

Aus dem Institut für Neuro- und Bioinformatik der Universität zu Lübeck

Direktor: Professor Dr. Thomas Martinetz

# Die intrinsische Dimension in der visuellen Informationsverarbeitung

Habilitationschrift

verfasst und der Technisch-Naturwissenschaftlichen Fakultät  
der  
Universität zu Lübeck

zur Erlangung der Venia legendi für das Fach  
Informatik

vorgelegt von

Dr.-Ing. Erhardt Barth

Lübeck, Mai 2004

## Liste der beigefügten Originalarbeiten

1. Barth, E. und Watson, A. B. (2000). A geometric framework for nonlinear visual coding. *Optics Express*, 7:155–85. <http://www.opticsexpress.org/oearchive/source/23045.htm>.
2. Barth, E. (1999). Bewegung als intrinsische Geometrie von Bildfolgen. In Förster, W., Buhmann, J. M., Faber, A., und Faber, P. (Hrsg.), *Musternererkennung 99*, Seiten 301–308, Bonn. Springer, Berlin.
3. Barth, E. (2000). The minors of the structure tensor. In Sommer, G. (Hrsg.), *Musternererkennung 2000*, Seiten 221–228. Springer, Berlin.
4. Mota, C. und Barth, E. (2000). On the uniqueness of curvature features. In Baratoff, G. und Neumann, H. (Hrsg.), *Dynamische Perzeption*, Band 9 der Reihe *Proceedings in Artificial Intelligence*, Seiten 175–8, Köln. Infix Verlag.
5. Barth, E., Stuke, I., und Mota, C. (2002). Analysis of motion and curvature in image sequences. In *Proc. IEEE Southwest Symp. Image Analysis and Interpretation*, Seiten 206–10, Santa Fe, NM. IEEE Computer Press.
6. Barth, E., Stuke, I., Aach, T., und Mota, C. (2003a). Spatio-temporal motion estimation for transparency and occlusion. In *Proc. IEEE Int. Conf. Image Processing*, Band III, Seiten 69–72, Barcelona, Spain. IEEE Signal Processing Soc.
7. Barth, E., Drewes, J., und Martinetz, T. (2003b). Dynamic predictions of tracked gaze. In *Seventh International Symposium on Signal Processing and its Applications*, Paris. Special Session on Foveated Vision in Image and Video Processing.
8. Mota, C., Dorr, M., Stuke, I., und Barth, E. (2004). Categorization of transparent-motion patterns using the projective plane. *International Journal of Computer & Information Science*, 5(2).

## Zusammenfassung

Zu verstehen, wie Menschen visuelle Information verarbeiten, ist eine große intellektuelle Herausforderung und ein Thema der Neuroinformatik. Der technischen Verarbeitung von Bildfolgen wird ein stark wachsender Markt vorausgesagt, jedoch sind derzeitige künstliche Sehsysteme dem menschlichen Sehen unterlegen. Deshalb werden in der Informatik beachtliche Anstrengungen unternommen, die biologischen Systeme zu verstehen, um daraus zu lernen, wie bessere technische Systeme gebaut werden können. Weiterhin wird zunehmend erkannt, dass die Technik aus der Perspektive der Nutzer nur dann sinnvoll ist, wenn sie in einer vernünftigen Weise mit dem Benutzer interagiert und an dessen Bedürfnisse angepasst wird. Auch dazu muss man sich als Informatiker mit der Funktionsweise der biologischen Systeme, speziell der menschlichen Wahrnehmung, auseinandersetzen. Schließlich bestimmt in den meisten Fällen allein die zum Teil subjektive Wahrnehmung den Wert der Dinge und die persönliche Lebensqualität. Ich denke, wir stehen am Anfang einer Entwicklungsphase, in der es möglich erscheint, diese Wahrnehmung mithilfe wissenschaftlicher Methoden mit zu gestalten. Dieses stellt allerdings eine weitaus größere Herausforderung dar als der traditionelle Umgang mit den sogenannten harten Tatsachen.

Wenn man versucht, die Informationsverarbeitung in biologischen Systemen zu verstehen, stößt man schnell auf komplexe Probleme; bekannte Theorien und Methoden erweisen sich oft als unzulänglich. Beim Sehen wird die visuelle Information in einer Art und Weise verarbeitet, die angesichts der gegebenen Umwelt optimal ist. Von Bedeutung ist dabei, dass sich die Helligkeitswerte in einem Bild über Ort und Zeit unterschiedlich ändern können. Trotz theoretisch fundierter Ansätze der mehrdimensionalen Signalverarbeitung lassen sich damit die Grundtypen dieser Änderungen, wie sie auch von visuellen Neuronen kodiert werden, formal nicht gut beschreiben. Wir haben deshalb den Begriff der *intrinsischen Dimension* geprägt. Diese beschreibt, inwieweit ein mehrdimensionales Signal die prinzipiell vorhandenen Freiheitsgrade nutzt. Werden von  $n$  Freiheitsgraden (lokal) nur  $m$  Freiheitsgrade genutzt, so ist das Signal intrinsisch  $m$ -dimensional, und wir sprechen von einem *imD*-Signal. Es zeigt sich, dass die Theorie der intrinsischen Dimension eine nichtlineare Theorie sein muss. Wir betrachten deshalb Bildfolgen als Hyperflächen und entwerfen damit eine differentialgeometrische Theorie der intrinsischen Dimension.

Daraus ergeben sich zunächst neuartige Methoden zur Bestimmung von einfachen und

mehrfachen Bewegungen in Bildfolgen. Zur Bestimmung einfacher Bewegungen wurden neue mathematische Beziehungen zwischen den Komponenten des Riemannschen Krümmungstensors und den Bewegungsparametern gefunden. Analog dazu konnten die Bewegungsparameter durch die Minoren des Strukturensors bestimmt werden. Der Krümmungstensor misst die Krümmung der als Hyperfläche interpretierten Bildfolge. Dazu werden Ableitungen zweiter Ordnung der Helligkeitsfunktion sowie nichtlineare Verknüpfungen dieser Ableitungen bestimmt. Der in der Bildverarbeitung bereits bekannte Strukturtensor kann als metrischer Tensor interpretiert werden und besteht aus den Ableitungen erster Ordnung. Zur Bestimmung mehrfacher Bewegungen wurde im Rahmen eines DFG-Projektes eine neue Theorie überlagelter Bewegungen entwickelt. Bemerkenswert ist, dass sich durch einen interdisziplinär motivierten Ansatz Erkenntnisse ergaben, die für die technische Bildverarbeitung neu und nützlich sind.

Weiterhin konnten wir bestimmte visuelle Neurone und Wahrnehmungsleistungen erklären. Damit konnte eine neue Interpretation bewegungsselektiver Neurone im Areal MT von Primaten gefunden werden, wonach diese Neurone eine optimale, weil weniger redundante Repräsentation der Umwelt leisten und zwar dadurch, dass sie i2D Signale kodieren. Weiterhin konnte die Wahrnehmung bestimmter Bewegungsreize, wie Gitter hinter Blenden und überlagerte Bewegungen, erklärt werden. Derartige Reize hatten in der Sehforschung einige offene Fragen hinterlassen.

Schließlich führten meine interdisziplinären Einsichten zu einigen, zum Teil größeren, von der Deutschen Forschungsgemeinschaft, dem BMBF, der EU und der mittelständischen Industrie geförderten Projekten, in denen die theoretischen Ergebnisse auch angewendet werden.

# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einleitung</b>  | <b>7</b>  |
| <b>2</b> | <b>Die intrinsische Dimension</b>                              | <b>9</b>  |
| 2.1      | Systemtheorie der intrinsischen Dimension . . . . .            | 10        |
| 2.2      | Differentialgeometrie und intrinsische Dimension . . . . .     | 10        |
| 2.2.1    | Riemannscher Krümmungstensor . . . . .                         | 11        |
| 2.2.2    | Strukturtenor . . . . .  | 12        |
| 2.3      | Informationsgehalt und intrinsische Dimension . . . . .        | 13        |
| 2.4      | Bewegung als intrinsische Geometrie von Bildfolgen . . . . .   | 14        |
| 2.5      | Bewegung und die Minoren des Strukturtenors . . . . .          | 15        |
| 2.6      | Überlagerte Bewegungen . . . . .                               | 15        |
| 2.6.1    | Mehrfache transparente Bewegungen . . . . .                    | 15        |
| 2.6.2    | Klassifikation von Bewegungsmustern . . . . .                  | 17        |
| 2.6.3    | Mehrfache verdeckende Bewegungen . . . . .                     | 18        |
| <b>3</b> | <b>Anwendungen</b>   | <b>19</b> |
| 3.1      | Bestimmung einfacher Bewegungen . . . . .                      | 19        |
| 3.1.1    | Krümmungstensor . . . . .                                      | 19        |
| 3.1.2    | Strukturtenor . . . . .  | 20        |
| 3.2      | Mehrfache überlagerte Bewegungen . . . . .                     | 20        |
| 3.2.1    | Transparente Überlagerungen . . . . .                          | 20        |
| 3.2.2    | Verdeckende Überlagerungen . . . . .                           | 21        |
| 3.3      | Modelle visueller Neurone . . . . .                            | 22        |
| 3.4      | Modellierung visueller Wahrnehmung . . . . .                   | 23        |
| 3.5      | Prädiktion der Blickrichtung . . . . .                         | 24        |
| <b>4</b> | <b>Diskussion</b>  | <b>25</b> |
| 4.1      | Die intrinsische Dimension in der Signalverarbeitung . . . . . | 25        |
| 4.2      | Die intrinsische Dimension in der Sehforschung . . . . .       | 26        |
| 4.3      | Bewegungsschätzung . . . . .                                   | 27        |
| 4.4      | Eigene, interdisziplinäre Forschungsprojekte . . . . .         | 28        |

|       |  |    |
|-------|--|----|
| 4.4.1 | Komplexe Bewegungen . . . . .                      | 28 |
| 4.4.2 | Lenkung der Aufmerksamkeit . . . . .               | 28 |
| 4.4.3 | KFZ-Assistenz-Systeme . . . . .                    | 29 |
| 4.4.4 | Computation by gaze interaction (COGAIN) . . . . . | 29 |

# 1 Einleitung

Dieses ist eine Habilitationsschrift im Rahmen einer kumulativen Habilitation. Demzufolge werden acht Originalarbeiten aus der Zeit nach meiner Promotion an der TU München abgedruckt und von einer Synopsis begleitet. Der Begriff *intrinsische Dimension* ist gewissermaßen der rote Faden durch die Themen und wird in Abschnitt 2 näher erläutert.

Die Arbeit *A geometric framework for nonlinear visual coding* [1] ist während meines Aufenthaltes bei der NASA entstanden und geht auf eigene Ideen zurück. Der dortige Kontext war zweifellos der Arbeit dienlich, einmal durch Hinweise auf noch unerklärte Daten aus der Sehforschung und weiterhin durch die intellektuell stimulierende Umgebung. Die San Francisco Bay Area ist einer der weltweit besten Orte für die interdisziplinäre Arbeit an Problemen des künstlichen und biologischen Sehens. Dieser Arbeit verdanke ich auch den Schloeffmann Preis.

Die Arbeit *Bewegung als intrinsische Geometrie von Bildfolgen* [2] entstand im Wesentlichen an der Universität zu Lübeck, wie die folgenden Arbeiten auch. Es ging mir dabei vor allem darum, die aus der vorigen Arbeit gewonnenen Erkenntnisse in der technischen Bildverarbeitung anzuwenden. Schließlich ergaben sich durch den differentialgeometrischen Zugang zum Problem der Bewegungsschätzung neue und einfache Ausdrücke, die gute praktische Ergebnisse lieferten und, trotz ausgiebiger Forschung auf diesem Gebiet, bis dahin übersehen wurden.

Die Arbeit *The minors of the structure tensor* [3] schließt an die vorige Arbeit an und wurde ebenfalls auf der Tagung der Deutschen Arbeitsgemeinschaft für Mustererkennung vorgestellt. Wiederum gelang es, neue Methoden der Bewegungsschätzung zu finden und deren Überlegenheit durch einen Vergleich mit gängigen Verfahren zu beweisen. Die Neuheit war dabei kein kleines Inkrement relativ zu bereits Vorhandenem, sondern bestand in neuen Ausdrücken für die Bewegungsvektoren anhand von Ableitungen erster Ordnung, sowie einem neuen Verfahren zur Kombination der Bewegungsschätzung mit einer Segmentierung. In dieser und der vorigen Arbeit ist der Begriff der intrinsischen Dimension eng verknüpft mit dem Problem der Bestimmung von Konfidenzmaßen für bestimmte Bewegungsmodelle.

Die Arbeit *On the uniqueness of curvature features* [4] beinhaltet ein Schlüsselergebnis bezüglich der intrinsischen Dimension. Es wird dort bewiesen, dass Signale mit int-

rinsischer Dimension kleiner als zwei redundant sind. Den Beweis eines entsprechenden Theorems hat Cicero Mota im Rahmen seiner Doktorarbeit ausgearbeitet. Die Anregung dazu erhielt er 1996 während meiner Vorlesungsreihe in Rio de Janeiro.

Bei *Analysis of motion and curvature in image sequences* [5] handelt es sich um eine eingeladene Originalarbeit, die eine Übersicht über die bis zu dem Zeitpunkt erarbeiteten Ergebnisse zur Bewegungsschätzung bietet. Erstmals werden dort theoretische und praktische Ergebnisse zu verdeckenden Bewegungen gezeigt.

*Spatio-temporal motion estimation for transparency and occlusion* [6] ist ebenfalls eine eingeladene und begutachtete Originalarbeit und wurde auf einer renommierten Tagung vorgestellt. Darin wird die Theorie der mehrfachen Bewegungen zusammengefasst und eine neue Bewegungsgleichung für verdeckende Bewegungen erstmalig vorgestellt. Die mathematische Ableitung dieses wichtigen Ergebnisses verdanken wir Cicero Mota. Inzwischen haben wir gemeinsam mehrere Arbeiten zum Thema mehrfache Bewegungen publiziert und das in einem effizienten und wachsenden Team. Zuerst kam Cicero Mota dazu und ließ alles mathematisch interessanter werden, dann verdanken wir Ingo Stuke den Großteil der Simulationen, sowie ihm und Til Aach einige neue Ideen zur Erweiterung des ursprünglichen Ansatzes.

