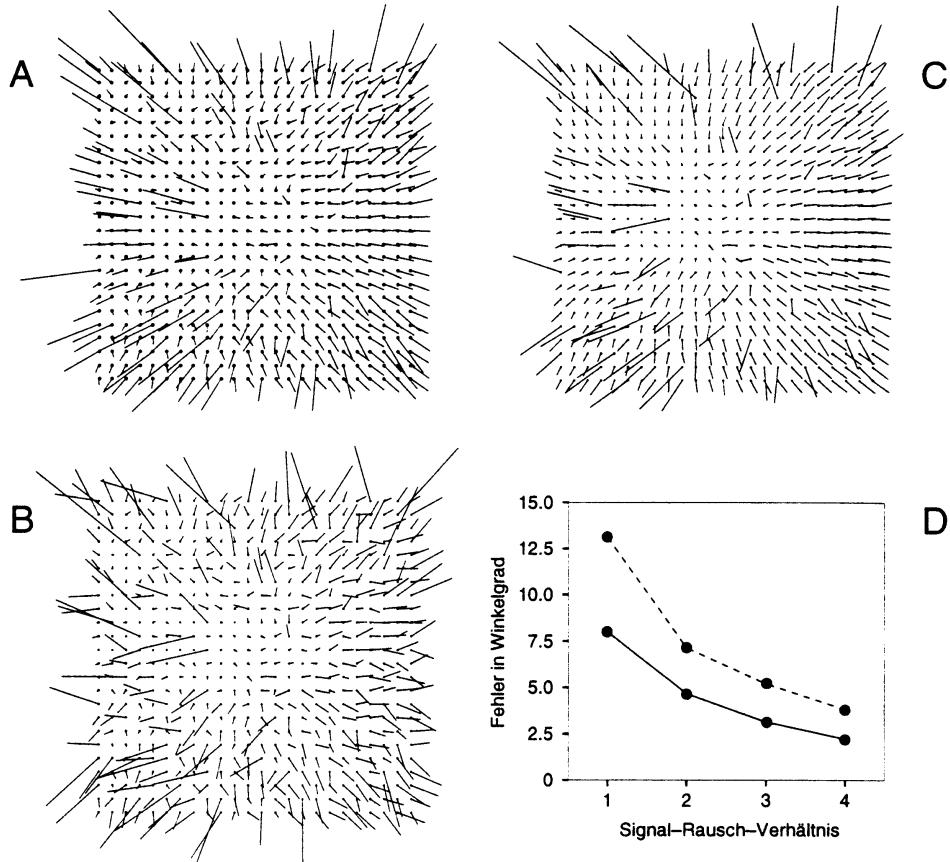


Abbildung 4: Einfluß disparitätssensitiver räumlicher Filterung des Flußfeldes auf die Robustheit der Eigenbewegungsschätzung. Dem Flußfeld einer Eigenbewegung durch eine randomisierte Umwelt (A) wurde additives Rauschen überlagert (B). C zeigt das verrauschte Flußfeld nach disparitätssensitiver Filterung. D zeigt den Fehler der Eigenbewegungsschätzung ohne (---) und mit (—) disparitätsbasierter Filterung.

und D zeigen. Die Eingänge entsprechen jeweils denen der Abbildungen A und B, jedoch weisen die Netzwerkausgänge in C und D in beiden Fällen zwei Lösungen auf. In A und B enthielt das Netzwerk jeweils einen Teil von Neuronen, die die oben beschriebenen ethologisch sinnvollen Randbedingungen an die Augenbewegungen implementieren. D.h. ein Teil der Neuronen erwartet eine mit der Eigenbewegung gekoppelte Augenbewegung. Diese Bedingung ist im Fall B für die inkorrekte Lösung nicht erfüllt. In C und D werden keine Randbedingungen an die Rotation gestellt, und auch nach Hinzufügen des motorischen Signals sind beide Lösungen möglich.

