Mehrauflösungsverfahren zur Detektion und Verfolgung von schnellen Gelenkbewegungen in Sport-Bildfolgen

Diplomarbeit
am Lehrstuhl für Kognitive Systeme
des Instituts für Informatik und Praktische Mathematik
der Christian-Albrechts-Universität Kiel

vorgelegt von
Stephanie Willemsen

betreut durch
Dr. Konstantinos Daniilidis
und
Prof. Dr. Gerald Sommer

Oktober 1997

Zusammenfassung

Die vorliegende Arbeit befaßt sich mit der visuellen Bewegungsanalyse des Menschen, wobei der Schwerpunkt auf der automatischen Verfolgung von schnellen Bewegungen liegt. Schnelle Bewegungen führen zum Problem des *zeitlichen Aliasings* und bedürfen daher einer speziellen Lösung.

Es handelt sich bei dem vorgestellten Verfahren um ein passives Verfahren, das auf Kennzeichnungen verzichtet und daher auch bereits aufgenommene Bildfolgen verarbeiten kann. Der vorgestellte Algorithmus gehört zu den in der zweidimensionalen Bildebene operierenden, modellfreien Verfahren. Seine Basis ist die *hierarchische* Bestimmung des *optischen Flusses*. Die Bewegungsanalyse erfolgt in zwei Schritten: Der eine besteht aus der Detektion der Person und der andere aus der Detektion und Verfolgung der Gelenke.

Bei der Detektion der Person werden keine restriktiven Bedingungen an den Hintergrund gestellt. Es wird lediglich vorausgesetzt, daß sich nur eine Person in der Szene befindet. Die Detektion erfolgt über die Bestimmung der Region, in der eine deutliche Bewegung stattfindet, durch sogenannte Änderungsdetektion. Daher ist man schon im Detektionsschritt mit dem Problem der schnellen Bewegungen konfrontiert. Dieses Problem wird durch einen hierarchischen Detektionsalgorithmus gelöst, der zur Detektion der Bewegung nicht allein die zeitliche Ableitung benutzt, sondern auf dem hierarchisch berechneten optischen Flußfeld operiert.

Zur Gelenkverfolgung wird die Verschiebung der Gelenke von Bild zu Bild über den optischen Fluß bestimmt. Durch die Applikation des optischen Flusses in der Verfolgung der Gelenke treten auch dort im allgemeinen Zuordnungsfehler auf, wenn die Geschwindigkeit der betrachteten Bewegung hoch ist. Die Verfolgung der Gelenke wird daher in dieser Arbeit durch einen hierarchischen Algorithmus realisiert, d.h. der zur Bestimmung der Verschiebungen notwendige optische Fluß wird auch hier auf hierarchischem Weg berechnet.

Der wesentliche Unterschied der hierarchischen Berechnungsalgorithmen liegt in den angenommenen Modellen für den optischen Fluß. Zur Detektion der Person wird das Modell des lokal konstanten optischen Flußfeldes angenommen, während bei der Verfolgung vorausgesetzt wird, daß die Bewegung eines jeden Körperglieds durch ein *affines Modell* des optischen Flusses dargestellt werden kann.

Bei einem hierarchischen Algorithmus ist es notwendig, die auf verschiedenen Auflösungsstufen berechneten Ergebnisse zu kombinieren. Man stößt dabei auf das Problem der Sensor-Fusion, welches durch die zwei in dieser Arbeit vorgestellten hierarchischen Verfahren zur differentiellen Berechnung des optischen Flusses gelöst wird. Diese Verfahren bedienen sich des sogenannten *Warpens*, welches dazu dient, daß man in einer feineren

Auflösungsstufe nur noch eine Verfeinerung der Ergebnisse berechnen muß. Außerdem liefern sie ein Gütemaß für die Zuverlässigkeit der Berechnungen, was bei einer Weiterverarbeitung der Daten, speziell bei der Fusion der Daten, unerläßlich ist.

Der methodische Beitrag dieser Arbeit ist die stochastische Behandlung der hierarchischen Flußberechnung sowohl im Falle des lokal konstanten als auch des affinen Modells. Da die zur Berechnung des optischen Flusses notwendigen Daten auf Messungen beruhen, ist es sinnvoll, die Meßungenauigkeiten stochastisch zu modellieren. Dazu werden in der vorliegenden Arbeit zwei unterschiedliche Rauschmodelle erläutert und die auf ihnen basierenden Berechnungsverfahren für den optischen Fluß erklärt. Die Verwendung des optischen Flusses bei der Analyse von Personenbewegungen, speziell im Zusammenhang mit schnellen Bewegungen, stellt den anwendungsbezogenen Beitrag dieser Arbeit dar.

Bei der Anwendung des Verfahrens auf Bildfolgen von Sportlern in der Biomechanik wird auf eine Kennzeichnung der Gelenke durch Marker, die der Sportler an sich trägt, verzichtet. Die Gelenke sind also allein durch ihren Grauwertverlauf definiert, und die Gelenkdetektion in der Initialisierungsphase der Verfolgung erfolgt per Hand.

Im Verlauf dieser Arbeit wird das System zur Detektion und Verfolgung der Bewegung vorgestellt und die zugrunde liegenden hierarchischen Varianten zur Bestimmung des optischen Flusses erläutert. Ferner wird einerseits ein analytischer Vergleich dieser beiden Berechnungsmethoden geliefert, und andererseits werden die damit erzielten experimentellen Ergebnisse sowohl in der üblichen Flußbestimmung als auch bei der Verfolgung und Detektion schneller Gelenkbewegungen präsentiert.

Inhaltsverzeichnis

Kapitel 1

Einführung

Das Thema dieser Arbeit ist die visuelle Analyse der Bewegungen eines Menschen. Welche Anwendungen es für die automatische Bewegungsanlyse gibt und wodurch sich die Analyse menschlicher Bewegungen als spezielles Gebiet auszeichnet, ist Gegenstand dieses Kapitels. Des weiteren wird die Entwicklung eines neuen Verfahrens durch eine kurze Darstellung der bisherigen Arbeiten zu diesem Thema und den dabei auftretenden Schwierigkeiten motiviert. Es folgt eine Darstellung des Beitrags, den die vorliegende Arbeit zum Fortschritt der visuellen Bewegungsanalyse liefert.

1.1 Problemstellung

In den letzten Jahren ist die Entwicklung von Systemen, die interaktiv auf ihre Umwelt reagieren, immer interessanter geworden. Damit der Maschine eine Interaktion möglich ist, muß sie aus den ihr zur Verfügung stehenden audiovisuellen Daten oder weiteren Sensordaten auf die Objekte und das Geschehen in der Umgebung schließen können. Zu den visuellen Daten gehören neben statischen Bildern auch Bildfolgen, aus denen man die Bewegung der Objekte ermitteln kann. Die Erkennung und Interpretation der Bewegungen von gegliederten Objekten, zu denen der menschliche Körper zählt, stellt dabei eine spezielle Herausforderung dar. Der Mensch selbst ist mit seinem visuellen System in der Lage, die Gestalt eines anderen Menschen zu erkennen und seine Bewegung zu verfolgen, ohne dabei auf zusätzliche Informationen angewiesen zu sein. Dies möchte man nun auch mit automatischen Systemen erreichen.

Die automatische Bewegungsanalyse des menschlichen Körpers hat einen großen Anwendungsbereich. Sie erlaubt eine Optimierung der folgenden Bereiche:

- Überwachungssysteme: In Gefahren- oder Sicherheitsbereichen (z.B. Fabriken, Museen) ist das Finden von sich bewegenden Personen und das Klassifizieren ihres Verhaltens wichtig, bei Fertigungsabläufen die Kontrolle von Robotern.

- Animation: In der Unterhaltungsindustrie wird die Bewegung von Modellen und Trickfiguren immer häufiger durch das Übertragen von Bewegungsmustern erzielt, die eine wirkliche Person ausführt (*'Actor to Cartoon'–Schnittstelle*).
- Biomechanik: Zum Bewerten der Bewegungsabläufe bei orthopädischen Patienten und für die Beurteilung und das Training sportlicher Übungen ist eine automatische Gelenkverfolgung sehr nützlich. So können Abweichungen von Musterbewegungen besser erkannt und genauer klassifiziert werden.
- Mensch–Maschine–Schnittstelle: Markante Bewegungen können als Steuerung von Maschinen dienen, z.b. die Hand als dreidimensionale Maus.
- Zeichensprachenübersetzung: Durch das Klassifizieren von Bewegungen ist es möglich automatische Übersetzungen von Zeichensprachen oder der Lippenbewegungen zu konstruieren.
- Virtual Reality: Damit die Displayparameter und die Animation an den Nutzer angepaßt werden können, ist die Kenntnis seiner Bewegung erforderlich.
- Kundenüberwachung: Die Interpretation der Bewegungen von Kunden ist eine nützliche Information für Werbefachleute.
- Verkehrsüberwachung: Das Beobachten von Fußgängern z.B. an Ampeln kann dazu dienen, Gefahrensituationen zu vermindern.
- Kamerasteuerung: Die Steuerung von Kameras durch Bewegung (z.B. bei Sportveranstaltungen, Konferenzen, Auftritten) macht die manuelle Steuerung durch einen Kameramann überflüssig.

1.2 Motivation

Da der menschliche Körper nicht starr ist, sondern aus mehreren sich verschieden bewegenden Gliedern besteht und er sich verbiegen, krümmen und drehen kann, stellt er in der Bewegungsanalyse ein eigenes herausforderndes Forschungsgebiet dar. Besondere Probleme werden bei der Analyse der menschlichen Bewegungen zusätzlich durch die Komplexität der Freiheitsgrade und speziell durch die Möglichkeit der Selbstverdeckung verursacht. Zudem kann es gerade an den Extremitäten durch Kopplung der einzelnen Bewegungen zu besonders schnellen Bewegungen kommen, die sich durch große Verschiebungen von Bild zu Bild bemerkbar machen, welche schwer zu verfolgen sind. Auch die Kleidung erschwert oft die Analyse, weil sie einerseits die Bewegung nicht in gleicher Weise ausführt wie der Mensch selbst und andererseits störende Reflexionen hervorrufen kann.

Die bisher üblichen Verfahren zur automatischen Bewegungsanalyse beruhen meist auf äußeren Hilfsmitteln (z.B. Marker, Datenhandschuh) und schränken so die Anwendbarkeit ein. Zum einen ist in vielen Anwendungen die Ausstattung der beobachteten Personen mit dieser technischen Ausrüstung undenkbar (z.B. bei Überwachungssystemen), weil die Personengruppe willkürlich ins Blickfeld der Kamera gerät, und zum anderen stellen die Hilfsmittel eine Einschränkung der Bewegungsfreiheit dar, was gerade bei Anwendungen in der Biomechanik von Bedeutung ist. Zudem haben psychologische Untersuchungen [10] ergeben, daß es dem menschlichen Gehirn schwerfällt, ungewohnte Bewegungsmuster allein aus Informationen über die Gelenke zu identifizieren, wodurch die Verwendung von Markern zu Schwierigkeiten führen kann. Man sucht also eine Methode zur Bewegungsanalyse, die auf zusätzliche Hilfsmittel verzichtet.

Ein solches System zur visuellen Bewegungsanalyse gliedert sich in zwei Teile: die Verfolgung der einzelnen Glieder bzw. Gelenke und die vorausgehenden Initialisierung der entsprechenden Körperteile. Zunächst müssen also die Bildbereiche detektiert werden, in denen der Mensch zu sehen ist, was über eine Bewegungsdetektion erreicht werden kann. Eine Aufspaltung in die sich einzeln bewegenden Glieder erfolgt im Anschluß daran über eine Modellanpassung oder die Verfeinerung der vorliegenden Silhouette durch Modellannahmen über die Struktur des Menschen. Bedingungen an die Lage der Gelenke spezifizieren dann ihre Position.

Die in der Literatur vorgestellten Systeme zur visuellen Bewegungsanalyse von artikulierten Objekten ohne besondere, kennzeichnende Hilfsmittel können in zwei Klassen eingeteilt werden: die modellbasierten und die modellfreien Verfahren.

Bei den modellbasierten Algorithmen erfolgt die Verfolgung einzelner Glieder durch wiederholtes Bestimmen der Körperhaltung, was wie in der Initialisierungsphase durch wiederholtes Anpassen des Körpermodells an die jeweiligen Bilddaten erreicht wird. Bei diesen Verfahren werden also alle möglichen Zustände des Modells generiert und derjenige Zustand als die aktuelle Haltung bestimmend gewählt, der bei der Anpassung an die Bilddaten ein Ähnlichkeitsmaß optimiert. Der dazu erforderliche Rechenaufwand ist wegen der umfangreichen Menge an Parametern hoch. Einige der modellbasierten Verfahren benutzen daher zusätzlich zu den Körpermodellen Bewegungsmuster, die die Freiheitsgrade der Bewegung und damit die Parametermenge einschränken. Der damit verbundene Nachteil ist die Beschränkung der Anwendung auf eine bestimmte Menge von möglichen Bewegungen.

Die modellfreien Verfahren nutzen direkt die durch die Änderung der Grauwerte charakterisierte Bildbewegung aus. Die Ansätze zur Bewegungsbestimmung entsprechen dabei denen bei der Behandlung von starren Objekten. Man bestimmt anhand der Grauwerte das sogenannte *Verschiebungsvektorfeld*, welches Schlüsse auf die ausgeführte Bewegung ermöglicht. Die Durchführung der durch die so ermittelten Bewegungsdaten definierten Bewegung an einem interessierenden Punkt führt dann an die Stelle im nächsten Bild, an der dieser Punkt wiederzufinden ist. Doch schon bei der Bewegungsanalyse von starren Objekten treten durch Beleuchtungsvariationen und Meßdatenungenauigkeiten gerade bei der Verarbeitung von realen Bildfolgen Schwierigkeiten auf, die bei der Betrachtung von artikulierten Objekten zu den bereits genannten Problemen hinzukommen. Außerdem erschweren die durch schnelle Bewegungen verursachten großen Verschiebungen von Grauwertstrukturen im Bild die Bestimmung des Verschiebungsvektorfeldes bzw. der Bildgeschwindigkeit.

Eine weiterer Unterschied zwischen den Verfahren zur Bewegungsanalyse besteht in der Dimension des Operationsbereiches. Man unterscheidet zwischen Verfahren, die unter Ausnutzung der Daten von mehreren starren Kameras im dreidimensionalen Raum arbeiten, und Verfahren, bei denen die Bewegungsanalyse in der zweidimensionalen Bildebene einer Kamera stattfindet. Bei den 3D-Verfahren muß zunächst aus den Kameradaten das dreidimensionale Modell des Objekts konstruiert werden, was ein eigenes Problem darstellt. Durch die räumliche Informationen, die bei diesen Verfahren zur Verfügung stehen, entfällt das Problem der Verdeckung. Die Verfolgung schneller Bewegungen wird durch ein dreidimensionales Modell des Körpers jedoch nicht gelöst, sondern die Schwierigkeiten werden eher noch verstärkt, da die Mehrdeutigkeit in der Zuordnung mit der Anzahl der Bilder für einen Zeitpunkt steigt.

Zu erwähnen ist hier auch die Möglichkeit, die Bewegungsanalyse auf Daten von Farbbildkameras zu stützen, wobei verschiedene Glieder durch unterschiedliche Farbwerte gekoppelt mit der relativen Lage zueinander charakterisiert werden. Dabei treten aber durch Verdeckung und schnelle Bewegungen die gleichen Probleme auf wie bei den auf Grauwertdaten basierenden Verfahren.

Es ist bei allen Methoden üblich, die Bewegungsschätzung durch ein Kalman–Filter robuster zu gestalten, was gerade bei abrupten Änderungen der Bewegungsrichtung und bei Verdeckung nützlich ist.

Die Verfolgung der Glieder bzw. Gelenke kann also durch verschiedene Strategien erreicht werden, die mit unterschiedlichen Schwierigkeiten konfrontiert sind. Die Leistung der modellbasierten Verfahren hängt direkt von der Güte des verwendeten Modells ab und erfordert einen hohen Rechenaufwand, während die modellfreien Verfahren Probleme beim Auftreten von schnellen Bewegungen zeigen.

1.3 Beitrag der Arbeit

Diese Arbeit konzentriert sich auf die Bewegungsanalyse des Menschen im Hinblick auf Anwendungen in der Biomechanik, insbesondere zur Untersuchung von Bewegungsabläufen bei Sportlern, wobei auf spezielle Hilfsmittel sowie vorgegebene Bewegungsmuster verzichtet wird. Das im folgenden vorgestellte Verfahren zur Detektion und Verfolgung gehört zu den modellfreien Methoden, welches als Hauptwerkzeug in beiden Teilbereichen den optischen Fluß benutzt. Ferner gehört es zu den passiven Verfahren. Es arbeitet auf bereits existierenden Bildfolgen, die mit einer starren Kamera aufgenommen worden sind. Daher erfolgen Detektion und Verfolgung der Bewegung in der 2D-Bildebene.

Die Detektion der Person wird gewöhnlich durch das Auffinden von deutlichen Änderungen der Grauwerte, die bei der Aufnahme durch die Bewegung der Person verursacht werden, erreicht. Man betrachtet dazu üblicherweise die zeitliche Ableitung der Grauwertfunktion, die durch die Bildfolge definiert wird.

In der Biomechanik werden u.a. die Bewegungen von Sportlern betrachtet. Gerade im Sport treten häufig hohe Geschwindigkeiten auf, die für die konventionelle Detektion über die zeitliche Ableitung und die bisherigen modellfreien Verfahren zur Verfolgung der Bewegung ein noch nicht gelöstes Problem darstellen. Zur Bewältigung dieses Problems wird daher in dieser Arbeit ein hierarchischer Ansatz vorgeschlagen.

Bei hierarchischen Verfahren stößt man auf das Problem der Sensor-Fusion, weil Berechnungsergebnisse aus verschiedenen Hierarchie-Ebenen kombiniert werden müssen. Daher ist es wichtig, sowohl robuste Messungen auf jeder Auflösungsstufe zu erzielen als auch eine zuverlässige Aussage über die Güte der Berechnungen machen zu können, so daß man die zu fusionierenden Daten entsprechend gewichten kann. Für die Definition eines Gütemasses ist eine stochastische Modellierung des Problems notwendig.

Das Verfahren stützt sich auf die Bewegungsinformation, die der optische Fluß liefert. Da die bekannten Verfahren zur Berechnung des optischen Flusses vernachlässigen, daß die verwendeten Meßdaten ungenau sind, und das Problem der Sensor-Fusion vorliegt, sind zwei differentielle Verfahren implementiert worden, die das Rauschen der Messungen berücksichtigen und ein Maß für die Güte der Berechnungen liefern. Zum einen ist das ein Verfahren von Simoncelli et al. ([31], [30]), das auf einem Modell basiert, welches das Rauschen in der zeitlichen Komponente umfaßt. Zum anderen handelt es sich um die Erweiterung eines Verfahrens von Kanatani [21], das zusätzlich das Rauschen in den örtlichen Komponenten mit einbezieht. Da Rauschen üblicherweise durch Zufallsvariablen charakterisiert wird, benutzen diese Verfahren Schätzmethoden der Stochastik.

Die Besonderheit des von Simoncelli et al. entwickelten Mehrauflösungsverfahrens ist neben der Berücksichtigung des Rauschens der Vorgang des *Warpens*. Warpen bedeutet, die Bewegungsdaten, die auf einer Hierarchie-Ebene berechnet worden sind, auf die nächste anzuwenden, d.h. die gefundene Bewegung auf der feineren Auflösungsstufe rückgängig zu machen, indem das Bild lokal entsprechend des berechneten optischen Flusses verschoben wird. So bleibt auf dieser Stufe nur noch eine kleine Restbewegung zu bestimmen.

In ähnlicher Weise ist das Verfahren von Kanatani auf eine hierarchische Verarbeitung erweitert worden, um so bei der Pyramidenverarbeitung von Bildsequenzen zur Berechnung des optischen Flusses das Rauschen in allen Komponenten zu berücksichtigen und ein darauf beruhendes Gütemaß benutzen zu können.

Diese Verfahren werden benutzt, um die Bewegung per Änderungsdetektion zu finden. Mit einem Schwellwertverfahren werden die Regionen ermittelt, an denen die Verschiebung von Bild zu Bild größer als ein vorgegebenes Minimum ist. Die anschließende Verfolgung ist auf bestimmte Gelenke, die nur durch ihren Grauwertverlauf definiert sind, beschränkt. Statt für jedes Bild eine Posenbestimmung durchzuführen, werden dabei die Bewegungsdaten jedes interessierenden Körperglieds pro Bild ermittelt, und eine gelenkspezifische Kombination der daraus resultierenden Verschiebungen wird auf die jeweilige aktuelle, bekannte Gelenkposition addiert.

Zur Berechnung der Geschwindigkeit der einzelnen Glieder wird ein affines Modell des optischen Flusses vorausgesetzt. Aus diesem Grund sind die oben genannten Berechnungsverfahren von Simoncelli et al. und Kanatani so abgewandelt worden, daß sie die Parameter des affinen Flusses bestimmen und dafür ebenfalls ein Gütemaß angeben. Eine spezielle Behandlung von besonders großen Geschwindigkeiten ist natürlich auch hier notwendig, weshalb diese Algorithmen als hierarchische Berechnung des affinen Flusses entwickelt worden sind.

Aus der so berechneten Geschwindigkeit erhält man die Verschiebung der Gelenke von einem Bild zum nächsten. Statt daraus die neue Position des Gelenks zu bestimmen, benutzt man die Verschiebung als Meßdatum für eine Kalman-Filterung, die eine robustere Schätzung der Verschiebung liefert, woraus die neue Gelenkposition berechnet wird.

In der Literatur wird bei der Präsentation von Verfahren zur Verfolgung von gegliederten Objekten zwar häufig der Ausblick gegeben, daß zur Lösung der Probleme, die bei der Bewegungsverfolgung durch schnelle Bewegungen entstehen, eine hierarchische Verarbeitung sinnvoll ist, doch ist bisher in diesem Zusammenhang kein hierarchischer Algorithmus vorgestellt worden. Die vorliegende Applikation der Mehrauflösungsverfahren in der Gelenkverfolgung ist die erste präsentierte Lösung, mit der man unter Verzicht auf ein Bewegungsmodell in der Bildebene die Verfolgung schneller Bewegungen realisieren kann.

1.4 Strukturierung der Arbeit

Im anschließenden Kapitel 2 werden zunächst die Grundlagen zur Berechnung des optischen Flusses erklärt. Anschließend werden in Abschnitt 3.1 die für das Verständnis der Berechnungsverfahren notwendigen Einblicke in Schätzmethoden der Stochastik gegeben.

Die Darstellung der Verfahren von Simoncelli et al. und Kanatani sowohl auf einer Auflösungsstufe als auch auf Bildpyramiden sowie deren Erweiterung zu Mehrauflösungsverfahren, mit denen der affine optische Fluß berechnet werden kann, findet man zusammen mit theoretischen und experimentellen Vergleichen in Abschnitt 3.2.

Ein Überblick über die bisherigen Arbeiten zur Verfolgung artikulierter Objekte sowie zur Anfangsschätzung der Körperstellung wird in Kapitel 4 gegeben. Außerdem werden in diesem Kapitel die Details des Verfolgungsverfahrens sowie die konkrete Implementierung vorgestellt. Am Ende des Kapitels werden die Ergebnisse der Experimente dargestellt, welche auf Bildsequenzen erzielt worden sind, die sportliche Übungen zeigen.

Verbleibende Probleme bei der Gelenkverfolgung und deren mögliche Lösungen, die nicht weiter verfolgt worden sind, werden in Kapitel 5 erläutert.

Abschließend gibt Kapitel 6 eine Zusammenfassung der Ergebnisse.

Kapitel 2

Berechnung des optischen Flusses

Um aus einer Bildfolge auf die Bewegung eines Objekts im dreidimensionalen Raum schließen zu können, muß man einen Zusammenhang zwischen der Bewegung des Objekts und der zeitlichen Veränderung der Grauwerte eines Bildes finden. Die relative Bewegung eines Objekts zu einer Kamera führt zu einer Verschiebung der zugehörigen Grauwertstruktur im Bild. Man nennt die Geschwindigkeit der durch die Grauwertänderungen verursachten scheinbaren Bewegung der Grauwertstruktur *optischer Fluß*. Die Detektion von Grauwertänderungen trägt zur Objektsegmentierung im Bild bei, während die Ermittlung der Bildbewegung zur Analyse der 3D-Objektbewegung dient.

Zur Darstellung der 3D-Objektbewegung benutzt man das sogenannte *Bewegungsfeld*, dessen Vektoren einander entsprechende Bild- bzw. Objektpunkte in zeitlich aufeinanderfolgenden Bildern verbinden. Es stellt somit die perspektivische Projektion der Bewegung von Objekten im Raum in die Bildebene dar und ist ein rein geometrischer Entwurf. Die Geschwindigkeit der projizierten Bewegung bezeichnet man als *Verschiebungsrate*. Das diskrete Bewegungsfeld entspricht demnach dem Produkt aus der Verschiebungsrate und dem Zeitintervall zwischen zwei Bildern. Das Produkt aus optischem Fluß und Zeitintervall wird hingegen *Verschiebungsvektorfeld* (VVF) genannt. Oft wird dieser Begriff in der Literatur auch für das Bewegungsfeld benutzt, was zu Verwirrungen führt.

Wie in [26] ausführlich erläutert wird, ist es notwendig, zwischen Bildbewegung und 3D-Objektbewegung, und daher auch zwischen optischem Fluß und Verschiebungsrate zu unterscheiden. Zur Veranschaulichung des Unterschieds betrachten wir die mathematische Definition der Verschiebungsrate.

Eine starre Bewegung eines Szeneobjekts kann durch eine Rotation um eine durch den Ursprung verlaufende Achse und eine anschließende Translation beschrieben werden. Wendet man diese linearen Transformationen auf die 3D-Darstellung eines Szenepunktes an und projiziert das Ergebnis in die Bildebene, erhält man die einem Vektor des Bewegungsfeldes entsprechende Verschiebung $\boldsymbol{d}(x, y, t + \tau)$. Die Verschiebungsrate ist definiert als zeitliche Ableitung der Verschiebung:

$$\boldsymbol{d}'(x, y, t) = \left. \frac{\partial \boldsymbol{d}(x, y, t + \tau)}{\partial \tau} \right|_{\tau = 0} .$$

Das bedeutet, daß die Verschiebungsrate die Steigung der Tangenten an die durch die Bewegung definierte Bahnkurve eines projizierten Szenepunktes ist.

Im Gegensatz dazu ist der optische Fluß direkt von der momentanen, lokalen ortszeitlichen Grauwertänderung bestimmt und wird daher von radiometrischen Effekten beeinflußt. Eine Veränderung der Lichtverhältnisse z.B. kann ohne das Vorliegen einer Bewegung die Ursache der Grauwertänderung sein. Außerdem kann im Falle von Grauwertplateaus eine Bewegung vorliegen, die sich nicht durch eine Änderung der Grauwerte bemerkbar macht. In diesem Fall ist die Korrespondenz von Bildpunkten nicht eindeutig.

Statt die Verschiebung $\boldsymbol{d}(x, y, t+\tau)$ direkt über das Zuordnungsproblem zu bestimmen, welchem Bildpunkt $\boldsymbol{g}(x, y, t+\tau)$ der Bildpunkt $\boldsymbol{g}(x, y, t)$ entspricht, kann man sie durch Integration über die Verschiebungsrate ermitteln. Dabei setzt man voraus, daß die Bilder mit genügend großer Bildrate aufgenommen werden.

Trotz der oben angesprochenen Schwierigkeiten geht man davon aus, daß die Bestimmung des optischen Flusses eine gute Näherung der Verschiebungsrate ist. Um mit dem optischen Fluß gute Schätzungen für die Verschiebungsrate zu erzielen, müssen dabei zusätzlich zu den geometrischen die radiometrischen Effekte berücksichtigt werden, welche die ortszeitlichen Variationen der Grauwerte beeinflussen.

Damit der optische Fluß als Näherung der Verschiebungsrate benutzt werden kann, ist demnach laut Nagel [26] vorauszusetzen, daß

- die Beleuchtung konstant ist.
- die Objekte das Licht diffus streuen (Lambert'sche Oberflächenreflexion).
- die Bewegung nur frontoparallel zur Bildebene stattfindet.

Ist die letzte Bedingung nicht erfüllt, verändert sich die Größe des betrachteten Oberflächenelements, wodurch die Merkmalsdichte im Bild ebenfalls geändert wird. Da der Grauwert durch die Diskretisierung keine Punkteigenschaft darstellt, sondern eine Dichte, weicht der optische Fluß in dem Fall von der Verschiebungsrate ab. Nagel schwächt diese Bedingung dennoch ab, und zwar mit der Begründung, daß die Effekte einer Tiefenbewegung bei Einhaltung der ersten zwei Bedingungen durch Terme der zeitlichen Ableitung kompensiert werden.

Die Verschiebung, also einen Vektor des Bewegungsfeldes, erhält man aus dem optischen Fluß, der üblicherweise mit $\boldsymbol{v} = (u, v)^T$ bezeichnet wird, und dem Zeitintervall $\Delta t = t_{i+1} - t_i$ über $(\Delta x, \Delta y) = (u\Delta t, v\Delta t)$, wobei man voraussetzt, daß die Bildrate so hoch ist, daß die Geschwindigkeit während des betrachteten Zeitintervalls konstant ist.

Zu weiteren Mehrdeutigkeiten bei der Bewegungsbestimmung, die unabhängig von der Grauwertverteilung sind, kommt es bei der Betrachtung von Objekten mit periodischer Textur (Wagenradeffekt), von deformierbaren Objekten und von mehreren identischen Teilchen, die sich unabhängig voneinander bewegen. In diesen Fällen braucht man zusätzliche, die Bewegung einschränkende Bedingungen, oder man muß die zeitliche Abtastrate der Bildaufnahme erhöhen.

Bei der Bestimmung des optischen Flusses betrachtet man nur die lokale Umgebung eines Bildpunktes, weil die Bewegung i.a. nicht örtlich konstant ist. Man kann sich die Maske der benutzten Filter vorstellen wie eine Blende, die man auf dem Bild verschiebt und die immer nur den begrenzten, sichtbaren Bereich des Bildes verarbeitet.

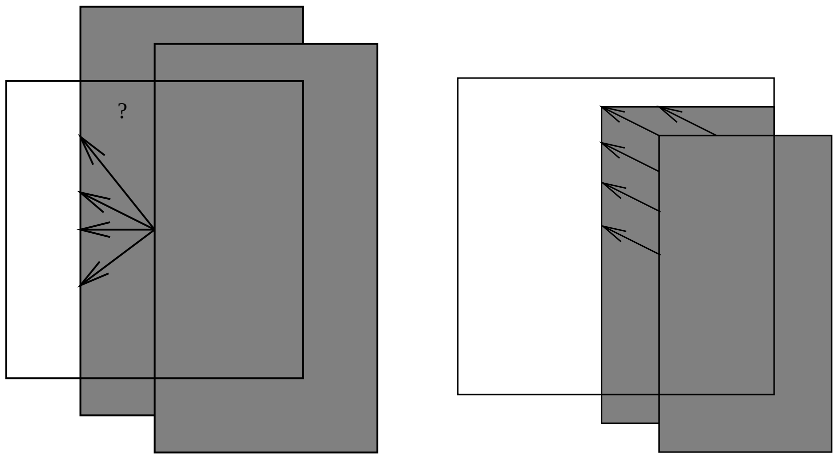


Abbildung 2.1: Beispiel zur Veranschaulichung des Aperturproblems: links: Mehrdeutigkeit der Verschiebungsvektoren an einer Kante; rechts: Eindeutigkeit des Verschiebungsvektors an einer Ecke

Daraus ergibt sich ein weiteres Problem: das Blenden- oder Aperturproblem. Wie in Abb. 2.1 gezeigt, kann man bei der Betrachtung eines Ausschnitts von der Größe der Maske der lokalen Filter die Bewegung einer lokalen Grauwertkante nicht bestimmen. Vergrößert man die Maske, so hat man die Chance als weitere Struktur eine Grauwertecke zu finden, an der die Bestimmung der Bewegungsrichtung wieder eindeutig wird. Falls man trotz einer Vergrößerung der Maske keine weiteren Strukturen findet oder ein Vergrößern zu kostenaufwendig ist, kann man nur den Fluß senkrecht zur Grauwertkante, den sogenannten *Normalfluß*, angeben.

Es gibt nun mehrere Möglichkeiten zur Berechnung des optischen Flusses benutzen kann. Man unterscheidet deshalb zwischen drei verschiedenen Arten von Verfahren, die in den folgenden Abschnitten vorgestellt werden:

1. *Differentielle Verfahren*
2. *Korrelationsverfahren*
3. *Frequenzbasierte Verfahren*

2.1 Differentielle Verfahren

Die grundlegende Annahme bei differentiellen Verfahren ist die Kontinuität des optischen Flusses (vgl. Strömungsphysik), d.h. daß durch eine Bewegung keine Grauwerte verloren gehen. Man geht hier also vom Prinzip der Grauwerterhaltung aus:

$$\boldsymbol{g}(x(t), y(t), t) = \text{const}., \tag{2.1}$$

wobei $\boldsymbol{g}(x, y, t)$ die Grauwertfunktion darstellt.