*Dynamic predictions of tracked gaze* [7] ist ebenfalls ein eingeladener Beitrag. Hier wird das Konzept der intrinsischen Dimension genutzt, um vorauszusagen, wo jemand beim Betrachten einer Bildfolge hinschauen wird. Dieses Thema ist Teil eines für uns hochaktuellen, vom BMBF geförderten Projektes, in dem neuartige Formen der visuellen Kommunikation und Interaktion entwickelt werden. Die ursprüngliche Idee zu diesem Projekt stammt von mir und wurde von Thomas Martinetz in einem frühen Stadium gefördert und ergänzt. Inzwischen setzen wir hier voll auf Teamarbeit, derzeit sind das hauptsächlich Michael Dorr, der schon lange dabei ist, sowie Martin Böhme und Christopher Krause.

Die Arbeit *Categorization of transparent-motion patterns using the projective plane* [8] befindet sich gerade im Druck und erklärt in interdisziplinärer Manier einige Phänomene des Bewegungssehens. Ein Großteil der Experimente wurde von Michael Dorr im Rahmen seiner Studienarbeit durchgeführt. Theoretisch erarbeitet wird erstmalig eine komplette Kategorisierung von Bewegungsmustern anhand des Rangs der generalisierten Strukturtenoren. Auch wird die projektive Ebene als Beschreibungsform eingeführt, eine Idee, die

in Gesprächen mit Cicero Mota geboren und von ihm später mathematisch ausgearbeitet wurde.

Weitere Bewertungen der ausgewählten Arbeiten und der dort berichteten Ergebnisse finden sich in der Zusammenfassung und der Diskussion.

Die folgende Synopsis erklärt in Abschnitt 2 zunächst das Konzept der intrinsischen Dimension und fasst die theoretischen Ergebnisse aller Arbeiten in diesem Kontext zusammen. Im Abschnitt 3 werden Anwendungen vorgestellt, die folgendermaßen unterteilt werden: Schätzung einfacher (3.1) und mehrfacher (3.2) Bewegungen, Modelle visueller Neurone (3.3), Modellierung visueller Wahrnehmung (3.4) und Prädiktion der Blickrichtung (3.5). Es folgen Diskussion, Danksagung und ein Literaturverzeichnis.

## 2 Die intrinsische Dimension

Der Begriff der intrinsischen Dimension unterscheidet zwischen den prinzipiellen Freiheitsgraden und den (lokal) tatsächlich genutzten Freiheitsgraden eines mehrdimensionalen Signals [9, 10]. Eine  $n$ -dimensionale Funktion  $f$  kann in  $k$  Richtungen konstant sein und ist damit durch eine  $m$ -dimensionale Funktion  $g$  vollständig bestimmt. Die intrinsische Dimension  $m = n - k$  ergibt sich daraus, dass die Koordinaten von  $g$  durch Rotation so gewählt werden, dass  $m$  minimiert wird. So sind z.B. im Falle statischer Bilder konstante Bereiche intrinsisch nulldimensional (i0D), gerade Kanten intrinsisch eindimensional (i1D) und Ecken intrinsisch zweidimensional (i2D).

Betrachtet man die Fourier-Transformierte der Signale, so transformieren sich die konstanten Richtungen zu Dirac-Delta Distributionen, d.h., die Energie von Signalen mit  $m < n$  ist auf Unterräume des Fourier-Raumes beschränkt. Dieser Umstand macht deutlich, dass es bei der Auswertung der intrinsischen Dimension im Wesentlichen darum geht, auf welche Unterräume die Energie des Signals durch Dirac-Delta Distributionen beschränkt wird. Damit wird auch klar, dass es um ein sehr grundsätzliches Problem geht. Dennoch wurde dieses Thema in der Vergangenheit vernachlässigt und es gibt immer noch keine zufriedenstellende Theorie der intrinsischen Dimension. Dieses liegt keinesfalls daran, dass eine solche Theorie nutzlos wäre, sondern eher daran, dass es eine nichtlineare Theorie sein muss. In der Signalverarbeitung wurden bisher mit Erfolg vor allem lineare Theorien eingesetzt.

## 2.1 Systemtheorie der intrinsischen Dimension

Aus systemtheoretischer Sicht stellt sich die Frage, welche Eigenschaften ein System haben muss, das in der Lage ist, auf Signale unterschiedlicher intrinsischer Dimension selektiv und differenziert zu reagieren. Ein solches System sollte zumindest in der Lage sein, Signale niedriger intrinsischer Dimension selektiv zu unterdrücken. Ein System, das Signale intrinsischer Dimension kleiner als  $m$  unterdrückt, nennen wir ein  $imD$ -System oder einen  $imD$ -Operator. Ein erstes Problem entsteht dadurch, dass lineare Systeme eine solche Leistung nicht erbringen können, wenn man vom einfachen Fall  $m = 1$  absieht. Bereits für  $m = 2$  muss das  $i2D$ -System nichtlinear sein, weil die Eigenfunktionen linearer Systeme intrinsisch eindimensional sind und lineare Systeme diese Eigenfunktionen lediglich gewichten und summieren. Somit können  $i1D$  Signale von linearen Systemen nur dadurch unterdrückt werden, dass alle Gewichte gleich Null gewählt werden. Die Auswertung der intrinsischen Dimension ist somit mithilfe linearer Systemtheorie grundsätzlich nicht möglich.

Für zweidimensionale Signale hatten wir zusammen mit meinen Kollegen eine nichtlineare Systemtheorie der intrinsischen Dimension entworfen [9, 11, 12, 13, 14, 15, 16, 17, 18, 19] und zur Modellierung des Sehens [20, 21, 22] sowie zur Bild-Kompression [23, 24, 25, 26] verwendet. Diese Theorie beinhaltete Methoden der linearen Systemtheorie, der Differentialgeometrie und die Volterra-Wiener Theorie nichtlinearer Systeme.

Eine Erweiterung auf mehr als zwei Dimensionen ist Thema der hier vorgestellten Arbeiten, beschränkt sich aber auf die Differentialgeometrie und die lineare Systemtheorie. Viele Ergebnisse werden lediglich mithilfe des sogenannten Strukturensors behandelt und nicht weiter generalisiert.

## 2.2 Differentialgeometrie und intrinsische Dimension

Ein guter Ausgangspunkt für eine nichtlineare Systemtheorie ist die Differentialgeometrie [27, 28, 29, 10]. Betrachtet man nämlich Bilder als Flächen, so findet man die zur intrinsischen Dimension analogen Flächentypen eben ( $i0D$ ), parabolisch ( $i1D$ ) und elliptisch/hyperbolisch (beide gekrümmt und  $i2D$ ). Demnach kann das Problem der intrinsischen Dimension in Analogie zur Krümmungsanalyse behandelt werden.

### 2.2.1 Riemannscher Krümmungstensor

Ist  $f(x, y, t)$  die Helligkeit am Ort  $(x, y)$  zum Zeitpunkt  $t$ , so beschreibt

$$S = (x, y, t, f(x, y, t)) \quad (1)$$

eine Hyperfläche. Die Krümmung dieser Hyperfläche wird durch den Riemannschen Krümmungstensor  $R$  gemessen: in Bereichen, wo die Hyperfläche flach ist, verschwindet der Tensor, d.h. alle Komponenten sind gleich Null, unabhängig vom gewählten Koordinatensystem. In 3D hat  $R$  sechs unabhängige Komponenten, die in kartesischen Koordinaten folgendermaßen durch die Ableitungen erster und zweiter Ordnung von  $f$  bestimmt sind:

$$\begin{aligned} R_{2121} &= (f_{yy}f_{xx} - f_{xy}^2)/(1 + \vec{\nabla}f^2) \\ R_{3131} &= (f_{tt}f_{xx} - f_{xt}^2)/(1 + \vec{\nabla}f^2) \\ R_{3232} &= (f_{tt}f_{yy} - f_{yt}^2)/(1 + \vec{\nabla}f^2) \\ R_{3121} &= (f_{yt}f_{xx} - f_{xt}f_{xy})/(1 + \vec{\nabla}f^2) \\ R_{3221} &= (f_{yt}f_{xy} - f_{yy}f_{xt})/(1 + \vec{\nabla}f^2) \\ R_{3231} &= (f_{tt}f_{xy} - f_{xt}f_{yt})/(1 + \vec{\nabla}f^2) \\ \text{mit: } 1 + \vec{\nabla}f^2 &= 1 + f_x^2 + f_y^2 + f_t^2 \end{aligned} \quad (2)$$

Die Krümmung ist in der Differentialgeometrie ein Maß für die Abweichung von der Flachheit und als solches sehr stark von der Dimension abhängig. Kurven sind immer flach, weshalb es in 1D kein Krümmungsmaß gibt ( $R$  verschwindet; die Krümmung von Kurven und die mittlere Krümmung  $H$  sind keine Krümmungsmaße im Sinne der Abweichung von der Flachheit). Flächen und Hyperflächen sind dann flach, wenn sie durch eine isometrische Transformation auf eine Ebene oder Hyperebene abzubilden sind (sie sind dann abwickelbar). In 2D hat  $R$  nur eine unabhängige Komponente (gleich  $R_{2121}$  mit  $f_t = 0$ ). Typische gekrümmte Merkmale in Bildern sind Ecken, Linienenden usw. In 2D, aber nur hier, sind  $R$  und die Gaußsche Krümmung äquivalent. Was aber sind gekrümmte Merkmale in Bildfolgen? Zur Beantwortung dieser Frage muss man  $R$  als Tensor (Gesamtheit der Komponenten) betrachten. Gilt  $R \neq 0$ , so ist die Bildfolge, d.h. die Hyperfläche (1), gekrümmt. In diesem Falle ist die intrinsische Dimension  $m \geq 2$ .

Zur weiteren Differenzierung können auch die mittlere Krümmung  $H$  und die Gaußsche Krümmung  $K$  der Hyperfläche (1) hinzugezogen werden. Gilt  $H \neq 0$ , so ist die intrinsische

Dimension eins oder höher, d.h.  $m \geq 1$ . Gilt  $K \neq 0$ , so ist  $m = 3$ . Signale mit  $m = 3$  entsprechen z.B. Diskontinuitäten im Flussfeld [30].

### 2.2.2 Strukturtensor

Ist  $f(x, y, t)$  wiederum die Helligkeitsfunktion, so ist der Strukturtensor definiert als:

$$\mathbf{J}(x, y, t) = h(x, y) * (\vec{\nabla} f)^T (\vec{\nabla} f) = h(x, y) * \begin{pmatrix} f_x^2 & f_x f_y & f_x f_t \\ f_x f_y & f_y^2 & f_y f_t \\ f_x f_t & f_y f_t & f_t^2 \end{pmatrix}. \quad (3)$$

Dabei bezeichnen Indizes die partiellen Ableitungen.  $h$  ist ein Faltungskern und  $*$  bezeichnet die Faltungsoperation, welche hier dazu genutzt wird, die nichtlinearen Komponenten des Tensors in einer gewissen Nachbarschaft gewichtet zu mitteln.

Wir betrachten nun die Minoren von  $\mathbf{J}$ , d.h. die Matrix

$$\mathbf{M} = \text{Minoren}(\mathbf{J}). \quad (4)$$

Die Elemente  $M_{ij}$ , ( $i, j = 1, 2, 3$ ) von  $\mathbf{M}$  sind die Determinanten der Untermatrizen, die man durch das Weglassen der Zeile  $4 - i$  und der Spalte  $4 - j$  erhält, d.h.,

$M_{11} = (h * f_x^2)(h * f_y^2) - (h * (f_x f_y))^2$ .  $\mathbf{J}$  ist eine symmetrische, positiv semidefinite Matrix und hat folgende Invarianten:

$$\begin{aligned} K &= \det \mathbf{J} = \lambda_1 \lambda_2 \lambda_3 \\ S &= (M_{11} + M_{22} + M_{33})/3 \\ &= (\lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_2 \lambda_3)/3 \\ H &= (\text{trace} \mathbf{J})/m = (\lambda_1 + \lambda_2 + \lambda_3)/3 \end{aligned} \quad (5)$$

$\lambda_i$  sind die Eigenwerte von  $\mathbf{J}$ .

Geometrisch ist der Strukturtensor  $\mathbf{J}$  (3) die Metrik der Einbettung  $F(x, y, t) = (f(x - x_1, y - y_1, t), \dots, f(x - x_s, y - y_s, t))$  und  $(x_s, y_s)$  sind dabei die Stellen, an denen  $f$  abgetastet wird. Die Invarianten  $K$ ,  $S$  und  $H$  sind metrische Invarianten dieser Einbettung und bestimmen die Krümmungen von  $F$ . Im Unterschied zu der vom Riemannschen Krümmungstensor gemessenen Krümmung sind diese Invarianten nicht unabhängig von der Einbettung, d.h. keine metrischen Invarianten.

Der oben dargestellte Zusammenhang zwischen Krümmungen und der intrinsischen Dimension gilt entsprechend für die Invarianten des Strukturtenors. Weiterhin gilt  $\text{Rang } \mathbf{J} = m$ . Man beachte dabei jedoch, dass der Strukturtenor lediglich eine spezifische Möglichkeit zur Bestimmung der intrinsischen Dimension darstellt.

### 2.3 Informationsgehalt und intrinsische Dimension

Von Bedeutung im Hinblick auf eine effiziente Bildkodierung ist, dass die Form der Bildflächen von den i2D Bereichen weitgehend bestimmt wird und sich somit Bilder allein aus den i2D Merkmalen auch gut rekonstruieren lassen [29, 4]. Weiterhin gibt es einen Bezug zu der Statistik der natürlichen visuellen Umwelt: i0D Merkmale kommen in natürlichen Bildern statistisch am häufigsten, i1D Merkmale weitaus seltener und i2D Merkmale am seltensten vor. Somit erlaubt die Auswertung der intrinsischen Dimension eine effizientere Kodierung [12].

Zum Thema Rekonstruktion hatten wir bereits in [29] gute Ergebnisse erzielt. Durch die Zusammenarbeit mit Cicero Mota erhielten wir später einen mathematischen Beweis dafür, dass Flächen und Hyperflächen, und somit Bilder und Bildfolgen, von den Bereichen intrinsischer Dimension  $m \geq 2$  vollständig bestimmt sind. Der Beweis ermöglichte auch neuartige Rekonstruktions-Algorithmen, die es z.B. erlaubten, das Bild eines Quadrates allein aus dessen Krümmungen zu rekonstruieren - siehe *Abbildung 1*.



**Abbildung 1:** Das Originalbild links kann aus den in der Mitte gezeigten i2D Merkmalen rekonstruiert werden. Das Ergebnis der Rekonstruktion ist rechts dargestellt und zeigt eine große Ähnlichkeit mit dem Original - aus [4].

## 2.4 Bewegung als intrinsische Geometrie von Bildfolgen

In [2] wurde gezeigt, wie die Krümmungseigenschaften von Bildfolgen mit dem Problem der Bewegungsschätzung zusammenhängen und wie sie zusätzlich die Detektion von Bewegung ermöglichen.