6.2 Disparitätsbasierte Filterung erlaubt robuste Flußfeldanalyse

Basierend auf Untersuchungen im Großhirnareal MT in Primaten findet im Modell eine funktionelle Verrechnung von komplexen Bewegungsreizen und Stereodisparität bereits in der ersten Schicht statt. Hier wird eine lokale räumliche Mittelung der Bewegungssignale nur innerhalb eines Disparitätsbereiches durchgeführt. In Flußfeldern, die aus einer Eigenbewegung durch eine dreidimensionale Umwelt entstehen, sind im allgemeinen Bewegungen von in der gleichen Tiefe gelegenen Punkten weitgehend gleichgerichtet, während Punkte in verschiedenen Tiefen starke Unterschiede in der Bewegungsrichtung aufweisen können. Die Mittelung innerhalb einer Tiefenebene erscheint also sinnvoll, zwischen verschiedenen Tiefenebenen jedoch nicht. Wie Abb. 4 zeigt, führt eine solche Mittelung zu einer deutlichen Verbesserung der Robustheit gegenüber Störungen. Abb. 4A zeigt ein synthetisches Flußfeld, das aus einer Vorwärtsbewegung durch eine randomisierte dreidimensionale Umgebung resultiert. Gleichzeitig ist der Vorwärtsbewegung eine Rotation überlagert. In Abb. 4B ist dieses Flußfeld durch additives Rauschen gestört. Dazu wurde zu jedem Flußvektor θ ein Störvektor der Länge $k\|\theta\|$ mit zufällig ausgewählter Richtung addiert. Die Konstante k beschreibt das Signal-Rausch-Verhältnis. Abb. 4C zeigt die Rekonstruktion des Flußfeldes nach disparitätsbasierter Filterung. In Abb. 4D

ist der Fehler bei der Eigenbewegungsschätzung in Abhängigkeit vom Signal-Rausch-Verhältnis aufgetragen. Die durchgezogene Linie gibt die mittleren Fehler nach disparitätsbasierter Filterung an, die gestrichelte Linie die Fehler ohne disparitätsbasierte Filterung. Man erkennt eine deutliche Verringerung des Fehlers bei Ausnutzung der Disparitätsinformation. Ein solches Verhalten ist auch in psychophysischen Experimenten bei Menschen beobachtet worden (van den Berg & Brenner, 1994).

7 Zusammenfassung

In dieser Arbeit wurde ein Modell der Eigenbewegungsbestimmung eines aktiven Beobachters beschrieben. Die aktive visuelle Exploration einer unbekannten Umwelt während einer Vorwärtsbewegung führt zu charakteristischen Veränderungen des optischen Flußfeldes. Mögliche Mechanismen zur Verarbeitung solcher Flußfelder lassen sich aus neurobiologischen Befunden ableiten. Dabei müssen verschiedene zusätzliche Signale und Randbedingungen in die Flußfeldverarbeitung integriert werden. Das auf dieser Grundlage entwickelte neurobiologische Modell läßt sich in Simulationen psychophysischer Experimente testen. Dabei zeigt sich, daß die Kombination verschiedener Informationsquellen zu einer robusteren und leistungsfähigeren Eigenbewegungsschätzung führt, die in vielen Bereichen der menschlichen Wahrnehmung entspricht.

Literatur

- Bruss, A. R., & Horn, B. K. P. (1983). Passive navigation. *Comp. Vis. Graph. Image Proc.*, 21, 3–20.
- Clark, J. J., & Yuille, A. L. (1990). *Data Fusion for Sensory Information Processing Systems*. Kluwer, Boston, MA.
- Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.*, 65(6), 1329–1345.
- Fermüller, C., & Aloimonos, Y. (1993). The role of fixations in visual motion analysis. *I.J. Computer Vision*, 11(2), 165–186.
- Grigo, A., & Lappe, M. (1996). Illusory optic flow transformation with binocular vision. *Perception*, 25 (suppl.), 66.
- Heeger, D. J., & Jepson, A. (1992). Subspace methods for recovering rigid motion I: Algorithm and implementation. *I.J. Computer Vision*, 7(2), 95–117.
- Lappe, M., Bremmer, F., Pekel, M., Thiele, A., & Hoffmann, K.-P. (1996). Optic flow processing in monkey STS: A theoretical and experimental approach. *J. Neurosci.*, 16, 6265–6285.
- Lappe, M., & Rauschecker, J. P. (1993). A neural network for the processing of optic flow from ego-motion in higher mammals. *Neural Comp.*, 5, 374–391.
- Lappe, M., & Rauschecker, J. P. (1995a). An illusory transformation in a model of optic flow processing. *Vision Res.*, 35, 1619–1631.
- Lappe, M., & Rauschecker, J. P. (1995b). Motion anisotropies and heading detection. *Biol. Cybern.*, 72, 261–277.
- Longuet-Higgins, H. C. (1984). The visual ambiguity of a moving plane. *Proc. Royal Soc. London B*, 223, 165–175.
- Mallot, H. A., Bülthoff, H. H., Little, J. J., & Bohrer, S. (1991). Inverse perspective mapping simplifies optical flow computation and obstacle detection. *Biol. Cybern.*, 64, 177–185.
- van den Berg, A. V., & Brenner, E. (1994). Why two eyes are better than one for judgements of heading. *Nature*, 371, 700–702.
- Warren, Jr., W. H., & Hannon, D. J. (1990). Eye movements and optical flow. *J. Opt. Soc. Am. A*, 7(1), 160–169.