Wegen der Abhängigkeit der Grauwerte von radiometrischen Effekten müssen die im vorherigen Abschnitt genannten Voraussetzungen an die Lichtverhältnisse (konstante Beleuchtung, Lambert'sche Oberflächenreflexion) erfüllt sein, damit der Erhaltungssatz gültig ist. In der Realität sind diese Voraussetzungen i.a. nicht gegeben. Man nimmt jedoch an, daß sie für lokale Bereiche des Bildes zutreffen. Der Grad des Zutreffens der Grauwerterhaltung in der lokalen Bildregion ist für die Korrektheit des berechneten Flusses ausschlaggebend.

Das Ableiten der Grauwerterhaltung (2.1) nach der Zeit liefert uns:

$$\frac{d}{dt}\boldsymbol{g}(x(t), y(t), t) = 0.$$

Das ist gleichbedeutend mit:

$$\frac{\partial \boldsymbol{g}}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial \boldsymbol{g}}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial \boldsymbol{g}}{\partial t} = 0.$$

Da der optische Fluß die Geschwindigkeit der Grauwertstruktur angibt, kann man ihn als $\boldsymbol{v} = (u, v)^T = (\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t})^T$ schreiben. Setzt man dies in die vorherige Gleichung ein, erhält man die sogenannte Kontinuitätsgleichung (OFCE (Optical Flow Constraint Equation) oder BCCE (Brightness Consistency Constraint Equation)):

$$\nabla_s \boldsymbol{g}^T \boldsymbol{v} + g_t = 0. \tag{2.2}$$

Mit ∇_s sei im folgenden stets der örtliche Gradient gemeint.
Der Unterschied zwischen dem optischen Fluß und den Verschiebungsvektoren läßt sich leicht veranschaulichen. Formuliert man die Grauwerterhaltung auf Verschiebungen, lautet sie:

$$\boldsymbol{g}(x, y, t) = \boldsymbol{g}(x - u\Delta t, y - v\Delta t, t - \Delta t). \tag{2.3}$$

Da der Verschiebungsvektor keine differentielle Größe ist, wird dabei vorausgesetzt, daß Δt so klein gewählt ist, daß der optische Fluß während der Zeit konstant ist. Mittels Taylorreihenentwicklung erster Ordnung nach Δt erhält man aus dieser Formulierung die Gleichung:

$$\boldsymbol{g}(x, y, t) = \boldsymbol{g}(x, y, t) - \frac{\partial \boldsymbol{g}}{\partial x}u\Delta t - \frac{\partial \boldsymbol{g}}{\partial y}v\Delta t - \frac{\partial \boldsymbol{g}}{\partial t}\Delta t + \mathcal{O}(\boldsymbol{v}^2 d\Delta t^2). \tag{2.4}$$

Die Terme zweiter Ordnung kommen in der OFCE (2.2) nicht vor, was bei einer Interpretation des Flusses als Verschiebung zu Fehlern führt, wenn der Fluß und das Zeitintervall zwischen den Bildern groß sind. Dividiert man (2.4) jedoch durch Δt und betrachtet den Grenzwert für Δt gegen null, geht die Verschiebung in die Geschwindigkeit über, $\lim_{\Delta t \to 0} \frac{u\Delta t}{\Delta t} \to u$ bzw. $\lim_{\Delta t \to 0} \frac{v\Delta t}{\Delta t} \to v$, und man erhält die korrekte Formulierung der Kontinuitätsgleichung.

An der Gleichung (2.2) läßt sich das Aperturproblem erkennen: Es steht lediglich eine Gleichung mit zwei Unbekannten zur Verfügung; das Gleichungssystem ist also unterbestimmt. Man kann daher nicht beide Komponenten des optischen Flusses gleichzeitig an einem Bildpunkt bestimmen. Nur der Normalfluß, die Komponente entlang des Grauwertgradienten, also senkrecht zur lokalen Grauwertkante, läßt sich angeben:

$$\boldsymbol{v}_\perp = \boldsymbol{v}^T \frac{\nabla_s \boldsymbol{g}}{\| \nabla_s \boldsymbol{g} \|} = -\frac{g_t}{\| \nabla_s \boldsymbol{g} \|}. \tag{2.5}$$

An der Gleichung sieht man, daß bei separater Betrachtung der Grauwertgradienten erster Ordnung auch der Normalfluß an einem Bildpunkt nur dann bestimmbar ist, wenn der Gradient nicht verschwindet, also genügend Grauwertstruktur vorhanden ist, d.h. kein Grauwertplateau vorliegt. Mittels differentialgeometrischer Modellierung des Grauwertverlaufs ([16]) ist es bei der Betrachtung von Grauwertextrema und Grauwertecken jedoch möglich, den vollen optischen Fluß zu bestimmen, obwohl der Gradient verschwindet und so der Normalfluß nicht bestimmbar ist. Da es genügend Grauwertstruktur gibt, so daß die Ableitung zweiter Ordnung nicht verschwindet, kann über ihn eine Lösung für den vollen optischen Flusses errechnet werden.

Basierend auf der Kontinuitätsgleichung (2.2) läßt sich eine quadratische Fehlerfunktion definieren:

$$\mathrm{err}(\boldsymbol{v}) = (\nabla_s \boldsymbol{g}^T \boldsymbol{v} + g_t)^2. \tag{2.6}$$

Man kann daraus mittels linearer Ausgleichsrechnung eine Berechnungsformel für den optischen Fluß angeben. Dazu setzt man die Ableitung der Fehlerfunktion (2.6) gleich null.

$$\tfrac{1}{2}\nabla_{\boldsymbol{v}}\, \mathrm{err}(\boldsymbol{v}) = \boldsymbol{M}\boldsymbol{v} + \boldsymbol{b} = \boldsymbol{0}, \tag{2.7}$$

mit

$$\boldsymbol{M} = \nabla_s \boldsymbol{g} \nabla_s \boldsymbol{g}^T = \begin{pmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{pmatrix}, \tag{2.8}$$

$$\boldsymbol{b} = \nabla_s \boldsymbol{g} g_t = \begin{pmatrix} g_x g_t \\ g_y g_t \end{pmatrix}. \tag{2.9}$$

Natürlich macht sich auch hier das Aperturproblem bemerkbar, da diese Gleichung wegen der Simgularität der Matrix $\boldsymbol{M}$ stets unterbestimmt ist. Man kann daher keine Bestimmungsgleichung für $\boldsymbol{v}$ angeben.

Eine zusätzliche Glattheitsbedingung an das Flußfeld löst das Aperturproblem. Man nimmt an, daß der optische Fluß an allen Bildpunkten innerhalb einer gewissen Umgebung des betrachteten Punktes konstant ist und die Kontinuitätsgleichung (2.2) an jedem dieser Bildpunkte $(x_i, y_i, t), i \in \{1, \ldots, N\}$, erfüllt ist:

$$\forall i \quad \nabla_s \boldsymbol{g}^T(x_i, y_i, t)\boldsymbol{v} + g_t(x_i, y_i, t) = 0.$$

Um eine genaue Lösung des optischen Flusses angeben zu können, braucht man ebenso viele Gleichungen wie Unbekannte. Hat man mehr Gleichungen als Unbekannte, ist das Gleichungssystem überbestimmt, und die Unbekannten können nur angenähert werden.

Das exakt lösbare Gleichungssystem

Das exakt lösbare Gleichungssystem erhält man, wenn man nur einen zusätzlichen Punkt betrachtet ($N = 2$):

$$\begin{pmatrix} g_{x_1} & g_{y_1} \\ g_{x_2} & g_{y_2} \end{pmatrix} \boldsymbol{v} = - \begin{pmatrix} g_{t_1} \\ g_{t_2} \end{pmatrix}.$$

Wenn die Determinante der Matrix aus den örtlichen Ableitungen ungleich null ist, erhält man die Lösung dieses Gleichungssystems durch Invertieren der Matrix:

$$\boldsymbol{v} = - \begin{pmatrix} g_{x_1} & g_{y_1} \\ g_{x_2} & g_{y_2} \end{pmatrix}^{-1} \begin{pmatrix} g_{t_1} \\ g_{t_2} \end{pmatrix}.$$

Die Bedingung dafür, daß die Determinante ungleich null ist, lautet $g_{x_1} g_{y_2} \neq g_{y_1} g_{x_2}$. Das bedeutet, daß weder der Betrag des Gradienten verschwinden darf (Grauwertplateau), noch dürfen die Richtungen der Gradienten an den beiden Punkten übereinstimmen (Grauwertkante).

Das überbestimmte Gleichungssystem

Im Falle des exakt lösbaren Gleichungssystems ist keine zuverlässige Aussage über den Fehler möglich, was bei verrauschten Bildern von Nachteil ist. Damit man eine Aussage über den Fehler machen kann, betrachtet man einen größeren Bereich von Bildpunkten ($N > 2$).
Der Fehler kann dann analog zu (2.6) formuliert werden:

$$\operatorname{err}(\boldsymbol{v}) = \sum_{i}^{N} [\nabla_s \boldsymbol{g}^T(x_i, y_i, t)\boldsymbol{v} + g_t(x_i, y_i, t)]^2. \tag{2.10}$$

Das durch die betrachteten N Bildpunkte gegebene überbestimmte Gleichungssystem lautet:

$$\begin{pmatrix} g_{x_1} & g_{y_1} \\ \vdots & \vdots \\ g_{x_N} & g_{y_N} \end{pmatrix} \boldsymbol{v} = \begin{pmatrix} -g_{t_1} \\ \vdots \\ -g_{t_N} \end{pmatrix}. \tag{2.11}$$

Die Lösung dieses Gleichungssystems erhält man über lineare Ausgleichsrechnung. Dazu muß der Fehler (2.10) nach $\boldsymbol{v}$ minimiert werden. Es gibt zwei Möglichkeiten, diese Fehlerminimierung zu erreichen. Bezeichnet man die Matrix, deren Elemente die örtlichen Ableitungen sind, mit $\boldsymbol{G}$ und den Vektor der zeitlichen Ableitungen mit $\boldsymbol{b}_t$, so läßt sich das Gleichungssystem (2.11) folgendermaßen schreiben:

$$\boldsymbol{G}\boldsymbol{v} = \boldsymbol{b}_t.$$

Die Lösung der Minimierung ist dann durch den Ausdruck

$$\boldsymbol{v} = \boldsymbol{G}^{+}\boldsymbol{b}_t \tag{2.12}$$

gegeben, wobei $\boldsymbol{G}^{+}$ die Pseudoinverse [14] der Matrix $\boldsymbol{G}$ ist. Falls die Matrix $\boldsymbol{G}$ nicht singulär ist, gilt für die Pseudoinverse

$$\boldsymbol{G}^{+} = (\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T. \tag{2.13}$$

Wie oben läßt sich das Ergebnis auch durch Nullsetzen der Ableitung der Fehlerfunktion berechnen. Die Ableitung von (2.10) bzgl. $\boldsymbol{v}$ ergibt:

$$\nabla_{\boldsymbol{v}}\, \mathrm{err}(\boldsymbol{v}) = \sum_{i=1}^{N}(\boldsymbol{M}_i\boldsymbol{v} + \boldsymbol{b}_i) \tag{2.14}$$

mit

$$\boldsymbol{M}_i = \boldsymbol{M}(x_i, y_i, t) \quad \text{und} \quad \boldsymbol{b}_i = \boldsymbol{b}(x_i, y_i, t).$$

Daraus erhält man die Schätzung:

$$\boldsymbol{v} = -\left(\sum_{i=1}^{N}\boldsymbol{M}_i\right)^{-1}\left(\sum_{i=1}^{N}\boldsymbol{b}_i\right). \tag{2.15}$$

Die Lösungen (2.12) und (2.15) unterscheiden sich lediglich in ihrer äußeren Form, während es keine quantitativen Unterschiede gibt. Es gelten die Gleichheiten:

$$\begin{aligned} (\boldsymbol{G}^T\boldsymbol{G})^{-1} &= \left(\sum_{i=1}^{N}\boldsymbol{M}_i\right)^{-1}, \\ \boldsymbol{G}^T\boldsymbol{b}_t &= -\sum_{i=1}^{N}\boldsymbol{b}_i. \end{aligned}$$

Zu beachten ist, daß die Matrizen $\boldsymbol{G}^T\boldsymbol{G}$ in Gleichung (2.13) bzw. $\boldsymbol{M}$ in (2.15) immer noch singulär werden können, z.B. beim Auftreten von Grauwerttableaus.

Um die Forderung nach der Konstanz des optischen Flusses in der Region, welche die N Punkte umfaßt, abzuschwächen, kann man eine Wichtungsfunktion in die Summe einfügen, die den Beitrag der Grauwerte vom Zentrum nach außen exponentiell abfallend gewichtet. Die Geschwindigkeit wird dann also aus geglätteten, quadratischen Funktionen der örtlichen und zeitlichen Ableitungen berechnet. Das führt dazu, daß der optische Fluß mit Subpixelgenauigkeit bestimmt werden kann.

Güte der Berechnung

Der Fehler, den man bei der Schätzung des optischen Flusses durch die Ungenauigkeit der Meßdaten erhält, wird auch mit Güte der Berechnung bezeichnet. Da die Meßgrößen

Momente erster Ordnung sind, kann man ihre Unsicherheit durch Momente zweiter Ordnung angeben. Man betrachtet also die *Kovarianz* σ^2 der Meßdaten, welche ein Maß für die Streuung der Meßgrößen $\boldsymbol{x}$ und $\boldsymbol{y}$ untereinander darstellt:

$$\sigma^2 = \mathrm{E}[(\boldsymbol{x} - \mathrm{E}[\boldsymbol{x}])(\boldsymbol{y} - \mathrm{E}[\boldsymbol{y}])].$$

Das Maß für die Streuung eines Meßwertes ergibt sich aus der obigen Formel mit $\boldsymbol{x} = \boldsymbol{y}$ und wird *Varianz* genannt. Die Unsicherheit in der Schätzung des optischen Flusses, die auf der Ungenauigkeit der Meßdaten beruht, wird dann durch die sogenannte *Kovarianzmatrix* angegeben, auf deren Diagonale die Varianzen stehen, während die Nichtdiagonalelemente von den Kovarianzen gebildet werden.

Die Matrizen $(\boldsymbol{G}^T\boldsymbol{G})^{-1}$ aus Gleichung (2.13) bzw. $(\sum_i \boldsymbol{M}_i)^{-1}$ aus Gleichung (2.15) sind proportional zur Kovarianzmatrix, also ein Maß für die Güte der Schätzung des optischen Flusses. Bemerkenswert ist, daß die Matrizen nicht von der zeitlichen Ableitung der Grauwertfunktion abhängen, was eine a priori Aussage über die Schätzung ermöglicht. Der Proportionalitätsfaktor ist gerade die Varianz σ^2 der Daten, die aus dem Fehlervektor (2.10) abgeschätzt werden kann, also ein Maß für die Güte des Modells:

$$\sigma^2 = \frac{1}{N-2}\,\mathrm{err}(\boldsymbol{v}).$$

Ausschlaggebend ist die Determinante der Matrix, da in ihr der Kreuzkorrelationsterm enthalten ist, der verschwindet, wenn die Gradienten gleichmäßig in alle Richtungen verteilt sind. Da dies gleichzeitig die Bedingung für eine Bestimmung des Verschiebungsvektorfeldes mit minimalem Fehler ist, reicht es, die Spur der Matrix als Gütemaß zu betrachten.

Affines Modell

Statt zu fordern, daß der optische Fluß in der betrachteten Region konstant ist, kann man davon ausgehen, daß er sich langsam, linear ändert. In dem Fall läßt sich der optische Fluß durch ein affines, parametrisches Modell darstellen:

$$\boldsymbol{v}_{\mathrm{aff}} = \begin{pmatrix} u_{\mathrm{aff}} \\ v_{\mathrm{aff}} \end{pmatrix} = \begin{pmatrix} u_0 + u_x\Delta x + u_y\Delta y \\ v_0 + v_x\Delta x + v_y\Delta y \end{pmatrix}. \tag{2.16}$$

Die Parameter u_0, v_0 sind Geschwindigkeiten, u_x, u_y, v_x, v_y partielle Ableitungen erster Ordnung, und $\Delta x, \Delta y$ geben den Abstand des Bildpunkts vom Bewegungszentrum des betrachteten Bereichs an. Der optische Fluß wird hier durch eine Überlagerung von Translation, Rotation, Dilatation und Scherung modelliert.

Die affine Modellierung ist nur gültig, wenn das Objekt vom Beobachter weit entfernt oder die betrachtete Bildregion sehr klein ist.

Die Kontinuitätsgleichung lautet analog (2.2) mit $\boldsymbol{v} = \boldsymbol{v}_{\mathrm{aff}}$. Es liegen in diesem Fall sechs statt zwei Unbekannte vor. Die Gleichung ist also ebenfalls unterbestimmt.

Zur Lösung betrachtet man wieder das durch die Punkte in einer ganzen Region gegebene Gleichungssystem:

$$\begin{pmatrix} g_{x_1} & g_{y_1} & g_{x_1}\Delta x_1 & g_{x_1}\Delta y_1 & g_{y_1}\Delta x_1 & g_{y_1}\Delta y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{x_N} & g_{y_N} & g_{x_N}\Delta x_N & g_{x_N}\Delta y_N & g_{y_N}\Delta x_N & g_{y_N}\Delta y_N \end{pmatrix} \begin{pmatrix} u_0 \\ v_0 \\ u_x \\ u_y \\ v_x \\ v_y \end{pmatrix} = \begin{pmatrix} -g_{t_1} \\ \vdots \\ -g_{t_N} \end{pmatrix}. \quad (2.17)$$

Für das exakte Gleichungssystem muß man in diesem Fall natürlich sechs Bildpunkte ($N = 6$) betrachten. Schreibt man das Gleichungssystem als $\hat{\boldsymbol{G}}\hat{\boldsymbol{v}}_{\text{aff}} = \boldsymbol{b}_t$, erhält man analog zu (2.12) die Lösung:

$$\hat{\boldsymbol{v}}_{\text{aff}} = \hat{\boldsymbol{G}}^{+}\boldsymbol{b}_t.$$

Setzt man $\boldsymbol{v} = \hat{\boldsymbol{v}}_{\text{aff}}$, so erhält man alternativ aus der Fehlerfunktion (2.10) auch hier wieder aus den Ableitungen bzgl. $\hat{\boldsymbol{v}}_{\text{aff}}$ die Lösung:

$$\hat{\boldsymbol{v}}_{\text{aff}} = \begin{pmatrix} u_0 \\ v_0 \\ u_x \\ u_y \\ v_x \\ v_y \end{pmatrix} = -\left(\sum_{i=1}^{N} \hat{\boldsymbol{M}}_i\right)^{-1} \left(\sum_i \hat{\boldsymbol{b}}_i\right) \quad (2.18)$$

mit

$$\sum_{i=1}^{N} \hat{\boldsymbol{M}}_i = \hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}} \quad \text{und} \quad \sum_{i=1}^{N} \hat{\boldsymbol{b}}_i = -\hat{\boldsymbol{G}}^T\boldsymbol{b}_t.$$

Der Fehler des Modells läßt sich erst nach der Schätzung durch (2.10) mit $\boldsymbol{v} = \hat{\boldsymbol{v}}_{\text{aff}}$ berechnen. Die Kondition der Meßdaten $\text{err}_{\text{aff}_2}$ kann man jedoch schon vorher an der Spur der Kovarianzmatrix, die bei der affinen Modellierung durch $(\sum_i \hat{\boldsymbol{M}}_i)^{-1}$ gegeben wird, oder alternativ am kleinsten Singulärwert der Messungsmatrix ablesen:

$$\text{err}_{\text{aff}_2} = \min \text{sing} \left(\begin{array}{cccccc|c} g_{x_1} & g_{y_1} & \Delta x g_{x_1} & \Delta y g_{x_1} & \Delta x g_{y_1} & \Delta y g_{y_1} & g_{t_1} \\ g_{x_2} & g_{y_2} & \Delta x g_{x_2} & \Delta y g_{x_2} & \Delta x g_{y_2} & \Delta y g_{y_2} & g_{t_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{x_N} & g_{y_N} & \Delta x g_{x_N} & \Delta y g_{x_N} & \Delta x g_{y_N} & \Delta y g_{y_N} & g_{t_N} \end{array}\right). \quad (2.19)$$

Weitere differentielle Verfahren sind in [6] beschrieben.

2.2 Frequenzbasierte Verfahren

Bei den frequenzbasierten Verfahren benötigt man eine längere Sequenz von Bildern. Man betrachtet daher den dreidimensionalen Raum der Bildfolgen, in dem eine allgemeine starre Bewegung eines Objekts durch eine Kurve und eine gleichförmige Bewegung durch eine geneigte Gerade beschrieben wird (Abb. 2.2). Die Tangente an die Kurve bzw. die Neigung der Geraden ist dann ein Maß für die Geschwindigkeit.

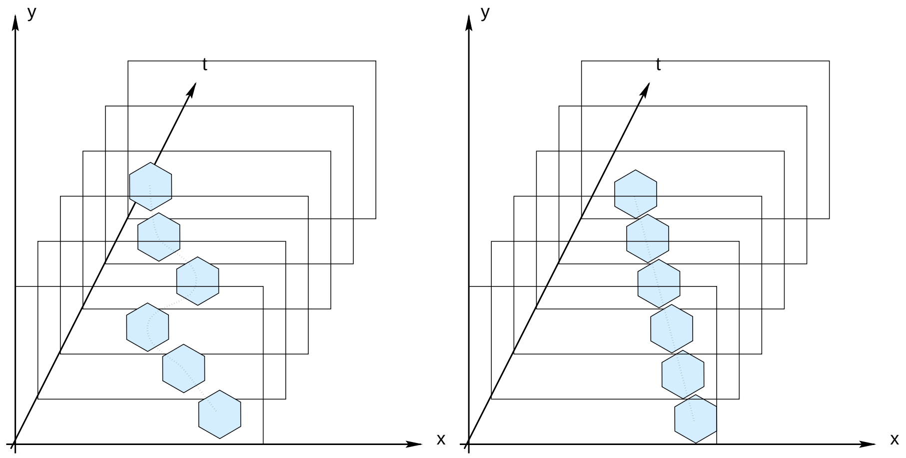


Abbildung 2.2: Darstellung von Bewegung im Bildfolgenraum: links: allgemeine Bewegung; rechts: gleichförmige Bewegung

Damit man die Bewegungsanalyse in dem von den örtlichen Frequenzen f_x, f_y und der Frequenz f_t aufgespannten Frequenzraum durchführen kann, muß man zunächst die Grauwertfunktion $\boldsymbol{g}(x, y, t)$ nach Fourier in den Frequenzraum transformieren. Zur Vereinfachung sei zunächst angenommen, daß sich das ganze Bild gleichmäßig mit der konstanten Geschwindigkeit $\boldsymbol{v} = (u, v)^T$ bewegt. Das Bild hat dann die Darstellung:

$$\boldsymbol{g}(x, y, t) = \boldsymbol{g}(x - ut, y - vt).$$

Aus dem Verschiebungssatz ergibt sich für die Fouriertransformierte:

$$\begin{aligned}
\boldsymbol{G}(f_x, f_y, f_t) &= \int_t \int_y \int_x \boldsymbol{g}(x, y, t)\, \mathrm{e}^{-\imath 2\pi(x f_x + y f_y + t f_t)}\, dx\, dy\, dt \\
&= \int_t \left\{ \int_y \int_x \boldsymbol{g}(x - ut, y - vt)\, \mathrm{e}^{-\imath 2\pi(x f_x + y f_y)}\, dx\, dy \right\} \mathrm{e}^{-\imath 2\pi t f_t}\, dt \\
&= \boldsymbol{G}(f_x, f_y) \int_t \mathrm{e}^{-\imath 2\pi(f_x u + f_y v + t f_t)}\, dt \\
&= \boldsymbol{G}(f_x, f_y) \delta(f_x u + f_y v + f_t). \qquad (2.20)
\end{aligned}$$

An der Fouriertransformierten kann man erkennen, daß im Frequenzraum nur eine Ebene besetzt ist, deren Ebenengleichung die Formulierung der OFCE (2.2) im Frequenzraum ist:

$$f_x u + f_y v + f_t = 0. \tag{2.21}$$

An der Gleichung läßt sich direkt ablesen, daß die Normale auf diese Ebene parallel zu dem Vektor $(u, v, 1)$ ist. Zur Berechnung des optischen Flusses muß man also die Neigung der Ebene bestimmen. Dies ist mit *geschwindigkeitsselektiven* Filtern möglich, zu denen die Gaborfilter gehören. Aus einer Kombination von mehreren Antworten einer Sammlung von Filtern wird die Neigung ermittelt. Eine Erklärung der Filter und der Bestimmung der Ebenenneigung findet man in [6] und in [16].

In dem Fall, daß keine konstante, globale Bildbewegung vorliegt, betrachtet man nur die Bewegung einer ortszeitlichen Region. Durch die Multiplikation mit einem Würfel im Bildfolgenraum entartet die Ebene im Frequenzraum zu einer flachen Wolke, was die Bestimmung der Neigung erschwert.

Auch bei den frequenzbasierten Verfahren macht sich das Aperturproblem bemerkbar. Ist als Struktur nur eine Grauwertkante vorhanden, entartet die Ebene zu einer Geraden, die sich durch Schneiden der Ebene $f_x u + f_y v + f_t = 0$ mit der Ebene, die parallel zur f_t-Achse und senkrecht zur Grauwertkante orientiert ist, ergibt.

2.3 Korrelationsverfahren

Stehen nur sehr kurze Bildfolgen zur Verfügung, so ist die numerische Differentiation bei den differentiellen Verfahren sehr ungenau. Auch bei den frequenzbasierten Verfahren braucht man für die dreidimensionale Fouriertransformation eine relativ große Anzahl von Bildern. Ein kleines Verhältnis von Signal zu Rauschen wirkt sich auf die Berechnung des optischen Flusses mit den beiden Verfahren ebenfalls schlecht aus. In diesen Fällen bieten sich korrelationsbasierte Zuordnungsverfahren an.

Korrelationsverfahren berechnen den Fluß, indem sie als Approximation der Geschwindigkeit die Verschiebung bestimmen, welche zur besten Übereinstimmung eines Bildausschnitts im ersten Bild mit einem entsprechenden Bildausschnitt im zweiten Bild führt. Sie berechnen also das Verschiebungsvektorfeld, das nur dann als Näherung des optischen Flusses gelten kann, wenn die Bildabtastrate hoch genug und damit die Zeitdifferenz zwischen zwei Bildern klein genug ist.

Als Indikator für die beste Übereinstimmung benutzt man oft ein Ähnlichkeitsmaß, das genau dann sein Maximum annimmt, wenn maximale Übereinstimmung vorliegt. Der Korrelationskoeffizient zwischen zwei eindimensionalen Funktionen beispielsweise ist definiert als

$$\int_D \boldsymbol{f}(x + \delta x)\boldsymbol{h}(x)dx.$$

Dasjenige δx, welches dieses Integral maximiert, ist die gesuchte Verschiebung, falls $\boldsymbol{f}(x + \delta x) = \boldsymbol{h}(x)$ gilt.

Für den Bildpunkt (x, y) mit Nachbarschaft D zum Zeitpunkt t_1 ist die Verschiebung $(u\Delta t, v\Delta t)$ der Region $\boldsymbol{g}(x, y, t_1)_{(x,y)\in D}$ zu finden, die zur besten Überlappung dieser Region mit einer Region gleicher Abmessung im folgenden Bild $\boldsymbol{g}(x+u\Delta t, y+v\Delta t, t_2)_{(x,y)\in D}$ führt. Eine Übersicht über Verfahren, die dieses Maß benutzen, findet man in [6].

Statt der Maximierung eines Ähnlichkeitsmaßes führt auch die Minimierung der Summe der quadrierten Abweichungen der Grauwerte (SSD (Sum of Squared Differences)) auf die gesuchte Verschiebung, wenn man davon ausgeht, daß der optische Fluß in der betrachteten Umgebung $D = \{(x+i, y+j) | i, j \in \{-N, \ldots, N\}\}$ konstant ist. Die SSD hat die Form:

$$\boldsymbol{S}(u, v) = \sum_{i=-N}^{N} \sum_{j=-N}^{N} [\boldsymbol{g}(x+i-u\Delta t, y+j-v\Delta t, t_1) - \boldsymbol{g}(x+i, y+j, t_2)]^2. \tag{2.22}$$

Der optische Fluß $\boldsymbol{v}$ ist dann also gegeben durch:

$$\boldsymbol{v} = \arg\min_{(u,v)} \boldsymbol{S}(u, v). \tag{2.23}$$

Zur Abschwächung der Glattheitsbedingung kann man auch in (2.22) wieder eine Wichtungsfunktion einfügen, die den Anteil der Bildpunkte radial exponentiell abfallend gewichtet.

Dieser Korrelationsansatz ist äquivalent zu den differentiellen Verfahren. Der Ausdruck in eckigen Klammern in (2.22) ergibt nach einer Taylorreihenentwicklung bzgl. Δt analog zu (2.4):

$$-\nabla_s \boldsymbol{g}(x+i, y+j, t_1)^T \begin{pmatrix} u \\ v \end{pmatrix} \Delta t - g_t(x+i, y+j, t_1)\Delta t + \mathcal{O}(\Delta t^2).$$

Das Minimieren des Quadrats ist daher äquivalent zum Minimieren der Fehlerfunktion (2.10).

Zu bemerken ist, daß das Minimum nicht immer eindeutig ist. Um das richtige lokale Minimum zu finden, betrachtet man die Krümmungen der durch die SSD definierten Oberfläche zweiter Ordnung. Sie stellen ein Maß für die Unsicherheit des optischen Flusses dar (siehe dazu [1]). Falls eine der Hauptkrümmungen ganz verschwindet, bedeutet dies, daß nur eine Komponente des optischen Flusses bestimmbar ist, was durch das Aperturproblem hervorgerufen wird. Vergrößert man in diesem Fall das Operatorfenster D, kann man, wie schon in Abb. 2.1 dargestellt, eventuell ein eindeutiges Minimum finden.

Außerdem kann man über diese Fläche die Schätzung des optischen Flusses verfeinern. Die Lösung aus (2.23) ist nur von Pixelgenauigkeit. Subpixelgenauigkeit erreicht man durch Anpassen der SSD-Fläche in unmittelbarer Umgebung des gefundenen Minimums. Eine neue Bestimmung des Minimums liefert die subpixelgenaue Lösung des optischen Flusses.

Ohne Vorwissen über die Länge der Bewegung kann sich der Bereich für die Suche des Minimums über das ganze Bild erstrecken. Die halbe Wellenlänge der Fouriertransformierten des Signals stellt in jedem Fall eine obere Schranke der Verschiebung und somit der

Suchregion dar. Doch auch dann ist der erforderte Rechenaufwand noch hoch. Um diesen zu verringern und um Zuordnungsfehler zu verhindern, die durch den Wagenradeffekt verursacht werden, führt man die Berechnung hierarchisch durch. Das heißt, man bildet eine Pyramide aus verschiedenen Auflösungsstufen des Bildes, sucht auf der gröbsten Auflösungsstufe das Minimum der SSD für jeden Bildpunkt und benutzt dieses Ergebnis auf der nächst feineren Stufe als Startwert. So läßt sich auf dieser Stufe der Suchraum weiter einschränken.

2.4 Zeitliches Aliasing

Wie schon in Kapitel 2 erwähnt, gibt es Mehrdeutigkeiten und dadurch bedingte Fehler in der Bewegungsanalyse, die nicht von der eigentlichen Grauwertverteilung abhängen. Durch schnelle Bewegungen des beobachteten Objekts relativ zur Bildrate der Kamera, kommt es zu zeitlichen Aliasing-Effekten. Der Mensch kann diese leicht erkennen, wie z.B. in dem klassischen Beispiel des Kutschenrades in einem Westernfilm: das Rad scheint sich falsch herum zu drehen.

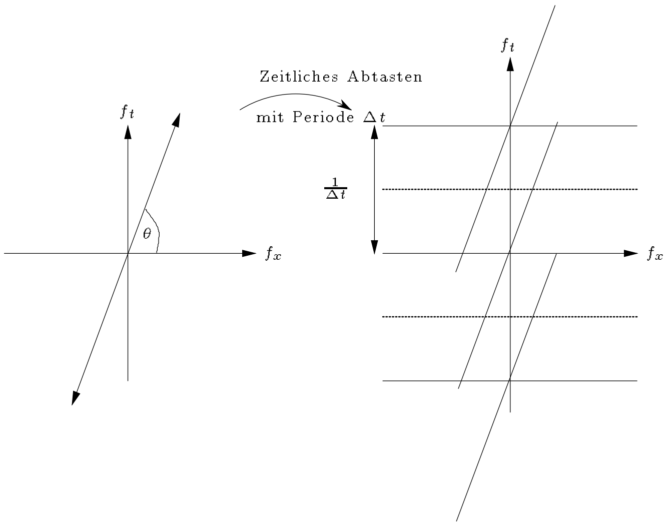


Abbildung 2.3: Das zeitliche Aliasingproblem im Frequenzraum: links: Fouriertransformierte eines eindimensionalen Signals mit konstanter Geschwindigkeit $u = -\tan\theta$; rechts: Die Abtastung erzeugt im Frequenzraum durch die Faltung mit dem Dirac-Kamm Kopien des Signals im Abstand der Abtastfrequenz $\nu_t = \frac{1}{\Delta t}$. Die punktierte Linie zeigt die nach dem Abtasttheorem einzuhaltende obere Grenzfrequenz $\nu_g = \frac{1}{2\Delta t}$ an.