Nehmen wir nun an, dass die Bildfolge durch Translation entsteht, d.h. durch eine gleichförmige Bewegung mit Geschwindigkeit  $\mathbf{v} = (v_x, v_y)$ .  $f$  erfüllt dann folgende Gleichung [31]:

$$f(x, y, t) = f(x + dx, y + dy, t + dt). \quad (6)$$

Daraus ergibt sich

$$\alpha(\mathbf{v})f = \vec{\nabla} f \mathbf{V} = 0 \quad (7)$$

mit  $\alpha(\mathbf{v}) = v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} + \frac{\partial}{\partial t}$ , der Ableitung in Richtung  $\mathbf{V} = (v_x, v_y, 1)^T$ . Die Lösung der Gleichung (7) ist dann:

$$f(x, y, t) = f(x - v_x t, y - v_y t). \quad (8)$$

Wird nun die Gleichung (8) in die Ausdrücke für die Komponenten (2) eingesetzt, so ergeben sich folgende Beziehungen zwischen den Komponenten von  $\mathbf{R}$ , s. [1]:

$$\begin{aligned} \mathbf{v} &= \mathbf{v}_1 = (R_{3221}, -R_{3121})/R_{2121} \\ \mathbf{v} &= \mathbf{v}_2 = (R_{3231}, -R_{3131})/R_{3121} \\ \mathbf{v} &= \mathbf{v}_3 = (R_{3232}, -R_{3231})/R_{3221}. \end{aligned} \quad (9)$$

Die Indizes von  $\mathbf{v}$  sollen lediglich zeigen, dass sich mehrere Ausdrücke für den Geschwindigkeitsvektor ergeben.

Die Gleichungen in (9) erlauben verschiedene Bewegungsschätzungen, die nur im Falle reiner Translation gleich sind. Der Mittelwert dieser verschiedenen Schätzungen ergibt eine robustere Schätzung der Translation; Differenzen zwischen unterschiedlichen Schätzungen sind Indikatoren für das Zutreffen der Gleichung (8). Festzustellen, ob das Bewegungsmodell (8) zutrifft, ist oft der schwierigere Teil der Bewegungsschätzung und hängt eng mit dem Problem der intrinsischen Dimension zusammen. Liegt nämlich eine Translation vor, so ist die Fourier-Transformierte des Signals auf eine Ebene beschränkt, was auch für

i2D Signale gilt. Eine Translation ist somit dadurch charakterisiert, dass der Riemannsche Tensor von Null verschieden, die Gaußsche Krümmung jedoch gleich Null ist, aber auch dadurch, dass die unterschiedlichen Bewegungsvektoren nach Gleichung (9) definiert und gleich sind.

## 2.5 Bewegung und die Minoren des Strukturensors

In Analogie zu den oben zusammengefassten Ergebnissen aus [2] wurden in [3] folgende Beziehungen zwischen den Minoren des Strukturensors und den Bewegungsvektoren gefunden:

$$\begin{aligned} \mathbf{v} &= \mathbf{v}_1 = (M_{31}, -M_{21})/M_{11} \\ \mathbf{v} &= \mathbf{v}_2 = (M_{23}, -M_{22})/M_{12} \\ \mathbf{v} &= \mathbf{v}_3 = (M_{33}, -M_{23})/M_{13}. \end{aligned} \tag{10}$$

Um diese Ergebnisse besser einordnen zu können, erinnern wir daran, dass die Bewegungsschätzung oft als Optimierungsproblem betrachtet und als solches dann auf ein Eigenwertproblem abgebildet wird, so dass der Bewegungsvektor als der minimale Eigenwert des Strukturensors bestimmt wird - s. [32, 33]. Der Bezug zu den obigen Beziehungen in (10) ist dadurch gegeben, dass der Eigenvektor zum minimalen Eigenwert grundsätzlich über die Minoren berechnet werden kann [34] und die beiden Methoden somit formal äquivalent sind. Was die praktische Anwendung angeht, wurde jedoch in [3] gezeigt, dass die Minoren-Methode schneller und genauer ist.

## 2.6 Überlagerte Bewegungen

### 2.6.1 Mehrfache transparente Bewegungen

Wir nehmen nun an, dass sich  $n$  Muster  $f_i$  additiv überlagern und dass sich diese Muster mit jeweils  $\mathbf{v}_i = (v_{ix}, v_{iy})$ , bewegen, d.h.

$$f(\mathbf{x}, t) = f_1(\mathbf{x} - \mathbf{v}_1 t) + \dots + f_n(\mathbf{x} - \mathbf{v}_n t). \tag{11}$$

mit  $\mathbf{x} = (x, y)$ .

In diesem Falle wird aus Gleichung (7)

$$\alpha(\mathbf{v}_1) \dots \alpha(\mathbf{v}_n) f = 0. \tag{12}$$

Es soll nun gezeigt werden, wie aus obiger Gleichung die Bewegungsvektoren berechnet werden können. Wir schreiben dazu zunächst Gl. (12) als

$$\sum_I c_I f_I = 0. \quad (13)$$

$I = (i_1, i_2, \dots, i_n)$  sind geordnete Sequenzen mit Elementen  $i_j \in (x, y, t)$  und  $f_I$  sind die partiellen Ableitungen von  $f$  anhand der Elemente in  $I$ . Damit definieren wir die gemischten Bewegungsparameter  $c_I$  und schreiben Gl. (13) als

$$\mathbf{L}\mathbf{V} = \mathbf{0} \quad (14)$$

mit  $\mathbf{L} = (f_I)$  und  $\mathbf{V} = (c_I)^T$ . Nun multiplizieren wir Gl. (14) mit  $\mathbf{L}^T$  und erhalten damit ein Gleichungssystem. Wir integrieren dann diese Gleichungen in einer Nachbarschaft durch eine Faltung mit einem Faltungskern  $h(x)$ ,

$$\int \mathbf{L}(x)^T \mathbf{L}(x) \mathbf{V}(x) h(x) dx = \mathbf{0} \quad (15)$$

um die Konditionierung zu verbessern. Weiterhin nehmen wir an, dass die Bewegungsvektoren in dieser Nachbarschaft konstant sind und können dadurch  $\mathbf{V}$  vor das Integral ziehen. Dadurch erhalten wir

$$\mathbf{J}_n \mathbf{V} = \mathbf{0} \quad (16)$$

mit dem somit definierten *generalisierten Strukturtenor für  $n$  Bewegungen*:

$$\mathbf{J}_n = \int \mathbf{L}(x)^T \mathbf{L}(x) h(x) dx. \quad (17)$$

Man beachte die Ähnlichkeit zwischen Gl. (16) und Gl. (7). Damit können die gemischten Bewegungsparameter genauso anhand von  $\mathbf{J}_n$  berechnet werden wie die einzelnen Bewegungen anhand von  $\mathbf{J}$ . Die Methode der Minoren ergibt dann  $l = \text{ord}(\mathbf{J}_n)$  verschiedene Ausdrücke für die gemischten Bewegungsparameter

$$\mathbf{V}_i \propto (M_{il}, -M_{il-1}, \dots, (-1)^l M_{i1}). \quad (18)$$

$M_{ij}, i = 1, \dots, l$  sind die Minoren von  $\mathbf{J}_n$  [34].

**Trennung der gemischten Bewegungsparameter** Nun müssen noch die Bewegungsvektoren  $\mathbf{v}_1, \dots, \mathbf{v}_n$  anhand der gemischten Bewegungsparameter  $c_I$  in  $\mathbf{V}$  bestimmt werden. Wir interpretieren dazu die  $\mathbf{v}_i$  als komplexe Zahlen, d.h.  $\mathbf{v}_i = v_{ix} + jv_{iy}$  mit  $j^2 = -1$ . Damit konnten wir in [35] zeigen, dass die gesuchten Bewegungsvektoren die Wurzeln des folgenden komplexen Polynoms  $Q_n(z)$  sind:

$$Q_n(z) = z^n - A_{n-1}z^{n-1} + \dots + (-1)^n A_0 \quad (19)$$

und dass die Koeffizienten des Polynoms durch  $c_I$  bestimmt sind.

**Konfidenzmaße** Wie bereits erwähnt, ist das Problem der Konfidenz für ein bestimmtes Bewegungsmodell oft schwieriger als die Bestimmung der Bewegungsparameter. Wie kann nun bestimmt werden, ob überlagerte Bewegungen vorliegen?

Die Konfidenz für nur eine Bewegung ist hoch, wenn ein Eigenwert des Strukturtenors klein und die anderen signifikant sind. Im Falle von  $n$  transparenten Bewegungen ist analog dazu die Konfidenz dann hoch, wenn  $\text{Rang}(\mathbf{J}_n) = l - 1$ . Wie kann nun dieser Zusammenhang ohne die explizite Bestimmung der Eigenwerte genutzt werden? Die in Gl. (5) definierten Invarianten  $K, S, H$  von  $\mathbf{J}$  können analog auch für  $\mathbf{J}_n$  definiert werden [34]. Mit diesen generalisierten Invarianten bekommen wir das Konfidenzkriterium  $K = 0$  UND  $S \neq 0$ . Um  $K$  und  $S$  quantitativ vergleichen zu können, nutzen wir die Beziehung  $K^{1/l} \leq S^{1/l-1} \leq H$ . Damit erhalten wir das Kriterium  $K^{1/l} \ll S^{1/l-1}$  oder äquivalent dazu  $K^{1/l} < \epsilon S^{1/l-1}$ .

### 2.6.2 Klassifikation von Bewegungsmustern

Der durch Gl. (17) definierte, generalisierte Strukturtenor erlaubt eine Klassifikation von Bewegungsmustern, die in der Tabelle 1 zusammengefasst ist. Der Zusammenhang zwischen dem Rang von  $\mathbf{J}_1$  und der intrinsischen Dimension wurde bereits in Abschnitt 2.2.2 dargestellt.

Wir hatten bereits gezeigt, dass die intrinsische Dimension sowohl damit zusammenhängt, dass die Energie der Signale auf Unterräume (Geraden, Ebenen, Hyperebenen) beschränkt ist als auch, dass eine solche Beschränkung mit Redundanzen verknüpft ist.

Hier untersuchen wir nun die Fälle, in denen das Signal auf mehrere Geraden oder Ebenen beschränkt ist, was gewissermaßen einer fraktalen intrinsischen Dimension ent-

spricht. Die in Tabelle 1 dargestellten Muster sind Überlagerungen von örtlichen i1D oder i2D Mustern, die sich transparent überlagert bewegen. Man beachte, dass durch die Strukturtenoren  $\mathbf{J}_2$  und  $\mathbf{J}_3$  die von der intrinsischen Dimension  $m$  gegebene Klassifikation weiter verfeinert wird.

| Bewegungsmuster | Projektive Repräsentation | $m = \text{Rang } \mathbf{J}_1$ | $\text{Rang } \mathbf{J}_2$ | $\text{Rang } \mathbf{J}_3$ |
|-----------------|---------------------------|---------------------------------|-----------------------------|-----------------------------|
| ○               | leere Menge               | 0                               | 0                           | 0                           |
|                 | ein Punkt                 | 1                               | 1                           | 1                           |
| +               | 2 Punkte                  | 2                               | 2                           | 2                           |
| +   +           | 3 Punkte                  | 3                               | 3                           | 3                           |
| •               | eine Linie                | 2                               | 3                           | 4                           |
| • +             | eine Linie + ein Punkt    | 3                               | 4                           | 5                           |
| • +   +         | eine Linie + 2 Punkte     | 3                               | 5                           | 6                           |
| • + •           | 2 Linien                  | 3                               | 5                           | 7                           |
| • + • +         | 2 Linien + ein Punkt      | 3                               | 6                           | 8                           |
| • + • + •       | 3 Linien                  | 3                               | 6                           | 9                           |
| andere          | andere                    | 3                               | 6                           | 10                          |

**Tabelle 1:** Unterschiedliche Bewegungsmuster (erste Spalte), deren projektive Repräsentation (zweite Spalte) und der Rang der generalisierten Strukturtenoren in den folgenden Spalten. Die Tabelle fasst die Ergebnisse bezüglich des Zusammenhangs zwischen Bewegungsmustern und dem Rang der Strukturtenoren zusammen. • symbolisiert bewegte i2D Muster und | bewegte i1D Muster; + bezeichnet deren additive Überlagerung.

### 2.6.3 Mehrfache verdeckende Bewegungen

Verdeckende Überlagerungen kommen in natürlichen Videosequenzen häufiger vor als transparente Überlagerungen. Gleichzeitig ist das Problem der Bestimmung verdeckender Bewegungen schwieriger. Wir hatten in [5] bereits einen Ansatz vorgestellt und die

endgültige Gleichung für verdeckende Bewegungen schließlich in [6] gefunden.

Verdeckende Bewegungen mit Bewegungsvektoren  $\mathbf{u}$  und  $\mathbf{v}$  werden durch folgende Gleichung beschrieben:

$$f(\mathbf{x}, t) = \chi(\mathbf{x} - t\mathbf{u})g_1(\mathbf{x} - t\mathbf{u}) + (1 - \chi(\mathbf{x} - t\mathbf{u}))g_2(\mathbf{x} - t\mathbf{v}). \quad (20)$$

Wird nun darauf der Operator  $\alpha(\mathbf{u})\alpha(\mathbf{v})$  angewandt, und werden die Ableitungen an den Diskontinuitäten der verdeckenden, binären Maske  $\chi$  im Sinne der Distributionen-Theorie ausgewertet, erhält man schließlich folgende Gleichung:

$$\alpha(\mathbf{u})\alpha(\mathbf{v})f = (\mathbf{v} - \mathbf{u}) \cdot \mathbf{N}\delta_B(\mathbf{x} - t\mathbf{u})\alpha(\mathbf{u})g_2(\mathbf{x} - t\mathbf{v}). \quad (21)$$

$B$  bezeichnet die Grenze zwischen den beiden Mustern  $g_1$  und  $g_2$ ,  $\delta_B$  ist eine Dirac-Delta Distribution auf  $B$ .  $\mathbf{N}$  ist der Normalenvektor zu  $B$ .

Aus Gleichung (21) wird zunächst klar, dass die Modelle für einfache (7) sowie transparent überlagerte Bewegungen (12) an den Verdeckungsgrenzen falsch sind. Weiterhin wird der Fehler quantifiziert: er steigt mit der Differenz der beiden Bewegungsvektoren relativ zur Normalen  $N$ , sowie mit dem Kontrast des Hintergrundes (den Ableitungen von  $g_2$ ). Abgesehen von  $B$  jedoch stimmt das Modell für zwei transparent überlagerte Bewegungen, weil die Dirac-Delta Distribution den Term  $(\mathbf{v} - \mathbf{u}) \cdot \mathbf{N}\delta_B(\mathbf{x} - t\mathbf{u})\alpha(\mathbf{u})g_2(\mathbf{x} - t\mathbf{v})$  außerhalb von  $B$  gleich Null setzt.