Der Effekt des Aliasing bei Filtern ist im Frequenzraum leicht zu veranschaulichen. Betrachtet man ein eindimensionales Signal, das sich mit einer konstanten Geschwindigkeit bewegt, liegt sein Energiespektrum auf einer Linie durch den Ursprung. Es sei angenommen, daß die örtliche Abtastrate so gewählt ist, daß räumliches Aliasing vermieden wird. Die zeitliche Abtastung bewirkt eine Kopie des Spektrums an den Stellen $\frac{1}{\Delta t}$ zeitlicher Frequenz, wobei Δt die Zeit zwischen zwei Bildern ist (siehe Abb. 2.3). Das liegt daran, daß die Abtastung im Orts-Zeit-Raum eine Multiplikation mit einer von der Zeit abhängigen Kammfunktion aus Dirac-Impulsen,

$$\mathbf{III}(t) = \sum_{n \in Z} \delta(t - n\Delta t),$$

der sogenannten Schah-Funktion, ist. Im Frequenzraum wird die Abtastung zu einer Faltung des Spektrums mit der Fouriertransformierten der Schah-Funktion, die wieder eine Schah-Funktion ist:

$$\boldsymbol{G}(f_x, f_y, f_t) = \boldsymbol{G}(f_x, f_y) \star \sum_{n \in Z} \delta(f_t - \frac{n}{\Delta t}). \tag{2.24}$$

Wenn nun $f_{x\,\max}$ und $f_{y\,\max}$ die maximalen örtlichen Frequenzen des analogen Signals sind, so muß nach dem Abtasttheorem für eine fehlerfreie Repräsentation gelten:

$$u f_{x\,\max} < \tfrac{1}{2\Delta t} \quad \wedge \quad v f_{y\,\max} < \tfrac{1}{2\Delta t} \quad \Rightarrow \quad u < \tfrac{\lambda_{x\,\min}}{2} \quad \wedge \quad v < \tfrac{\lambda_{y\,\min}}{2}. \tag{2.25}$$

Gilt dies nicht, kommt es zu zeitlichen Aliasing-Effekten.

Bei der Berechnung des optischen Flusses mittels differentieller Verfahren führt das zeitliche Aliasing zu falschen Ergebnissen, da das Differenzieren durch Multiplikation mit einem Filter im Frequenzraum erreicht wird und die Anteile der kopierten Signale innerhalb des Intervalls $[-\frac{1}{2\Delta t}, \frac{1}{2\Delta t}]$ zu Fehlern in den zeitlichen Ableitungen führen. Auch bei den anderen Verfahren kommt es durch zeitliches Aliasing zu falschen Berechnungen.

Eine wichtige Beobachtung ist in diesem Zusammenhang, daß das zeitliche Aliasing nur Auswirkungen auf hohe örtliche Frequenzen hat. Naheliegend ist in diesem Fall eine Vorverarbeitung mit einem niederfrequenten Filter, der höhere Frequenzen vernachlässigt. Da so ein Filter eine große örtliche Ausdehnung hat, berechnet er im Prinzip die mittlere Geschwindigkeit einer ganzen Bildregion.

In dem Fall, daß es nur eine globale Bewegung gibt oder die Bewegung sehr langsam ist (keine örtlich hochfrequenten Anteile), liefert diese Filterung den gesuchten optischen Fluß. Meistens ist dies jedoch nicht der Fall, was eine Verfeinerung der Ergebnisse durch die Berücksichtigung von Bändern höherer Frequenz verlangt. Durch die Anwendung von örtlich kleineren Filtern kann man so eine genauere Berechnung erzielen. Man benutzt also eine Sammlung von Filtern verschiedener Größe, um beginnend mit dem größten Filter die Berechnung des optischen Flusses von einer Filtergröße zur nächst kleineren zu verfeinern.

2.5 Hierarchische Berechnungsverfahren

Statt für die Berechnung des optischen Flusses eine Kollektion von Ableitungs- und Glättungsfiltern verschiedener Größen zu berechnen, damit die Aliasing-Effekte nicht das Ergebnis beeinträchtigen, ist es möglich, das Verfahren mit einer Pyramide von Bildern und Filtern einer einzigen Größe durchzuführen. Die Bildpyramide wird durch rekursive Tiefpaßfilterung und Unterabtastung konstruiert, so daß Bilder verschiedener Größe Bildern verschiedener Abtastung entsprechen. Eine Gaußpyramide erfüllt die Anforderungen für das hierarchische Berechnungsverfahren.

Hierarchische Berechnungsansätze gibt es in allen drei Arten von Verfahren zur Flußberechnung. Im weiteren wird jedoch nur die differentielle Methode vertieft, da die in dieser Arbeit betrachteten Verfahren mit dieser Methode arbeiten. Ein kurzer Einblick in den hierarchischen Berechnungsansatz bei Korrelationsverfahren wurde oben bereits gegeben. In [6] werden weitere hierarchische Ansätze behandelt.

Man beginnt damit, auf der gröbsten (also kleinsten) Ebene der Pyramide den optischen Fluß mittels eines der differentiellen Verfahren zu ermitteln. Das Ergebnis dieser Berechnung stellt den Ausgangswert für die nächst feinere Stufe dar. Statt jedoch das erlangte Ergebnis nur als a priori Schätzung zu betrachten, kann man es zum Verfolgen der Bewegung benutzen.

Der Mensch verfolgt beim Betrachten einer Bewegung das bewegte Objekt mit seinen Augen, so daß die relative Bewegung kleiner wird. Bei der Berechnung des optischen Flusses kann man das Verfolgen der Bewegung dadurch simulieren, daß man die Anfangsschätzung dazu benutzt, diese Bewegung zurückzurechnen. Bevor man also den Fluß auf der nächsten Stufe bestimmt, kann man die auf das feinere Gitter interpolierte, grob berechnete Bewegungsschätzung dazu benutzen, das Bild so zu verschieben, zu *warpen*, daß diese Bewegung nicht mehr vorliegt. Die so auf dieser Auflösungsstufe verbleibende Bewegung ist nur noch sehr klein. Sie stellt eine Korrektur der vorausgegangenen Berechnung dar. Das Ergebnis der nun folgenden Flußberechnung ergibt kombiniert mit dem vorherigen Ergebnis eine genauere Bestimmung des optischen Flusses.

Das Warpen des Bildes läßt sich dabei wie folgt berechnen:

$$\mathcal{W}\{\boldsymbol{g}, \boldsymbol{v}\}(x, y, t + \Delta t) = g(x - u\Delta t, y - v\Delta t, t + \Delta t),$$

wobei $\boldsymbol{v}$ die auf das entsprechende Gitter interpolierte Version des grob berechneten Flusses ist.

Die Verschiebung braucht dabei nur in einem die zeitliche Ausdehnung des Ableitungsoperators umfassenden Bereich Δt durchgeführt zu werden. Dieses Verfahren wird für immer feinere Stufen wiederholt, bis man die unterste Ebene der Pyramide erreicht hat.

Dieses Vorgehen birgt eine Gefahr in sich: Falls die Schätzung des optischen Flusses in grober Auflösung falsch ist, kann sie nicht durch Berechnungen in feineren Auflösungen korrigiert werden. Deshalb ist es wichtig, daß man für die Berechnung in jeder Stufe ein Gütemaß, z.B. die Kovarianzmatrix, angibt, damit bei der Fusion der Berechnungen von zwei Stufen eine der Zuverlässigkeit der Schätzung entsprechende Gewichtung der Anteile vorgenommen werden kann.

Der Vorgang des Warpens läßt noch einige weitere Fragen offen. Zur Berechnung der zeitlichen Ableitung für ein Bild der Bildfolge reicht es, wie schon gesagt, aus, die Bilder in der zeitlichen Umgebung, die der Größe der Maske entspricht, zu warpen. Die Verschiebung wird hierbei durch den interpolierten optischen Fluß der entsprechend gröberen Auflösung gegeben. Diese Bedingung reicht jedoch noch nicht aus, um das Warpen durchzuführen. Es ist einerseits möglich, jedes Bild mit dem dafür berechneten Fluß zu warpen, andererseits kann man auch jedes Bild mit einem Vielfachen des optischen Flusses des betrachteten Bildes warpen. Ersteres geht jedoch mit einem großen Speicherbedarf einher, so daß die Glattheitsbedingung für die Geschwindigkeit bei kleinen zeitlichen Masken eine sinnvolle Alternative darstellt.

Ferner ist zu bemerken, daß die Ergebnisse des optischen Flusses, die man zum Warpen verwendet, meistens nicht ganzzahlig sind, was bedeutet, daß man die Grauwerte an den gesuchten Stellen interpolieren muß.

Kapitel 3

Schätzungstheoretische Berechnung

Die hierarchische Berechnung des optischen Flusses ist wegen des Problems der Fusion von Daten aus verschiedenen Auflösungsebenen und der dafür notwendigen Formulierung eines Gütemaßes auf stochastische Schätzungsmethoden gestützt. Auch die Berücksichtigung des Rauschens in der BCCE erfordert Methoden der Stochastik.

Bei der Fusion der Ergebnisse für den optischen Fluß, die auf den einzelnen Hierarchie-Ebenen berechnet werden, ist die Güte der Berechnungen zu beachten. Ein zuverlässiges Gütemaß läßt sich jedoch nur formulieren, wenn man ein stochastisches Modell zugrunde legt. Auch die Komposition der Daten selbst beruht bei Berücksichtigung der Güte der Daten auf stochastischen Strategien.

Zudem sind auf jeder Hierarchie-Ebene die Meßdaten, die für die Berechnung des optischen Flusses benutzt werden, gleichermaßen verrauscht. Durch die stochastische Modellierung des Rauschens können die Auswirkungen des Rauschens auf das Ergebnis der Berechnung berücksichtigt werden, und die Berechnungsmethode kann optimiert werden. Das auf einer Pyramidenstufe berechnete Ergebnis für den optischen Fluß wird auf der nächst gröberen Stufe benutzt, um über den Warpvorgang diese bereits bestimmte Bewegung von der Gesamtbewegung abzuziehen. Der Vorgang wird auf der Eingabesequenz vor der Berechnung des optischen Flusses auf dieser Stufe durchgeführt. Es ist daher wichtig, daß die Berechnungen trotz Rauschens gute Ergebnisse liefern.

Da das Problem der schnellen Bewegungen in dieser Arbeit durch die hierarchische Berechnung des optischen Flusses bewältigt wird, ist die Betrachtung der schätzungstheoretischen Methoden daher von großer Bedeutung. Die Wahl der Berechnungsmethode des optischen Flusses ist ausschlaggebend für das Ergebnis der Verfolgung.

In diesem Kapitel werden zunächst die zum Verständnis notwendigen stochastischen Grundlagen der Schätzungstheorie erklärt und mit Beispielen illustriert. Die Anwendungen der Schätzungsmethoden zur Berücksichtigung des Rauschens bei der Berechnung des optischen Flusses und zur Definition eines Gütemaßes folgen im zweiten Teil des Kapitels. Dort werden ebenfalls verschiedene Strategien zur Kombination der Daten aus verschiedenen Hierarchie-Ebenen erklärt. Am Ende des Kapitels werden nach einem theoretischen Vergleich die experimentellen Ergebnisse der verschiedenen Berechnungsverfahren gegenübergestellt.

3.1 Schätzungsgrundlagen

Bei der Schätzung von Parametern, die einen dynamischen stochastischen Prozeß beschreiben, stößt man auf das Problem, daß die zur Verfügung stehenden Informationen (Systemmodelle, Messungen) ungenau sind. Damit man trotzdem eine zuverlässige Aussage über die gesuchten Parameter machen kann, benutzt man zu deren Schätzung stochastische Verfahren.
Das Problem der Parameterschätzung hat folgende formale Beschreibung. Gegeben sind die Meßdaten $\boldsymbol{z}(j)$:

$$\boldsymbol{z}(j) = \boldsymbol{h}(j, \boldsymbol{x}, \boldsymbol{n}(j)),$$

wobei $\boldsymbol{x}$ die Unbekannte ist, j die Nummer der Messung angibt und $\boldsymbol{n}(j)$ das Meßrauschen darstellt.
Gesucht ist eine Schätzung für $\boldsymbol{x}$ nach k Messungen:

$$\hat{\boldsymbol{x}}(k) = \hat{\boldsymbol{x}}(k, \mathcal{Z}^k),$$

wobei $\mathcal{Z}^k$ die Menge aller vorangegangenen Messungen ist, $\mathcal{Z}^k = \{\boldsymbol{z}(j), j = 1, \dots, k\}$. Im folgenden bezeichne das Dach, "$\hat{}$", das Ergebnis einer Schätzung.

Abhängig von der Kenntnis der a priori Dichtefunktion der Unbekannten und je nach Optimalitätskriterium unterscheidet man verschiedene Schätzverfahren: die *Maximum Likelihood*–Methode, Methode des *kleinsten mittleren quadratischen Fehlers*, die *Maximum A Posteriori*–Methode und das *Kalman–Filter*.

Die erwähnten Verfahren werden im folgenden kurz vorgestellt. Ausführlichere Erklärungen der einzelnen Strategien sind in [3], [32] und [8] zu finden.

3.1.1 Maximum Likelihood

Dieser Ansatz beruht auf der Forderung, daß alle Werte für die unbekannten Parameter mit gleicher Wahrscheinlichkeit auftreten. Es wird demnach keine a priori Information über die Wahrscheinlichkeitsdichte benutzt. Die ML–Methode erzielt eine Schätzung, die für unendlich viele Messungen gegen den wahren Wert des Parameters konvergiert.
Als Maß für die Gleichheit von Parameter und Messung dient hier die Dichtefunktion $p(\mathcal{Z}^k|\boldsymbol{x})$, die angibt, wie groß die Wahrscheinlichkeit für das Auftreten der Messungen $\mathcal{Z}^k$ ist, falls die Unbekannte den Wert $\boldsymbol{x}$ hat.

Als Ergebnis liefert die ML–Strategie den Parametervektor $\hat{\boldsymbol{x}}^{\mathrm{ML}}$, welcher die bedingte Wahrscheinlichkeit $p(\mathcal{Z}^k|\boldsymbol{x})$ maximiert:

$$\hat{\boldsymbol{x}}^{\mathrm{ML}}(k) = \arg\max_{\boldsymbol{x}} p(\mathcal{Z}^k|\boldsymbol{x}). \tag{3.1}$$

Man kann sagen, daß die ML–Schätzung denjenigen Wert für $\boldsymbol{x}$ erzielt, dessen Messung am wahrscheinlichsten ist.

Setzt man voraus, daß das Meßrauschen $\boldsymbol{n}$ additiv, $\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{n}$, und mit Mittelwert null und Varianz σ^2 normalverteilt ist, $\boldsymbol{n} \sim N(0, \sigma_v^2)$, ist es ausreichend, den Logarithmus der Wahrscheinlichkeitsfunktion zu betrachten. Um $p(\mathcal{Z}^k|\boldsymbol{x})$ zu maximieren, wählt man daher dasjenige $\boldsymbol{x}$, das $\ln p(\mathcal{Z}^k|\boldsymbol{x})$ maximiert.

Eine notwendige Bedingung für $\hat{\boldsymbol{x}}^{\mathrm{ML}}(k)$ ist

$$\forall i \quad \frac{\partial}{\partial \boldsymbol{x}_i} \ln p(\mathcal{Z}^k|\hat{\boldsymbol{x}}^{\mathrm{ML}}(k)) = 0. \tag{3.2}$$

Beispiel:

Betrachten wir als Beispiel die Dichtefunktion für M unabhängige Messungen einer gaußverteilten Größe mit unbekanntem Mittelwert μ und unbekannter Varianz σ_v^2, die für alle Messungen konstant sein soll:

$$p(\mathcal{Z}^M|x) = p(\mathcal{Z}^M|\mu, \sigma_v^2) = (2\pi\sigma_v^2)^{-\frac{M}{2}} \exp\{-\frac{1}{2\sigma_v^2}\sum_{i=1}^{M}(z_i - \mu)^2\}.$$

Der Logarithmus der Dichtefunktion ist dann:

$$\ln p(\mathcal{Z}^M|\mu, \sigma_v^2) = -\frac{M}{2}\ln 2\pi\sigma_v^2 - \frac{1}{2\sigma_v^2}\sum_{i=1}^{M}(z_i - \mu)^2.$$

Die zu lösende Wahrscheinlichkeitsgleichung lautet in diesem Fall:

$$\frac{\partial}{\partial \mu}\ln p(\mathcal{Z}^M|\hat{\mu}_{\mathrm{ML}}, \hat{\sigma}_{\mathrm{ML}}^2) = \frac{1}{\hat{\sigma}_{\mathrm{ML}}^2}\sum_{i=1}^{M}(z_i - \hat{\mu}_{\mathrm{ML}}) = 0.$$

Daraus folgt als Schätzwert der einfache Mittelwert

$$\hat{\mu}_{\mathrm{ML}} = \frac{1}{M}\sum_{i=1}^{M} z_i. \tag{3.3}$$

Die Schätzung der Fehlervarianz erhält man aus folgender Bedingung:

$$\frac{\partial}{\partial \sigma^2}\ln p(\mathcal{Z}^M|\hat{\mu}_{\mathrm{ML}}, \hat{\sigma}_{\mathrm{ML}}^2) = -\frac{M}{2}\frac{1}{\hat{\sigma}_{\mathrm{ML}}^2} + \frac{1}{2}\frac{1}{\hat{\sigma}_{\mathrm{ML}}^4}\sum_{i=1}^{M}(z_i - \hat{\mu}_{\mathrm{ML}})^2 = 0.$$

Löst man diese Gleichung nach $\hat{\sigma}_{\mathrm{ML}}^2$ auf, so findet man

$$\hat{\sigma}_{\mathrm{ML}}^2 = \frac{1}{M}\sum_{i=1}^{M}(z_i - \hat{\mu}_{\mathrm{ML}})^2. \tag{3.4}$$

Es ist zu bemerken, daß die Varianz σ^2 keinen Einfluß auf das Ergebnis hat, weil sie für alle Messungen als konstant angenommen wird.

3.1.2 Maximum A Posteriori

Um mittels der MAP–Methode eine Schätzung für den Parameter $\boldsymbol{x}$ zu bestimmen, betrachtet man die bedingte a posteriori Wahrscheinlichkeit $p(\boldsymbol{x}|\mathcal{Z}^k)$. Diese läßt sich nach Anwendung der Bayes–Regel folgendermaßen berechnen:

$$p(\boldsymbol{x}|\mathcal{Z}^k) = \frac{p(\mathcal{Z}^k|\boldsymbol{x})p(\boldsymbol{x})}{p(\mathcal{Z}^k)}. \tag{3.5}$$

Man muß also zur Lösung Kenntnisse über die a priori Wahrscheinlichkeit $p(\boldsymbol{x})$ haben.

Das Ergebnis der MAP–Schätzung ist der Parametervektor, der die a posteriori Wahrscheinlichkeit (3.5) maximiert. Da der Nenner unabhängig von $\boldsymbol{x}$ ist, ist er für die Maximierung der Wahrscheinlichkeit irrelevant. Das Ergebnis ergibt sich demnach zu:

$$\hat{\boldsymbol{x}}^{\text{MAP}}(k) = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x}|\mathcal{Z}^k) = \arg\max_{\boldsymbol{x}}[p(\mathcal{Z}^k|\boldsymbol{x})p(\boldsymbol{x})]. \tag{3.6}$$

Unter der Annahme, daß die Wahrscheinlichkeitsdichte $p(\boldsymbol{x})$ gleichmäßig über einen Bereich von Werten für $\boldsymbol{x}$ verteilt ist, in dem $p(\mathcal{Z}^k|\boldsymbol{x})$ größer als null ist, ist das Ergebnis der MAP–Schätzung dasselbe wie das der ML–Schätzung (3.3).

Setzt man auch hier wieder voraus, daß das Meßrauschen additiv und mit Mittelwert null und Varianz $\boldsymbol{R}_k$ normalverteilt ist, ist es nützlich, auch von einem normalverteilten Fehler des prädizierten Wertes auszugehen. Wir nehmen ferner an, daß die Meßfunktion linear ist, $\boldsymbol{z}_k = \boldsymbol{h}_k(\boldsymbol{x}) + \boldsymbol{n}$, die Varianz der a priori Dichtefunktion durch $\boldsymbol{P}_k$ gegeben ist und die a priori Schätzung den Wert $\hat{\boldsymbol{x}}^-$ hat. Dann läßt sich die a posteriori Dichtefunktion durch eine Exponentialfunktion ausdrücken:

$$p(\boldsymbol{x}|\mathcal{Z}^k) = \frac{1}{c_k}\exp\{-\Phi(\boldsymbol{x})\}$$

mit

$$\Phi(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{z}_k - \boldsymbol{h}_k(\boldsymbol{x}))^T \boldsymbol{R}_k^{-1}(\boldsymbol{z}_k - \boldsymbol{h}_k(\boldsymbol{x})) + \frac{1}{2}(\boldsymbol{x} - \hat{\boldsymbol{x}}^-)^T \boldsymbol{P}_k^{-1}(\boldsymbol{x} - \hat{\boldsymbol{x}}^-). \tag{3.7}$$

Um die a posteriori Wahrscheinlichkeit zu maximieren, minimiert man den Betrag des Exponenten, $\Phi(\boldsymbol{x})$ aus (3.7), was äquivalent ist zum Maximieren des Logarithmus' der Dichtefunktion.

Das Optimalitätskriterium lautet daher:

$$\forall i \quad \frac{\partial}{\partial \boldsymbol{x}_i}\ln p(\mathcal{Z}^k|\hat{\boldsymbol{x}}^{\text{MAP}}) + \frac{\partial}{\partial \boldsymbol{x}_i}\ln p(\hat{\boldsymbol{x}}^{\text{MAP}}) = 0. \tag{3.8}$$

Ist sowohl das Meßrauschen als auch die a priori Wahrscheinlichkeitsdichte normalverteilt und die Meßfunktion linear, so gibt es eine lineare Lösung für $\hat{\boldsymbol{x}}^{\text{MAP}}$. Ansonsten kann man keine analytische Lösung für den Parameter $\boldsymbol{x}$ angeben, sondern muß sich numerischer Lösungsverfahren bedienen (vgl. [8]).

Beispiel:

Wir wollen als Beispiel analog zu dem in Abschnitt (3.1.1) M unabhängige Messungen einer normalverteilten Zufallsvariablen μ, $N(\mu, \sigma_v^2)$, mit der a priori Verteilung $N(\bar{\mu}, \sigma_\mu^2)$ betrachten. Unsere Annahme ist, daß die Varianzen σ_v^2 und σ_μ^2 bekannt sind.

Es gilt dann:

$$\begin{aligned} p(\mu|\mathcal{Z}^M) &= p(\mathcal{Z}^M|\mu)p(\mu) \qquad (3.9) \\ &= \left[(2\pi\sigma_v^2)^{-\frac{M}{2}} \exp\left\{-\frac{\sum_{i=1}^M (z_i-\mu)^2}{2\sigma_v^2}\right\}\right] [(2\pi\sigma_\mu^2)^{-\frac{1}{2}} \exp\{-\frac{(\mu-\bar{\mu})^2}{2\sigma_\mu^2}\}]. \end{aligned}$$

Die notwendige Wahrscheinlichkeitsbedingung lautet daher

$$\frac{\partial}{\partial\mu} \ln p(\mathcal{Z}^M|\mu) + \frac{\partial}{\partial\mu} \ln p(\mu) = \sum_{i=1}^M \frac{(z_i - \hat{\mu}_{\text{MAP}})}{\sigma_v^2} - \frac{(\bar{\mu} - \hat{\mu}_{\text{MAP}})}{\sigma_\mu^2} = 0.$$

Löst man diese Gleichung nach $\hat{\mu}_{\text{MAP}}$ auf, so erhält man

$$\hat{\mu}_{\text{MAP}} = \bar{\mu} + \frac{\sigma_\mu^2}{\sigma_v^2 + M\sigma_\mu^2} \sum_{i=1}^M (z_i - \bar{\mu}). \qquad (3.10)$$

Man sieht leicht, daß bei fehlendem Wissen über die a priori Wahrscheinlichkeitsdichte, d.h. $\sigma_\mu^2 \to \infty$, das Ergebnis der MAP–Strategie mit dem der ML–Strategie übereinstimmt:

$$\begin{aligned} \lim_{\sigma_\mu^2\to\infty} \left(\bar{\mu} + \frac{\sigma_\mu^2}{\sigma_v^2 + M\sigma_\mu^2} \sum_{i=1}^M (z_i - \bar{\mu})\right) &= \lim_{\sigma_\mu^2\to\infty} \left(\bar{\mu} + \frac{1}{M + \frac{\sigma_v^2}{\sigma_\mu^2}} \sum_{i=1}^M (z_i - \bar{\mu})\right) \\ &= \frac{1}{M} \sum_{i=1}^M z_i. \end{aligned}$$

Das bedeutet ebenfalls, daß der Einfluß der a priori Wahrscheinlichkeit mit zunehmender Anzahl von Messungen abnimmt. Dann entspricht die MAP–Schätzung der ML–Schätzung.

Betrachten wir nun den Einfluß des relativen Verhältnisses zwischen a priori Varianz und Meßvarianz auf die Schätzung. Ist die Meßvarianz gegenüber der a priori Varianz sehr groß, so konvergiert der Wichtungsfaktor $(M + \sigma_v^2/\sigma_\mu^2)^{-1}$ aus (3.10) gegen null, und das MAP–Verfahren liefert als Ergebnis den Mittelwert der a priori Verteilung $\bar{\mu}$. Überwiegt jedoch die a priori Varianz, dann nähert sich der Wichtungsfaktor $\frac{1}{M}$ an, und das Ergebnis ist der Mittelwert der Messungen $\frac{1}{M}\sum_{i=1}^M z_i$.

Die a posteriori Dichte ist direkt bestimmbar. Durch Ergänzung im Exponenten erlangt sie die Form einer Gaußverteilung:

$$p(\mu|\mathcal{Z}^M) = (2\pi\sigma_p^2)^{-\frac{1}{2}} \exp\{-\frac{(\mu-\hat{\mu})^2}{2\sigma_p^2}\}, \qquad (3.11)$$

wobei

$$\hat{\mu} = \bar{\mu} + \frac{\sigma_p^2 \sum_{i=1}^{M}(z_i - \bar{\mu})}{\sigma_v^2} = \bar{\mu} + \frac{\sigma_\mu^2}{\sigma_v^2 + M\sigma_\mu^2}\sum_{i=1}^{M}(z_i - \bar{\mu}) = \hat{\mu}_{\mathrm{MAP}}$$

und

$$\frac{1}{\sigma_p^2} = (\frac{1}{\sigma_\mu^2} + \frac{M}{\sigma_v^2}) = \frac{\sigma_v^2 + M\sigma_\mu^2}{\sigma_\mu^2\sigma_v^2}.$$

Die Lösung für $\hat{\mu}_{\mathrm{MAP}}$ ist beim Betrachten von Normalverteilungen und einer linearen Meßfunktion also gleich dem Erwartungswert $\mathrm{E}[\mu|\mathcal{Z}^k]$, weil der Mittelwert die höchste Dichte besitzt. In diesem Fall entspricht die MAP–Schätzung auch der MMSE–Schätzung (siehe Abschnitt 3.1.3).
Zu beachten ist jedoch, daß bei allgemeiner Form der Meßfunktion, $\boldsymbol{z} = h(\mu)$, die Dichte $p(\mu|\mathcal{Z}^M)$ nicht gaußverteilt ist.

3.1.3 Kleinster mittlerer quadratischer Fehler

Der Fehler einer Schätzung von unbekannten Parametern, $\mathrm{err}_{\boldsymbol{x}} = \hat{\boldsymbol{x}} - \boldsymbol{x}$, ist für ein Schätzverfahren von größtem Interesse. Ziel eines Schätzverfahrens ist es, den Schätzungsfehler so klein wie möglich zu halten. Ein weiteres Verfahren dazu ist die Methode des kleinsten mittleren quadratischen Fehlers (MMSE (Minimum Mean Square Error)).

Der mittlere quadratische Fehler ist definiert als

$$\mathrm{err}_{\mathrm{MS}}(k) = \mathrm{E}[(\hat{\boldsymbol{x}} - \boldsymbol{x})^T(\hat{\boldsymbol{x}} - \boldsymbol{x})|\mathcal{Z}^k]. \tag{3.12}$$

Die Minimierung dieses Fehlers liefert die Schätzung:

$$\hat{\boldsymbol{x}}^{\mathrm{MMSE}}(k) = \arg\min_{\hat{\boldsymbol{x}}} \mathrm{E}[(\hat{\boldsymbol{x}} - \boldsymbol{x})^2|\mathcal{Z}^k]. \tag{3.13}$$

Durch Differenzieren der rechten Seite von (3.13) bekommt man einen Ausdruck für das Optimalitätskriterium:

$$\frac{\partial}{\partial\hat{\boldsymbol{x}}}\mathrm{E}[(\hat{\boldsymbol{x}} - \boldsymbol{x})^2|\mathcal{Z}^k] = \mathrm{E}[2(\hat{\boldsymbol{x}} - \boldsymbol{x})|\mathcal{Z}^k] = 2(\hat{\boldsymbol{x}} - \mathrm{E}[\boldsymbol{x}|\mathcal{Z}^k]) = 0. \tag{3.14}$$

Oft ist die benötigte Information über die Verteilung zur Berechnung des bedingten Erwartungswertes nicht verfügbar. Doch auch bei vorhandenen Kenntnissen über die Wahrscheinlichkeitsdichte können die Berechnungen sehr kompliziert sein.

Nimmt man nun an, daß der Parametervektor $\boldsymbol{x}$ und die Messung $\mathcal{Z}^k$ die bedingte Dichte $p(\boldsymbol{x}|\mathcal{Z}^k)$ haben, wird die Schätzung des kleinsten mittleren quadratischen Fehlers durch den bedingten Mittelwert

$$\hat{\boldsymbol{x}}^{\mathrm{MMSE}} = \mathrm{E}[\boldsymbol{x}|\mathcal{Z}^k] = \int \boldsymbol{x}p(\boldsymbol{x}|\mathcal{Z}^k)d\boldsymbol{x} \tag{3.15}$$

gegeben.

Die gesuchte Lösungsstrategie sollte die Schätzung als lineare Funktion der Messung liefern und wenig zusätzliche Information voraussetzen. Solch eine Strategie beruht auf dem Orthogonalitätsprinzip ([3], [32]). Man nennt die Schätzung $\hat{\boldsymbol{x}}^{\text{MMSE}}$ auch Schätzung nach der kleinsten Varianz, da der mittlere quadratische Fehler gerade die Varianz angibt.

Setzt man zusätzlich voraus, daß das System– und das Meßrauschen normalverteilt sind und die Meßfunktion $\boldsymbol{h}$ linear ist, ist die a posteriori Wahrscheinlichkeitsdichte $p(\boldsymbol{x}|\mathcal{Z}^k)$ normal. Somit stellt der Mittelwert dieser Verteilung den gesuchten Schätzwert der MMSE–Methode dar. Mit dem später beschriebenen Kalman–Filter kann man diesen optimal bestimmen.

Beispiel:

Als Beispiel betrachten wir wieder die Schätzung eines unbekannten Parameters μ aus M unabhängigen Messungen. Das Meßrauschen sei als additiv und mit Mittelwert null und Varianz σ_v^2 normalverteilt vorausgesetzt. Die a priori Dichte von μ sei ebenfalls normalverteilt mit Mittelwert $\bar{\mu}$ und Varianz σ_μ.

Die Verteilung aus (3.9) ist die a posteriori Dichtefunktion. Da sie die Form einer Normalverteilung (3.11) hat, ist das Ergebnis der MMSE–Schätzung gerade der Mittelwert dieser Normalverteilung:

$$\hat{\boldsymbol{x}}^{\text{MMSE}} = \hat{\mu} = \hat{\mu}_{\text{MAP}}.$$

Das ML– und das MAP–Verfahren haben gegenüber dem MMSE–Verfahren den Vorteil, daß das jeweilige Optimalitätskriterium, das die Schätzung erfüllen muß (3.2), (3.8), durch einfaches Differenzieren der Wahrscheinlichkeitsgleichung und Nullsetzen der Ableitung gefunden werden kann.
Trotzdem ist es nicht immer möglich, eine analytische Lösung für die Schätzer anzugeben, da die resultierende Verteilung der Schätzung nicht unweigerlich normalverteilt sein muß. Oft müssen zur Lösung numerische Verfahren benutzt werden.

3.1.4 Das Kalman–Filter

Angewendet auf ein durch ein stochastisches dynamisches Signal (z. B. eine Bildfolge) gebildetes lineares System schätzt das Kalman–Filter rekursiv die Zustände dieses Systems aus den diskreten, verrauschten Daten.

Betrachten wir zunächst den Fall, daß die gesuchte Unbekannte $\boldsymbol{x}$ konstant ist, die Messungen aber nicht alle gleichzeitig zur Verfügung stehen. Zur Lösung dieses Problems kann man das rekursive MAP–Verfahren benutzen.

Der aus der Bayes–Formel hergeleitete Ausdruck für die a posteriori Dichtefunktion (3.5) läßt sich folgendermaßen umformulieren:

$$p(\boldsymbol{x}|\mathcal{Z}^k) = \frac{p(\boldsymbol{z}_k|\mathcal{Z}^{k-1}, \boldsymbol{x})p(\mathcal{Z}^{k-1}|\boldsymbol{x})p(\boldsymbol{x})}{p(\mathcal{Z}^k)}.$$

Nimmt man zusätzlich an, daß die Messungen voneinander unabhängig sind, ergibt sich für den Aktualisierungsschritt:

$$p(\boldsymbol{x}|\mathcal{Z}^k) = \frac{1}{c_k} p(\boldsymbol{z}_k|\boldsymbol{x}) p(\boldsymbol{x}|\mathcal{Z}^{k-1}), \tag{3.16}$$

wobei c_k die totale Wahrscheinlichkeit der Messung angibt:

$$c_k = p(\boldsymbol{z}_k|\mathcal{Z}^{k-1}) = \int p(\boldsymbol{z}_k|\boldsymbol{x}) p(\boldsymbol{x}|\mathcal{Z}^{k-1}) d\boldsymbol{x}. \tag{3.17}$$

Man erkennt, daß der Faktor $p(\boldsymbol{x}|\mathcal{Z}^{k-1})$ in die Berechnung der Aktualisierung genauso eingeht wie $p(\boldsymbol{x})$ in (3.5) und betrachtet ihn als a priori Wissen zur Messung $\boldsymbol{z}_k$. Da c_k nicht von dem gesuchten $\boldsymbol{x}$ abhängt, kann dieser Faktor zur Lösungsbestimmung vernachlässigt werden.

Möchte man nun jedoch Aussagen über Zustände von dynamischen Systemen machen, d.h. die betrachteten Zustände sind nicht konstant, sondern verändern sich von Schritt zu Schritt, benutzt man das Kalman–Filter.

Das Kalman–Filter ist ein rekursiver Schätzer. Ohne eine präzise Kenntnis der Natur des modellierten Systems erzielt es eine effiziente Berechnung der Methode der kleinsten Quadrate. Die Schätzung des Signals wird also derart gebildet, daß die Abweichung zwischen wahrem Signal und Schätzung schrittweise minimiert wird. Das bedeutet, daß das Filter den stochastischen Prozeß des Signals synthetisiert.

Modell des linearen dynamischen Systems

Ein diskreter Prozeß erster Ordnung läßt sich durch seine Zustände $\boldsymbol{x} \in \mathbb{R}^n$ charakterisieren, die über eine lineare Zustandsgleichung ineinander übergehen:

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}_k \boldsymbol{x}_k + \boldsymbol{w}_k. \tag{3.18}$$

Dabei stellt $\boldsymbol{A}_k \in \mathbb{R}^{n \times n}$ die Transitionsmatrix und $\boldsymbol{w}_k$ das additive Systemrauschen dar. Eine Erweiterung ist durch die Modellierung von kontrollierten Systemen möglich ([3],[8],[32]), die aber in der Bildfolgenanalyse nicht relevant ist. Die Zustände $\boldsymbol{x}_k$ entsprechen den exakten Größen, auf die man jedoch keinen direkten Zugriff hat. Man verfügt lediglich über Messungen $\boldsymbol{z}_k \in \mathbb{R}^m$ der Daten, also durch additives Meßrauschen verfälschte Werte, die aber über eine Messungsmatrix $\boldsymbol{H}_k \in \mathbb{R}^{m \times n}$ in Relation zu den exakten Zuständen stehen:

$$\boldsymbol{z}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{n}_k. \tag{3.19}$$

Man geht dabei davon aus, daß sowohl das System– als auch das Meßrauschen unkorreliert und normalverteilt ist und außerdem beide voneinander unabhängig sind, $\mathrm{E}[\boldsymbol{w}_k, \boldsymbol{n}_k] = 0$. Man kann sie daher durch ihre Kovarianzmatrizen $\boldsymbol{Q}$ und $\boldsymbol{R}$ darstellen, die die Form von Diagonalmatrizen haben:

$$\boldsymbol{w} \sim N(0, Q), \quad \boldsymbol{n} \sim N(0, R).$$

Als Voraussetzung für das Meß– und das Systemrauschen wird gefordert, daß es zeitlich unkorreliert ist, d.h.

$$\mathrm{E}[\boldsymbol{w}_k \boldsymbol{w}_l^T] = \boldsymbol{Q}_k \delta_{kl}$$

bzw.

$$\mathrm{E}[\boldsymbol{n}_k \boldsymbol{n}_l^T] = \boldsymbol{R}_k \delta_{kl},$$

was schon direkt aus der Normalverteilung und der Unabhängigkeit folgt.

Berechnungsgrundlagen

In einem Schritt k des Kalman–Filters wird aus vorhergehendem Wissen, dem sogenannten *a priori Zustand* $\hat{\boldsymbol{x}}_k^-$, und der aktuellen Messung $\boldsymbol{z}_k$ der geschätzte Zustand $\hat{\boldsymbol{x}}_k$ berechnet, den man *a posteriori Zustand* nennt, und der wiederum zur Prädiktion des a priori Zustands im Schritt $k+1$ dient.

Man definiert daraufhin den a priori bzw. a posteriori Schätzungsfehler wie folgt:

$$\boldsymbol{e}_k^- \equiv \boldsymbol{x}_k - \hat{\boldsymbol{x}}^-, \quad \boldsymbol{e}_k \equiv \boldsymbol{x}_k - \hat{\boldsymbol{x}}_k.$$

Folglich gilt für die a priori bzw. a posteriori Fehlerkovarianzmatrix:

$$\boldsymbol{\Lambda}_k^- = \mathrm{E}[\boldsymbol{e}_k^- \boldsymbol{e}_k^{-T}], \tag{3.20}$$
$$\boldsymbol{\Lambda}_k = \mathrm{E}[\boldsymbol{e}_k \boldsymbol{e}_k^T]. \tag{3.21}$$

Man sucht jetzt nach dem a posteriori Zustand $\hat{\boldsymbol{x}}_k$ als Linearkombination aus der a priori Schätzung und einer gewichteten Differenz zwischen der aktuellen Messung und der prädizierten Messung $\boldsymbol{H}_k \hat{\boldsymbol{x}}_k^-$:

$$\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^- + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_k^-). \tag{3.22}$$

Die Differenz $(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_k^-)$ nennt man *Innovation* oder *Residuum*. Sie gibt die Abweichung zwischen Messung und Prädiktion an. Ein Residuum von null bedeutet also vollständige Übereinstimmung.

Die *Verstärkungsmatrix* oder der *Gain*, $\boldsymbol{K}_k$, dient zur Minimierung der a posteriori Fehlervarianz, d.h. der a posteriori Zustand $\hat{\boldsymbol{x}}_k$ stellt nach der Methode des kleinsten mittleren quadratischen Fehlers die optimale Schätzung dar.

Um die Verstärkung zu bestimmen, muß man also die Fehlervarianz $\mathrm{E}[\boldsymbol{e}_k^T \boldsymbol{e}_k]$ minimieren. Berechnet man aus der Linearkombination (3.22) für den a posteriori Zustand den a posteriori Schätzungsfehler und setzt diesen in die Kovarianzmatrix (3.21) ein, so erhält man für die Fehlervarianz:

$$\begin{aligned}\mathrm{E}[\boldsymbol{e}_k^T \boldsymbol{e}_k] &= \mathrm{spur}[\boldsymbol{\Lambda}_k] \\ &= \mathrm{spur}[\mathrm{E}[(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^- - \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_k^-))(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^- - \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_k^-))^T]].\end{aligned}$$

Leitet man diesen Ausdruck nach $\boldsymbol{K}_k$ ab und setzt die Ableitung gleich null, so erhält man als Lösung für die Verstärkung:

$$\boldsymbol{K}_k = \boldsymbol{\Lambda}_k^- \boldsymbol{H}_k^T (\boldsymbol{H}_k \boldsymbol{\Lambda}_k^- \boldsymbol{H}_k^T + \boldsymbol{R}_k)^{-1}. \tag{3.23}$$

Ohne im voraus die Linearität des Kalman–Filters zu verlangen, findet man $\hat{\boldsymbol{x}}_k$ alternativ durch Maximieren der a posteriori Dichte nach k Messungen:

$$\hat{\boldsymbol{x}}_k = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x}_k|\mathcal{Z}^k) = \arg\max_{\boldsymbol{x}} p(\boldsymbol{z}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\mathcal{Z}^{k-1}), \tag{3.24}$$

wobei c_k die totale Wahrscheinlichkeit der Messung $\boldsymbol{z}_k$ angibt (3.17). Die gesamte Information der Vergangenheit wird über die Normalverteilung

$$\frac{1}{c} \exp\{-(\boldsymbol{x} - \hat{\boldsymbol{x}}_k^-) \boldsymbol{\Lambda}_k^{-1} (\boldsymbol{x} - \hat{\boldsymbol{x}}_k^-)^T\}$$

weitergegeben. Da man außerdem von einer Normalverteilung des Rauschens ausgeht,

$$\frac{1}{c_R} \exp\{-(\boldsymbol{z} - \boldsymbol{H}_k \boldsymbol{x}_k) \boldsymbol{R}_k^{-1} (\boldsymbol{z} - \boldsymbol{H}_k \boldsymbol{x}_k)^T\},$$

ist das Ergebnis ebenfalls normalverteilt:

$$\begin{aligned} p(\boldsymbol{x}_k|\mathcal{Z}^k) &\sim N(\mathrm{E}[\boldsymbol{x}_k], \mathrm{E}[(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)^T]) \\ &= N(\hat{\boldsymbol{x}}_k, \boldsymbol{\Lambda}_k). \end{aligned} \tag{3.25}$$

Dieser Ansatz liefert ebenfalls das Ergebnis (3.23). Man sieht an (3.25), daß die Kovarianzmatrix der Schätzung monoton fällt, da die Schätzung sich dem wahren Wert immer weiter annähert.

Betrachtet man die Schätzung eines skalaren Parameters $\hat{\boldsymbol{x}}_k = \mu$, so werden die a priori Varianz $\boldsymbol{\Lambda}_k^-$ und das Meßrauschen $\boldsymbol{R}_k$ ebenfalls durch einen Skalar gegeben:

$$\begin{aligned} \boldsymbol{\Lambda}_k^- &= \sigma_\mu^2, \\ \boldsymbol{R}_k = \boldsymbol{R} &= \sigma_v^2. \end{aligned}$$

Die Messungsmatrix $\boldsymbol{H}_k$ ist in diesem Fall durch die Identität dargestellt. Somit ergibt sich für die Verstärkung:

$$\boldsymbol{K}_k = \frac{\sigma_\mu^2}{\sigma_v^2 + \sigma_\mu^2}.$$

Also gilt für die Schätzung:

$$\hat{\boldsymbol{x}}_k = (1 - \frac{\sigma_\mu^2}{\sigma_v^2 + \sigma_\mu^2}) \hat{\boldsymbol{x}}_k^- + \frac{\sigma_\mu^2}{\sigma_v^2 + \sigma_\mu^2} \boldsymbol{z}_k.$$

Wie bei dem MAP–Schätzer sieht man auch hier, daß bei einem Überwiegen der Rauschvarianz gegenüber der a priori Varianz die Verstärkung gegen null konvergiert und die

Schätzung durch den a priori Zustand bestimmt wird. Andererseits grenzt bei Überwiegen der a priori Varianz die Verstärkung an den Wert eins, und das Schätzergebnis ist der gemessene Wert.

Entsprechend der stochastischen Modellierung der Korrektur der Schätzung nach einer Messung (3.22) benötigt man auch eine stochastische Betrachtung der Prädiktion des Zustands zwischen zwei Messungen. Für den Prädiktionsschritt gilt:

$$p(\boldsymbol{x}_{k+1}|\mathcal{Z}^k) = \int p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\mathcal{Z}^k)d\boldsymbol{x}_k.$$

Die Rolle des Meßrauschens übernimmt hier das Systemrauschen, das das Vertrauen gegenüber dem dynamischen Modell

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}_k\boldsymbol{x}_k$$

widerspiegelt.

Man betrachtet also die bedingte Wahrscheinlichkeit

$$p(\boldsymbol{x}_{k+1}^-|\hat{\boldsymbol{x}}_k) = \frac{1}{c_Q}\exp\{-\frac{1}{2}(\boldsymbol{x}_{k+1}^- - \boldsymbol{A}_k\hat{\boldsymbol{x}}_k)^T\boldsymbol{Q}^{-1}(\boldsymbol{x}_{k+1} - \boldsymbol{A}_k\hat{\boldsymbol{x}}_k)\}.$$

Die optimale Prädiktion nach dem MMSE– oder dem MAP–Kriterium lautet dann

$$\hat{\boldsymbol{x}}_{k+1}^- = \boldsymbol{A}_k\hat{\boldsymbol{x}}_k. \tag{3.26}$$

Ist $\hat{\boldsymbol{x}}_k$ normalverteilt, dann ist auch der Zustand vor der Messung normalverteilt mit Mittelwert $\hat{\boldsymbol{x}}_k^-$ und Kovarianzmatrix

$$\boldsymbol{\Lambda}_{k+1}^- = \boldsymbol{A}_k\boldsymbol{\Lambda}_k\boldsymbol{A}_k^T + \boldsymbol{Q}_k. \tag{3.27}$$

Jeweils beide Schritte der Prädiktion und der Korrektur lassen sich aus Schema 3.1 ablesen.

Das Filter synthetisiert den stochastischen Prozeß durch eine Rückkopplung über die Messungen. Es schätzt den Zustand zu einer Zeit k und korrigiert mittels verrauschter Meßdaten.

Die Kovarianzmatrizen $\boldsymbol{Q}$ und $\boldsymbol{R}$ können vor der Anwendung des Filters gemessen werden. Die Meßfehlervarianz läßt sich leicht durch Beispielmessungen abschätzen. Die Wahl der Systemvarianz $\boldsymbol{Q}$ ist jedoch weniger eindeutig. Sie ist zudem der einzige Faktor, der das stetige Abfallen der Fehlervarianz $\boldsymbol{\Lambda}$ bremsen kann, so daß Messungen stärker gewichtet werden.

Unter der Bedingung, daß $\boldsymbol{Q}_k$ und $\boldsymbol{R}_k$ über die Berechnung konstant gewählt sind, tritt schon bald eine Stabilisierung der Fehlerkovarianzmatrix $\boldsymbol{\Lambda}_k$ und der Verstärkungsmatrix $\boldsymbol{K}_k$ ein, die dadurch ebenfalls konstante Werte $\boldsymbol{\Lambda}$ bzw. $\boldsymbol{K}$ annehmen.

Die Wahl von $\boldsymbol{Q}$ und $\boldsymbol{R}$, das sogenannte Tuning ist ausschlaggebend für den Erfolg des Kalman–Filter–Verfahrens.

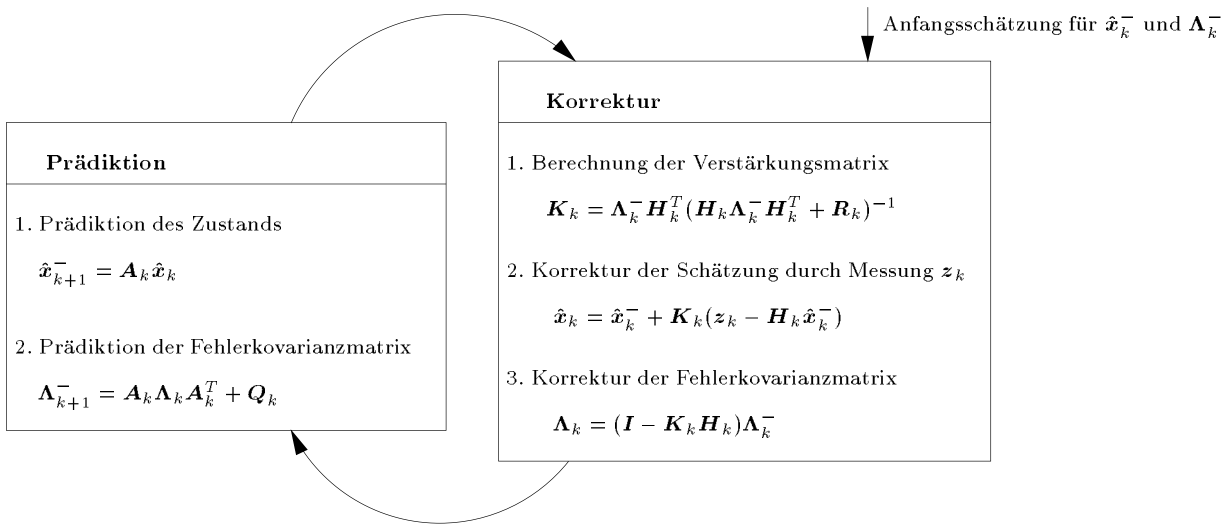


Abbildung 3.1: Schema über die Arbeitsweise des Kalman–Filters

Anwendung des Kalman–Filters

In dem in dieser Arbeit vorgestellten Verfahren zur Bewegungsanalyse wird das Kalman–Filter zur Berechnung der Bewegungszustände der einzelnen Gelenke benutzt. Der betrachtete Prozeß ist eine gleichförmig beschleunigte Bewegung ($\boldsymbol{a} = \text{const.}$). Durch Integration der Beschleunigung findet man die Gleichungen für Geschwindigkeit und zurückgelegten Weg dieser Bewegung:

$$\boldsymbol{v} = \boldsymbol{a}t + \boldsymbol{v}_0, \tag{3.28}$$

$$\boldsymbol{s} = \frac{1}{2}\boldsymbol{a}t^2 + \boldsymbol{v}_0 t + \boldsymbol{s}_0, \tag{3.29}$$

wobei $\boldsymbol{v}$ die Geschwindigkeit, $\boldsymbol{a}$ die Beschleunigung, $\boldsymbol{s}$ den zurückgelegten Weg, t die Zeit ab dem Startpunkt t_0 und $\boldsymbol{v}_0$, $\boldsymbol{s}_0$ die jeweiligen Anfangswerte zur Zeit t_0 darstellen.

Der den Prozeß charakterisierende Zustand $\boldsymbol{x}$ sieht folgendermaßen aus:

$$\boldsymbol{x} = \begin{pmatrix} x \\ y \\ u \\ v \\ a_x \\ a_y \end{pmatrix}. \tag{3.30}$$

Die (3.18) entsprechende Formulierung der Transitionsvorschrift ist aus den Bewegungsgleichungen (3.28) und (3.29) abzuleiten:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}_k \boldsymbol{x}_k + \boldsymbol{w} \\ &= \begin{pmatrix} 1 & 0 & t & 0 & \frac{1}{2}t^2 & 0 \\ 0 & 1 & 0 & t & 0 & \frac{1}{2}t^2 \\ 0 & 0 & 1 & 0 & t & 0 \\ 0 & 0 & 0 & 1 & 0 & t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{x}_k + \boldsymbol{w}. \end{aligned} \tag{3.31}$$

Die Zeit t ist hier der Zeitraum, der zwischen der Messung zum Zeitpunkt k und der zum Zeitpunkt $k+1$ liegt.

Das Systemrauschen $\boldsymbol{w}$ beruht darauf, daß die zu beobachtende Bewegung nicht wirklich mit konstanter Beschleunigung erfolgt. Daher hat die das Systemrauschen charakterisierende Varianzmatrix $\boldsymbol{Q}$ nur zwei von null verschiedene Einträge, und zwar für die Beschleunigung:

$$\boldsymbol{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{a_x} & 0 \\ 0 & 0 & 0 & 0 & 0 & q_{a_y} \end{pmatrix}. \tag{3.32}$$

Als Messung wird die Gelenkpositionen im Bild benutzt, d.h. die Messungsmatrix $\boldsymbol{H}_k$ verändert sich von Schritt zu Schritt nicht. Sie hat die Form:

$$\boldsymbol{H} = \boldsymbol{H}_k = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Also wird die Messung modelliert durch:

$$\boldsymbol{z}_k = \begin{pmatrix} \tilde{x}_k \\ \tilde{y}_k \end{pmatrix} = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{n}. \tag{3.33}$$

Das Meßrauschen wird durch die Diagonalmatrix $\boldsymbol{R} \in \mathbb{R}^{2\times 2}$ gegeben, die auf ihrer Diagonalen die Einträge r_x, r_y hat.

Beispiel:

Als Beispiel gehen wir auf die Beobachtung einer kreisförmigen Bewegung ein. Die Modellierung des Systems entspricht (3.30), (3.31) und (3.33). Gegeben sind die verrauschten Daten dieser Bewegung, die in Abb. 3.1.4 veranschaulicht sind.

Die Parameter des Meßrauschens $\boldsymbol{R}$ haben die Werte $r_x = r_y = 0.01$. Als Eingabe benötigt das Filter zusätzlich Anfangsschätzungen für den a priori Zustand $\hat{\boldsymbol{x}}_1^-$ und die a priori Varianz $\boldsymbol{\Lambda}_1^-$.

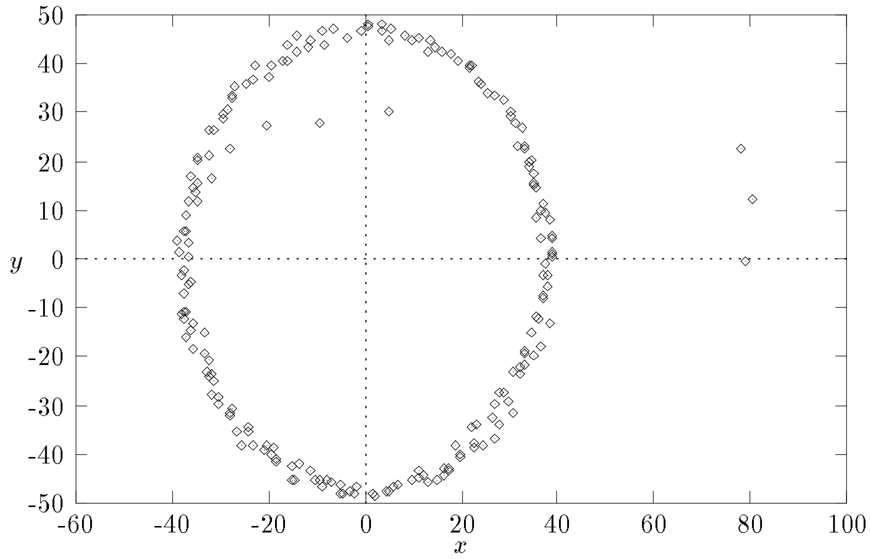


Abbildung 3.2: Gemessene Bahnpunkte (x, y) einer kreisförmigen Bewegung

Der a priori Zustand $\hat{\boldsymbol{x}}_1^-$ sei der Nullvektor. Falls dies eine absolut sichere Wahl wäre, würde man die a priori Varianz ebenfalls auf null setzen, doch dann wäre eine Änderung des Zustands unmöglich, d.h. die Schätzung wäre stets der Nullvektor. Stattdessen wählt man $\boldsymbol{\Lambda}_1^-$ ungleich Null. Wir wollen für $\boldsymbol{\Lambda}_1^-$ von dem Wert eins ausgehen, so daß die Kovarianzmatrix Diagonalform hat und die Diagonalelemente den Wert eins.

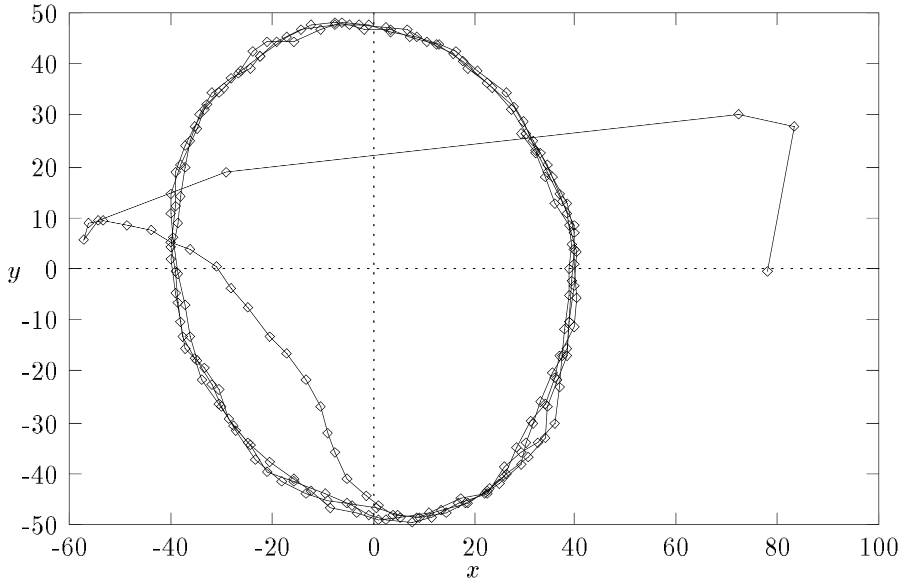


Abbildung 3.3: Schätzung der Kreisbewegung bei Systemvarianz $q_{a_x} = q_{a_y} = 0.00001$

Den Einfluß des Tunings durch die Vorgabe des Systemrauschens verdeutlichen wir an zwei unterschiedlichen Belegungen der Kovarianzmatrix $\boldsymbol{Q}$. Geht man davon aus, daß das

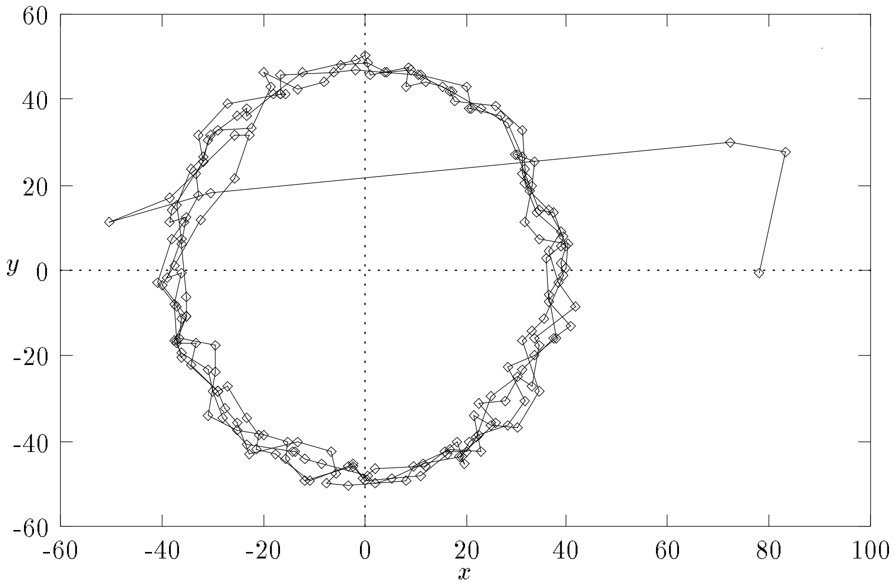


Abbildung 3.4: Schätzung der Kreisbewegung bei Systemvarianz $q_{a_x} = q_{a_y} = 0.01$

Systemrauschen gering ist, $q_{a_x} = q_{a_y} = 0.00001$, erhält man das in Abb. 3.1.4 gezeigte Ergebnis. Bei Vorgabe eines größeren Systemrauschens ist die Schätzung unzuverlässiger, so daß die Meßdaten stärker berücksichtigt werden, wie in Abb. 3.1.4 für $q_{a_x} = q_{a_y} = 0.01$ zu sehen ist.

3.2 Berücksichtigung des Rauschens in der BCCE

Das differentielle Verfahren zur Berechnung des optischen Flusses basiert auf der Kontinuitätsgleichung (2.2). Diese geht von idealisierten Werten für die örtliche und zeitliche Ableitung aus.

Zum einen verfügt man in der Realität jedoch nur über Messungen der Grauwertableitungen, die durch Rauschen, Diskretisierung und Quantisierung des Originalbildes sowie durch Ungenauigkeiten in der Filterantwort fehlerbehaftet sind. Zum anderen ist der Unterschied zwischen dem optischen Fluß und dem gesuchten Verschiebungsvektorfeld zu berücksichtigen, die einerseits durch die Annahme verursacht wird, daß der Fluß im betrachteten Bereich konstant ist, und andererseits durch Vernachlässigung der Terme quadratischer Ordnung in der BCCE herrührt.

Wenn man den optischen Fluß durch lineare Ausgleichsrechnung ermittelt (vgl. Kapitel 2), behandelt man stochastisch gesehen nur das Rauschen in der zeitlichen Ableitung, wobei die Rauschvarianz für alle Punkte gleich ist.

Für eine Fusion von verschiedenen Auflösungsstufen ist diese einfache Rauschbehandlung unzulänglich. Um auch die weiteren Abweichungen in die Berechnung der Bildbewegung mit einzubeziehen, ist eine stochastische Formulierung der Kontinuitätsgleichung notwendig. Dazu modelliert man die oben genannten Typen von Rauschen durch Gaußverteilungen.

Man definiert folgenden Zusammenhang zwischen den gemessenen Ableitungen g_x, g_y, g_t und den exakten $\tilde{g}_x, \tilde{g}_y, \tilde{g}_t$:

$$g_t = \boldsymbol{h}(\tilde{g}_t) + \boldsymbol{n} = \tilde{g}_t + \delta g_t, \tag{3.34}$$

$$\nabla_s \boldsymbol{g} = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \begin{pmatrix} \boldsymbol{h}(\tilde{g}_x) + \boldsymbol{n} \\ \boldsymbol{h}(\tilde{g}_y) + \boldsymbol{n} \end{pmatrix} = \begin{pmatrix} \tilde{g}_x + \delta g_x \\ \tilde{g}_y + \delta g_y \end{pmatrix}. \tag{3.35}$$

Diesem Rauschmodell entsprechend kann man nun den optischen Fluß mit einem der oben angegebenen Schätzverfahren bestimmen. Möchte man den Fluß lediglich auf einer Auflösungsstufe berechnen, bietet sich hierzu das ML–Verfahren an. Bei hierarchischer Berechnung verfügt man jedoch für feinere Auflösungsstufen über Kenntnisse der a priori Wahrscheinlichkeitsdichte, so daß in dem Fall die MAP–Strategie das geeignetere Verfahren darstellt.

Im folgenden werden die Ergebnisse der ML– und der MAP–Strategie angewendet auf zwei unterschiedliche Rauschmodelle von Simoncelli et al. und Kanatani vorgestellt. Das Rauschmodell von Simoncelli et al. berücksichtigt explizit das Rauschen in der zeitlichen Komponente des Grauwertgradienten. Außerdem enthält das Modell eine normalverteilte Zufallsvariable, welche die Abweichung des tatsächlichen optischen Flusses von dem, der die BCCE erfüllt, widerspiegelt. Das Rauschmodell von Kanatani bezieht dahingegen explizit das Rauschen in den örtlichen Komponenten des Grauwertgradienten mit ein. Bei beiden Rauschmodellen wird die Annahme gemacht, daß das Rauschen unkorreliert ist. Am Ende des Abschnitts wird daher ein kurzer Einblick in ein Rauschmodell von Nagel gegeben, welches diese Annahme nicht macht.

Auf den Rauschmodellen aufbauend werden im folgenden zudem die hierarchischen Varianten zur Berechnung des optischen Flusses vorgestellt, die zur Detektion der Person im ersten Schritt der Bewegungsanalyse dienen. Da die Geschwindigkeiten der einzelnen Körperteile bei der Verfolgung durch ein affines Modell des optischen Flusses modelliert werden, wird ferner die Übertragung der hierarchischen Algorithmen auf solch ein affines Modell erklärt.

3.2.1 Rauschen in der zeitlichen Ableitung – ML–Strategie

In ihren Arbeiten ([30],[31]) modellieren Simoncelli et al. zum einen das Meßrauschen in der zeitlichen Ableitung durch (3.34) mit

$$\delta g_t \sim N(0, \lambda_t)$$

und zusätzlich die Abweichung zwischen dem tatsächlichen optischen Fluß und dem, der die BCCE erfüllt, durch eine unabhängige, additive Normalverteilung. Wenn $\tilde{\boldsymbol{v}}$ den tatsächlichen optischen Flußvektor darstellt und $\boldsymbol{v}$ den über den betrachteten Bereich gemittelten Flußvektor, der die BCCE erfüllt, so ist die Differenz gegeben durch

$$\boldsymbol{v} = \tilde{\boldsymbol{v}} + \delta\boldsymbol{v} \quad \delta\boldsymbol{v} \sim N(\mathbf{0}, \boldsymbol{\Lambda}_{\boldsymbol{v}}).$$

Sie machen jedoch die Annahme, daß in der örtlichen Ableitung kein Rauschen vorliegt, $\delta g_x = \delta g_y = 0$, damit sie eine analytische Lösung für den optischen Fluß $\boldsymbol{v}$ angeben können. Ihre stochastische Formulierung der Kontinuitätsgleichung (2.2) lautet also:

$$\begin{aligned} 0 &= \nabla_s\tilde{\boldsymbol{g}}^T\tilde{\boldsymbol{v}} + \tilde{g}_t \\ &= \nabla_s\boldsymbol{g}^T(\boldsymbol{v} - \delta\boldsymbol{v}) + g_t - \delta g_t \\ \text{bzw. } \nabla_s\boldsymbol{g}^T\boldsymbol{v} + g_t &= \nabla_s\boldsymbol{g}^T\delta\boldsymbol{v} + \delta g_t. \end{aligned} \tag{3.36}$$

Um mittels der ML–Strategie die Lösung $\hat{\boldsymbol{v}}$ für den optischen Fluß $\boldsymbol{v}$ berechnen zu können, muß man (3.1) zufolge die bedingte Wahrscheinlichkeit $p(g_t|\boldsymbol{v}, \nabla_s\boldsymbol{g})$ kennen. Diese läßt sich aus Gleichung (3.36) ablesen.