## 3 Anwendungen

### 3.1 Bestimmung einfacher Bewegungen

#### 3.1.1 Krümmungstensor

In [2] wurden Bildfolgen als Hyperflächen betrachtet und anhand des Riemannschen Krümmungstensors dieser Hyperflächen wurden neuartige Methoden zur Bewegungsschätzung gefunden. Insbesondere wurde gezeigt, wie mithilfe der Krümmungseigenschaften und der intrinsischen Dimension der Bildfolge das Vorliegen einer Translation und somit die Konfidenz der Bewegungsschätzung beurteilt werden kann. In Anwendungsbeispielen wurde schließlich anhand synthetischer und natürlicher Bildfolgen veranschaulicht, wie

falsche Bewegungsvektoren vermieden werden können, die typischerweise durch Verdeckungen oder Rauschen entstehen.

*Abbildung 2* (aus [2] entnommen) zeigt Ergebnisse für eine synthetische Sequenz - siehe Bildunterschrift. *Abbildung 3* zeigt Ergebnisse für eine Verkehrsszene (Taxi-Sequenz), in der sich ein helles und zwei dunkle Autos sowie ein Fußgänger bewegen. Nach dem Vorbild des menschlichen Gehirns (s. Modelle in [1]) wurden die Bewegungsvektoren lediglich an wenigen Stellen mit hoher Konfidenz ausgewertet. Diese Vektoren wurden dann über den Ort integriert und für die zusammenhängenden Bereiche zu einem einzelnen Vektor zusammengefaßt. Das Ergebnis liefert eine gute, symbolisch wirkende Beschreibung der Bewegungen in der Szene.

### 3.1.2 Strukturtensor

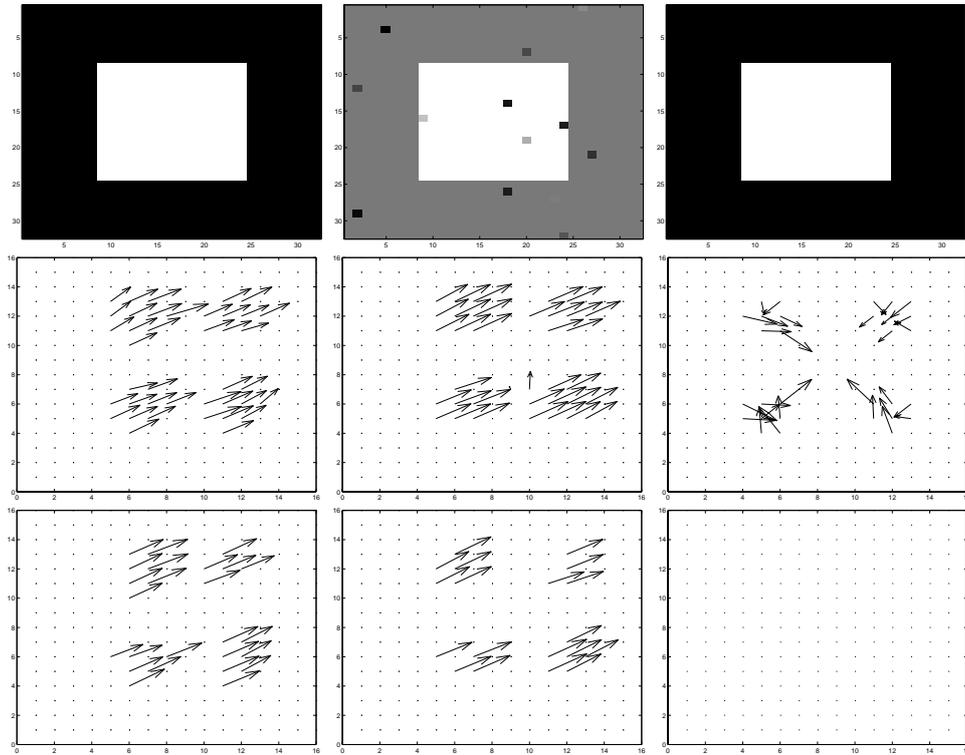
In [33] und weiteren Arbeiten der gleichen Gruppe um Bernd Jähne wurde die Nützlichkeit des Strukturensors zur Bewegungsschätzung eindringlich demonstriert. In [3] wurde dann von mir gezeigt, dass durch den Bezug zur Differentialgeometrie und der intrinsischen Dimension die Bewegungsschätzung mithilfe des Strukturensors noch weiter verbessert werden konnte.

In *Abbildung 4* wird die Robustheit unterschiedlicher Verfahren mithilfe einer synthetischen Sequenz und zusätzlichem Rauschen untersucht. In *Abbildung 5* werden Ergebnisse für eine Verkehrsszene gezeigt. Weitere hier nicht reproduzierte Ergebnisse aus [3] zeigen, dass das Minoren-Verfahren auch schneller und genauer ist als die mit untersuchten Standardverfahren zur Bewegungsschätzung.

## 3.2 Mehrfache überlagerte Bewegungen

### 3.2.1 Transparente Überlagerungen

Ergebnisse zur Bestimmung transparent überlagelter Bewegungen wurden in [34, 36, 37, 38, 39] publiziert. In *Abbildung 6* wird das erste Ergebnis aus [34] dargestellt und erläutert. In [36] wurde zusätzlich auch gezeigt, wie sich die überlagerten Muster mithilfe der geschätzten Bewegungsparameter trennen lassen und in [37], wie die Schätzung der Bewegung durch zusätzliche Regularisierung robuster gemacht werden kann. In [38] wurden die

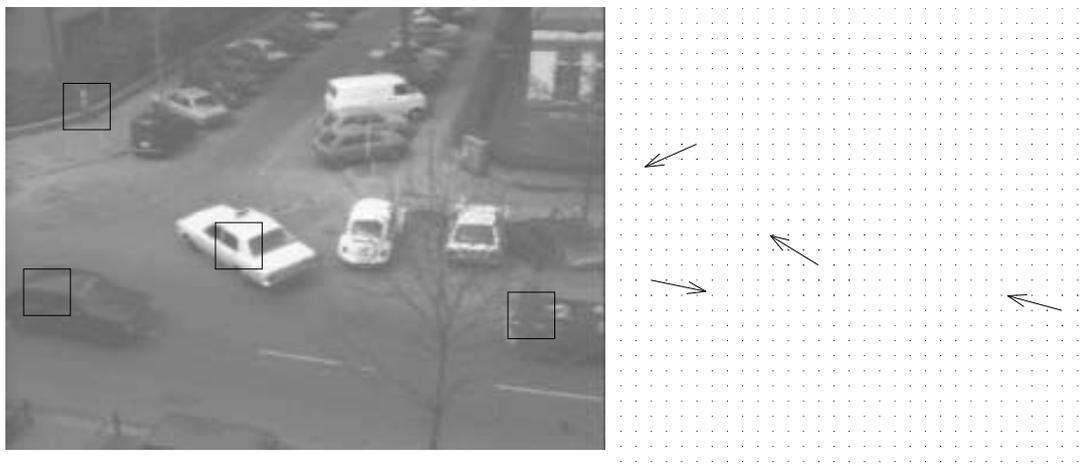


**Abbildung 2:** Selektion von Bewegungsvektoren dargestellt für unterschiedliche Testfilme: links bewegt sich ein Rechteck nach links oben, in der Mitte kommt Rauschen dazu, rechts taucht das Rechteck auf. Die mittlere Reihe zeigt den Mittelwert der vier Bewegungsvektoren, die durch die Gleichungen (9) definiert sind. In der untersten Reihe wurden Vektoren ausgeschlossen, wenn die Varianz der vier unterschiedlichen Bewegungsvektoren groß war. Bemerkenswert ist, dass dadurch die falschen Vektoren beseitigt werden.

Methoden durch ein sogenanntes block-matching Verfahren und in [39] durch statistische Modelle erweitert. In [40] wurde die Theorie der überlagerten Bewegungen schließlich auf das Problem der Bestimmung mehrfacher Orientierungen in Bildern angewandt.

### 3.2.2 Verdeckende Überlagerungen

Ergebnisse zu verdeckenden Bewegungen wurden zuerst in [5] publiziert und werden hier in *Abbildung 7* reproduziert und erläutert. In *Abbildung 8* werden Ergebnisse aus [6] dargestellt. Im Unterschied zu den zuerst genannten Ergebnissen handelt es sich hier um natürliche Texturen. Weiterhin wurden auch Ergebnisse für der Fall eines stationären



**Abbildung 3:** Einzelnes Bild aus der Taxi-Sequenz (links) und symbolisch wirkendes Ergebnis der Berechnung von Bewegungsvektoren (rechts).

Hintergrundes berechnet. Alle Ergebnisse für verdeckende Überlagerungen wurden mit folgendem hierarchischen Algorithmus berechnet, der die in Abschnitt 2.6.1 beschriebenen Konfidenzmaße nutzt. Der hierarchische Algorithmus bestimmt zunächst die Konfidenz für das einfache Modell nur einer Bewegung und bestimmt diese eine Bewegung, falls die Konfidenz gut ist. Andernfalls werden abhängig von einer entsprechenden Konfidenz zwei transparente Bewegungen bestimmt. Die dabei nicht behandelten Pixel werden im nächsten Iterationsschritt behandelt und zwar mit einem Faltungskern, der schrittweise größer wird, dabei aber die Pixel mit schlechter Konfidenz nicht integriert. Die Ergebnisse werden in den *Abbildungen 7* und *8* gezeigt.

### 3.3 Modelle visueller Neurone

Visuelle Neurone wurden von meinen Münchner Kollegen und mir in früheren Arbeiten modelliert, z.B. in [9, 12, 22, 21]. Es handelte sich dabei um sogenannte *endstopped* Neurone, die im primären und sekundären visuellen Kortex vorkommen.

Bei den in [1] vorgestellten Modellen hingegen handelt es sich um bewegungsselektive Neurone des visuellen Areals MT, welches auf die Verarbeitung bewegter Reize spezialisiert ist. Die weit verbreitete Sichtweise bezüglich der Funktion dieser Neurone ist die, dass sie die Bewegung von Objekten kodieren. Alternativ dazu haben wir in [1] vorgeschla-

---

**Algorithm 1** Algorithmus zur hierarchischen Bewegungsschätzung

---

```
1: Berechne  $\mathbf{J}_n$ 
2: if  $K^{1/l} < \epsilon_n S^{1/l-1}$  (hohe Konfidenz) then
3:   Bestimme die gemischten Bewegungsparameter anhand von  $\mathbf{J}_n$ 
4:   if  $n = 1$  then
5:      $\mathbf{v} = (V_x, V_y)$ 
6:   else
7:     Bestimme  $\mathbf{u}, \mathbf{v}$  als die Wurzeln von  $Q_2(z)$ 
8:     Trage Pixel  $\mathbf{x}_0$  in Liste  $L$  ein
9:   for all  $\mathbf{x}_0 \notin L$  do
10:    Wiederhole Schritte 1 bis 8 mit  $h(\mathbf{x} - \mathbf{x}_0) = 0, \forall \mathbf{x} \notin L$ . Vergrößere den Faltungskern  $h$  s.d.  $M$  Pixel die nicht in  $L$  sind in die Mittelung eingehen.
```

---

gen, dass diese Neurone die intrinsische Dimension kodieren und dadurch eine effiziente, weil weniger redundante Repräsentation leisten. Um diese Hypothese zu testen, wurden Experimente betrachtet, in denen die MT Neurone mit anderen als den einfachen Bewegungsreizen gemessen wurden. Beim ersten Experiment [41] wurde die Selektivität dieser Neurone auf die Orientierung eines geblitzten Balkens gemessen. Die Daten und unsere Simulationsergebnisse werden in *Abbildung 9* dargestellt. Beim zweiten in [42] beschriebenen Experiment wurden die MT Neurone nicht nur wie üblich mit einem bewegten Lichtpunkt, sondern auch mit zwei Lichtpunkten und unterschiedlicher relativer Bewegung gemessen. Die Daten und unsere Simulationsergebnisse werden in *Abbildung 10* dargestellt. Bemerkenswert ist, dass die Simulationen sehr gut die Daten erklären und das, obwohl die Simulationsergebnisse rein analytisch berechnet wurden. Details dieser Berechnung finden sich in [1].

### 3.4 Modellierung visueller Wahrnehmung

Bereits während meiner Promotion hatte ich bestimmte Aspekte der visuellen Wahrnehmung modelliert. Es ging dabei hauptsächlich um die Texturwahrnehmung [20] und die Wahrnehmung topologischer Merkmale [43, 44].

Im Unterschied dazu geht es nun um die Modellierung der Wahrnehmung von Bewe-

gung. In [1] konnten einige Aspekte der klassischen Experimente von Wallach [45] erklärt werden. Darauf wird hier nicht weiter eingegangen, sondern auf die online Publikation [1] verwiesen, welche die für das Verständnis sehr hilfreichen Filme enthält.

Die Arbeit [8] beruht auf Ergebnissen, die teilweise bereits in einigen Konferenzbeiträgen publiziert waren [46, 47, 48] und hauptsächlich die Wahrnehmung mehrfacher, transparent überlagerter Bewegungen behandelt. Ich greife daraus lediglich ein Beispiel heraus, welches in *Abbildung 11* dargestellt und erklärt wird. Weiterhin haben wir in dieser Arbeit den *entrainment effect* experimentell nachgewiesen und eine neue Täuschung entdeckt. Der Effekt besteht darin, dass ein örtliches i2D Muster die Wahrnehmung der Bewegung eines örtlichen i1D Musters beeinflusst, wenn beide überlagert werden. Die Täuschung besteht darin, dass der obige Effekt erhalten bleibt, auch wenn das i2D Muster lediglich an den Randbereichen des i1D Musters überlagert wird. Beides hatten wir vorher anhand der Theorie vorausgesagt und konnten es somit auch gut erklären. Damit knüpfen wir an die traditionsreichen Arbeiten von Wallach [45] an.

### 3.5 Prädiktion der Blickrichtung

Das Sehen ist ein aktiver, komplexer Vorgang, bei dem nur sehr selektiv Information aus der Umwelt aufgenommen wird. Ein wichtiger Mechanismus dieser Selektion sind die Augenbewegungen. Trotz der selektiven und aktiven Aufnahme visueller Information ist die subjektive Wahrnehmung die einer vollständigen, stabilen Welt. Dieses ist gewissermaßen eine Täuschung, die durch Aufmerksamkeit, Erwartungen und Erfahrung bestimmt ist.