Die rechte Seite der Gleichung kann man zu einer normalverteilten Zufallsvariablen zusammenfassen, deren Varianz $\lambda = \nabla_s\boldsymbol{g}^T\boldsymbol{\Lambda}_{\boldsymbol{v}}\nabla_s\boldsymbol{g} + \lambda_t$ entspricht. Damit gilt für die Messung der zeitlichen Ableitung an einem Bildpunkt i bei gegebenem Fluß $\boldsymbol{v}$ und gegebener örtlicher Ableitung $\nabla_s\boldsymbol{g}$:

$$\begin{aligned} p_i(g_t|\boldsymbol{v}, \nabla_s\boldsymbol{g}) &= \frac{1}{\sqrt{2\pi}\lambda}\exp\{-\frac{(\nabla_s\boldsymbol{g}^T\boldsymbol{v} + g_t)^2}{2\lambda^2}\} \\ &\propto \exp\{-\frac{1}{2}(\nabla_s\boldsymbol{g}^T\boldsymbol{v} + g_t)(\nabla_s\boldsymbol{g}^T\boldsymbol{\Lambda}_{\boldsymbol{v}}\nabla_s\boldsymbol{g} + \lambda_t)^{-1}(\nabla_s\boldsymbol{g}^T\boldsymbol{v} + g_t)\}. \end{aligned} \tag{3.37}$$

Betrachtet man nur einen Bildpunkt, so tritt auch hier wieder das Aperturproblem auf. Unter der Annahme, daß der Fluß in einer Region konstant und das Meßrauschen an den verschiedenen Bildpunkten dieser Region unabhängig voneinander ist, berücksichtigt man daher wieder alle Punkte einer gewissen Umgebung. Die Wahrscheinlichkeitsdichte aller Messungen in dieser Umgebung wird dann durch das Produkt der Einzelwahrscheinlichkeiten (3.37) gegeben:

$$p(g_t|\boldsymbol{v}, \nabla_s\boldsymbol{g}) \propto \exp\{-\sum_i \frac{(\nabla_s\boldsymbol{g}_i^T\boldsymbol{v} + g_{t\,i})^2}{2\lambda_i}\}, \tag{3.38}$$

wobei $\lambda_i = \frac{1}{2(\nabla_s\boldsymbol{g}_i^T\boldsymbol{\Lambda}_{\boldsymbol{v}}\nabla_s\boldsymbol{g}_i + \lambda_t)}$ ist.

Das Maximieren dieser Wahrscheinlichkeit erreicht man durch die Minimierung der gewichteten Summe $\sum_i \frac{1}{2\lambda_i}(\nabla_s\boldsymbol{g}_i^T\boldsymbol{v} + g_{t\,i})^2$, die der Fehlerfunktion aus (2.10) ähnelt. Als Ergebnis erhält man den Lösungsvektor:

$$\hat{\boldsymbol{v}}_{\mathrm{ML}} = -\left(\sum_i \lambda_i\boldsymbol{M}_i\right)^{-1}\left(\sum_i \lambda_i\boldsymbol{b}_i\right), \tag{3.39}$$

wobei $\boldsymbol{M}_i$ und $\boldsymbol{b}_i$ wie in (2.14) zu wählen sind.
Die Kovarianzmatrix $\boldsymbol{\Lambda}_{\hat{v}}$ hat die Form:

$$\boldsymbol{\Lambda}_{\hat{v}} = \left(\sum_i \boldsymbol{M}_i\right)^{-1}. \tag{3.40}$$

Das Übereinstimmen der Ergebnisse der ML–Strategie mit denen der linearen Ausgleichsrechnung beruht auf der Normalverteilung der gegebenen Messungen.

3.2.2 Rauschen in der zeitlichen Ableitung – MAP–Strategie

Bei der MAP–Strategie ist zur Schätzung von $\boldsymbol{v}$ nach (3.6) die a posteriori Wahrscheinlichkeitsdichte $p(\boldsymbol{v}|\nabla\boldsymbol{g})$ zu maximieren. Diese bedingte Wahrscheinlichkeit wird bei gegebenen Messungen aller Ableitungen $\nabla\boldsymbol{g}$ an einem Bildpunkt nach der Bayes–Formel (3.5) berechnet:

$$p(\boldsymbol{v}|\nabla\boldsymbol{g}) = \frac{p(g_t|\boldsymbol{v}, \nabla_s\boldsymbol{g})p(\boldsymbol{v})}{p(g_t)}. \tag{3.41}$$

Simoncelli et al. gehen davon aus, daß die a priori–Wahrscheinlichkeit $p(\boldsymbol{v})$ eine Normalverteilung mit Mittelwert null und Kovarianz $\boldsymbol{\Lambda}_p$ ist. Dies ist eine starke Voraussetzung, weil sie dadurch eine Fusion der a priori Schätzung mit einer neuen Schätzung verhindern. Durch die Annahme, daß der Mittelwert der Verteilung den Wert null hat, wird das Ergebnis der Berechnung auf einer gröberen Stufe nicht für die neue Berechnung berücksichtigt. Die Wahrscheinlichkeit $p(g_t)$ ist von $\boldsymbol{v}$ unabhängig und trägt nichts zur Maximierung bei. Die so mit dem Ergebnis aus (3.37) resultierende Verteilung stellt eine Normalverteilung mit mittlerem Vektor $\boldsymbol{\mu}_v$ und Varianz $\boldsymbol{\Lambda}_{\hat{v}}$ dar:

$$\begin{aligned}
p(\boldsymbol{v}|\nabla\boldsymbol{g}) &\propto \exp\{-\frac{1}{2}(\nabla_s\boldsymbol{g}^T\boldsymbol{v} + g_t)^T(\nabla_s\boldsymbol{g}^T\boldsymbol{\Lambda}_v\nabla_s\boldsymbol{g} + \lambda_t)^{-1}(\nabla_s\boldsymbol{g}^T\boldsymbol{v} + g_t)\}\exp\{-\frac{1}{2}\boldsymbol{v}^T\boldsymbol{\Lambda}_p^{-1}\boldsymbol{v}\} \\
&= \exp\{-\frac{1}{2}\boldsymbol{v}^T[\nabla_s\boldsymbol{g}(\nabla_s\boldsymbol{g}^T\boldsymbol{\Lambda}_v\nabla_s\boldsymbol{g} + \lambda_t)^{-1}\nabla_s\boldsymbol{g}^T + \boldsymbol{\Lambda}_p^{-1}]\boldsymbol{v} \\
&\quad -g_t(\nabla_s\boldsymbol{g}^T\boldsymbol{\Lambda}_v\nabla_s\boldsymbol{g} + \lambda_t)^{-1}\nabla_s\boldsymbol{g}^T\boldsymbol{v} \\
&\quad -\frac{1}{2}g_t(\nabla_s\boldsymbol{g}^T\boldsymbol{\Lambda}_v\nabla_s\boldsymbol{g} + \lambda_t)^{-1}g_t\} \qquad (3.42) \\
&=: \exp\{-\frac{1}{2}(\boldsymbol{\mu}_v - \boldsymbol{v})^T\boldsymbol{\Lambda}_{\hat{v}}^{-1}(\boldsymbol{\mu}_v - \boldsymbol{v})\}. \qquad (3.43)
\end{aligned}$$

Nach der MAP–Strategie ist nun dasjenige $\boldsymbol{v}$ gesucht, welches die bedingte Wahrscheinlichkeit $p(\boldsymbol{v}|\nabla\boldsymbol{g})$ maximiert. Durch Minimierung des Betrags des Exponenten, kann man als Lösung den mittleren Vektor $\boldsymbol{\mu}_v$ und die Kovarianzmatrix $\boldsymbol{\Lambda}_{\hat{v}}$ der resultierenden Normalverteilung bestimmen:

$$\begin{aligned}
\boldsymbol{\Lambda}_{\hat{v}} &= [\nabla_s\boldsymbol{g}(\nabla_s\boldsymbol{g}^T\boldsymbol{\Lambda}_v\nabla_s\boldsymbol{g} + \lambda_t)^{-1}\nabla_s\boldsymbol{g}^T + \boldsymbol{\Lambda}_p^{-1}]^{-1}, \\
\hat{\boldsymbol{v}}_{\text{MAP}} = \boldsymbol{\mu}_v &= -\boldsymbol{\Lambda}_{\hat{v}}\nabla_s\boldsymbol{g}(\nabla_s\boldsymbol{g}^T\boldsymbol{\Lambda}_v\nabla_s\boldsymbol{g} + \lambda_t)^{-1}g_t.
\end{aligned}$$

Simoncelli et al. gehen davon aus, daß die Varianz in beiden Komponenten des VVF gleich ist und setzen daher die Kovarianzmatrix $\mathbf{\Lambda}_{\boldsymbol{v}}$ einer Diagonalmatrix mit den Einträgen $\lambda_{\boldsymbol{v}}$ gleich.

Man erhält so mit $\boldsymbol{M}$ aus (2.8) und $\boldsymbol{b}$ aus (2.9):

$$\mathbf{\Lambda}_{\hat{\boldsymbol{v}}} = \left[\frac{\boldsymbol{M}}{(\lambda_{\boldsymbol{v}} \parallel \nabla_s \boldsymbol{g} \parallel^2 + \lambda_t)} + \mathbf{\Lambda}_p^{-1}\right]^{-1}, \tag{3.44}$$

$$\hat{\boldsymbol{v}}_{\mathrm{MAP}} = \boldsymbol{\mu}_{\boldsymbol{v}} = -\mathbf{\Lambda}_{\hat{\boldsymbol{v}}} \frac{\boldsymbol{b}}{(\lambda_{\boldsymbol{v}} \parallel \nabla_s \boldsymbol{g} \parallel^2 + \lambda_t)}. \tag{3.45}$$

Der Vektor (3.45) gibt wegen des Aperturproblems den Normalfluß an. Bei der MAP–Strategie hat man eigentlich zwar zwei den optischen Flußvektor bestimmende Gleichungen, nämlich die BCCE und $\boldsymbol{v} = \boldsymbol{v}_{\text{a priori}}$, aber die Annahme, daß die a priori Schätzung des Flusses den Mittelwert null hat, reduziert die bestimmenden Gleichungen auf die BCCE.

Unter der Annahme, daß das Rauschen an verschiedenen Punkten unabhängig voneinander ist, kann man den vollen optischen Fluß auch hier durch Betrachtung der Umgebungspunkte bestimmen:

$$\mathbf{\Lambda}_{\hat{\boldsymbol{v}}} = \left[\sum_i \frac{\omega_i \boldsymbol{M}(x_i, y_i, t)}{(\lambda_{\boldsymbol{v}} \parallel \nabla_s \boldsymbol{g}(x_i, y_i, t) \parallel^2 + \lambda_t)} + \mathbf{\Lambda}_p^{-1}\right]^{-1}, \tag{3.46}$$

$$\hat{\boldsymbol{v}}_{\mathrm{MAP}} = \boldsymbol{\mu}_{\boldsymbol{v}} = -\mathbf{\Lambda}_{\hat{\boldsymbol{v}}} \sum_i \frac{\omega_i \boldsymbol{b}(x_i, y_i, t)}{(\lambda_{\boldsymbol{v}} \parallel \nabla_s \boldsymbol{g}(x_i, y_i, t) \parallel^2 + \lambda_t)}. \tag{3.47}$$

Dabei stellt ω_i einen Wichtungsfaktor dar, der die Beiträge der Punkte einer Region gewichtet.

3.2.3 Rauschen in der zeitlichen Ableitung – hierarchisches Verfahren

Wie schon in Kapitel 2.5 erwähnt, ist es zur Vermeidung von Berechnungsfehlern durch zeitliches Aliasing sinnvoll, den optischen Fluß hierarchisch zu berechnen. Auch das Verfahren von Simoncelli et al. ist auf Bildpyramiden anwendbar. Im folgenden wird erst die Fusion der Berechnungen der verschiedenen Pyramidenstufen erklärt und anschließend der gesamte Algorithmus der Pyramidenverarbeitung. Zuletzt wird eine Abwandlung des Verfahrens vorgestellt, mit der man den affinen Fluß hierarchisch bestimmen kann.

Die Fusion der Ergebnisse wird hier durch ein Kalman–Filter realisiert. Das Warpen der einzelnen Pyramidenstufen, abgesehen von der obersten, führt dazu, daß man die hierarchische Berechnung des optischen Flusses als ein dynamisches System ansehen kann. Dabei werden die Messungen jedoch nicht in zeitlichen Schritten gemacht, sondern in Schritten, die von der Auflösung der Messung abhängen. Das heißt, der Abstieg von der höchsten zur tiefsten Pyramidenstufe wird für das Kalman-Filter als zeitliche Abfolge angesehen.

Wie bei dem Kalman–Filter definiert man analog zu (3.18) eine Zustandsgleichung, die die Zustände, hier den optischen Fluß, von zwei übereinander liegenden Pyramidenstufen zueinander in Relation setzt:

$$\boldsymbol{v}(l+1) = \boldsymbol{L}(l)\boldsymbol{v}(l) + \boldsymbol{w}(l), \tag{3.48}$$

wobei l die Stufe der Pyramide angibt, $\boldsymbol{L}(l)$ die Matrixdarstellung eines linearen Interpolationsoperators ist und $\boldsymbol{w} \sim N(0, \boldsymbol{Q})$ das Systemrauschen darstellt, das nach Voraussetzung von der Stufe unabhängig und somit unkorreliert sein soll. Je größer der Wert für l ist, desto feiner ist das Gitter der Pyramidenstufe.

Die Messungsgleichung lautet analog zu (3.19) und basierend auf der stochastischen Formulierung der Kontinuitätsgleichung (3.36):

$$-g_t(l) = \nabla_s \boldsymbol{g}^T(l)\boldsymbol{v}(l) - (\nabla_s \boldsymbol{g}^T(l)\delta\boldsymbol{v} + \delta g_t),$$

wobei wieder $\delta\boldsymbol{v} \sim N(0, \boldsymbol{\Lambda_v})$ und $\delta g_t \sim N(0, \lambda_t)$ gilt.

Der optimale Schätzer für den optischen Fluß auf der Stufe $l+1$ bei gegebener a priori Schätzung als Resultat der Stufe l und Messungen der Gradienten auf der Stufe $l+1$ hat dann analog zu (3.22) die Form:

$$\hat{\boldsymbol{v}}(l+1) = \boldsymbol{L}(l)\hat{\boldsymbol{v}}(l) + \boldsymbol{K}(l+1)\nu(l+1). \tag{3.49}$$

Dabei steht $\nu(l+1)$ für den Innovationsprozeß, der unter anderem von der zeitlichen Ableitung des Bildes auf Stufe $l+1$ abhängt:

$$\begin{aligned} \nu(l+1) &= -g_t(l+1) - \nabla_s \boldsymbol{g}^T(l+1)\boldsymbol{L}(l)\hat{\boldsymbol{v}}(l) \\ &= -\frac{\partial}{\partial t}\boldsymbol{g}(\boldsymbol{x} + (t_0+t)\boldsymbol{L}(l)\hat{\boldsymbol{v}}(l), t_0+t). \end{aligned}$$

Berücksichtigt man nun, daß vor dem Bilden der zeitlichen Ableitung das Bild mit der unterabgetasteten, interpolierten Version des zuvor berechneten optischen Flusses gewarpt werden soll, so kann man den Ausdruck für die Innovation zum Zeitpunkt t_0 folgendermaßen approximieren:

$$\nu(l+1) \approx -\frac{\partial}{\partial t}\mathcal{W}\{\boldsymbol{g}(l+1), \boldsymbol{L}(l)\hat{\boldsymbol{v}}(l)\}(x, y, t_0), \tag{3.50}$$

wobei das Warpen als fehlerfrei vorausgesetzt wird.

Für die Verstärkungsmatrix $\boldsymbol{K}(l+1)$ gilt analog zu (3.23):

$$\boldsymbol{K}(l+1) = \boldsymbol{\Lambda}^-(l+1)\nabla_s\boldsymbol{g}(l+1)[\nabla_s\boldsymbol{g}^T(l+1)(\boldsymbol{\Lambda}^-(l+1) + \boldsymbol{\Lambda_v})\nabla_s\boldsymbol{g}(l+1) + \lambda_t]^{-1}, \tag{3.51}$$

wobei die Prädiktion der Fehlerkovarianzmatrix durch den folgenden Ausdruck gegeben ist:

$$\boldsymbol{\Lambda}^-(l+1) = \boldsymbol{L}(l)\boldsymbol{\Lambda}(l)\boldsymbol{L}(l)^T + \boldsymbol{Q}.$$

Die Korrektur der Fehlerkovarianzmatrix der Schätzung lautet:

$$\boldsymbol{\Lambda}(l+1) = (\boldsymbol{I} - \boldsymbol{K}(l+1)\nabla_s \boldsymbol{g}^T(l+1))\boldsymbol{\Lambda}^-(l+1).$$

Unter Ausnutzung der Matrizengleichung

$$[\boldsymbol{B} - \boldsymbol{B}\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}^T + \boldsymbol{A})^{-1}\boldsymbol{C}\boldsymbol{B}]^{-1} = \boldsymbol{B}^{-1} + \boldsymbol{C}^T\boldsymbol{A}^{-1}\boldsymbol{C}$$

erhält man aus der Inversen der korrigierten Kovarianzmatrix mit

$$\boldsymbol{B} = \boldsymbol{\Lambda}^-(l+1), \quad \boldsymbol{C} = \nabla_s \boldsymbol{g}^T \text{ und } \boldsymbol{A} = (\nabla_s \boldsymbol{g}^T \boldsymbol{\Lambda}_{\boldsymbol{v}} \nabla_s \boldsymbol{g} + \lambda_t)$$

die Korrektur der Kovarianzmatrix:

$$\begin{aligned} \boldsymbol{\Lambda}(l+1) &= \left[\boldsymbol{\Lambda}^-(l+1)^{-1} + \nabla_s \boldsymbol{g}(\nabla_s \boldsymbol{g}^T \boldsymbol{\Lambda}_{\boldsymbol{v}} \nabla_s \boldsymbol{g} + \lambda_t)^{-1} \nabla_s \boldsymbol{g}^T\right]^{-1} \\ &= \left[\boldsymbol{\Lambda}^-(l+1)^{-1} + \frac{\boldsymbol{M}}{(\lambda_{\boldsymbol{v}} \|\nabla_s \boldsymbol{g}\|^2 + \lambda_t)}\right]^{-1}. \end{aligned} \tag{3.52}$$

Setzt man nun die Verstärkung (3.51) und die Innovation (3.50) in die Schätzung (3.49) ein, so erhält man:

$$\begin{aligned} \hat{\boldsymbol{v}}(l+1) &= \boldsymbol{L}(l)\hat{\boldsymbol{v}}(l) + \boldsymbol{\Lambda}(l+1)\nabla_s \boldsymbol{g}(\nabla_s \boldsymbol{g}^T \boldsymbol{\Lambda}_{\boldsymbol{v}} + \lambda_t)^{-1}\nu(l+1) \\ &= \boldsymbol{L}(l)\hat{\boldsymbol{v}}(l) + \boldsymbol{\Lambda}(l+1)\frac{\nabla_s \boldsymbol{g}\nu(l+1)}{(\lambda_{\boldsymbol{v}} \|\nabla_s \boldsymbol{g}\|^2 + \lambda_t)}. \end{aligned} \tag{3.53}$$

Vergleicht man die beiden Ausdrücke mit (3.44) bzw. (3.45), stellt man fest, daß die Gleichungen große Ähnlichkeit besitzen. Die einzigen Unterschiede liegen darin, daß beim hierarchischen Verfahren das Warpen der Bilder der eigentlichen Berechnung vorausgeht. Daher wird die auf das feinere Gitter interpolierte Schätzung des optischen Flusses von der gröberen Stufe addiert und der Vektor $\boldsymbol{b}$ aus (3.45) aus der zeitlichen Ableitung der gewarpten Version gewonnen.

Man beginnt bei dem hierarchischen Berechnungsalgorithmus damit, daß man eine Bildpyramide für jedes Bild der Sequenz generiert. Hierzu eignet sich die Gaußpyramide, die durch Tiefpaßfilterung mit einem Binomialfilter und Unterabtasten entsteht, wobei jeder zweite Bildpunkt einer Zeile bzw. Spalte weggelassen wird. Auf der gröbsten und damit kleinsten Stufe der Pyramide wird nun der optische Fluß gemäß (3.47) und die Fehlerkovarianzmatrix gemäß (3.46) berechnet. Simoncelli et al. benutzen für die gröbste Stufe, also die erste Berechnung des optischen Flusses, das MAP–Verfahren, wobei sie als inverse a priori Varianz einen wählbaren Parameter λ_p benutzen, so daß die a priori Kovarianzmatrix $\boldsymbol{\Lambda}^-(0)$ die Form einer Diagonalmatrix mit den Elementen $\frac{1}{\lambda_p}$ hat. Sie gewährleisten damit die Invertierbarkeit der Kovarianzmatrix.

Das Ergebnis wird dann auf das feinere Gitter interpoliert, wobei beachtet werden muß, daß der grob geschätzte Fluß auf dem feineren Gitter doppelt so groß ist und die Varianz sogar den vierfachen Wert hat:

$$\begin{aligned} \boldsymbol{v}_{\text{fein}} &= 2\boldsymbol{v}_{\text{grob}}, \\ \boldsymbol{\Lambda}_{\text{fein}} &= 4\boldsymbol{\Lambda}_{\text{grob}}. \end{aligned}$$

Der interpolierte Fluß wird dann dazu benutzt, das Bild der feinen Auflösung zu warpen. Als nächstes wird der optische Fluß auf dieser Auflösungsstufe durch (3.53) berechnet. Dadurch stellt der Fluß auf dieser Stufe eine Korrektur des bisher berechneten Flusses dar. Man wiederholt den Prozeß bestehend aus Warpen und Korrektur solange, bis auch die feinste Stufe, also das Originalbild, verarbeitet wurde.

Übertragung der hierarchischen Berechnung auf das affine Modell

Geht man vom affinen Modell des optischen Flusses (2.16) aus, dann sind einige Änderungen des hierarchischen Verfahrens vorzunehmen. Das Prinzip des Algorithmus' bleibt erhalten, jedoch ändern sich bei den Berechnungen in jeder Pyramidenebene die Komponenten.

Simoncelli et al. betrachten den Modellfehler separat, indem sie den gemessenen Fluß in exakten Fluß und Modellfehler aufteilen, $\boldsymbol{v} = \tilde{\boldsymbol{v}} + \delta\boldsymbol{v}$. In der Literatur zur Parameterschätzung werden Meßmodellfehler nicht durch eine Modellierung von additivem Zustandsrauschen, sondern als Bestandteil des Meßrauschens betrachtet. Wir wollen daher im folgenden diese Trennung aufheben und den Modellfehler mit dem Rauschen in der zeitlichen Ableitung in dem zeitlichen Rauschparameter vereinigt betrachten. Das hat zur Folge, daß in allen Berechnungen, die auf die Rauschparameter zurückgreifen, der Ausdruck $\lambda_{\boldsymbol{v}}\|\nabla_s \boldsymbol{g}\|^2 + \lambda_t$ durch den beide Rauschquellen umfassenden Ausdruck λ'_t ersetzt werden muß.

Wenn in die bedingte Wahrscheinlichkeit (3.38) außerdem für den lokal konstanten Fluß $\boldsymbol{v}$ der affine Fluß $\boldsymbol{v}_{\text{aff}}$ eingesetzt wird, liefert das Maximieren der a posteriori Wahrscheinlichkeit über Ableiten nach jeder Komponente von $\hat{\boldsymbol{v}}_{\text{aff}}$ aus (2.18) und Nullsetzen jeder dieser Ableitungen die (3.46) und (3.47) entsprechenden Ergebnisse, wobei die Matrix $\boldsymbol{M}$ und der Vektor $\boldsymbol{b}$ abgewandelt sind. Der Vektor $\boldsymbol{b}$ wird durch $\hat{\boldsymbol{b}}$ aus (2.18) ersetzt, während die Matrix $\boldsymbol{M}$ durch $\hat{\boldsymbol{M}}$ aus (2.18) substituiert wird. Durch das affine Modell geht eine Art Gewichtung der Bildpunkte der betrachteten Region über ihre relative Lage in die Matrix $\boldsymbol{M}$ ein. Die Applikation einer speziellen Wichtungsfunktion ω verfälscht daher den affinen Fluß und wird daher weggelassen. Man erhält demnach insgesamt für die Schätzung mittels der MAP–Strategie:

$$\boldsymbol{\Lambda}_{\hat{\boldsymbol{v}}_{\text{aff}}} = \left[\sum_i \frac{\hat{\boldsymbol{M}}(x_i, y_i, t)}{\lambda'_{t_i}} + \boldsymbol{\Lambda}_p^{-1}\right]^{-1}, \tag{3.54}$$

$$\hat{\boldsymbol{v}}_{\text{aff}} = -\boldsymbol{\Lambda}_{\hat{\boldsymbol{v}}_{\text{aff}}} \sum_i \frac{\hat{\boldsymbol{b}}(x_i, y_i, t)}{\lambda'_{t_i}}. \tag{3.55}$$

Die oben erwähnten Ersetzungen sind auch bei der Kombination der Ergebnisse von zwei Pyramidenstufen zu beachten. Die Gleichungen (3.52) und (3.53) werden zu

$$\boldsymbol{\Lambda}_{\text{aff}}(l+1) = \left[\boldsymbol{\Lambda}_{\text{aff}}(l+1)^{-1} + \frac{\hat{\boldsymbol{M}}}{\lambda'_t}\right]^{-1} \tag{3.56}$$

und

$$\hat{\boldsymbol{v}}_{\text{aff}}(l+1) = \boldsymbol{L}(l)\hat{\boldsymbol{v}}_{\text{aff}}(l) + \boldsymbol{\Lambda}_{\text{aff}}(l+1)\frac{\boldsymbol{b}_{\text{aff}}\nu(l+1)}{\lambda_t'} \tag{3.57}$$

mit

$$\boldsymbol{b}_{\text{aff}} = \begin{pmatrix} \boldsymbol{g}_x \\ \boldsymbol{g}_y \\ \Delta x \boldsymbol{g}_x \\ \Delta y \boldsymbol{g}_x \\ \Delta x \boldsymbol{g}_y \\ \Delta y \boldsymbol{g}_y \end{pmatrix}. \tag{3.58}$$

In die Matrix $\hat{\boldsymbol{M}}$, die in der Kovarianzmatrix (3.55) bzw. (3.56) enthalten ist, gehen Gittergrößen $\Delta\boldsymbol{x}$ ein. Die Anpassung der Kovarianzmatrix an das nächst feinere Gitter kann daher nicht durch eine Vervierfachung erreicht werden. Statt dessen werden alle Elemente, die unabhängig von der Lage sind (z.B. $\boldsymbol{g}_x^2, \boldsymbol{g}_x\boldsymbol{g}_y$), mit vier, alle Elemente, in die die relativen Lage linear eingeht (z.B. $\Delta x\boldsymbol{g}_x^2, \Delta y\boldsymbol{g}_x\boldsymbol{g}_y$), mit zwei multipliziert, während alle anderen unverändert bleiben. Da auch der affine Fluß selbst über Gittergrößen definiert ist, gilt für das Anpassen des Flusses vom groben Gitter an das feine folgender Zusammenhang, der bei der Interpolation berücksichtigt werden muß:

$$\boldsymbol{v}_{\text{aff}_{l+1}} = 2\boldsymbol{v}_{0_l} + \boldsymbol{v}_{x_l}\Delta x_{l+1} + \boldsymbol{v}_{y_l}\Delta y_{l+1}. \tag{3.59}$$

Abhängig von seiner relativen Lage zum gewählten Zentrum der Bewegung, $(x, y) = (0, 0)$, wird dann jeder Bildpunkt mit dem an das feinere Gitter angepaßten Fluß verschoben.

Betrachtet man dabei nicht das ganze Bild als eine Region, in der der optische Fluß linearen Verlauf hat, stößt man auf die Schwierigkeit, daß der Einzugsbereich der betrachteten Region, die den Ausschnitt des linearen Verlaufs bestimmt, für jedes Bild in der Umgebung der Maskengröße der zeitlichen Ableitung vorgegeben sein muß. Dieses Einbringen von Vorwissen sollte möglichst vermieden werden. Als Alternative kann man die den Einzugsbereich des betrachteten Bildes bestimmende Blende ebenfalls mit dem berechneten Fluß warpen.

3.2.4 Rauschen in allen Ableitungen – ML–Strategie

Das stochastische Rauschmodell von Simoncelli et al. stellt eine Erweiterung der herkömmlichen Flußberechnung dar. Es entspricht jedoch keinesfalls realen Verhältnissen, daß nur die zeitliche Ableitung verrauscht ist. Es ist daher sinnvoll, auch das Rauschen in den örtlichen Ableitungen in die Berechnung mit einzubeziehen.

Ein auf stochastischen Betrachtungen von Kanatani [21] aufbauendes Berechnungsmodell berücksichtigt das Rauschen in allen Komponenten, wobei die Exaktheit der BCCE vorausgesetzt wird. Kanatani nimmt an, daß das Meßrauschen in allen Ableitungen durch

Normalverteilungen gegeben wird. Sein Rauschmodell lautet analog zu (3.34) und (3.35):

$$\nabla \boldsymbol{g} = \nabla \tilde{\boldsymbol{g}} + \boldsymbol{e} \tag{3.60}$$

mit $\boldsymbol{e} = (\delta g_x, \delta g_y, \delta g_t)^T$ und $\boldsymbol{e} \sim N(0, \boldsymbol{\Lambda}_e)$.

Wählt man für den Fluß $\boldsymbol{v}$ eine dreidimensionale Darstellung $\boldsymbol{v}_{\mathbf{3D}} = (u, v, 1)^T$, so erhält man folgende Formulierung der Kontinuitätsgleichung:

$$\nabla \tilde{\boldsymbol{g}}^T \boldsymbol{v}_{\mathbf{3D}} = (\nabla \boldsymbol{g}^T - \boldsymbol{e}^T) \boldsymbol{v}_{\mathbf{3D}} = 0, \tag{3.61}$$

wobei $\boldsymbol{e}^T = (\delta g_x, \delta g_y, \delta g_t)$.
Für die Diskrepanz zwischen optischem Fluß und VVF wird der Fehlerterm $\delta_{\boldsymbol{v}}$ in (3.61) eingeführt:

$$\nabla \boldsymbol{g}^T \boldsymbol{v}_{\mathbf{3D}} = \delta_{\boldsymbol{v}}.$$

Die bedingte Wahrscheinlichkeitsdichte $p(\nabla \boldsymbol{g} | \boldsymbol{v}_{\mathbf{3D}})$ an jedem einzelnen Bildpunkt i ist analog zu (3.37) definiert durch:

$$\begin{aligned} p_i(\nabla \boldsymbol{g} \mid \boldsymbol{v}_{\mathbf{3D}}) &= \frac{1}{\sqrt{2\pi}\lambda} \exp\left\{ -\frac{(\nabla \boldsymbol{g}^T \boldsymbol{v}_{\mathbf{3D}})^2}{2\lambda^2} \right\} \\ &\propto \exp\left\{ -\frac{(\nabla \boldsymbol{g}^T \boldsymbol{v}_{\mathbf{3D}})^2}{2\lambda^2} \right\}. \end{aligned} \tag{3.62}$$

Dabei gilt nach (3.61) für die Varianz:

$$\begin{aligned} \lambda^2 &= \mathrm{E}[(\boldsymbol{e}^T \boldsymbol{v}_{\mathbf{3D}})^2] \\ &= \boldsymbol{v}_{\mathbf{3D}}^T \boldsymbol{\Lambda}_e \boldsymbol{v}_{\mathbf{3D}} = \boldsymbol{v}_{\mathbf{3D}}^T \begin{pmatrix} \lambda_s^2 & 0 & 0 \\ 0 & \lambda_s^2 & 0 \\ 0 & 0 & \lambda_t^2 \end{pmatrix} \boldsymbol{v}_{\mathbf{3D}}. \end{aligned} \tag{3.63}$$

Wegen des Aperturproblems macht man wiederum die Annahme, daß der Fluß in einer gewissen Region konstant ist. Die a priori Wahrscheinlichkeit für die Region lautet als Produkt der Einzelwahrscheinlichkeiten (3.62):

$$p(\nabla \boldsymbol{g} \mid \boldsymbol{v}_{\mathbf{3D}}) \propto \exp\left\{ -\sum_i \frac{(\nabla \boldsymbol{g}_i^T \boldsymbol{v}_{\mathbf{3D}})^2}{2\lambda_i^2} \right\}. \tag{3.64}$$

Man sucht nach (3.2) also nun denjenigen Vektor, der (3.64) maximiert. Da man von Normalverteilungen und einer linearen Meßfunktion ausgeht, ist dies gerade der Vektor, der den Betrag des Exponenten minimiert:

$$\hat{\boldsymbol{v}}_{\mathbf{3D}} = \arg\min_{\boldsymbol{v}_{\mathbf{3D}}} \sum_i \frac{(\nabla \boldsymbol{g}_i^T \boldsymbol{v}_{\mathbf{3D}})^2}{(\boldsymbol{v}_{\mathbf{3D}}^T \boldsymbol{\Lambda}_{e_i} \boldsymbol{v}_{\mathbf{3D}})} =: \arg\min_{\boldsymbol{v}_{\mathbf{3D}}} J. \tag{3.65}$$

Unter der Voraussetzung, daß an allen Bildpunkten die Varianz gleich ist, $\boldsymbol{\Lambda}_{e_i} = \boldsymbol{\Lambda}_e$, kann man das Minimierungsproblem umformulieren. Man zieht die Summe in den Zähler und setzt $\boldsymbol{A} = \sum_i \nabla \boldsymbol{g}_i \nabla \boldsymbol{g}_i^T$, so daß man als zu minimierende Funktion den verallgemeinerten Rayleighschen Quotienten erhält:

$$\hat{\boldsymbol{v}}_{\mathbf{3D}} = \arg\min_{\boldsymbol{v}_{\mathbf{3D}}} \frac{\boldsymbol{v}_{\mathbf{3D}}^T \boldsymbol{A} \boldsymbol{v}_{\mathbf{3D}}}{\boldsymbol{v}_{\mathbf{3D}}^T \boldsymbol{\Lambda}_e \boldsymbol{v}_{\mathbf{3D}}}. \tag{3.66}$$

Auch hier kann man zur Gewichtung der Beiträge der einzelnen Bildpunkte wieder eine Wichtungsfunktion $\boldsymbol{\omega}$ einführen, so daß die Matrix $\boldsymbol{A}$ folgende Form annimmt:

$$\boldsymbol{A} = \sum_i (\omega_i \boldsymbol{g}_i \boldsymbol{g}_i^T).$$

Da die Kovarianzmatrix $\boldsymbol{\Lambda}_e$ die Form einer Diagonalmatrix hat, kann man durch Wurzelziehen leicht eine Diagonalmatrix $\boldsymbol{B}$ finden, so daß $\boldsymbol{\Lambda}_e = \boldsymbol{B}^T \boldsymbol{B}$ gilt. Sinnvollerweise sind die Varianzwerte λ_s und λ_t ungleich null zu wählen, weswegen ebenfalls $\boldsymbol{B}^{-1}$ existiert, so daß man $\boldsymbol{v}_{\mathbf{3D}}$ durch $\boldsymbol{v}_{\mathbf{3D}} = \boldsymbol{B}^{-1} \boldsymbol{v}'_{\mathbf{3D}} = \boldsymbol{B}^{-1}(\boldsymbol{B} \boldsymbol{v}_{\mathbf{3D}})$ ersetzen kann. So wird das Problem zur Minimierung des verallgemeinerten symmetrischen Eigenwertproblems vereinfacht:

$$\hat{\boldsymbol{v}}_{\mathbf{3D}} = \boldsymbol{B}^{-1} \arg\min_{\boldsymbol{v}'_{\mathbf{3D}}} \frac{\boldsymbol{v}'^T_{\mathbf{3D}} \boldsymbol{B}^{-1} \boldsymbol{A} \boldsymbol{B}^{-1} \boldsymbol{v}'_{\mathbf{3D}}}{\boldsymbol{v}'^T_{\mathbf{3D}} \boldsymbol{v}'_{\mathbf{3D}}}.$$

Setzt man nun $\boldsymbol{v}''_{\mathbf{3D}} = \frac{\boldsymbol{v}'_{\mathbf{3D}}}{\|\boldsymbol{v}'_{\mathbf{3D}}\|}$ und $\boldsymbol{C} = \boldsymbol{B}^{-1} \boldsymbol{A} \boldsymbol{B}^{-1}$ erhält man als ein zu (3.65) äquivalentes Minimierungsproblem das Eigenwertminimierungsproblem:

$$\hat{\boldsymbol{v}}_{\mathbf{3D}} \propto \boldsymbol{B}^{-1} \arg\min_{\boldsymbol{v}''_{\mathbf{3D}}} \boldsymbol{v}''_{\mathbf{3D}} \boldsymbol{C} \boldsymbol{v}''_{\mathbf{3D}}. \tag{3.67}$$

Gesucht wird also der normierte Eigenvektor zum kleinsten Eigenwert von Matrix $\boldsymbol{C}$. Diesen findet man wegen der symmetrischen Form von $\boldsymbol{C}$ über Minimierung der Eigenwertgleichung mit Nebenbedingung:

$$\boldsymbol{v}''_{\mathbf{3D}} = \arg\min_{\boldsymbol{x}} \{\boldsymbol{x}^T \boldsymbol{C} \boldsymbol{x} + \lambda(1 - \boldsymbol{x}^T \boldsymbol{x})\}. \tag{3.68}$$

Die Lösung $\hat{\boldsymbol{v}}_{\mathbf{3D}}$ des optischen Flusses erhält man aus $\boldsymbol{v}''_{\mathbf{3D}}$, indem man die dritte Komponente des Vektors auf den Wert eins normiert und den resultierenden Vektor anschließend mit $\boldsymbol{B}^{-1}$ multipliziert.