In der immer wichtiger werdenden visuellen Kommunikation entsteht dadurch ein Problem: man zeigt ein Bild oder einen Film und meint, damit sei die zu übermittelnde Nachricht definiert. Tatsächlich aber können vom gleichen Bild oder Film ganz unterschiedliche Nachrichten übermittelt werden, abhängig von z.B. den Erwartungen des Betrachters und vor allem auch abhängig von seinen Augenbewegungen. Deshalb haben wir vorgeschlagen, über die Messung und Beeinflussung der Augenbewegungen, diese Teil der Kommunikation werden zu lassen, sowie damit eine Schnittstelle zwischen biologischen und technischen Sehsystemen zu schaffen [49, 50]. Ersteres soll zu einer besseren Kommunikation führen, indem besser definiert werden kann, was die visuell zu übermittelnde Nachricht ist. Letzteres soll z.B. dazu führen, dass die Aufmerksamkeit eines Autofahrers auf einen von

ihm ansonsten übersehenen Fußgänger gelenkt wird, der von einer Videokamera entdeckt wurde. Dass wir als Menschen tatsächlich vieles einfach übersehen, ist durch eine beeindruckende Reihe von Experimenten gezeigt worden [51, 52, 53].

Als ersten wichtigen Schritt in diesem zukunftsorientierten Szenario haben wir Modelle entwickelt, mithilfe derer die Augenbewegungen vorausgesagt werden können [54, 7, 55]. Die Idee dabei ist, während der Darbietung eines Films einige Orte im gerade aktuellen Bild vorauszusagen, wo der Betrachter am ehesten hinschauen würde (Kandidaten, die Augenbewegungen auf sich ziehen könnten). Vor der Darbietung des nächsten Bildes aus der Bildfolge würde deren Darbietung so verändert, dass einer der vorher bestimmten Kandidaten wahrscheinlicher wird als die übrigen. Dieses kann durch Veränderung der Kontraste und Dynamik, sowie durch zusätzliche Reize geschehen.

Bereits in den ersten Arbeiten zu diesem Thema [54, 7] haben wir die Invarianten des Strukturtenors - s. Gl. (5) - genutzt, um für die Steuerung der Augenbewegungen signifikante Bereiche in Bildfolgen zu bestimmen. Neueste in *Abbildung 12* gezeigte Ergebnisse bestätigen unsere Vermutung, dass Bildbereiche mit höherer intrinsischer Dimension eher Augenbewegungen auf sich ziehen.

## 4 Diskussion

### 4.1 Die intrinsische Dimension in der Signalverarbeitung

Der Begriff *intrinsische Dimension* wurde in [9] eingeführt und später auch als kontinuierliche Eigenschaft definiert [56]. Bewährt hat sich das Konzept zunächst in der Bilddaten-Kompression, weil die intrinsische Dimension eng mit der Statistik von natürlichen Bildern zusammenhängt [12].

Die Bestimmung der intrinsischen Dimension hängt mit dem Problem der lokalen Orientierungs-Analyse zusammen, welches über die Eigenwert-Analyse des Strukturtenors auf eine spezifische Art gelöst werden konnte [57, 32]. Alternativ zu der von uns verfolgten Strategie, eine Theorie der intrinsischen Dimension über eine Synthese aus Differentialgeometrie und linearer Systemtheorie aufzubauen, erscheint die Clifford Algebra als eine mächtige und dafür geeignete Theorie [58, 59, 60]. Vielleicht gelingt in Zukunft eine Synthese der beiden Ansätze.

Von Bedeutung ist, dass Signale mit intrinsischer Dimension kleiner als zwei redundant sind [29, 61, 4]. Damit ergibt sich ein Zusammenhang zwischen Information und Geometrie. Aus der Praxis war schon länger bekannt, dass bestimmte i2D Merkmale, z.B. Ecken, für die visuelle und technische Bildverarbeitung von Bedeutung sind [62, 63, 12, 64].

Zusammenfassend kann man sagen, dass die *intrinsische Dimension* eine grundlegende Eigenschaft mehrdimensionaler Signale ist, die dennoch in der Vergangenheit nur wenig untersucht wurde.

## 4.2 Die intrinsische Dimension in der Sehforschung

Schon lange sind in der Sehforschung Ergebnisse bekannt, die mit der intrinsischen Dimension der dargebotenen Reize zusammenhängen. So antworten z.B. Ganglienzellen in der Netzhaut von Fröschen nur auf i2D Signale wie dunkle Ecken und Flecken [65, 66]. Später wurden zunehmend viele Neurone im visuellen Kortex von Säugetieren gefunden, die i2D Signale bevorzugen [67, 68, 69]. Weiterhin gibt es in der visuellen Psychophysik zahlreiche Hinweise auf die Existenz von Mechanismen, die i2D spezifisch sind [70, 71, 72, 73, 20]. Ein erster Versuch, diese Phänomene zu modellieren [74], scheiterte teilweise daran, dass eine Theorie der intrinsischen Dimension fehlte [9]. Inzwischen wurden die örtlichen Eigenschaften von i2D Neuronen mehrfach modelliert [9, 11, 75, 12, 16, 22, 76]. Die Arbeiten von Koenderink und Kollegen, z.B. [77], haben unseren differentialgeometrischen Ansatz zu diesem Thema beeinflusst.

In [1] werden psychophysische und neurophysiologische Ergebnisse erklärt, die sich auf dynamische Reize beziehen. Dabei passten die von Hans Wallach gefundene Ergebnisse, vgl. [45], sowie Daten zu Neuronen aus dem visuellen Hirnareal MT sehr gut zu unserer Theorie. Damit konnten wir die Rolle bewegungselektiver Neurone (diese spielen in der Hirnforschung eine wichtige Rolle) neu interpretieren und bestimmte Datensätze erklären [42, 78] - s. auch Abschnitt 3.3. Es erscheint mir in diesem Kontext sinnvoll, kurz auf die Gutachten zu [1] einzugehen. Einer der Gutachter kritisierte die Interpretation von Bildfolgen als Hyperflächen. Kritikpunkt war, dass die Dimensionen Ort, Zeit und Helligkeit nicht gleicher Natur sind. Dieses ist ein berechtigter Einwand, der jedoch den Nutzen der geometrischen Interpretation nur bedingt einschränkt, weil es im Kontext der intrinsischen Dimension hauptsächlich um die Vorzeichen der Krümmung geht. Vielleicht

gelingt in Zukunft auch eine Verbindung der Koordinaten Ort und Zeit, ähnlich der in der Relativitätstheorie gefundenen Ankopplung über die imaginäre Zeitachse. Die Kritik wurde von mir erwidert und das Manuskript in der ursprünglichen Form gedruckt. Daraufhin publizierte der (namhafte) Gutachter einen Kommentar in der gleichen Zeitschrift [79], der das Interesse am Thema weiter verstärkt hat. Das zweite Gutachten war sehr positiv.

Unsere Theorie der mehrfachen Bewegungen [8] kann die Wahrnehmung mehrfacher Bewegungen (deren Erkennung als Funktion der Anzahl von Bewegungen und des Winkels zwischen den Bewegungsrichtungen) erklären. Bisherige Ergebnisse auf diesem Gebiet waren eher heuristischer Natur [80, 81].

Wie in Abschnitt 3.5 gezeigt werden konnte, besteht ein Zusammenhang zwischen der intrinsischen Dimension einer Bildfolge und den auf dieser Bildfolge ausgeführten Augenbewegungen. Bisherige Modelle zur Voraussage und Interpretation von Augenbewegungen beschränkten sich weitgehend auf statische Bilder [82, 83, 84].

### 4.3 Bewegungsschätzung

Die Bewegungsschätzung hat viele Anwendungen in der Bildverarbeitung und es gibt entsprechend viele Verfahren - gute Übersichten dazu finden sich in [85, 33]. Das Problem bleibt aber unterbestimmt [86] und kann nur unter weiteren Annahmen und Modellen gelöst werden [87]. Deshalb ist es wichtig, Konfidenzmaße für die Modelle berechnen zu können. In [2, 3] konnten derartige Konfidenzmaße als Krümmungsmaße definiert, sowie neue Ausdrücke für die Bewegungsvektoren gefunden werden. Es konnte weiterhin gezeigt werden, dass die sich daraus ergebenden Verfahren einigen Standardverfahren [88, 89, 33] überlegen sind.

Transparent und verdeckend überlagerte Bewegungen kommen in natürlichen Bildfolgen sowie in einigen speziellen Anwendungen, z.B. in der medizinischen Bildgebung, häufig vor. Eine Übersicht zum Thema überlagerte Bewegungen findet sich in [90]. Eine erste Lösung für zwei Bewegungen ergab sich mit [91] und Analysen des Problems im Fourier Raum finden sich in [92, 93], sowie weitere Lösungsansätze in [94, 95, 96, 97]. Unsere Theorie der überlagerten Bewegungen [34, 46, 47, 5, 36, 48, 6, 38, 37, 39, 8] bietet jedoch die erste analytische Lösung für bis zu vier überlagerte Bewegungen und die Möglichkeit einer numerischen Lösung für beliebig viele Bewegungen. Weiterhin konnten

wir mit Gl. (21) erstmalig die Bewegung an verdeckenden Grenzen beschreiben. Diese Gleichung ist bemerkenswert und wurde bisher nicht gefunden, obwohl es sich hier um ein sehr altes und häufig auftretendes Problem der Bildverarbeitung handelt.

## 4.4 Eigene, interdisziplinäre Forschungsprojekte

Abschließend möchte ich einige aktuelle Projekten kurz vorstellen, die auf der Grundlage des von mir langfristig verfolgten interdisziplinären Ansatzes entstanden sind.

### 4.4.1 Komplexe Bewegungen

Ziel des LOCOMOTOR (Nonlinear analysis of multi-dimensional signals: LOcal adaptive estimation of COmplex MOTion and ORientation patterns) Projektes ist es, die Grundlagen für neue Verfahren zur Detektion und exakten Quantifizierung von Bewegung, Orientierung und Symmetrie in Bildfolgen zu schaffen und damit die technische und wissenschaftliche Anwendbarkeit von Techniken der Bildfolgenanalyse zu erhöhen. Mein Schwerpunkt liegt auf der Schätzung transparenter und verdeckter Bewegungen, der Untersuchung des Informationsgehaltes von unterschiedlichen dynamischen Merkmalen, sowie der Modellierung des Sehens.

Dieses Vorhaben wird von der Deutschen Forschungsgemeinschaft gefördert, und zwar im Rahmen des Schwerpunktprogramms 1114: *Mathematische Methoden der Zeitreihenanalyse und digitalen Bildverarbeitung*.

### 4.4.2 Lenkung der Aufmerksamkeit

Die Rolle der Aufmerksamkeit und die aktive Komponente des Sehens werden durch die derzeitigen Kommunikationssysteme nicht berücksichtigt. Die von einem Bild übermittelte Nachricht hängt in hohem Maße davon ab, wie durch Augenbewegungen die Information im Bild ausgelesen wird. Um diese aktive Rolle des Sehens zu berücksichtigen, entwickeln wir (i) interaktive Displays, die Bildfolgen abhängig von der Blickrichtung darstellen und damit das Blickmuster des Beobachters gezielt verändern und (ii) eine mobile Sehhilfe mithilfe derer die Aufmerksamkeit auf bestimmte Merkmale und Ereignisse gelenkt werden kann. Dass die Lenkung der Aufmerksamkeit prinzipiell möglich ist, konnten wir bereits nachweisen [98, 99].

Unser Projekt *Theoretische und technologische Grundlagen einer neuen Kommunikations und Interaktionstechnik auf Basis des aktiven Sehens* ist Teil des vom Bundesministerium für Bildung und Forschung geförderten Projektvorhabens *Neue Verfahren der Informationsverarbeitung auf der Basis neurokognitiver Modellierung*.

#### **4.4.3 KFZ-Assistenz-Systeme**

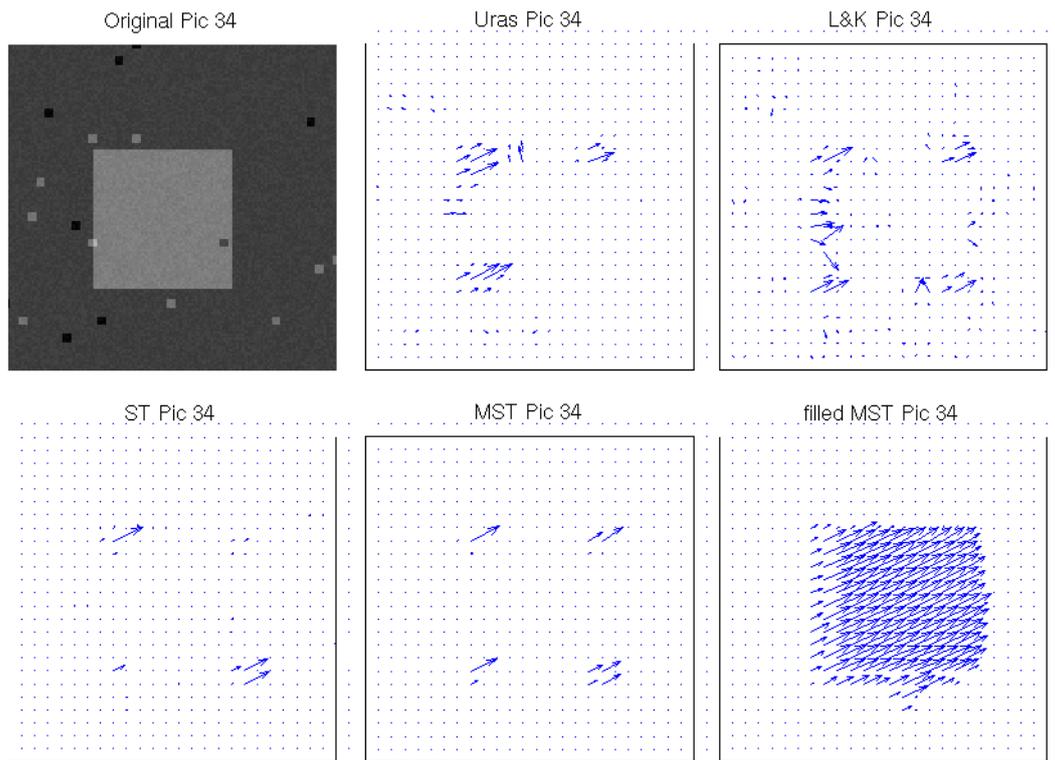
Assistenzsysteme im Fahrzeug werden derzeit durch die wachsende Komplexität des Cockpits, aber auch durch wachsende Anforderungen an die Sicherheit, von den Automobilherstellern verstärkt entwickelt und evaluiert. Ich bin an der Entwicklung von Systemen beteiligt, welche die Müdigkeit und Aufmerksamkeit des Fahrers messen (WakeUp), sowie die Auslösung des Airbags in Abhängigkeit von der Sitzbelegung steuern sollen (OoP: Out of Position).

Diese Projekte werden von einer mittelständischen Firma gefördert.

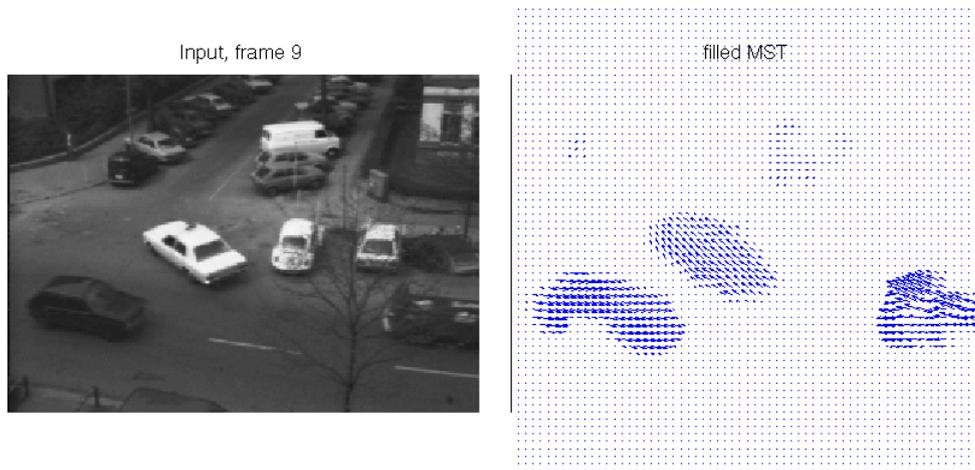
#### **4.4.4 Computation by gaze interaction (COGAIN)**

COGAIN soll die technologischen Möglichkeiten der Interaktion von Benutzern mit Rechnern und weiteren Geräten einer breiteren Masse und vor allem auch behinderten Menschen zugänglich machen.

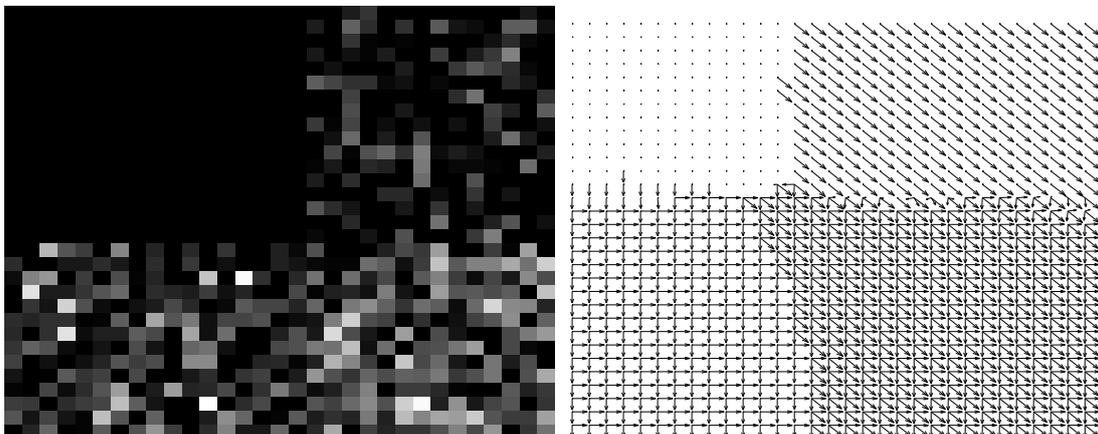
Hierbei handelt es sich um ein *Network of Excellence*, welches aus Mitteln der Europäischen Union ab September 2004 gefördert werden wird.



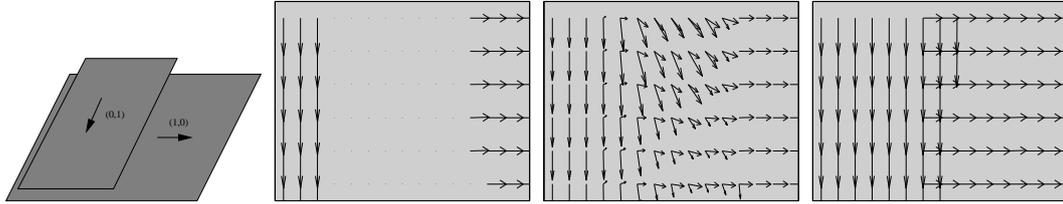
**Abbildung 4:** Vergleich unterschiedlicher Verfahren zur Bewegungsschätzung. Oben links wird ein Einzelbild aus einer Sequenz gezeigt, in der sich ein Rechteck nach oben rechts bewegt. Zusätzlich erscheinen und verschwinden an zufällig gewählten Orten kleine Punkte mit zufälligen Helligkeitswerten. Die Bewegungsvektoren oben wurden mit zwei Standardverfahren berechnet. Die Ergebnisse unten wurden durch die Eigenwertanalyse des Strukturtenors (links), mithilfe der Minoren des Strukturtenors (mitte) und schließlich mit einem zusätzlichen Füllmechanismus gerechnet. Die beiden letzten Verfahren wurden in [3] erstmalig vorgestellt.



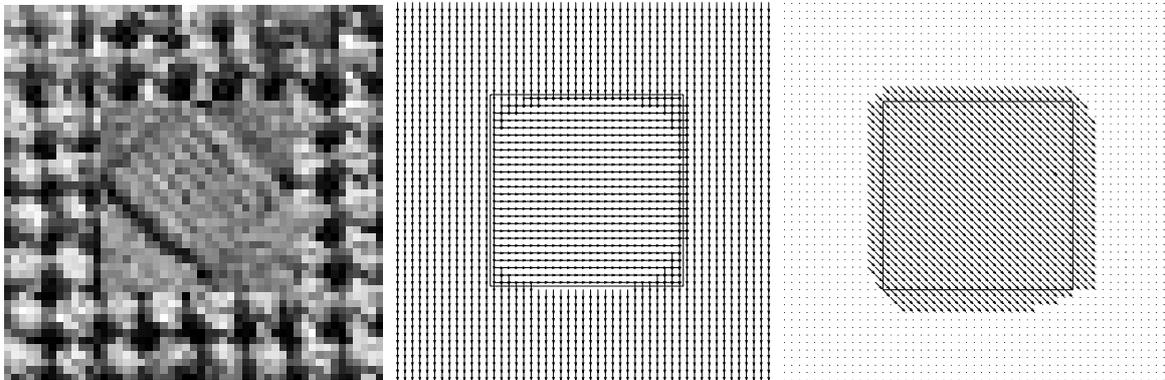
**Abbildung 5:** Links wird wie in Abbildung 3 ein Bild der Taxi-Sequenz dargestellt. Die Ergebnisse rechts zeigen ein dichtes Flussfeld mit nur wenigen Ausreißern.



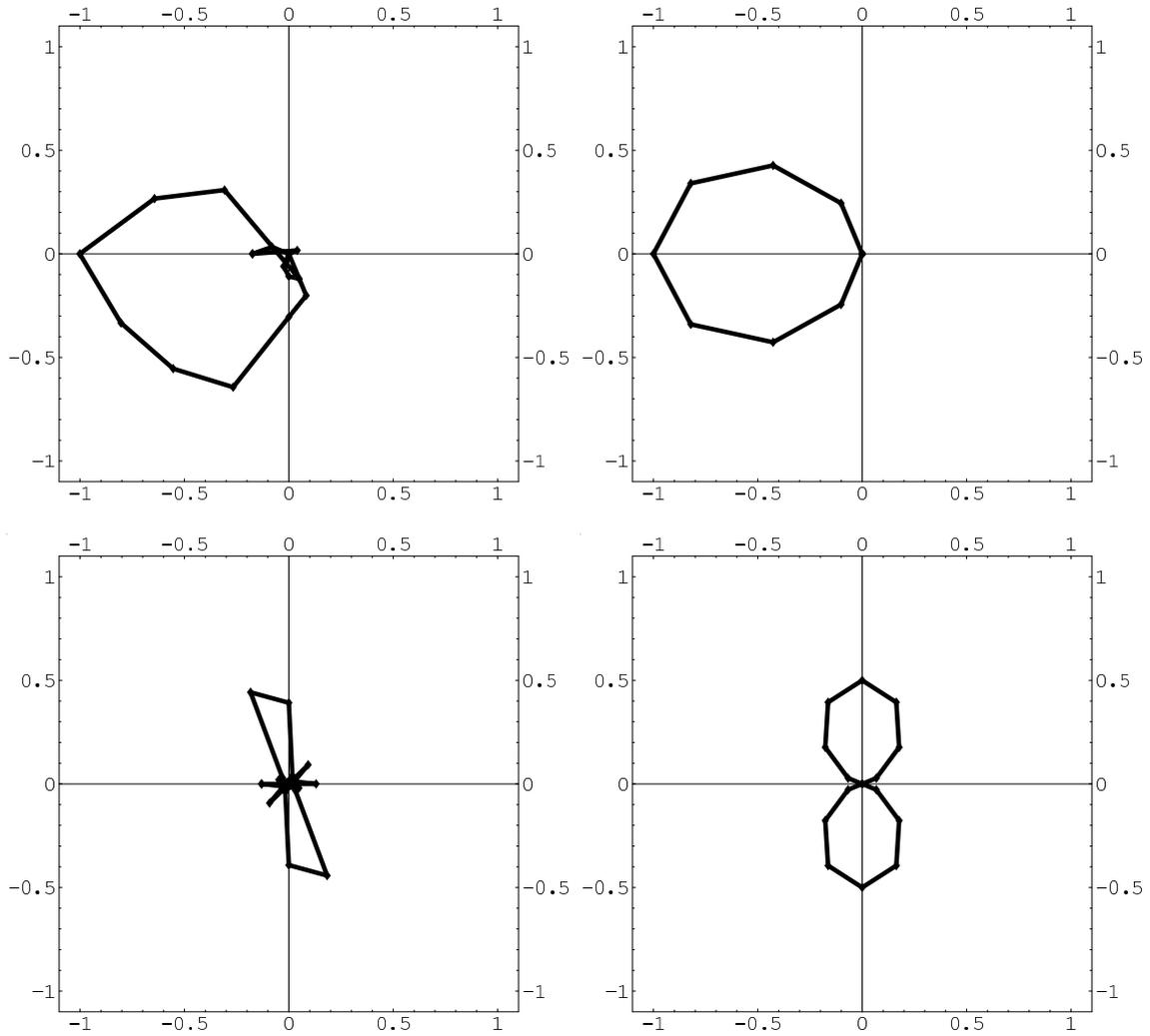
**Abbildung 6:** Links wird ein Bild einer Sequenz gezeigt, in der sich eine unterschiedliche Anzahl von transparenten Schichten bewegt und zwar im ersten Quadranten keine, im zweiten eine, im dritten zwei und im vierten drei. Die rechts dargestellten Ergebnisse zeigen, dass die Parameter dieser überlagerten Bewegungen sehr gut berechnet werden können.



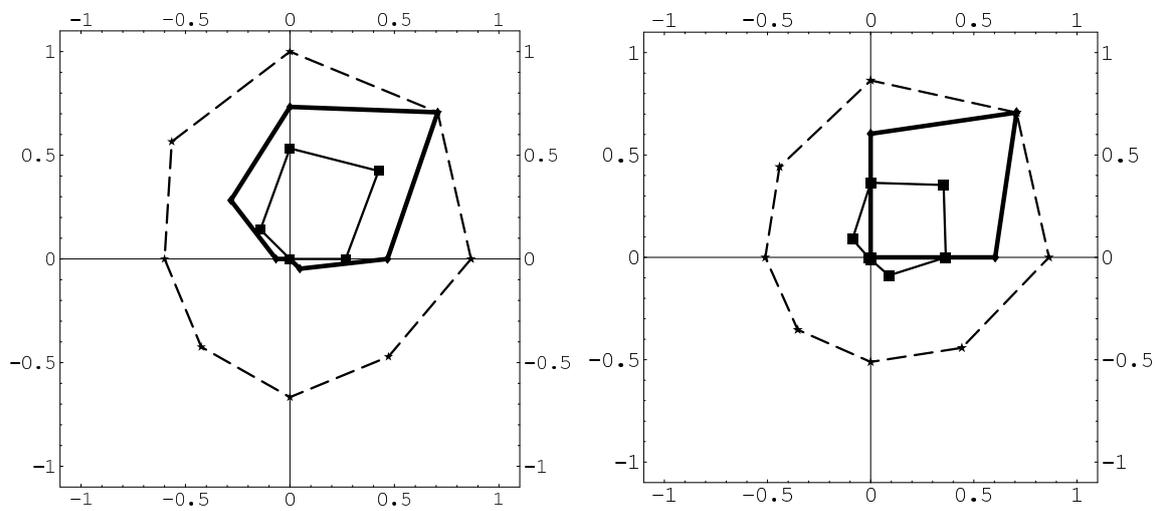
**Abbildung 7:** Ergebnisse der Bestimmung mehrfacher Bewegungen von Schichten, die sich verdecken. Links wird schematisch die Anordnung und Bewegung der Schichten dargestellt und rechts davon ein Zwischenergebnis des hierarchischen Algorithmus. Das nächste Ergebnis bekommt man, wenn die Konfidenzmaße ignoriert werden - entsprechend falsch sind die Bewegungsvektoren. Das endgültige Ergebnis rechts entstand durch die Lösung der Gleichung für mehrfache Bewegungen in einer Umgebung, welche die verdeckende Grenze ausspart - siehe hierarchischen Algorithmus.