Im folgenden wird nun die Beziehung zwischen (3.65) und den frequenzbasierten Verfahren [16] geklärt. Man geht davon aus, daß man eine lokale spektrale Repräsentation der Bildfolge gefunden hat, mit der man eine Approximation des Spektrums $\boldsymbol{G}(f_x, f_y, f_t)$ erreicht hat. Aus (2.20) weiß man, daß die Energie bei lokal konstantem Fluß in der Ebene $f_x u + f_y v + f_t = 0$ konzentriert ist. Es ist naheliegend, die Neigung dieser Ebene mittels folgender Anpassung zu finden:

$$\boldsymbol{n} = \arg\min_{\boldsymbol{n}, \|\boldsymbol{n}\|=1} \sum_{f_x} \sum_{f_y} \sum_{f_t} \left(\boldsymbol{n}^T \begin{pmatrix} f_x \\ f_y \\ f_t \end{pmatrix} \right)^2 |\boldsymbol{G}(f_x, f_y, f_t)|^2, \tag{3.69}$$

wobei $\boldsymbol{n} = (u, v, 1)^T$ die gesuchte Ebenennormale ist.
Das Parseval-Theorem stellt einen Zusammenhang zwischen der Summe der Betragsquadrate im Orts- und der im Frequenzraum her. Für ein Grauwertsignal $\boldsymbol{h}(x, y, t)$ und seine Darstellung im Fourierraum $\boldsymbol{H}(f_x, f_y, f_t)$ besagt es:

$$\int_x \int_y \int_t |\boldsymbol{h}(x,y,t)|^2 dx\, dy\, dt = \int_{f_x} \int_{f_y} \int_{f_t} |\boldsymbol{H}(f_x, f_y, f_t)|^2 df_x df_y df_t. \tag{3.70}$$

Wir setzen

$$\begin{array}{ccccccc}
\boldsymbol{h}(x,y,t) & = & g_x u & + & g_y v & + & g_t \\
\circ & & \circ & & \circ & & \circ \\
| & & | & & | & & | \\
\bullet & & \bullet & & \bullet & & \bullet \\
\boldsymbol{H}(f_x, f_y, f_t) & = & \imath 2\pi f_x \boldsymbol{G} u & + & \imath 2\pi f_y \boldsymbol{G} v & + & \imath 2\pi f_t \boldsymbol{G}.
\end{array}$$

Nähert man die Integrale in (3.70) im diskreten Fall durch Summen an, liefert das Parseval-Theorem nur noch Proportionalität:

$$\sum_x \sum_y \sum_t |\boldsymbol{h}(x,y,t)|^2 dx\, dy\, dt \propto \sum_{f_x} \sum_{f_y} \sum_{f_t} |\boldsymbol{H}(f_x, f_y, f_t)|^2 df_x df_y df_t.$$

Der Proportionalitätsfaktor wird durch die Integrationsbereiche bzw. Summationsbereiche bestimmt.

Da eine Minimierung von proportionalen Ausdrücken das gleiche Ergebnis liefert, wenn der Proportionalitätsfaktor nicht von der Minimierungsvariablen abhängt, stellt man fest, daß das oben erklärte Verfahren von Kanatani zum Verfahren der Ebenenanpassung im Frequenzraum äquivalent ist, falls man im Bildfolgenraum zusätzlich eine zeitliche Nachbarschaft mit berücksichtigt.

Da es wichtig ist, eine Aussage über die Güte des berechneten optischen Flusses machen zu können, ist es notwendig, die Kovarianzmatrix $\boldsymbol{\Lambda_v}$ des geschätzten Flusses zu bestimmen. Dazu führt man, wie in Kapitel 7.1 in [21] nachzulesen ist, folgende Schritte aus:

1. Approximation der Funktion J aus (3.65) durch eine Quadrik–Funktion $\tilde{J}$, die von Verteilungen $\boldsymbol{v}_{\mathbf{3D}}$ und $\nabla \boldsymbol{g}$ um ihre exakten Werte $\tilde{\boldsymbol{v}}_{\mathbf{3D}}$ und $\nabla \tilde{\boldsymbol{g}}$ abhängt.
2. Bestimmung des Wertes $\Delta \boldsymbol{v}_{\mathbf{3D}}$, wobei $\boldsymbol{v}_{\mathbf{3D}} = \tilde{\boldsymbol{v}}_{\mathbf{3D}} + \Delta \boldsymbol{v}_{\mathbf{3D}}$ gilt, der als lineare Funktion von $\nabla \boldsymbol{g}$ die Funktion $\tilde{J}$ minimiert.
3. Berechnung der Kovarianzmatrix durch $\boldsymbol{\Lambda_v} = \mathrm{E}[\Delta \boldsymbol{v}_{\mathbf{3D}} \Delta \boldsymbol{v}_{\mathbf{3D}}^T]$.

Es läßt sich beweisen [21], daß man als Ergebnis folgende Pseudoinverse erhält:

$$\tilde{\boldsymbol{\Lambda}}_{\hat{v}} = \left(\sum_{i=1}^{N} \frac{\nabla_s \tilde{\boldsymbol{g}}_i \nabla_s \tilde{\boldsymbol{g}}_i^T}{\tilde{\boldsymbol{v}}_{\mathbf{3D}}^T \boldsymbol{\Lambda}_e \tilde{\boldsymbol{v}}_{\mathbf{3D}}} \right)^+ . \tag{3.71}$$

Bei einer konkreten Berechnung werden die exakten Werte in (3.71) durch die geschätzten Werte des optischen Flusses und die gemessenen Grauwertgradienten angenähert:

$$\boldsymbol{\Lambda}_{\hat{\boldsymbol{v}}} = \left(\sum_{i=1}^{N} \frac{(\boldsymbol{P}\nabla \boldsymbol{g}_i)(\boldsymbol{P}\nabla \boldsymbol{g}_i)^T}{(\hat{\boldsymbol{v}}_{\mathbf{3D}}^T \boldsymbol{\Lambda}_e \hat{\boldsymbol{v}}_{\mathbf{3D}})} \right)^{+}, \tag{3.72}$$

wobei $\boldsymbol{P} = \operatorname{diag}(1,1,0)$ gilt.
Daher kann man (3.72) auch folgendermaßen ausdrücken:

$$\boldsymbol{\Lambda}_{\hat{\boldsymbol{v}}} = \begin{pmatrix} \boldsymbol{\Lambda}_{(\hat{u},\hat{v})} & 0 \\ 0 & 0 \end{pmatrix}, \tag{3.73}$$

wobei $\boldsymbol{\Lambda}_{(\hat{u},\hat{v})}$ eine 2×2-Matrix ist, die die folgende, Gleichung (12.12) aus Kapitel 12.1.2 in [21] entsprechende, Form hat:

$$\boldsymbol{\Lambda}_{(\hat{u},\hat{v})} = \hat{\boldsymbol{v}}_{\mathbf{3D}}^T \boldsymbol{\Lambda}_e \hat{\boldsymbol{v}}_{\mathbf{3D}} \left(\sum_{i=1}^{N} \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T \right)^{-1}. \tag{3.74}$$

Die Unabhängigkeit der Gewichtung der Matrixelemente von der Summe resultiert aus der Annahme, daß das Rauschen in den örtlichen Gradienten isotrop und homogen und das Rauschen im zeitlichen Gradienten vom Rauschen in den örtlichen unabhängig ist. Die Inverse existiert nicht in jedem Fall, sondern nur, wenn die Gradientenmatrix $\sum_{i=1}^{N} \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T$ den Rang zwei hat, was uns auf das Aperturproblem verweist.

Das Aperturproblem

Das Aperturproblem kann formal beschrieben werden durch:

$$\det\Big(\sum_i \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T\Big) = 0. \tag{3.75}$$

Welche Auswirkungen hat diese Bedingung auf das Verfahren von Kanatani? Wie sieht in diesem Fall der Lösungsraum für $\boldsymbol{v}_{\mathbf{3D}}$ aus?
Da der kleinste Eigenwert der Matrix $\boldsymbol{C}$ aus (3.67) im exakt lösbaren Fall immer den Wert null annimmt, erwarten wir, daß beim Auftreten des Aperturproblems der Eigenwert null entartet ist. Dann ist keine eindeutige Aussage über den Lösungsvektor $\boldsymbol{v}_{\mathbf{3D}}$ möglich.

Dazu muß der Kern der Matrix im exakten Fall mehr als einen nichttrivialen Vektor enthalten, d.h. die Determinante der Matrix $\boldsymbol{C}$ mit $N > 2$ muß null werden. Die Matrix $\boldsymbol{C}$ aus (3.67) hat die Form:

$$\begin{aligned} \boldsymbol{C} &= \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}^{-1} \\ &= \begin{pmatrix} \frac{1}{\sqrt{\lambda_s}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\lambda_s}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\lambda_t}} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{N} \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T & \sum_{i=1}^{N} g_{t_i} \nabla_s \boldsymbol{g}_i \\ \sum_{i=1}^{N} g_{t_i} \nabla_s \boldsymbol{g}_i^T & \sum_{i=1}^{N} g_{t_i}^2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\lambda_s}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\lambda_s}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\lambda_t}} \end{pmatrix}. \end{aligned}$$

Wegen der Eigenschaften der Determinanten (vgl. [11]) gilt:

$$\begin{aligned}\det \boldsymbol{C} &= \frac{1}{\det \boldsymbol{B}} \det \boldsymbol{A} \frac{1}{\det \boldsymbol{B}} \\ &= \frac{1}{\lambda_s \sqrt{\lambda_t}} \det \boldsymbol{A} \frac{1}{\lambda_s \sqrt{\lambda_t}}.\end{aligned}$$

Deswegen reicht es zu beweisen, daß $\det \boldsymbol{A} = 0$ gilt.

Per Induktion über den Summationsindex zeigt man die intuitive Folgerung, daß Bedingung (3.75) bedeutet, daß die Grauwertgradienten parallel ausgerichtet sind oder ihr Betrag verschwindet:

$$\det(\sum_i \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T) = 0 \quad \Rightarrow \quad \forall i\, \exists \alpha_i : \nabla \boldsymbol{g}_i = \alpha_i \boldsymbol{n}, \quad \|\boldsymbol{n}\| = 1. \tag{3.76}$$

Beweis: (per Induktion über N)

I.A.: $(N = 1)$*:*
Es gelte $\det(\nabla_s \boldsymbol{g}_1 \nabla_s \boldsymbol{g}_1^T) = 0$.
Setze $\alpha_1 = \|\nabla_s \boldsymbol{g}_1\|$ und $\boldsymbol{n} = \frac{1}{\alpha_1} \nabla_s \boldsymbol{g}_1$.
Offensichtlich gilt dann: $\nabla_s \boldsymbol{g}_1 = \alpha_1 \boldsymbol{n}$ mit $\|\boldsymbol{n}\| = 1$.

I.V.: $(N \in I\!N)$: Es gelte $\forall k \leq N$:
$\det(\sum_{i=1}^k \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T) = 0 \quad \Rightarrow \quad \forall i \leq k\, \exists \alpha_i : \nabla_s \boldsymbol{g}_i = \alpha_i \boldsymbol{n} \quad \|\boldsymbol{n}\| = 1.$

I.S.: $(k \to k+1)$:
Es gelte: $\det(\sum_{i=1}^{k+1} \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T) = 0$.
Dann existiert $\boldsymbol{u} \neq \boldsymbol{0}$ mit $(\sum_{i=1}^{k+1} \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T)\boldsymbol{u} = \boldsymbol{0}$.
Man kann die Summe wie folgt trennen:

$$\sum_{i=1}^{k+1} \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T = \underbrace{(\sum_{i=1}^{k} \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T)}_{\text{nach I.V.: } \sum_{i=1}^k \alpha_i \boldsymbol{n}\boldsymbol{n}^T} + \nabla_s \boldsymbol{g}_{k+1} \nabla_s \boldsymbol{g}_{k+1}^T.$$

Setze nun $\alpha' = \sum_{i=1}^k \alpha_i \boldsymbol{n}^T \boldsymbol{u}$ und $\alpha'' = \nabla_s \boldsymbol{g}_{k+1}^T \boldsymbol{u}$, dann gilt:
$\alpha' \boldsymbol{n} + \alpha'' \nabla_s \boldsymbol{g}_{k+1} = \boldsymbol{0}$.
Da o.B.d.A. $\alpha'' \neq 0$ gilt, folgt somit:
$\nabla_s \boldsymbol{g}_{k+1} = -\frac{\alpha'}{\alpha''} \boldsymbol{n} =: \alpha_{k+1} \boldsymbol{n}$.

Nach dem Prinzip der vollständigen Induktion folgt die Behauptung.

$\square$

Die Determinante berechnet sich nun durch Streichen der dritten Zeile und Spalte zu:

$$\begin{aligned}
\det \boldsymbol{A} &= \det \begin{pmatrix} \sum_i \nabla_s \boldsymbol{g}_i \nabla_s \boldsymbol{g}_i^T & \sum_i g_{t_i} \nabla_s \boldsymbol{g}_i \\ \sum_i g_{t_i} \nabla_s \boldsymbol{g}_i^T & \sum_i g_{t_i}^2 \end{pmatrix} \\
&= \sum_i (g_{t_i} g_{x_i}) \begin{vmatrix} \sum_i (g_{x_i} g_{y_i}) & \sum_i (g_{t_i} g_{x_i}) \\ \sum_i g_{y_i}^2 & \sum_i (g_{t_i} g_{y_i}) \end{vmatrix} \\
&- \sum_i (g_{t_i} g_{y_i}) \begin{vmatrix} \sum_i g_{x_i}^2 & \sum_i (g_{t_i} g_{x_i}) \\ \sum_i (g_{x_i} g_{y_i}) & \sum_i (g_{t_i} g_{y_i}) \end{vmatrix} \\
&+ \sum_i g_{t_i}^2 \left| \sum_i \nabla \boldsymbol{g}_i \nabla \boldsymbol{g}_i^T \right| .
\end{aligned}$$

Aus der obigen Folgerung (3.76) erhalten wir:

$$\begin{aligned}
\begin{vmatrix} \sum_i (g_{x_i} g_{y_i}) & \sum_i g_{y_i}^2 \\ \sum_i (g_{t_i} g_{x_i}) & \sum_i (g_{t_i} g_{y_i}) \end{vmatrix} &= \begin{vmatrix} n_1 n_2 \sum_i \lambda_i^2 & n_2^2 \sum_i \lambda_i^2 \\ n_1 \sum_i (\lambda_i g_{t_i}) & n_2 \sum_i (\lambda_i g_{t_i}) \end{vmatrix} \\
&= n_1 n_2^2 \sum_i \lambda_i^2 \sum_i (\lambda_i g_{t_i}) - n_1 n_2^2 \sum_i \lambda_i^2 \sum_i (\lambda_i g_{t_i}) = 0.
\end{aligned}$$

Analog folgt:

$$\begin{vmatrix} \sum_i g_{x_i}^2 & \sum_i (g_{t_i} g_{x_i}) \\ \sum_i (g_{x_i} g_{y_i}) & \sum_i (g_{t_i} g_{y_i}) \end{vmatrix} = 0.$$

Daraus ergibt sich mit (3.75): $\det \boldsymbol{A} = 0 - 0 + 0 = 0$.

Damit ist bewiesen, daß beim Auftreten des Aperturproblems das Verfahren von Kanatani keine Lösung für den optischen Fluß liefert. Der Eigenwert null der Matrix $\boldsymbol{C}$ ist in diesem Fall entartet, so daß es keine eindeutige Lösung gibt.

3.2.5 Rauschen in allen Ableitungen – MAP–Strategie

Um die *MAP*–Strategie auf das Modell von Kanatani anwenden zu können, muß man einen diesem Modell entsprechenden Ausdruck für die a posteriori Wahrscheinlichkeit $p(\boldsymbol{v}_{\mathbf{3D}}|\nabla \boldsymbol{g})$ angeben:

$$\begin{aligned}
p(\boldsymbol{v}_{\mathbf{3D}} \mid \nabla \boldsymbol{g}) &\propto p(\nabla \boldsymbol{g} \mid \boldsymbol{v}_{\mathbf{3D}}) p(\boldsymbol{v}_{\mathbf{3D}}) \\
&= \exp\{-\frac{1}{2} \sum_i \frac{(\nabla \boldsymbol{g}_i^T \boldsymbol{v}_{\mathbf{3D}})^2}{\boldsymbol{v}_{\mathbf{3D}}^T \boldsymbol{\Lambda}_e \boldsymbol{v}_{\mathbf{3D}}}\} \exp\{-\frac{1}{2}(\boldsymbol{v}_{\mathbf{3D}}^T \boldsymbol{\Lambda}_p^{-1} \boldsymbol{v}_{\mathbf{3D}})\}.
\end{aligned}$$

Man geht davon aus, daß die a priori Wahrscheinlichkeit mit Mittelwert null und Varianz $\boldsymbol{\Lambda}_p$ normalverteilt ist. Kanatani macht zudem die Annahme, daß das Meßrauschen $\boldsymbol{e}$ durch eine dreidimensionale Normalverteilung mit Mittelwert null und Varianz $\boldsymbol{\Lambda}_e$ dargestellt werden kann.

Im allgemeinen ist die bedingte Wahrscheinlichkeitsdichte $p(\nabla\boldsymbol{g}|\boldsymbol{v}_{3\mathrm{D}})$ von unbestimmter Form. Man kann jedoch davon ausgehen, daß $\boldsymbol{x} - \boldsymbol{h}^{-1}(\boldsymbol{z})$ um den Mittelwert null annähernd normalverteilt ist. Hier entspricht der Zustand $\boldsymbol{x}$ dem optischen Fluß, und die Meßfunktion $\boldsymbol{h}$ ist implizit über die BCCE gegeben, was eine direkte Inversion der Meßfunktion verhindert. Trotzdem nimmt man an, daß die bedingte Wahrscheinlichkeit $p(\nabla\boldsymbol{g}|\boldsymbol{v}_{3\mathrm{D}})$ mit Kovarianzmatrix $\hat{\boldsymbol{\Lambda}}_{\hat{v}}$ aus (3.72) und Mittelwert $\hat{\boldsymbol{v}}_{3\mathrm{D}}$ aus (3.67) normalverteilt ist.

Entsprechend erhält man mit $\boldsymbol{v}_0$ als a priori Schätzung des Flusses für die a posteriori Dichte:

$$p(\boldsymbol{v}_{3\mathrm{D}} \mid \nabla\boldsymbol{g}) \quad \propto \quad \exp\{-\frac{1}{2}[(\boldsymbol{\mu}_v - \hat{\boldsymbol{v}}_{3\mathrm{D}})^T \boldsymbol{\Lambda}_{\hat{v}}^{-1}(\boldsymbol{\mu}_v - \hat{\boldsymbol{v}}_{3\mathrm{D}}) + (\boldsymbol{\mu}_v - \boldsymbol{v}_0)^T \boldsymbol{\Lambda}_p^{-1}(\boldsymbol{\mu}_v - \boldsymbol{v}_0)]\}. \tag{3.77}$$

Die Minimierung des Exponenten zur Maximierung der Wahrscheinlichkeit erreicht man, indem man die Ableitung des Exponenten nach $\boldsymbol{\mu}_v$ gleich null setzt und nach den Unbekannten auflöst:

$$(\boldsymbol{\Lambda}_{\hat{v}}^{-1} + \boldsymbol{\Lambda}_p^{-1})\boldsymbol{\mu}_v = \boldsymbol{\Lambda}_{\hat{v}}^{-1}\hat{\boldsymbol{v}}_{3\mathrm{D}} + \boldsymbol{\Lambda}_p^{-1}\boldsymbol{v}_0. \tag{3.78}$$

Daraus erhalten wir für die Kovarianzmatrix $\boldsymbol{P}$ der a posteriori Dichte:

$$\boldsymbol{P} \quad = \quad (\boldsymbol{\Lambda}_{\hat{v}}^{-1} + \boldsymbol{\Lambda}_p^{-1})^{-1}. \tag{3.79}$$

Das bedeutet, der Vektor $\hat{\boldsymbol{v}}_{\mathrm{MAP}}$, der die a posteriori–Wahrscheinlichkeit maximiert, entspricht $\boldsymbol{\mu}_v$ und ergibt sich als Summe aus dem geschätzten Vektor $\hat{\boldsymbol{v}}_{3\mathrm{D}}$ aus (3.67) und dem über die Kovarianzmatrix $\boldsymbol{P}^{-1}$ abgebildeten a priori–Fluß $\boldsymbol{v}_0$:

$$\hat{\boldsymbol{v}}_{\mathrm{MAP}} = \boldsymbol{\mu}_v = \boldsymbol{P}(\boldsymbol{\Lambda}_{\hat{v}}^{-1}\hat{\boldsymbol{v}}_{3\mathrm{D}} + \boldsymbol{\Lambda}_p^{-1}\boldsymbol{v}_0). \tag{3.80}$$

3.2.6 Rauschen in allen Ableitungen – hierarchisches Verfahren

Basierend auf dem von Kanatani vorgestellten Rauschmodell wird die Berechnung des optischen Flusses hierarchisch mit Warpen der Bilder durchgeführt. Dazu braucht man eine Berechnungsvorschrift für die Komposition der Ergebnisse zweier Ebenen der Bildpyramide. Im folgenden werden dazu zwei mögliche Varianten der Kombination vorgeschlagen, die aufgrund verschiedener Annahmen über den Warpvorgang die neuen Berechnungen unterschiedlich stark gewichten.

Im Anschluß an die Darstellung und den skalaren Vergleich dieser Kompositionsmöglichkeiten wird der gesamte hierarchische Algorithmus beschrieben. Zuletzt folgt eine Erläuterung der Möglichkeit, den affinen Fluß hierarchisch zu bestimmen.

Eine Möglichkeit der Komposition besteht darin, wie Simoncelli et. al. eine dem Kalman–Filter entsprechende Lösungsstrategie anzugeben. Die Messung des nach dem Warpen der Sequenz verbleibenden optischen Flusses auf einer Pyramidenstufe kann man als eine verschobene Messung

$$\boldsymbol{z} = \boldsymbol{v} - \boldsymbol{a} \tag{3.81}$$

darstellen, wobei $\boldsymbol{a}$ ein deterministischer Vektor ist.
Die Zustandsgleichung in diesem Fall entspricht (3.48). Der nach dem Kalman–Filter–Verfahren optimale Schätzer sieht dann analog zu (3.22) folgendermaßen aus:

$$\hat{\boldsymbol{v}}(l+1) = \boldsymbol{L}(l)\hat{\boldsymbol{v}}(l) + \boldsymbol{K}(l+1)(\boldsymbol{z} - \hat{\boldsymbol{v}}(l) + \boldsymbol{a}). \tag{3.82}$$

Es gibt nun zwei Möglichkeiten: Entweder geht man wie Simoncelli et al. von fehlerfreiem Warpen und einer guten Anfangsschätzung aus ($\boldsymbol{a} = \boldsymbol{L}(l)\hat{\boldsymbol{v}}(l)$), oder man nimmt im Gegensatz dazu an, daß die Anfangsschätzung und damit das Warpen fehlerbehaftet ist ($\boldsymbol{a} \approx \boldsymbol{L}(l)\hat{\boldsymbol{v}}(l)$). Im ersten Fall ist die a posteriori Kovarianzmatrix gegeben durch (3.79):

$$\boldsymbol{\Lambda}(l+1) = [\boldsymbol{\Lambda}_{\hat{v}}^{-1}(l+1) + (\boldsymbol{L}(l)\boldsymbol{\Lambda}(l)\boldsymbol{L}^T(l))^{-1}]^{-1}. \tag{3.83}$$

Andernfalls gilt für die Varianz der Schätzung des Restflusses $\boldsymbol{v}(l+1) - \boldsymbol{L}(l)\boldsymbol{v}(l)$:

$$\boldsymbol{\Lambda}_{\tilde{v}}(l+1) = \boldsymbol{\Lambda}_{\hat{v}}(l+1) + \boldsymbol{\Lambda}_{\boldsymbol{a}}(l+1).$$

Man erhält damit aus der zu (3.77) analogen a posteriori Dichte auf der feineren Stufe die a posteriori Kovarianzmatrix, nämlich:

$$\boldsymbol{\Lambda}(l+1) = [(\boldsymbol{\Lambda}_{\hat{v}}(l+1) + \boldsymbol{\Lambda}_{\boldsymbol{a}}(l+1))^{-1} + (\boldsymbol{L}(l)\boldsymbol{\Lambda}(l)\boldsymbol{L}^T(l))^{-1}]^{-1}. \tag{3.84}$$

Die Messung resultiert direkt aus der Eigenvektorbestimmung (3.67), was bedeutet, daß die Messungsmatrix $\boldsymbol{H}(l)$ die Identitätsmatrix ist.

Damit ergibt sich für die Verstärkungsmatrix:

$$\boldsymbol{K}(l+1) = \boldsymbol{\Lambda}(l+1)\boldsymbol{H}^T(l+1)\boldsymbol{R}^{-1} = \boldsymbol{\Lambda}(l+1)\boldsymbol{R}^{-1}.$$

Bezeichnet man mit $\boldsymbol{v}_{\text{warped}}$ den mittels (3.67) auf der Pyramidenstufe $l+1$ berechneten Restfluß, so gilt im Fall des fehlerfreien Warpens:

$$\hat{\boldsymbol{v}}(l+1) = \boldsymbol{L}(l)\boldsymbol{v}(l) + [\boldsymbol{\Lambda}_{\hat{v}}^{-1}(l+1) + (\boldsymbol{L}(l)\boldsymbol{\Lambda}(l)\boldsymbol{L}^T(l))^{-1}]^{-1}\boldsymbol{\Lambda}_{\hat{v}}^{-1}(l+1)\boldsymbol{v}_{\text{warped}}. \tag{3.85}$$

Im fehlerbehafteten Fall gilt entsprechend:

$$\begin{aligned} \hat{\boldsymbol{v}}(l+1) &= \boldsymbol{L}(l)\boldsymbol{v}(l) \\ &+ [(\boldsymbol{\Lambda}_{\hat{v}}(l+1) + \boldsymbol{\Lambda}_{\boldsymbol{a}}(l+1))^{-1} + (\boldsymbol{L}(l)\boldsymbol{\Lambda}(l)\boldsymbol{L}^T(l))^{-1}]^{-1} \\ &\quad [\boldsymbol{\Lambda}_{\hat{v}}(l+1) + \boldsymbol{\Lambda}_{\boldsymbol{a}}(l+1)]^{-1}\boldsymbol{v}_{\text{warped}}. \end{aligned} \tag{3.86}$$

Da das Warpen von der richtigen Berechnung des Flusses auf der vorherigen Stufe abhängt, setzt man

$$\boldsymbol{\Lambda}_{\boldsymbol{a}}(l+1) = \boldsymbol{L}(l)\boldsymbol{\Lambda}(l)\boldsymbol{L}^T(l).$$

Das gleiche Ergebnis wie (3.86) erhält man, wenn man die MAP–Strategie zur Schätzung von $\boldsymbol{z}_{l+1} = \boldsymbol{v}_{l+1} + \boldsymbol{L}(l)\boldsymbol{v}$ verfolgt. Das Ergebnis hat dann folgende Form:

$$\begin{aligned}\hat{\boldsymbol{v}}(l+1) &= [(\boldsymbol{\Lambda}_{\hat{v}}(l+1) + \boldsymbol{\Lambda}_{\boldsymbol{a}}(l+1))^{-1} + (\boldsymbol{L}(l)\boldsymbol{\Lambda}(l)\boldsymbol{L}^T(l))^{-1}]^{-1}\\ &\quad [(\boldsymbol{\Lambda}_{\hat{v}}(l+1) + \boldsymbol{\Lambda}_{\boldsymbol{a}}(l+1))^{-1}(\boldsymbol{v}_{\text{warped}} + \boldsymbol{L}(l)\boldsymbol{v}_l)\\ &+ \quad (\boldsymbol{L}(l)\boldsymbol{\Lambda}(l)\boldsymbol{L}^T(l))^{-1}\boldsymbol{L}(l)\hat{\boldsymbol{v}}(l)].\end{aligned}$$

Für einen Vergleich der beiden Kombinationsverfahren wird die Schätzung derart vereinfacht, daß man nur von einem eindimensionalen Fluß u, also der Bewegung eines eindimensionalen Signals ausgeht. Dann sind die einzelnen Varianzen durch Skalare gegeben:

$$\boldsymbol{\Lambda}_{\hat{v}}(l+1) \sim p_{\hat{v}}(l+1), \quad \boldsymbol{\Lambda}_{\boldsymbol{a}}(l+1) \sim p_{\boldsymbol{a}}(l+1) \text{ und } \boldsymbol{\Lambda}(l) \sim p(l).$$

Wegen der Übersichtlichkeit repräsentiere im folgenden jeder Skalar bereits die entsprechende Interpolation des zugehörigen Wertes. Man setzt nun die Varianzen gleich einem Wert r:

$$p_{\hat{v}}(l+1) = p_{\boldsymbol{a}}(l+1) = p(l) = r.$$

Es ergibt sich daraus bei einem a priori Fluß $u(l)$ und einer a priori Varianz $p(l)$ für die zwei nächst feineren Pyramidenstufen:

fehlerfreies Warpen	fehlerbehaftetes Warpen
$u(l+1) = u(l) + \frac{1}{2}u_{\text{warped}}$ $p(l+1) = (\frac{2}{r})^{-1} = \frac{1}{2}r$	$u(l+1) = u(l) + \frac{1}{3}u_{\text{warped}}$ $p(l+1) = (\frac{1}{2r} + \frac{1}{r})^{-1} = \frac{2}{3}r$
$u(l+2) = u(l) + \frac{1}{2}u(l+1) + \frac{1}{3}u_{\text{warped}}$ $p(l+2) = (\frac{2}{r} + \frac{1}{r})^{-1} = \frac{1}{3}r$	$u(l+2) = u(l) + \frac{1}{3}u(l+1) + \frac{5}{8}u_{\text{warped}}$ $p(l+2) = (\frac{3}{2r} + 1r)^{-1} = \frac{2}{5}r$

Der Tabelle ist zu entnehmen, daß im Fall des fehlerbehafteten Warpens die neuen Messungen stärker gewichtet werden und die Kovarianz langsamer sinkt. Das bedeutet, daß diese Schätzung pessimistischer ist als die auf der Fehlerfreiheit beruhende, zu der auch die von Simoncelli et al. vorgestellte Methode der hierarchischen Berechnung gehört.