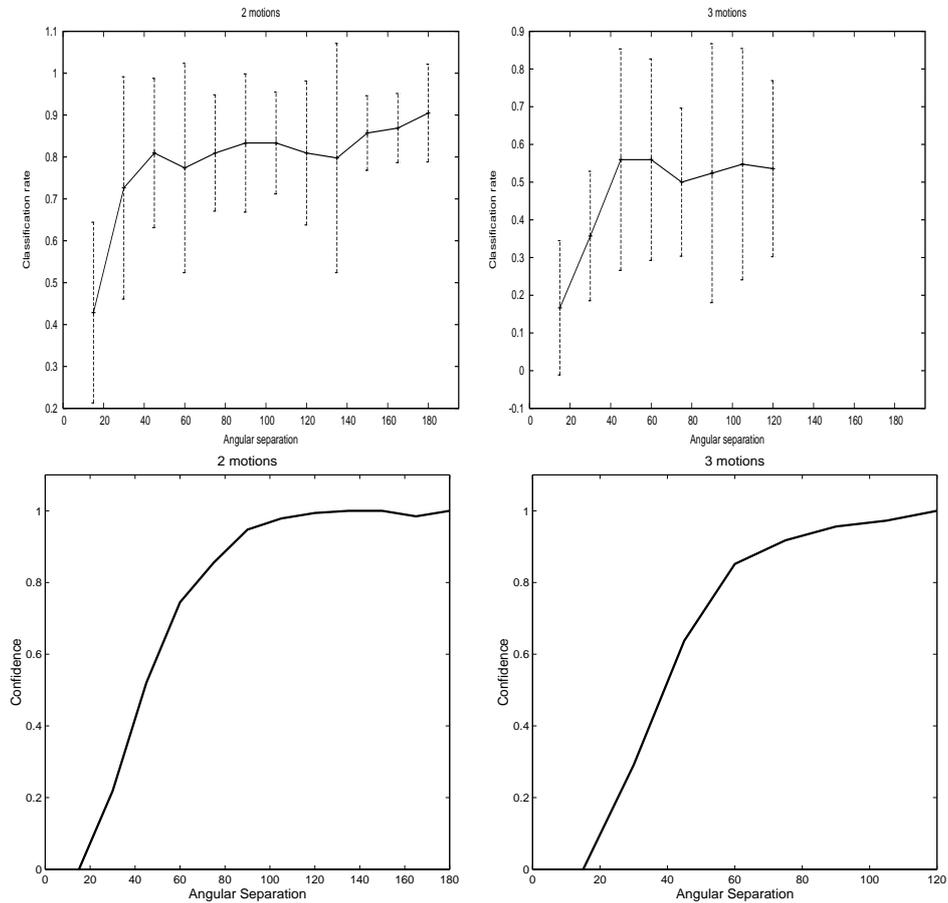
**Abbildung 8:** Ergebnisse der Berechnung eines Flussfeldes mit Verdeckungen. Rechts wird ein Einzelbild einer Sequenz gezeigt, in der sich ein Quadrat bewegt und dabei den Hintergrund verdeckt. Das Ergebnis in der Mitte wurde für eine Sequenz gerechnet, in der sich das Quadrat nach rechts und der Hintergrund nach unten bewegen. Das Ergebnis rechts erhielten wir für eine Bewegung des Quadrates nach unten rechts gegen einen stationären Hintergrund. In beiden Fällen werden die Bewegungsvektoren richtig berechnet.



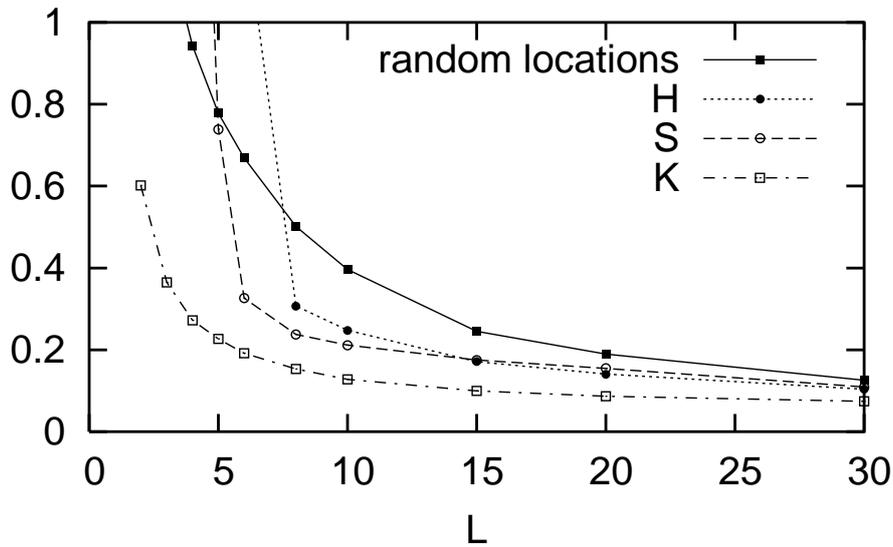
**Abbildung 9:** Simulation von Neuronen des Areals MT von Affen. Die Neurone antworten bevorzugt auf eine bestimmte Bewegungsrichtung (links oben) aber auch auf eine bestimmte Orientierung im Bild (links unten). Die Simulationen (rechts) entsprechen sehr gut den gemessenen Daten (Einzelzell-Ableitungen in wachen Makaken).



**Abbildung 10:** Simulation von Neuronen des Areals MT von Affen. Die Aktivität der Neurone ist links dargestellt. Gemessen wurde mit folgenden Stimuli: einem (mittlere fette Kurve), zwei gleichgerichteten (äußere gestrichelte Kurve) und zwei entgegengesetzten (innere dünne Kurve) Bewegungsreizen. Bewegt wurden dabei kleine Lichtpunkte. Die Simulationen (rechts) entsprechen sehr gut den gemessenen Daten (Einzelzell-Ableitungen in wachen Makaken).



**Abbildung 11:** Daten und Simulationsergebnisse zur Wahrnehmung überlagerter Bewegungen. Probanden mussten angeben, wie viele Bewegungen in einem kurzen Film zu sehen waren. Die zugelassenen Antworten waren 1 bis 5. Gezeigt wurden Überlagerungen von 1 bis 4 Schichten, die sich in verschiedene Richtungen bewegten. Der Winkel zwischen den Bewegungsrichtungen wurde verändert und ist als Parameter auf der Abszisse aufgetragen. Die Ordinate gibt an, wie oft 2 (linkes Bild oben) oder 3 (rechtes Bild oben) Bewegungen richtig erkannt wurden. Die entsprechenden Simulationsergebnisse werden darunter dargestellt. Es zeigt sich, dass ein Konfidenzmaß, welches die Konditionierung des Strukturtenors beschreibt, die Daten sehr gut erklärt.



**Abbildung 12:** Prädiktionsfehler für unterschiedliche Signifikanz-Maße. Mehrere Probanden betrachteten einen kurzen Videofilm, der eine Straßenszene zeigte. Dabei wurde gemessen, wo sie hinsahen. Danach wurden für den gleichen Videofilm signifikante Pixel bestimmt, und zwar so, dass diese Pixel idealerweise diejenigen waren, die tatsächlich angeschaut wurden.  $L$  solcher Kandidaten zur Auslösung von Augenbewegungen wurden für jedes Bild bestimmt. Ausgewertet wurde dann der Abstand zwischen den tatsächlichen Blickpunkten und demjenigen Kandidaten, der am nächsten am tatsächlichen Blickpunkt lag. Dieser Abstand wurde durch die mittlere Sakkadenlänge dividiert und auf der Ordinate aufgetragen. Für zufällig gewählte Kandidaten ist der Fehler am größten (durchgezogene Linie). Die übrigen Kandidaten wurden anhand der Invarianten des Strukturensors bestimmt, und zwar so, dass ein Kandidat wahrscheinlich wurde, wenn die entsprechende Invariante einen hohen Wert hatte [55]. Man beachte, dass die Wahrscheinlichkeit dafür, dass die Augenbewegungen an einem bestimmten Ort landen, mit der intrinsischen Dimension an dem Ort steigt (Zur Erinnerung:  $H, S, K \neq 0$  implizieren eine intrinsische Dimension von mindestens 1, 2, 3 - s. Abschnitt 2.2.2).

## Danksagung

Die Arbeiten sind teilweise am NASA Ames Forschungszentrum in Kalifornien entstanden und wurden damals von der DFG unter Ba 1176/4-1 und der NASA gefördert. Ich danke meinen dortigen Kollegen und Freunden Beau Watson, Al Ahumada und Jeff Mulligan.

Die Arbeiten wurden an der Universität zu Lübeck fortgesetzt. Ich danke Til Aach und Thomas Martinetz, die mich an ihren Instituten aufgenommen und unterstützt haben. Auch möchte ich mich bei Jürgen Jost vom MPI in Leipzig für seine Unterstützung bedanken.

Ebenfalls von der DFG wurden die Arbeiten zu den komplexen Bewegungsmustern gefördert, und zwar unter Ba 1176/7-1 sowie Ba 1176/7-2. Ergebnisse zu diesem Thema entstanden jedoch bereits im Jahre 2000 und wurden damals durch ein Stipendium des DAAD an Cicero Mota gefördert. Der Kontakt zu Cicero Mota wiederum entstand während meines Aufenthaltes am Instituto de Matematica Pura e Applicada in Rio de Janeiro, Brasilien, der von der Gesellschaft für Mathematik und Datenverarbeitung finanziert wurde. Mitgewirkt haben im Bereich Bewegungsschätzung weiterhin Til Aach, meine Kollegen vom ISIP Ingo Stuke und Daniel Toth, sowie die damaligen Studenten Michael Dorr, Martin Haker, Amir Madany und Thomas Otto.

Die Arbeiten zur Voraussage und Lenkung von Augenbewegungen werden derzeit vom BMBF, Projekt ModKog-Itap, gefördert. Ich danke dem Itap Team am INB bestehend aus Michael Dorr, Martin Böhme, Christopher Krause und Thomas Martinetz, sowie unseren Itap Partnern Karl Gegenfurtner und der SensoMotorik Instruments GmbH.

Michael Dorr und Cicero Mota danke ich für die Hilfe bei der Durchsicht des Manuskripts.

Die Arbeit widme ich meinen Mädchen.

## Literatur

- [1] Barth, E. und Watson, A. B. (2000). A geometric framework for nonlinear visual coding. *Optics Express*, 7:155–85. <http://www.opticsexpress.org/oearchive/source/23045.htm>.
- [2] Barth, E. (1999). Bewegung als intrinsische Geometrie von Bildfolgen. In Förster, W., Buhmann, J. M., Faber, A., und Faber, P. (Hrsg.), *Mustererkennung 99*, Seiten 301–308, Bonn. Springer, Berlin.
- [3] Barth, E. (2000). The minors of the structure tensor. In Sommer, G. (Hrsg.), *Mustererkennung 2000*, Seiten 221–228. Springer, Berlin.
- [4] Mota, C. und Barth, E. (2000). On the uniqueness of curvature features. In Baratoff, G. und Neumann, H. (Hrsg.), *Dynamische Perzeption*, Band 9 der Reihe *Proceedings in Artificial Intelligence*, Seiten 175–8, Köln. Infix Verlag.
- [5] Barth, E., Stuke, I., und Mota, C. (2002). Analysis of motion and curvature in image sequences. In *Proc. IEEE Southwest Symp. Image Analysis and Interpretation*, Seiten 206–10, Santa Fe, NM. IEEE Computer Press.
- [6] Barth, E., Stuke, I., Aach, T., und Mota, C. (2003a). Spatio-temporal motion estimation for transparency and occlusion. In *Proc. IEEE Int. Conf. Image Processing*, Band III, Seiten 69–72, Barcelona, Spain. IEEE Signal Processing Soc.
- [7] Barth, E., Drewes, J., und Martinetz, T. (2003b). Dynamic predictions of tracked gaze. In *Seventh International Symposium on Signal Processing and its Applications*, Paris. Special Session on Foveated Vision in Image and Video Processing.
- [8] Mota, C., Dorr, M., Stuke, I., und Barth, E. (2004). Categorization of transparent-motion patterns using the projective plane. *International Journal of Computer & Information Science*, 5(2).
- [9] Zetsche, C. und Barth, E. (1990). Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Research*, 30:1111–7.
- [10] Barth, E., Zetsche, C., und Krieger, G. (1998). Curvature Measures in Visual Information Processing. *Open Systems and Information Dynamics*, 5:25–39.
- [11] Zetsche, C. und Barth, E. (1990). Image surface predicates and the neural encoding of two-dimensional signal variation. In Rogowitz, B. (Hrsg.), *Human Vision and Electronic Imaging: Models, Methods, and Applications*, Band SPIE 1249, Seiten 160–177.
- [12] Zetsche, C., Barth, E., und Wegmann, B. (1993). The importance of intrinsically two-dimensional image features in biological vision and picture coding. In Watson, A. B. (Hrsg.), *Digital Images and Human Vision*, Seiten 109–38. MIT Press.
- [13] Zetsche, C. und Barth, E. (1992). Models of end-stopping and curvature detection: towards a general mathematical theory. *Perception*, 21:99.
- [14] Barth, E., Zetsche, C., Ferraro, M., und Rentschler, I. (1993). Fractal properties from 2D-curvature on multiple scales. In Vemuri, B. (Hrsg.), *Geometric Methods in Computer Vision II*, Band SPIE 2031, Seiten 87–99.

- [15] Barth, E. (1995). *Geometrische Analyse und iterative Synthese von Bildern*, Band 361 der Reihe 10:Informatik/Kommunikationstechnik. VDI Verlag, Düsseldorf.
- [16] Krieger, G., Zetzsche, C., und Barth, E. (1995). Nonlinear image operators for the detection of local intrinsic dimensionality. In *Proc. IEEE Workshop Nonlinear Signal and Image Processing*, Seiten 182–185.
- [17] Krieger, G., Zetzsche, C., und Barth, E. (1996). Die Bedeutung krümmungsselektiver Informationsverarbeitung für die visuelle Wahrnehmung. In *Proc. 40. Kongress der Deutschen Gesellschaft für Psychologie*. München.
- [18] Barth, E., Zetzsche, C., und Krieger, G. (1994). Lokalisation informationstragender Bildmerkmale durch 2D-Neurone. In *Präattentive and attentive Prozesse bei der visuellen Wahrnehmung*, Göttingen. MPI für Biophysikalische Chemie.
- [19] Krieger, G., Zetzsche, C., und Barth, E. (1997). Higher-order statistics of natural images and their exploitation by operators selective to intrinsic dimensionality. In *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics*, Band PR08005, Seiten 147–151. IEEE Computer Society, Los Alamitos, CA.
- [20] Barth, E., Zetzsche, C., und Rentschler, I. (1998). Intrinsic two-dimensional features as textons. *J. Opt. Soc. Am. A*, 15(7):1723–1732.
- [21] Zetzsche, C., Barth, E., Krieger, G., und Wegmann, B. (1997). Neural network models and the visual cortex: The missing link between cortical orientation selectivity and the natural environment. *Neuroscience Letters*, 228(3):155–158.
- [22] Barth, E. und Zetzsche, C. (1998). Endstopped operators based on iterated nonlinear center-surround inhibition. In Rogowitz, B. und Pappathomas, T. (Hrsg.), *Human Vision and Electronic Image Processing*, Band 3299 der Reihe *Proc. SPIE*, Seiten 67–78, Bellingham, WA.
- [23] Wegmann, B. und Zetzsche, C. (1992). Efficient image sequence coding by vector quantization of spatiotemporal bandpass outputs. In Maragos, P. (Hrsg.), *Visual Communications and Image Processing '92*, Band SPIE-1818, Seiten 1146–1154. SPIE, Bellingham, WA.
- [24] Zetzsche, C., Barth, E., und Wegmann, B. (1993). Nonlinear aspects of primary vision: entropy reduction beyond decorrelation (Invited Address). In Morreale, J. (Hrsg.), *SID International Symposium - Digest of Technical Papers*, Band XXIV, Seiten 933–936. Soc. Inform. Display, Playa del Ray, CA.
- [25] Wegmann, B. (1994). *Bildsequenzcodierung auf der Basis des menschlichen visuellen Systems*. Nummer 283 der Reihe Reihe 10: Informatik/Kommunikationstechnik. VDI Verlag, Düsseldorf.
- [26] Wegmann, B. und Zetzsche, C. (1996). Feature-specific vector quantization of images. [*Special Issue: Vector Quantization*] - *IEEE Trans. Image Processing*, 5:274–288.
- [27] do Carmo, M. P. (1976). *Differential Geometry of Curves and Surfaces*. Prentice-Hall, Englewood Cliffs, NJ.

- [28] Spivak, M. (1970/75). *A Comprehensive Introduction to Differential Geometry*, Band 1 to 5. Publish or Perish, Boston, MA.
- [29] Erhardt Barth, Terry Caelli, C. Z. (1993). Image Encoding, Labeling, and Reconstruction from Differential Geometry. *CVGIP: Graphical Model and Image Processing*, 55(6):428–46.
- [30] Zetsche, C. und Barth, E. (1991). Direct detection of flow discontinuities by 3D curvature operators. *Pattern Recognition Letters*, 12:771–9.
- [31] Ballard, D. und Brown, C. (1982). *Computer Vision*. Prentice Hall, Englewood Cliffs, New Jersey.
- [32] Granlund, G. H. und Knutsson, H. (1995). *Signal Processing for Computer Vision*. Kluwer.
- [33] Haußecker, H. und Spies, H. (1999). Motion. In Jähne, B., Haußecker, H., und Geißler, P. (Hrsg.), *Handbook of Computer Vision and Applications*, Band 2, Seiten 309–96. Academic Press.
- [34] Mota, C., Stuke, I., und Barth, E. (2001). Analytic solutions for multiple motions. In *Proc. IEEE Int. Conf. Image Processing*, Band II, Seiten 917–20, Thessaloniki, Greece. IEEE Signal Processing Soc.
- [35] Mota, C., Stuke, I., und Barth, E. (2002). Analytic solutions for multiple motions. Poster presented at the First Annual Meeting of the DFG SPP 1114.
- [36] Stuke, I., Aach, T., Mota, C., und Barth, E. (2003a). Estimation of multiple motions: regularization and performance evaluation. In Vasudev, B., Hsing, T. R., Tescher, A. G., und Ebrahimi, T. (Hrsg.), *Image and Video Communications and Processing 2003*, Band 5022 der Reihe *Proceedings of SPIE*, Seiten 75–86.
- [37] Stuke, I., Aach, T., Mota, C., und Barth, E. (2003b). Linear and regularized solutions for multiple motion. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Band III, Seiten 157–60, Hong Kong. IEEE Signal Processing Soc.
- [38] Stuke, I., Aach, T., Barth, E., und Mota, C. (2003c). Estimation of multiple motions by block matching. In Dosch, W. und Lee, R. Y. (Hrsg.), *Proc. ACIS 4th Int. Conf. Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, Seiten 358–62, Lübeck, Germany.
- [39] Stuke, I., Aach, T., Barth, E., und Mota, C. (2004). Estimation of multiple motions using block-matching and Markov random fields. In Panchanathan, S. und Vasudev, B. (Hrsg.), *Visual Communications and Image Processing 2004, IS&T/SPIE 16th Annual Symposium Electronic Imaging*, San Jose, California. To appear.
- [40] Aach, T., Stuke, I., Mota, C., und Barth, E. (2004). Estimation of Multiple Local Orientations in Image Signals. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*. IEEE Signal Processing Soc.
- [41] Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the Macaque. *The American Physiological Society*, 52(6):1106–1130.
- [42] Recanzone, G. H., Wurtz, R. H., und Schwarz, U. (1997). Responses of MT and MST Neurons to One and Two Moving Objects in the Receptive Field. *J Neurophysiology*, 78:2904–15.

- [43] Barth, E., Ferraro, M., Zetzsche, C., und Rentschler, I. (1993). Computational models for the topological selectivity in early and primitive vision systems. *OSA Annual Meeting Technical Digest*, 16:186.
- [44] Barth, E., Ferraro, M., und Zetzsche, C. (2001). Global topological properties of images derived from local curvature features. In *4th International Workshop on Visual Form*, Capri.
- [45] Wuerger, S., Shapley, R., und Rubin, N. (1996). On the visually perceived direction of motion by Hans Wallach: 60 years later. *Perception*, 25:1317–67.
- [46] Dorr, M., Stuke, I., Mota, C., und Barth, E. (2001). Mathematical and perceptual analysis of multiple motions. In Bülhoff, H. H., Gegenfurtner, K. R., Mallot, H. A., und Ulrich, R. (Hrsg.), *TWK 2001 Beiträge zur 4. Tübinger Wahrnehmungskonferenz*, Seite 174.
- [47] Barth, E., Dorr, M., Stuke, I., und Mota, C. (2001). Theory and some data for up to four transparent motions. *Perception*, 30 (Supplement):36.
- [48] Mota, C., Dorr, M., Stuke, I., und Barth, E. (2003). Categorization of transparent-motion patterns using the projective plane. In Dosch, W. und Lee, R. Y. (Hrsg.), *Proc. ACIS 4th Int. Conf. Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, Seiten 633–9.
- [49] Barth, E. (2001). Information technology for active perception. First GRP Symposium, Sehen und Aufmerksamkeit im Alter, Benediktbeuren.
- [50] Barth, E. und Martinetz, T. (2002). Information technology for active perception. Poster presented at the 8th Annual German-American Beckman Frontiers of Science Symposium.
- [51] O’Regan, J. K., Rensink, R. A., und Clark, J. J. (1999). Change-blindness as a result of ‘mudsplashes’. *Nature*, 398:34.
- [52] O’Regan, J. K. und Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):939–1031.
- [53] Simons, D. J. und Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, 28:1059–74.
- [54] Barth, E., Drewes, J., und Martinetz, T. (2003). Individual predictions of eye-movements with dynamic scenes. In Rogowitz, B. und Pappas, T. (Hrsg.), *Electronic Imaging 2003*, Band 5007. SPIE.
- [55] Böhme, M., Krause, C., Barth, E., und Martinetz, T. Eye Movement Predictions Enhanced By Saccade Detection. In *Brain Inspired Cognitive Systems 2004*.
- [56] Krüger, N. und Felsberg, M. (2003). A continuous formulation of intrinsic dimension. In *Proceedings of the British Machine Vision Conference*.
- [57] Knutsson, H. (1989). Representing Local Structure Using Tensors. In *Proc 6th Scand Conf Image Analysis Oulu, Finland, 1989*, Seiten 248–51.
- [58] Sommer, G., Bayro-Corrochano, E., und Bülow, T. (1997). Geometric Algebra as a Framework for the Perception-Action Cycle. In Solina, F., Kropatsch, W., Klette, R., und Bajcsy, R. (Hrsg.), *Advances in Computer Vision*. Springer.

- [59] Sommer, G. und Koenderink, J. (Hrsg.) (1997). *Algebraic Frames for the Perception–Action Cycle*, Band 1315 der Reihe *Lecture Notes in Computer Science*. Springer–Verlag, Heidelberg.
- [60] Sommer, G. (Hrsg.) (2001). *Geometric Computing with Clifford Algebras*. Springer-Verlag, Heidelberg.
- [61] Barth, E., Caelli, T., und Zetsche, C. (1991). Efficient visual representation and reconstruction from generalized curvature measures. In Vemuri, B. (Hrsg.), *Geometric Methods in Computer Vision*, Band SPIE 1570, Seiten 86–95.
- [62] Adelson, E. H. und Bergen, J. R. (1991). The Plenoptic Function and the Elements of Early Vision. In Landy, M. S. und Movshon, J. A. (Hrsg.), *Computational Models of Visual Processing*, Seiten 3–20. MIT Press, Cambridge, MA.
- [63] Rohr, K. und Schnörr, C. (1993). An efficient approach to the identification of characteristic intensity variations. *image and vision computing*, 11:273–277.
- [64] Shi, J. und Tomasi, C. Good Features to Track. In *Proc IEEE Conference on Computer Vision and Pattern Recognition (CVPR94) Seattle, 1994*.
- [65] Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., und Pitts, W. H. (1959). What the frog’s eye tells the frog’s brain. *Proceedings of the Institute of Radio Engineers*, 47:1950–61.
- [66] Lettvin, J. Y., Maturana, H. R., Pitts, W. H., und McCulloch, W. S. (1961). Two remarks on the visual system of the frog. In Rosenblith, W. A. (Hrsg.), *Sensory Communications*, Seiten 757–76. John Wiley.
- [67] aki Saito, H., Tanaka, K., Fukada, Y., und Oyamada, H. (1988). Analysis of Discontinuity in Visual Contours in Area 19 of the Cat. *J Neuroscience*, 8(4):1131–43.
- [68] Orban, G. A. (1984). *Neuronal operations in the visual cortex*. Springer, Heidelberg.
- [69] von der Heydt, R. und Peterhans, E. (1989). Mechanisms of Contour Perception in Monkey Visual Cortex. *J Neuroscience*, 9(5):1731–48.
- [70] Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61:183–193.
- [71] Kanisza, G. (1976). Subjective Contours. *Scientific American*, 234:48–52.
- [72] Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*.
- [73] Biederman, I. (1985). Human Image Understanding: Recent Research and a Theory. *Computer Vision, Graphics, and Image Processing*, 32:29–73.
- [74] Dobbins, A., Zucker, S. W., und Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329:438–41.
- [75] Heitger, F., Rosenthaler, L., von der Heydt, R., Peterhans, E., und Kübler, O. (1992). Simulation of Neural Contour Mechanisms: from Simple to End-stopped Cells. *Vision Research*, 32(5):63–981.

- [76] Rao, R. P. N. und Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.
- [77] Koenderink, J. J. und van Doorn, A. J. (1987). Representation of Local Geometry in the Visual System. *Biol Cybernetics*, 55:367–75.
- [78] Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the Macaque. *The American Physiological Society*.
- [79] Brill, M. H. (2001). Riemann tensor of motion vision revisited. *Optics Express*, 9(1):7–8.
- [80] Mulligan, J. B. (1993). Nonlinear Combination Rules and the Perception of Visual Motion Transparency. *Vision Research*, 33(14):2021–30.
- [81] Braddick, O. und Quian, N. (2001). The Organization of Global Motion and Transparency. In Zanker, J. M. und Zeil, J. (Hrsg.), *Motion Vision - Computational, Neural, and Ecological Constraints*, Seiten 86–111. Springer Verlag, Berlin Heidelberg New York.
- [82] Noton, D. und Stark, L. (1971). Eye Movements and Visual Perception. *Scientific American*, 224(6):34–43.
- [83] Zetzsche, C., Schill, K., H.Deubel, G.Krieger, Umkehrer, E., und Beinlich, S. (1998). Investigation of a sensorimotor system for saccadic scene analysis: an integrated approach. In Pfeifer, R., Blumenberg, B., Meyer, J., und Wilson, S. (Hrsg.), *Proc. 5th Intl. Conf. Soc. Adaptive Behavior*, Band 5, Seiten 120–126. MIT Press, Cambridge, MA.
- [84] Itti, L. und Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203.
- [85] Barron, J. L., Fleet, D. J., und Beauchemin, S. S. (1994). Performance of optical flow techniques. *IJCV*, 12(1):43–77.
- [86] Bertero, M. A., Poggio, T., und Torre, V. (1988). Ill-Posed Problems in Early Vision. *Proceedings of IEEE*, 76(8):869–89.
- [87] Horn, B. und Schunck, B. (1981). Determining Optical Flow. *Artificial Intelligence*, 17(1–3):185–203.
- [88] Lucas, B. und Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proc. DARPA Image Understanding Workshop*, Seiten 121–30.
- [89] Uras, S., Giosi, F., Verri, A., und Torre, V. (1988). A computational approach to motion computation. *Biological Cybernetics*, 60(5):79–97.
- [90] Black, M. J. und Anandan, P. (1996). The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104.
- [91] Shizawa, M. und Mase, K. (1990). Simultaneous multiple optical flow estimation. In *IEEE Conf. Computer Vision and Pattern Recognition*, Band I, Seiten 274–8, Atlantic City, NJ. IEEE Computer Press.

- [92] Beauchemin, S. S. und Barron, J. L. (2000). The Frequency Structure of 1D Occluding Image Sequences. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(2):200–6.
- [93] Yu, W., Sommer, G., Beauchemin, S., und Daniilidis, K. (2002). Oriented structure of the occlusion distortion: is it reliable? *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(9):1286–90.
- [94] Darrell, T. und Simoncelli, E. (1993). Nulling Filters and the Separation of Transparent Motions. In *IEEE Conf. Computer Vision and Pattern Recognition*, Seiten 738–9, New York. IEEE Computer Press.
- [95] Irani, M., Rousso, B., und Peleg, S. (1994). Computing Occluding and Transparent Motions. *International Journal of Computer Vision*, 12(1):5–16.
- [96] Vernon, D. (1998). Decoupling Fourier components of dynamic image sequences: a theory of signal separation, image segmentation and optical flow estimation. In Burkhardt, H. und Neumann, B. (Hrsg.), *Computer Vision - ECCV'98*, Band 1407/II der Reihe *LNCS*, Seiten 68–85. Springer Verlag.
- [97] Yu, W., Daniilidis, K., Beauchemin, S., und Sommer, G. (1999). Detection and Characterization of Multiple Motion Points. In *IEEE Conf. Computer Vision and Pattern Recognition*, Band I, Seiten 171–7, Fort Collins, CO. IEEE Computer Press.
- [98] Dorr, M. (2004). Effects of Gaze-Contingent Stimuli on Eye Movements. Diplomarbeit, Universität zu Lübeck. <http://www.inb.uni-luebeck.de/~dorr/diplthesis.html>.
- [99] Dorr, M., Martinetz, T., Gegenfurtner, K., und Barth, E. (2004). Effects of gaze-contingent stimulation on eye movements with natural videos. Zur *European Conference on Visual Perception* angenommener Beitrag.