Der gesamte hierarchische Algorithmus beginnt nun wiederum damit, eine Gaußpyramide für jedes Bild der Bildfolge zu bilden. Auf der gröbsten Stufe wird dann der optische Fluß $\hat{\boldsymbol{v}}_0$ gemäß (3.67) berechnet, was der Berechnung (3.80) mit a priori Schätzung null und inverser a priori Varianz $\boldsymbol{\Lambda}_p^{-1} = \mathbf{0}$ entspricht. Die Kovarianzmatrix $\boldsymbol{\Lambda}_{\hat{v}_0}$ berechnet sich nach (3.72). Mit der interpolierten Version dieses Fluß-Ergebnisses wird das Bild der nächst feineren Stufe verschoben. Auch hier muß folgendes beachtet werden:

$$\begin{aligned}\boldsymbol{v}_{\text{fein}} &= 2\boldsymbol{v}_{\text{grob}},\\ \boldsymbol{\Lambda}_{\text{fein}} &= 4\boldsymbol{\Lambda}_{\text{grob}}.\end{aligned}$$

Man setzt nun $\boldsymbol{v}_0 = \hat{\boldsymbol{v}}_0$ und $\boldsymbol{\Lambda}(0) = \boldsymbol{\Lambda}_{\hat{v}_0}$ und berechnet den neuen kombinierten Fluß je nach Annahme über den Warpvorgang nach (3.85) oder nach (3.86). Entsprechend aktualisiert man die Kovarianzmatrix gemäß (3.83) bzw. (3.84). Den Vorgang des Warpens und der jeweiligen Berechnung wiederholt man solange, bis auch das Originalbild verarbeitet worden ist.

Übertragung der hierarchischen Berechnung auf das affine Modell

Möchte man den affinen Fluß mittels dieses hierarchischen Verfahrens berechnen, steht man vor dem Problem, daß der Nenner von (3.65) direkt vom zu berechnenden Fluß abhängt. Bei dem affinen Modell geht laut (2.16) für jeden Bildpunkt über die Varianz die relative Lage zum zentralen Bildpunkt ein, was eine Vereinfachung, wie sie in (3.66) durchgeführt wird, verhindert.

Entsprechend zur dreidimensionalen Darstellung des optischen Flusses wählt man für die hierarchische Berechnung des affinen Flusses ein siebendimensionales Modell:

$$\hat{\boldsymbol{v}}_{\mathbf{7D}\,\mathrm{aff}} = \begin{pmatrix} u_0 \\ v_0 \\ u_x \\ u_y \\ v_x \\ v_y \\ 1 \end{pmatrix}. \tag{3.87}$$

Um dieses hierarchische Verfahren trotzdem für die Berechnung des affinen optischen Flusses einsetzen zu können, muß man nach einer Möglichkeit suchen, ein Maß für die Varianz zu finden, das eine (3.66) gleichkommende Umformung zuläßt.

Die Varianz der Berechnung des affinen Flusses ist definiert durch

$$\begin{aligned} \lambda_i^2 &= \mathrm{E}[(\boldsymbol{\delta}\hat{\boldsymbol{b}}_{\mathrm{aff}_i}^T \hat{\boldsymbol{v}}_{\mathbf{7D}\,\mathrm{aff}})^2] \\ &= \hat{\boldsymbol{v}}_{\mathbf{7D}\,\mathrm{aff}}^T \,\mathrm{E}[\boldsymbol{\delta}\hat{\boldsymbol{b}}_{\mathrm{aff}_i} \boldsymbol{\delta}\hat{\boldsymbol{b}}_{\mathrm{aff}_i}^T] \hat{\boldsymbol{v}}_{\mathbf{7D}\,\mathrm{aff}}, \end{aligned}$$

wobei $\hat{\boldsymbol{b}}_{\mathrm{aff}_i}^T = (\boldsymbol{b}_{\mathrm{aff}_i}, g_{t\,i})$ gilt, mit $\boldsymbol{b}_{\mathrm{aff}_i}$ aus (3.58).

Für die Summe über die Matrizen $E[\boldsymbol{\delta}\hat{\boldsymbol{b}}_{\mathrm{aff}_i} \boldsymbol{\delta}\hat{\boldsymbol{b}}_{\mathrm{aff}_i}^T]$ kann man nun im allgemeinen keinen Ausdruck so finden, daß man wieder auf die Eigenwertgleichung gelangt. Approximativ kann man jedoch die Fehlermatrix durch ihre Spur ersetzen:

$$E[\boldsymbol{\delta}\hat{\boldsymbol{b}}_{\mathrm{aff}_i} \boldsymbol{\delta}\hat{\boldsymbol{b}}_{\mathrm{aff}_i}^T] \quad \rightsquigarrow_{\mathrm{spur}} \quad s_i = (1 + \Delta x^2 + \Delta y^2) 2\lambda_s^2 + \lambda_t^2.$$

Damit man darauf den oben beschriebenen Algorithmus zur Berechnung des affinen Flusses anwenden kann, muß man wieder einige Komponenten in den einzelnen Berechnungen ersetzen. Substituiert man in (3.64) den optischen Fluß $\boldsymbol{v}_{\mathbf{3D}}$ durch die dreidimensionale Darstellung seines affinen Modells $\boldsymbol{v}_{\mathbf{3D}} \rightsquigarrow (\boldsymbol{v}_{\mathbf{3D}\,\mathrm{aff}}, 1)^T$ (vgl.(2.16)) und die Fehlermatrix durch die Spur s_i, so liefert die Maximierung der bedingten Wahrscheinlichkeit unter Berücksichtigung des Rauschens in der zeitlichen Ableitung das Minimierungsproblem:

$$\hat{\boldsymbol{v}}_{\mathbf{7D}\,\mathrm{aff}} = \arg \min_{\boldsymbol{v}_{\mathbf{7D}\,\mathrm{aff}}} \frac{\sum_i \frac{1}{s_i} (\hat{\boldsymbol{b}}_{\mathrm{aff}_i}^T \boldsymbol{v}_{\mathbf{7D}\,\mathrm{aff}})^2}{\boldsymbol{v}_{\mathbf{7D}\,\mathrm{aff}}^T \boldsymbol{v}_{\mathbf{7D}\,\mathrm{aff}}}. \tag{3.88}$$

Zur Berechnung des affinen Flusses nach (3.80) mit a priori Schätzung $\boldsymbol{v}_0$ auf der gröbsten Stufe muß man also analog zu (3.68) den zum kleinsten Eigenwert gehörenden normierten Eigenvektor der mit der Spur der Fehlermatrix gewichteten Gradientenmatrix $\sum_{i=1}^{N}(\frac{1}{s_i}\hat{\boldsymbol{b}}_{\text{aff}_i}\hat{\boldsymbol{b}}_{\text{aff}_i}^T)$ bestimmen.

Die Kovarianzmatrix berechnet sich nach (3.72), wobei jedoch hier ebenfalls die Gewichtung $\boldsymbol{v}_{\mathbf{3D}}^T\boldsymbol{\Lambda}_e\boldsymbol{v}_{\mathbf{3D}}$ durch $\sum_{i=1}^{N} s_i\|\boldsymbol{v}_{\mathbf{7D\,aff}}\|$ ersetzt werden muß, wie auch $\nabla\boldsymbol{g}_i$ durch $\hat{\boldsymbol{b}}_{\text{aff}_i}$.
Die Anpassung der Elemente der Kovarianzmatrix an das feinere Gitter erfolgt in gleicher Weise wie schon bei der Erweiterung des Verfahrens von Simoncelli et al. beschrieben. Ebenso gilt (3.59) für die Transformation des optischen Flusses auf das nächst feinere Gitter.

Mit dem interpolierten, der Lage des jeweiligen Bildpunktes entsprechenden Fluß verschiebt man die Bilder der nächst feineren Pyramidenstufe in der Umgebung der Größe der zeitlichen Ableitungsmaske und berechnet auf der gewarpten Sequenz erneut den affinen Fluß. Die Ergebnisse kombiniert man dann gemäß (3.85) oder (3.86), und die Kovarianzmatrizen erhält man aus (3.83) bzw. (3.84). Das wiederholt man solange, bis der verbleibende Fluß der Originalsequenz in die Berechnung mit einbezogen worden ist.

3.2.7 Beziehungen zwischen den unkorrelierten Rauschmodellen

Die Betrachtung der möglichen Ansätze zur Modellierung des Rauschens in der Berechnung des optischen Flusses lassen die Frage aufkommen, in welchem Verhältnis die untersuchten Verfahren auf den unterschiedlichen Rauschmodellen zueinander stehen.
Dazu wollen wir zunächst einen kurzen Einblick in ein weiteres Schätzverfahren geben, an dem der Zusammenhang zwischen den Berechnungen aus Kapitel 3.2.1 und Kapitel 3.2.4 gut zu veranschaulichen ist.

Zur Berechnung des optischen Flusses auf differentiellem Weg benutzen Weber und Malik in [34] die totale lineare Ausgleichsrechnung, TLS (Total Least Squares). Sie unterscheidet sich, wie in Abb. 3.5 zu sehen ist, von der einfachen linearen Ausgleichsrechnung dadurch, daß man den zu minimierenden Fehler nicht auf nur eine Komponente, sondern auf alle Komponenten bezieht.

Bei der Vorgabe von rauschfreien Daten kann die BCCE (2.2) für ein n Bildpunkte umfassendes Gebiet in folgender Form geschrieben werden:

$$\begin{pmatrix} g_{x1} & g_{y1} \\ g_{x2} & g_{y2} \\ \vdots & \vdots \\ g_{xn} & g_{yn} \end{pmatrix} \boldsymbol{v} = \begin{pmatrix} -g_{t1} \\ -g_{t2} \\ \cdots \\ -g_{tn} \end{pmatrix}.$$

Ersetzt man die aus den örtlichen Ableitungen bestehende Matrix durch $\boldsymbol{G}$ und den Vektor der zeitlichen Ableitungen durch $\boldsymbol{b}$, so erhält man die bei der totalen linearen

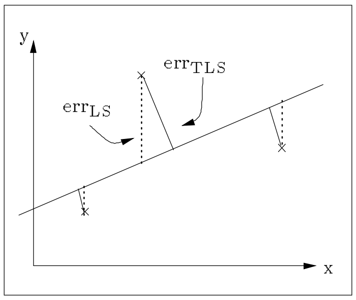


Abbildung 3.5: Veranschaulichung des Unterschieds zwischen der totalen und der einfachen linearen Ausgleichsrechnung am Beispiel der Geradenanpassung aus [34]

Ausgleichsrechnung benutzte vereinfachte Darstellung :

$$[\boldsymbol{G}|\boldsymbol{b}_t]\begin{pmatrix}\boldsymbol{v}\\1\end{pmatrix} = \boldsymbol{0}. \tag{3.89}$$

Die kombinierte Matrix $[\boldsymbol{G}|\boldsymbol{b}_t]$ wird Messungsmatrix genannt, weil jedes ihrer Elemente rauschanfällig ist [14]. Als Ergebnis liefert die TLS den Vektor

$$\boldsymbol{v} = -(\boldsymbol{G}^T\boldsymbol{G} - \sigma_3^2\boldsymbol{I})^{-1}\boldsymbol{G}^T\boldsymbol{b}_t, \tag{3.90}$$

wobei $\sigma_3 = \sqrt{\alpha_{\min}([\boldsymbol{G}|\boldsymbol{b}_t]^T[\boldsymbol{G}|\boldsymbol{b}_t])}$ der kleinste Singulärwert der Messungsmatrix ist. Setzt man $\sigma_3 = 0$, dann erhält man die Lösung der einfachen linearen Ausgleichsrechnung.

Zur vergleichsweisen Betrachtung des von Kanatani vorgestellten Verfahrens ist es sinnvoll, von einem Gaußschen Fehler, d.h. einer normierten Normalverteilung, in den einzelnen Komponenten der Grauwertableitungen auszugehen, d.h. $\boldsymbol{\Lambda}_e = \boldsymbol{I}$. Damit ist die erzielte Schätzung des Verfahrens gerade der zum kleinsten Eigenwert gehörende, normierte Eigenvektor der Matrix $\boldsymbol{A}$ aus (3.66). Für nur einen Bildpunkt läßt sich ein direkter Zusammenhang zwischen der Messungsmatrix $[\boldsymbol{G}|\boldsymbol{b}]$ und der Matrix $\boldsymbol{A}$ formulieren:

$$[\boldsymbol{G}|\boldsymbol{b}_t]^T[\boldsymbol{G}|\boldsymbol{b}_t] = \boldsymbol{A}.$$

Das Quadrat des kleinsten Singulärwerts der Messungsmatrix σ_3^2 ist daher gerade der kleinste Eigenwert der Matrix $\boldsymbol{A}$. Nach der speziellen Eigenwertgleichung gilt dann für die Schätzung $\boldsymbol{v}_{3\mathrm{D}} = (\boldsymbol{v}, 1)^T$:

$$(\boldsymbol{A} - \sigma_3^2\boldsymbol{I}_{3\times3})\boldsymbol{v}_{3\mathrm{D}} = 0.$$

Verzichtet man nun auf die durch die letzte Zeile des obigen Gleichungssystems gegebene Gleichung und formt linear um, kann man schreiben:

$$\begin{aligned} &(\boldsymbol{G}^T\boldsymbol{G} - \sigma_3^2\boldsymbol{I}_{2\times2})\boldsymbol{v} = -\boldsymbol{G}^T\boldsymbol{b}_t \\ \text{bzw.}\quad &\boldsymbol{v} = -(\boldsymbol{G}^T\boldsymbol{G} - \sigma_3^2\boldsymbol{I}_{2\times2})^{-1}\boldsymbol{G}^T\boldsymbol{b}_t \end{aligned} \tag{3.91}$$

Das erzielte Ergebnis stimmt also direkt mit dem Ergebnis der totalen linearen Ausgleichsrechnung (3.90) überein.

Ersetzt man im Verfahren von Simoncelli et al. den Nenner des Ergebnisses durch β und setzt die Kovarianzmatrix (3.44) in (3.45) ein, läßt sich das Ergebnis wie folgt darstellen:

$$\boldsymbol{v} = -(\frac{1}{\beta}\boldsymbol{M} + \boldsymbol{\Lambda}_p^{-1})^{-1}\frac{1}{\beta}\boldsymbol{b} \tag{3.92}$$

Wählt man in (3.45) die Parameter $\lambda_{\boldsymbol{v}} = 0$ und $\lambda_t = 1$, so gilt $\beta = 1$. Diese Wahl der Parameter ist gleichbedeutend damit, daß man beim Rauschen in der zeitlichen Ableitung wieder von einem Gaußschen Fehler ausgeht und zusätzlich annimmt, daß es keinen Fehler zwischen geschätztem und wirklichem optischen Fluß gibt. An der Definition von $\boldsymbol{M}$ aus (2.8) und $\boldsymbol{b}$ aus (2.9) erkennt man, daß für einen betrachteten Bildpunkt ($n = 1$) folgende Gleichheiten gelten:

$$\begin{aligned} \boldsymbol{G}^T\boldsymbol{G} &= \boldsymbol{M} \\ \text{und} \quad \boldsymbol{b}_t &= \boldsymbol{G}^T\boldsymbol{b}. \end{aligned}$$

Damit läßt sich das Ergebnis (3.92) umformulieren:

$$\boldsymbol{v} = -(\boldsymbol{G}^T\boldsymbol{G} + \boldsymbol{\Lambda}_p)^{-1}\boldsymbol{G}^T\boldsymbol{b}_t. \tag{3.93}$$

Vergleicht man (3.90) mit (3.91) bzw. (3.93), so stellt man fest, daß die Ergebnisse sich nur dadurch unterscheiden, daß bei dem Verfahren von Simoncelli et al. die a priori Varianz addiert und bei der totalen linearen Ausgleichsrechnung bzw. bei der von Kanatani beschriebenen Berechnung ein der Varianz ähnlicher Ausdruck subtrahiert wird.

3.2.8 Korreliertes Rauschen in allen Komponenten

Das von Nagel in [27] vorgestellte Rauschmodell hat eine ähnliche Form wie (3.34) und (3.35). Dabei berücksichtigt er im Gegensatz zu den bisher vorgestellten Ansätzen, daß das Meßrauschen an den einzelnen Bildpunkten zueinander korreliert sein kann.
Der Unterschied im Ansatz besteht darin, daß er das Rauschen nicht in den Messungen der Ableitungen modelliert, sondern direkt in den Messungen der Grauwerte:

$$\boldsymbol{g} = \tilde{\boldsymbol{g}} + \delta\boldsymbol{g}, \tag{3.94}$$

wobei das Rauschen in den Grauwerten unabhängig und normalverteilt und daher auch unkorreliert sein soll:

$$\delta\boldsymbol{g} \sim N(0, \lambda_{\boldsymbol{g}}^2), \quad \mathrm{E}[\boldsymbol{g}_l\boldsymbol{g}_k] = 0.$$

Die Ableitungen der Grauwerte nach Ort bzw. Zeit werden mittels diskreter Faltung mit der entsprechenden Ableitung der dreiwertigen Gaußverteilung $\boldsymbol{w}(\boldsymbol{x})$ berechnet,

$$\boldsymbol{w}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{|\boldsymbol{\Lambda}_{\boldsymbol{w}}|}}\exp\{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Lambda}_{\boldsymbol{w}}^{-1}\boldsymbol{x}\},$$

wobei $\boldsymbol{x} = (x, y, t)^T \in \{(1, 1, 1)^T, \ldots, (n_x, n_y, n_t)^T\}$ gilt. Sie sind dann also gegeben durch:

$$\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial x_j} = \boldsymbol{w}_j(\boldsymbol{x})^T \boldsymbol{g}, \quad j = x, y, t. \tag{3.95}$$

Die Kontinuitätsgleichung hat demnach die Form:

$$(\boldsymbol{w}_x^T \tilde{\boldsymbol{g}})u + (\boldsymbol{w}_y^T \tilde{\boldsymbol{g}})v + (\boldsymbol{w}_t^T \tilde{\boldsymbol{g}}) = 0. \tag{3.96}$$

Für das Schätzungsproblem bildet man nun einen Vektor $\boldsymbol{c}_{\text{patch}}(\boldsymbol{g}, \boldsymbol{v}, \boldsymbol{w})$ aus der BCCE (3.96) für jeden Punkt in der Umgebung des betrachteten Bildpunkts, so daß die BCCE für die ganze Region gegeben ist durch:

$$\boldsymbol{c}_{\text{patch}}(\boldsymbol{g}, \boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{0}. \tag{3.97}$$

Auch die Messungen der Grauwerte in dem betrachteten Gebiet lassen sich entsprechend (3.94) als Vektor darstellen:

$$\boldsymbol{g}_{\text{patch}} = \tilde{\boldsymbol{g}}_{\text{patch}} + \boldsymbol{\delta g}_{\text{patch}}, \tag{3.98}$$

wobei der Rauschvektor $\boldsymbol{\delta g}_{\text{patch}}$ wie folgt verteilt ist:

$$p(\boldsymbol{\delta g}_{\text{patch}}) = \frac{1}{(2\pi\lambda_{\boldsymbol{g}}^2)^{\frac{n}{2}}} \exp\{-\frac{1}{2} \boldsymbol{\delta g}_{\text{patch}}^T \boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1} \boldsymbol{\delta g}_{\text{patch}}\}$$

mit

$$\boldsymbol{\Lambda}_{\boldsymbol{g}} = \lambda_{\boldsymbol{g}}^2 \boldsymbol{I}_{n\times n}.$$

Das Schätzungsproblem besteht nun darin, die Verbundwahrscheinlichkeit $p(\boldsymbol{\delta g}_{\text{patch}})$ unter der Nebenbedingung (3.97) zu maximieren. Die dazu äquivalente, aus (3.97) und (3.98) resultierende Minimierung lautet dann:

$$\boldsymbol{\delta g}_{\text{patch}}^T \boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1} \boldsymbol{\delta g}_{\text{patch}} + 2\boldsymbol{\alpha}^T (\boldsymbol{c}_{\text{patch}}(\boldsymbol{g}, \boldsymbol{v}, \boldsymbol{w})) \to \min. \tag{3.99}$$

Es sind also Werte für den Zustandsvektor $\boldsymbol{\alpha}$ und die Komponenten von $\boldsymbol{v}$ zu bestimmen, die (3.99) minimieren.
Da also sowohl der optische Fluß $\boldsymbol{v} = (u, v)^T$ als auch die exakten Grauwerte $\tilde{\boldsymbol{g}}_{\text{patch}}$ unbekannt sind, ist das Optimierungsproblem aus (3.97) nicht mehr linear, sondern bilinear. Man kann die Lösung daher nur noch iterativ bestimmen. Ferner beachtet Nagel in seiner Modellierung nicht, daß die in (3.95) gebildeten Ableitungen selbst für exakte Grauwerte $\tilde{\boldsymbol{g}}$ nicht mehr exakt sind, da die abgetasteten Filterantworten der FIR–Gaußfilter nur Näherungen der Ableitungen darstellen.

3.3 Experimenteller Vergleich in Testsequenzen

Nach den theoretischen Einblicken in die verschiedenen Verfahren zur Berechnung des optischen Flusses wird in diesem Kapitel die Leistung der Verfahren in Experimenten getestet. Außerdem werden das Verfahren von Simoncelli et al. und das auf den Ansätzen von Kanatani beruhende Verfahren erstmalig in der Literatur experimentell verglichen. Es werden dazu folgende Tests durchgeführt:

1. Simoncelli-Verfahren: nicht hierarchisch ↔ hierarchisch
2. Eigenwertverfahren: nicht hierarchisch ↔ hierarchisch
3. hierarchisches Eigenwertverfahren: Gegenüberstellung der Kompositionsmethoden
4. Gegenüberstellung der nicht hierarchischen Verfahren: Simoncelli ↔ Kanatani
5. Gegenüberstellung der hierarchischen Verfahren: Simoncelli ↔ Eigenwertverfahren
6. Variation der a priori Varianz bei den hierarchischen Verfahren
7. Variation der Rauschparameter bei den nicht hierarchischen Verfahren

Für einen quantitativen Vergleich der Verfahren werden im Laufe dieses Abschnittes die erzielten experimentellen Ergebnisse der verschiedenen Verfahren bei Variation der Eingabeparameter vorgestellt. Es werden dazu die in den vorigen Kapiteln beschriebenen Algorithmen auf künstliche Bildsequenzen angewendet, von denen der exakte optische Fluß bekannt ist. Die dazu benutzten Sequenzen sind die Yosemite-Fly-Through-Sequenz, die Insight-Sequenz und die Diverging-Tree-Sequenz, von denen je ein Bild und der zugehörige exakte optische Fluß in Abb. 3.6 zu sehen ist.

Die Yosemite-Fly-Through-Sequenz zeigt einen künstlich erzeugten Flug durch den Yosemite-Nationalpark. Der exakte optische Fluß liegt zwischen null bis zu ca. 5.5 Pixeln pro Bild. Die in der Sequenz sichtbaren Wolken deformieren sich bei der Bewegung, so daß dort das erwähnte Korrespondenzproblem zum Tragen kommt. In der Insight-Sequenz sind verschiedene marmorierte Säulen zu sehen, die auf einem stark texturierten Untergrund stehen. Die exakte Bewegung ist eine Translation der Größe zwischen ca. 0.4 und 2.5 Pixeln pro Bild, bei der die gegenseitige Verdeckung der Säulen eine Schwierigkeit darstellt. Die Diverging-Tree-Sequenz stellt einen Baum dar. Der exakte Fluß ist durch eine Bewegung entlang der optischen Achse gegeben, die die Geschwindigkeit von 1.4 bis zu 2.0 Pixeln pro Bild hat.

Zur Darstellung der Ergebnisse werden zwei Fehlermaße benutzt, die zunächst kurz beschrieben werden. In der Literatur ist die Darstellung der Leistung durch das von Barron et al. [2] hervorgebrachte Winkelfehlermaß üblich. Der Winkelfehler zwischen geschätztem Fluß $\hat{\boldsymbol{v}} = (\hat{u}, \hat{v})$ und exaktem Fluß $\tilde{\boldsymbol{v}} = (\tilde{u}, \tilde{v})$ ist definiert durch den in Abb. 3.7 abgebildeten Winkel γ zwischen den dreidimensionalen, normierten Darstellungen der Vektoren $\boldsymbol{v}_{\text{3D}} = (|\boldsymbol{v}|^2 + 1)^{-\frac{1}{2}}(u, v, 1)$. Aus dem Kosinussatz erhält man die Fehlerformel:

$$\gamma = \arccos \frac{\hat{\boldsymbol{v}}\tilde{\boldsymbol{v}} + 1}{\sqrt{(\hat{\boldsymbol{v}}^2 + 1)}\sqrt{(\tilde{\boldsymbol{v}}^2 + 1)}}. \tag{3.100}$$

Testsequenzen

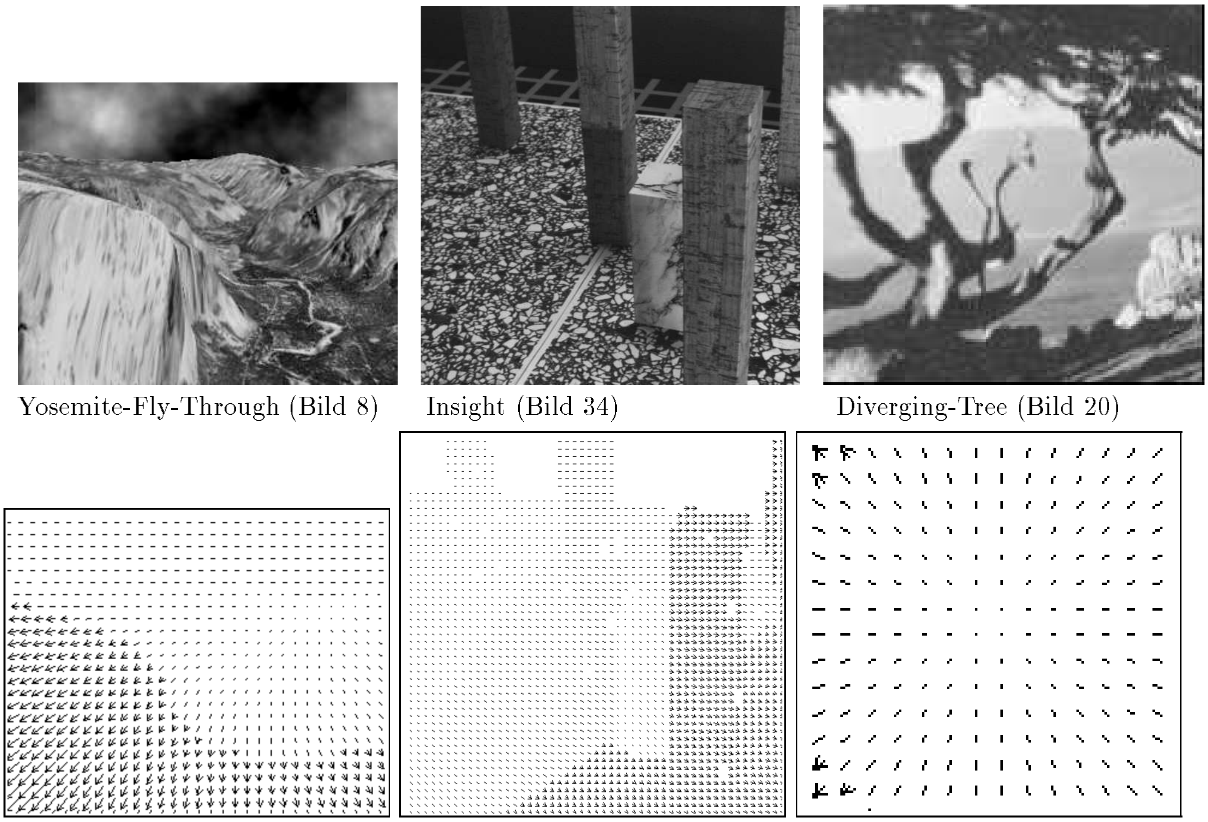


Abbildung 3.6: Ausschnitte aus den Testsequenzen und Darstellung des zugehörigen exakten, optischen Flusses (von links nach rechts): Yosemite-Fly-Through-, Insight- und Diverging-Tree-Sequenz

Als weitere Fehlermaße benutzen wir den relativen Fehler

$$\mathrm{err}_{\mathrm{rel}} = 100\frac{\|\hat{\boldsymbol{v}} - \tilde{\boldsymbol{v}}\|}{\|\tilde{\boldsymbol{v}}\|}.$$

Da es bei den meisten Anwendungen des optischen Flusses nicht darauf ankommt, seinen Wert an jedem Bildpunkt zu kennen, ist es nicht sinnvoll, den Fehler über alle Bildpunkte zu mitteln. Statt dessen betrachtet man nur den über die besten Schätzungen gemittelten Fehler und stellt diesem den Anteil der sicheren Schätzungen am Gesamtbild, also die Dichte des optischen Flußfeldes, gegenüber.

Die Kovarianzmatrix des jeweiligen Verfahrens liefert uns eine Abschätzung des Fehlers der Berechnung. Betrachten wir den kleinsten Singulärwert der inversen Kovarianzmatrix, haben wir ein Gütemaß, das steigt, je besser die Schätzung konditioniert ist. Wir betrachten nun diejenigen Schätzungen als sicher, an denen dieses Gütemaß einen Schwellwert nicht unterschreitet. Damit wir den mit verschiedenen Verfahren erzielten Fehler bei gleicher Dichte angeben können, legen wir im voraus die Dichte fest und mitteln den Fehler über die durch die Dichte festgelegte Menge von Bildpunkten, an denen der kleinste Singulärwert der inversen Kovarianzmatrix am höchsten ist.

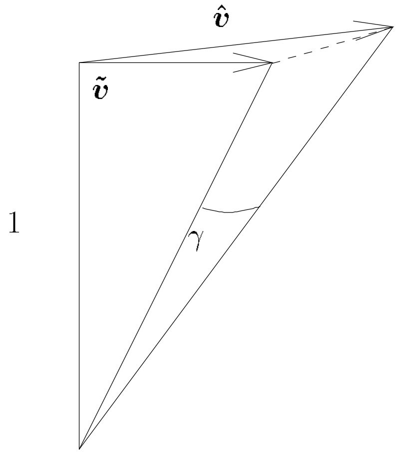


Abbildung 3.7: Das Winkelfehlermaß nach Barron et. al. ist durch den Winkel γ definiert, und der absolute Fehler ist die Länge der absoluten Vektordifferenz (gestrichelte Linie)

In den folgenden Versuchen benutzen wir jeweils Ableitungsfilter der Größe 5×5. Außerdem nehmen wir zur Lösung des Aperturproblems an, daß der optische Fluß in einem Bereich von 7×7 Pixeln konstant ist, d.h. die Pixel dieses Gebiets tragen alle gleichgewichtet zur Berechnung des optischen Flusses an einem Bildpunkt bei.

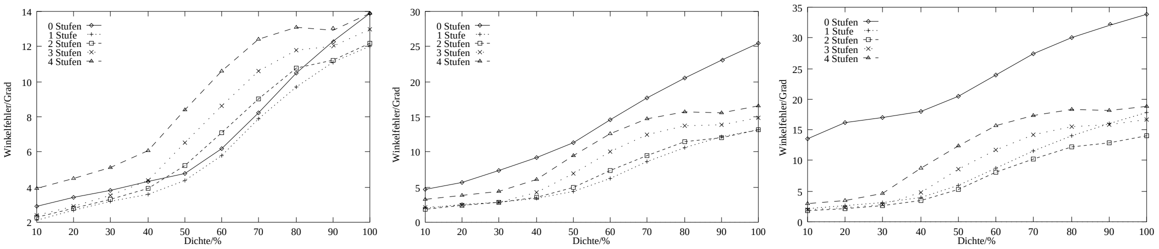
Test 1: Simoncelli-Verfahren: nicht hierarchisch $\leftrightarrow$ hierarchisch

Zunächst soll untersucht werden, ob die hierarchische Berechnung des lokal konstanten optischen Flusses bei schnellen Bewegungen eine Verbesserung der Ergebnisse gegenüber der direkten Berechnung ohne Unterabtastung liefert. Dazu betrachten wir das Verfahren von Simoncelli et al. und das neu entwickelte hierarchische Eigenwertverfahren, das auf dem Verfahren von Kanatani basiert, gesondert. Da die Yosemite-Fly-Through-Sequenz, die Insight-Sequenz und die Diverging-Tree-Sequenz nur relativ kleine Bewegungen zeigen und gerade die Ergebnissen beim Auftreten von großen Geschwindigkeiten interessant sind, sind die Sequenzen zusätzlich zeitlich unterabgetastet worden, und zwar in den Verhältnissen $1:2$ und $1:3$. Für den Vergleich mit dem exakten Fluß geht man davon aus, daß dieser durch die entsprechende Multiplikation der oben gezeigten Flußfelder mit 2 bzw. 3 gegeben ist. Auf die unterabgetasteten Sequenzen werden nun zunächst das Verfahren von Simoncelli et. al. und anschließend das hierarchische Eigenwertverfahren unter Variation der Stufenzahl der zu bearbeitenden Bildpyramide angewendet.

Damit dabei nur die Auswirkungen der hierarchischen Verarbeitung deutlich werden, geht man von einer Gaußkurve bzw. -glocke für das Rauschen in den Messungen aus, d.h. von einem Mittelwert null und einer Varianz von eins. Bei dem Verfahren von Simoncelli et al. wird also $\lambda_t = 1$ gesetzt, und es wird angenommen, daß das Modellrauschen in dem Rauschen der zeitlichen Ableitung enthalten ist ($\lambda_v = 0$). Außerdem wird auf eine a priori Varianz verzichtet ($\lambda_p = 0$).

In Abb. 3.8 sind die erzielten Ergebnisse des Simoncelli-Verfahrens angewendet auf die Yosemite-Fly-Through-Sequenz zu sehen, während in Abb. 3.9 die auf der Insight-Sequenz erzielten Ergebnisse zu sehen sind und Abb. 3.10 die Ergebnisse bei Anwendung des Verfahrens auf die Diverging-Tree-Sequenz zeigt. Die Bezeichnung “0 Stufen” bedeutet, daß es sich um die nicht hierarchische Berechnung handelt.

Hierarchisches Simoncelli-Verfahren - Yosemite-Fly-Through-Sequenz



Originalsequenz | zeitliche Unterabtastung (1 : 2) | zeitliche Unterabtastung (1 : 3)

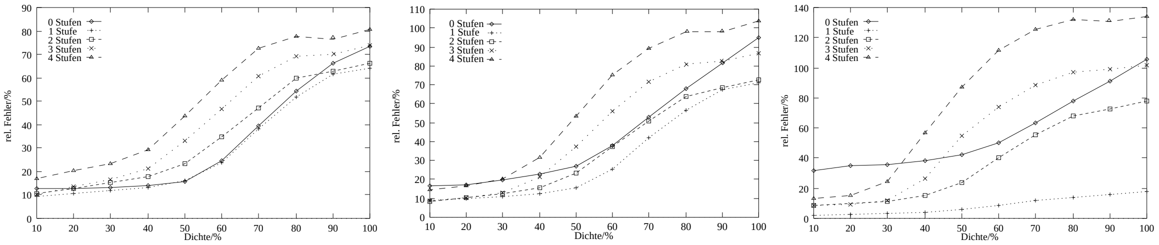


Abbildung 3.8: Winkelfehler (oben) und relativer Fehler (unten) bei Anwendung des Verfahrens von Simoncelli ($\lambda_t = 1, \lambda_v = \lambda_p = 0$) auf die Yosemite-Fly-Through-Sequenz unter Variation der Zahl von Pyramidenstufen

Hierarchisches Simoncelli-Verfahren - Insight-Sequenz

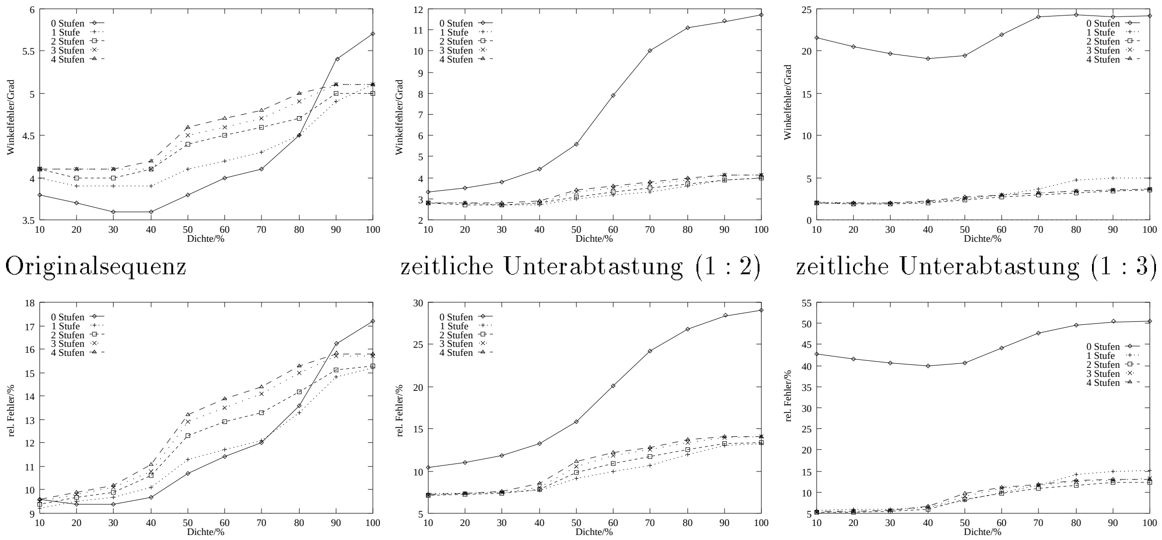


Abbildung 3.9: Winkelfehler (oben) und relativer Fehler (unten) bei Anwendung des Verfahrens von Simoncelli ($\lambda_t = 1, \lambda_{\boldsymbol{v}} = \lambda_p = 0$) auf die Insight-Sequenz unter Variation der Zahl von Pyramidenstufen

Hierarchisches Simoncelli-Verfahren - Diverging-Tree-Sequenz

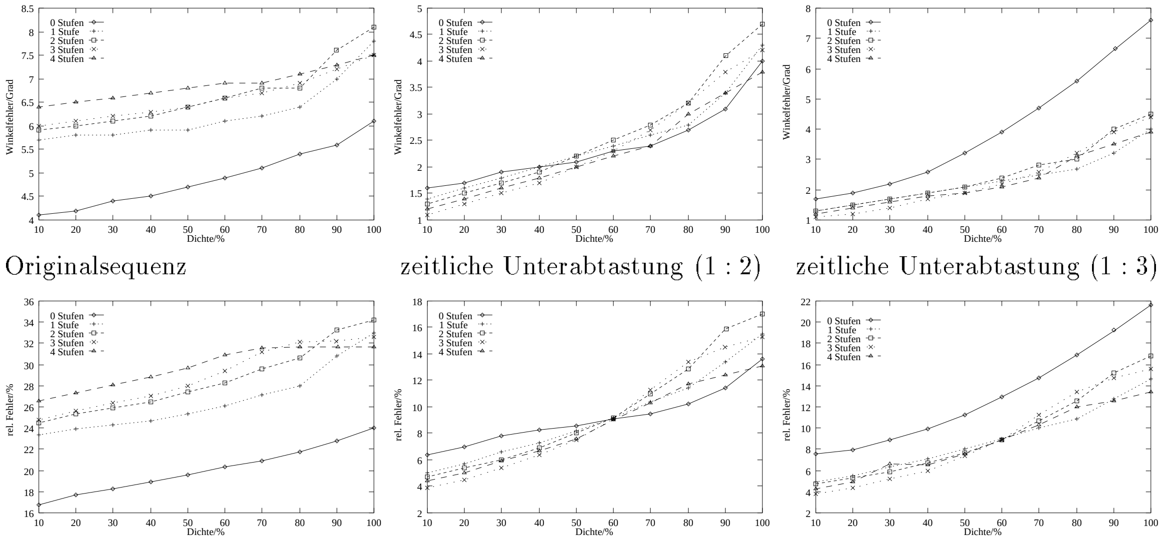


Abbildung 3.10: Winkelfehler (oben) und relativer Fehler (unten) bei Anwendung des Verfahrens von Simoncelli ($\lambda_t = 1, \lambda_{\boldsymbol{v}} = \lambda_p = 0$) auf die Diverging-Tree-Sequenz unter Variation der Zahl von Pyramidenstufen

Der mittlere Winkelfehler ist jeweils in den oberen Diagrammen dargestellt und der mittlere relative Fehler in den unteren. Der Unterschied im Verlauf der Kurven der unterschiedlichen Fehlermaße läßt sich durch das Auftreten von sehr kleinen Geschwindigkeiten erklären, die durch die Division das Fehlermaß verfälschen können.

Man macht bei allen drei Sequenzen die Beobachtung, daß sich bei einer günstig gewählten Pyramidenstufenzahl mit der hierarchischen Verarbeitung bei hohen Geschwindigkeiten (zeitliche Abtastung 1 : 2 und 1 : 3) deutlich bessere Ergebnisse erzielen lassen als mit der einfachen Verarbeitung. Wie erwartet steigt außerdem der Fehler mit zunehmender Dichte des optischen Flußfeldes an, was bedeutet, daß das berechnete Gütemaß zuverlässig ist.

Bei den Originalsequenzen fällt auf, daß die einfache Verarbeitung bei der Insight-Sequenz und der Diverging-Tree-Sequenz sowohl einen kleineren mittleren Winkelfehler als auch einen kleineren relativen Fehler erzielt als der hierarchische Algorithmus. Dies ist damit zu erklären, daß die in den Sequenzen vorkommende mittlere Geschwindigkeit bei 1.3 bzw. 0.9 Pixeln pro Bild liegt. Da beim Warpen die subpixelgenauen Werte nur durch bilineare Interpolation weiterverarbeitet werden, tritt dadurch bei einer Pyramidenverarbeitung, bei der die Geschwindigkeit von einer Auflösungsstufe zur nächst gröberen halbiert wird, in diesen Fällen ein nicht zu korrigierender Fehler auf. Bei der Yosemite-Fly-Through-Sequenz liegt die mittlere Geschwindigkeit bei 1.5 Pixeln pro Bild, weswegen die Verarbeitung mit einer aus zwei Stufen bestehenden Pyramide (Originalauflösung und nächst gröbere Auflösung) einen kleineren mittleren Fehler liefert.

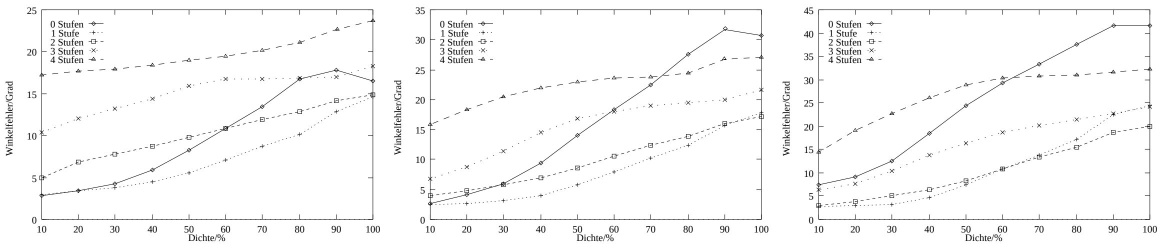
Zu bemerken ist ebenfalls, daß bei der Yosemite-Fly-Through-Sequenz und bei der Insight-Sequenz die Berechnung des optischen Flusses über eine vierstufige Pyramide den größten Fehler liefert. Zum einen liegt dies daran, daß der exakte optische Fluß nicht örtlich konstant ist, wodurch die Anfangsschätzung auf der gröbsten Auflösungsstufe schon fehlerhaft ist. Durch das Warpen mit der interpolierten Version dieses groben Flußfeldes auf der feineren Stufe kommt es zu einer Fortsetzung des Fehlers bis in die feinste Berechnungsstufe. Zum anderen geht sowieso über jeden Warpvorgang ein kleiner Fehler in die Berechnung ein, der sich durch Wiederholung des Vorgangs auf einer feineren Stufe verstärkt.

Test 2: Eigenwertverfahren: nicht hierarchisch $\leftrightarrow$ hierarchisch

Als nächstes betrachten wir die Ergebnisse des Eigenwertverfahrens, welches bei der nicht hierarchischen Verarbeitung dem Algorithmus von Kanatani entspricht. Bei der Entwicklung des hierachischen Eigenwert-Algorithmus' gibt es, wie in Kapitel 3.2.6 erläutert wird, zwei Möglichkeiten für die Komposition von Daten aus verschiedenen Auflösungsebenen. Einerseits wird der Warpvorgang als fehlerfrei betrachtet und andererseits als fehlerbehaftet. Im folgenden sind die Ergebnisse für beide Strategien dargestellt.

Um auch hier nur die Auswirkungen der hierarchischen Verarbeitung zu veranschaulichen, sind die Parameter für das Rauschen in der örtlichen und zeitlichen Komponente gleich eins gesetzt ($\lambda_s = \lambda_t = 1$), und es wird ebenfalls auf eine a priori Varianz verzichtet ($\lambda_p = 0$).

Hierarchisches Eigenwertverfahren-Verfahren ("fehlerfrei") - Yosemite-Fly-Through-Sequenz



Originalsequenz | zeitliche Unterabtastung (1 : 2) | zeitliche Unterabtastung (1 : 3)

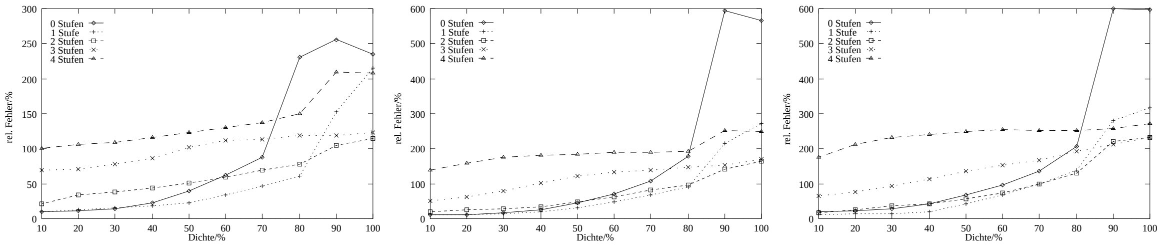
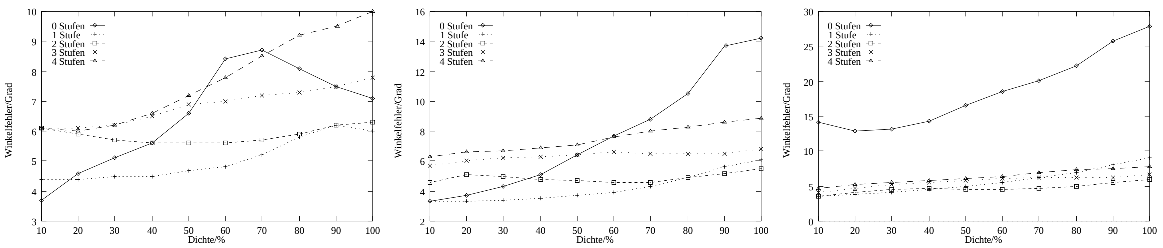


Abbildung 3.11: Winkelfehler (oben) und relativer Fehler (unten) bei Anwendung des Eigenwertverfahrens ($\lambda_s = \lambda_t = 1, \lambda_p = 0$) auf die Yosemite-Fly-Through-Sequenz unter Variation der Zahl von Pyramidenstufen, wobei das Warpen als fehlerfrei vorausgesetzt wird.

Hierarchisches Eigenwertverfahren ("fehlerfrei") - Insight-Sequenz



Originalsequenz | zeitliche Unterabtastung (1 : 2) | zeitliche Unterabtastung (1 : 3)

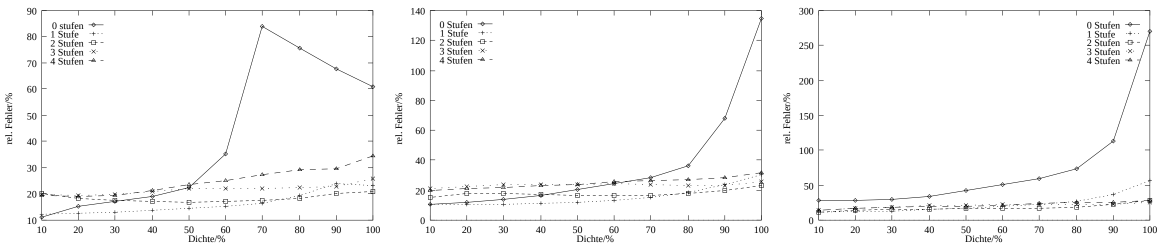
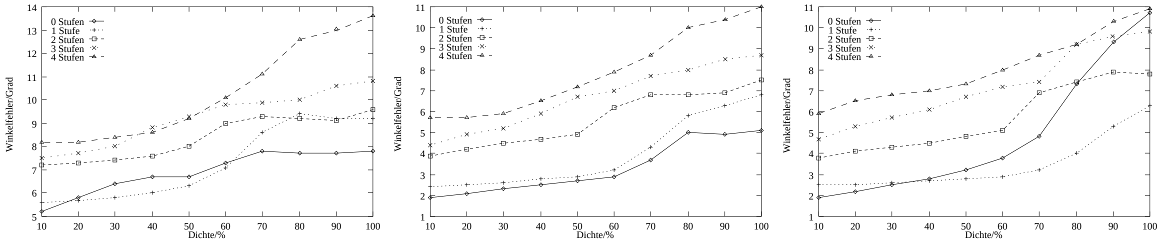


Abbildung 3.12: Winkelfehler (oben) und relativer Fehler (unten) bei Anwendung des Eigenwertverfahrens ($\lambda_s = \lambda_t = 1, \lambda_p = 0$) auf die Insight-Sequenz unter Variation der Zahl von Pyramidenstufen, wobei das Warpen als fehlerfrei vorausgesetzt wird.

Hierarchisches Eigenwertverfahren ("fehlerfrei") - Diverging-Tree-Sequenz



Originalsequenz | zeitliche Unterabtastung (1 : 2) | zeitliche Unterabtastung (1 : 3)

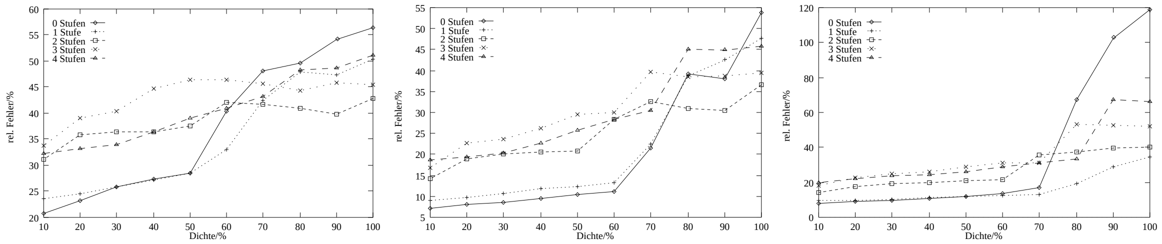
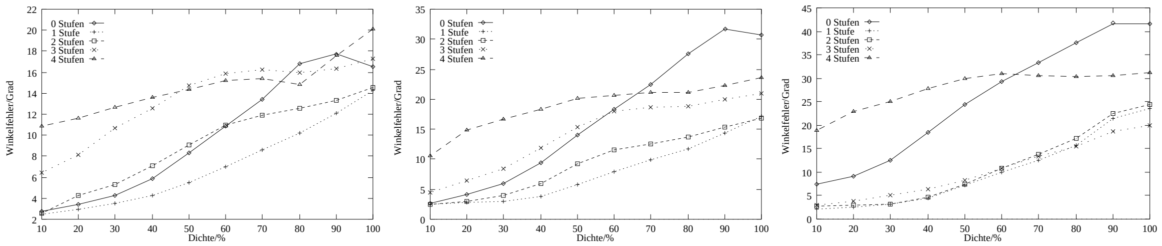


Abbildung 3.13: Winkelfehler (oben) und relativer Fehler (unten) bei Anwendung des Eigenwertverfahrens ($\lambda_s = \lambda_t = 1, \lambda_p = 0$) auf die Diverging-Tree-Sequenz unter Variation der Zahl von Pyramidenstufen, wobei das Warpen als fehlerfrei vorausgesetzt wird.

Hierarchisches Eigenwertverfahren-Verfahren ("fehlerbehaftet") - Yosemite-Fly-Through



Originalsequenz | zeitliche Unterabtastung (1 : 2) | zeitliche Unterabtastung (1 : 3)

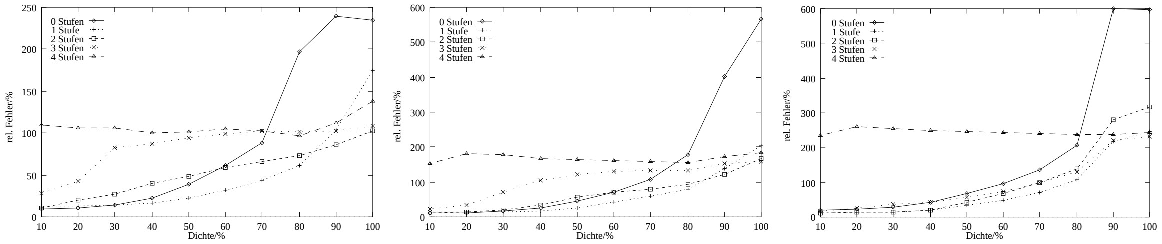


Abbildung 3.14: Winkelfehler (oben) und relativer Fehler (unten) bei Anwendung des Eigenwertverfahrens ($\lambda_s = \lambda_t = 1, \lambda_p = 0$) auf die Yosemite-Fly-Through-Sequenz unter Variation der Zahl von Pyramidenstufen, wobei das Warpen als fehlerbehaftet vorausgesetzt wird.

Hierarchisches Eigenwertverfahren-Verfahren ("fehlerbehaftet") - Insight-Sequenz

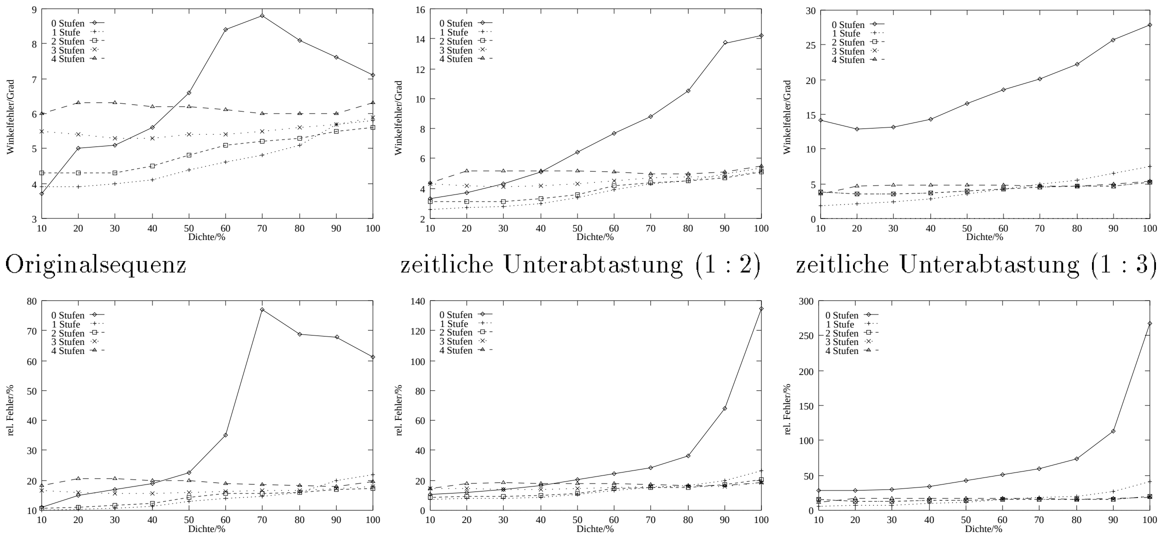


Abbildung 3.15: Winkelfehler (oben) und relativer Fehler (unten) bei Anwendung des Eigenwertverfahrens ($\lambda_s = \lambda_t = 1, \lambda_p = 0$) auf die Insight-Sequenz unter Variation der Zahl von Pyramidenstufen, wobei das Warpen als fehlerbehaftet vorausgesetzt wird.

Hierarchisches Eigenwertverfahren-Verfahren ("fehlerbehaftet") - Diverging-Tree-Sequenz

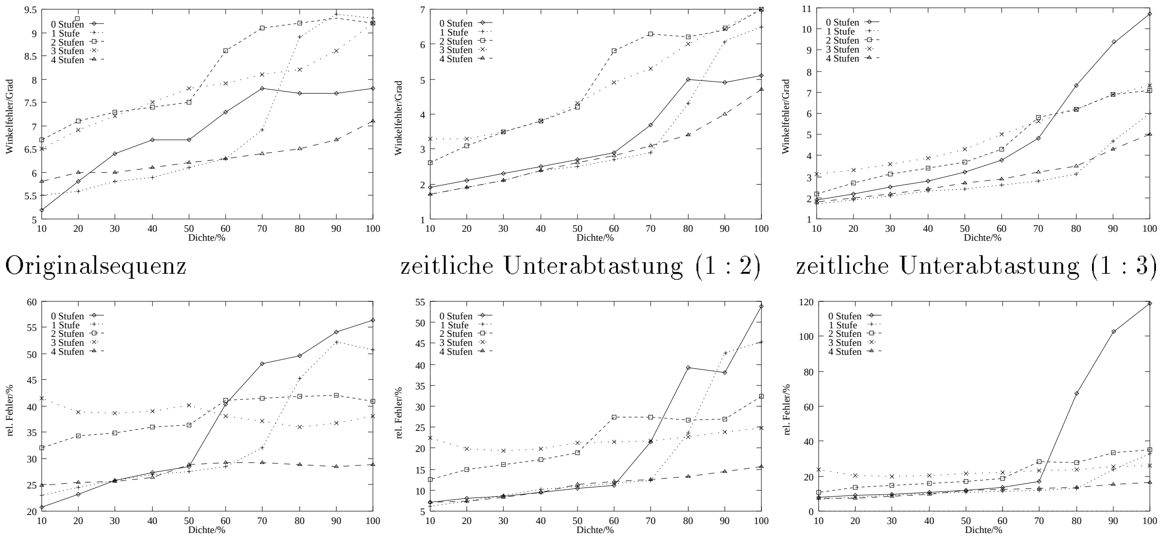


Abbildung 3.16: Winkelfehler (oben) und relativer Fehler (unten) bei Anwendung des Eigenwertverfahrens ($\lambda_s = \lambda_t = 1, \lambda_p = 0$) auf die Diverging-Tree-Sequenz unter Variation der Zahl von Pyramidenstufen, wobei das Warpen als fehlerbehaftet vorausgesetzt wird.

In den Abbildungen 3.11, 3.12 und 3.13 sind die Ergebnisse der Berechnung zu sehen, bei der der Warpvorgang als fehlerfrei vorausgesetzt wird, während in den Abbildungen 3.14, 3.15 und 3.16 die Ergebnisse der fehlerbehafteten Variante dargestellt sind. Auch hier ist zu erkennen, daß eine hierarchische Verarbeitung bessere Ergebnisse liefert, wenn die Stufenzahl der Bildpyramide geeignet gewählt wird. Auffällig ist, daß schon bei der Berechnung des optischen Flusses der Originalsequenz mit der hierarchischen Verarbeitung bessere Ergebnisse erzielt werden.

Im allgemeinen steigt auch bei diesem Verfahren der Fehler mit der Dichte des Flußfeldes. Nur bei einer sehr großen bzw. sehr kleinen Dichte ist das Gütemaß nicht monoton, was dadurch zu erklären ist, daß Ausreißer in der Berechnung auftreten. Des weiteren ist zu bemerken, daß der kleinste Fehler, der durch hierarchische Berechnung auf den Testsequenzen erzielt werden kann, bei steigender Geschwindigkeit sinkt.

Bei dem fehlerfreien Eigenwertverfahren kann man beobachten, daß die Verwendung einer vierstufigen Pyramide nicht sinnvoll ist. Dies läßt sich auch hier wieder damit erklären, daß durch das wiederholte Warpen der dabei gemachte Fehler verstärkt wird. Geht man wie bei der alternativen Strategie davon aus, daß das Warpen fehlerbehaftet ist, so liefert auch die Berechnung mit einer vierstufigen Pyramide bis auf bei Anwendung auf die Yosemite-Fly-Through-Sequenz (s. Abb. 3.14) bessere Ergebnisse als die konventionelle Berechnung.

Der Fehler bei der Yosemite-Fly-Through-Sequenz entsteht wiederum dadurch, daß der exakte optische Fluß im Vergleich zu den anderen Sequenzen am wenigsten örtlich konstant ist. Ferner stellt man fest, daß bei der fehlerbehafteten Variante die Streuung der Fehlers bei verschiedener Anzahl von verwendeten Pyramidenstufen kleiner ist.

Insgesamt läßt sich aus den Ergebnissen folgern, daß gerade bei schnellen Geschwindigkeiten eine hierarchische Berechnung des optischen Flusses vorteilhaft ist. Im allgemeinen liefert die hierarchische Berechnung bessere Ergebnisse als die konventionelle, wobei jedoch zu beachten ist, daß die Zahl der Pyramidenstufen zwar mit steigender Geschwindigkeit der beobachteten Bewegung ebenfalls zunehmen darf, jedoch eine höhere Pyramide nicht unweigerlich bessere Ergebnisse liefert. Der Zusammenhang, je höher die Pyramide ist, desto besser ist die Berechnung, gilt nicht.

Test 3: Hierarchisches Eigenwertverfahren: Gegenüberstellung der Kompositionsmethoden

Bei dem hierarchischen Eigenwertverfahren ist nun von Interesse, wie sich die beiden möglichen Strategien der Komposition von Daten aus verschiedenen Hierarchieebenen unterscheiden. Für einen quantitativen Vergleich der Strategien ist es sinnvoll, den jeweiligen Fehler bei einer vorgegebenen Dichte zu betrachten. Die Ergebnisse der fehlerfreien (*ffW*) und der fehlerbehafteten (*fbW*) Strategie sind in der folgenden Tabelle 3.1 gegenübergestellt. Dabei ist jedes Mal das beste Ergebnis der hierarchischen Berechnung und die Zahl der dabei verwendeten Pyramidenstufen angegeben. Zum Vergleich der Strategien sind die Rauschparameter λ_s, λ_t wieder gleich eins gesetzt und die inverse a priori Varianz λ_p auf null.

Hierarchisches Eigenwertverfahren: *fbW* $\longleftrightarrow$ *ffW*

Yosemite-Fly-Through-Sequenz					
Strategie	*zeitl. Abtastrate*	*Hierarchiestufen*	*Dichte*	*Winkelfehler*	*Stdabw.*
fbW	1 : 1	1	40	4.3	9.8
ffW	1 : 1	1	40	4.5	9.6
fbW	1 : 2	1	40	3.8	9.1
ffW	1 : 2	1	40	3.9	9.3
fbW	1 : 3	1	30	3.1	7.3
ffW	1 : 3	1	30	3.2	6.7
Insight-Sequenz					
Strategie	*zeitl. Abtastrate*	*Hierarchiestufen*	*Dichte*	*Winkelfehler*	*Stdabw.*
fbW	1 : 1	1	60	4.6	4.3
ffW	1 : 1	1	60	4.8	4.6
fbW	1 : 2	1	50	3.4	3.8
ffW	1 : 2	1	50	3.7	4.0
fbW	1 : 3	1	40	2.9	3.6
ffW	1 : 3	1	40	4.5	6.1
Diverging-Tree-Sequenz					
Strategie	*zeitl. Abtastrate*	*Hierarchiestufen*	*Dichte*	*Winkelfehler*	*Stdab.*
fbW	1 : 1	1	50	6.1	2.2
ffW	1 : 1	1	50	6.3	3.3
fbW	1 : 2	1	50	2.5	2.1
ffW	1 : 2	1	50	2.9	2.3
fbW	1 : 3	1	50	2.4	1.9
ffW	1 : 3	1	40	2.8	2.1

Tabelle 3.1: Vergleich der verschiedenen Kompositionsstrategien bei der hierarchischen Erweiterung des Verfahrens von Kanatani ($\lambda_s = \lambda_t = 1, \lambda_p = 0$): zum einen die Strategie des fehlerbehafteten Warpens (*fbW*), zum anderen die des fehlerfreien Warpens (*ffW*).

Da, wie man schon bei der Darstellung der exakten Flußfelder (s. Abb. 3.6) gesehen hat, wegen der Bildstruktureigenschaften die Dichte des berechneten Flußfeldes nicht bei jeder der drei Testsequenzen gleich hoch zu wählen ist, sind die Ergebnisse für unterschiedliche Dichten dargestellt. Der relative Fehler hat, wie in den Abbildungen (3.11) bis (3.16) gezeigt, einen ähnlichen Verlauf wie der Winkelfehler. Daher ist nur der Winkelfehler zusammen mit der Standardabweichung dargestellt.

Wie man Tabelle 3.1 entnehmen kann, werden unter der Annahme, daß das Warpen mit einem Fehler verbunden ist, deutlich bessere Ergebnisse erzielt. Auch die Standardabweichung ist bei der fehlerbehafteten Variante im allgemeinen geringer als bei der fehlerfreien. Die schwächere Gewichtung der feineren Messungen führt demnach bei der hierarchischen Berechnung zu einer genaueren Schätzung des optischen Flusses.

Test 4: Gegenüberstellung der nicht hierarchischen Verfahren: Simoncelli ↔ Kanatani

Als nächstes folgt der experimentelle Vergleich des Verfahrens von Simoncelli et al. ohne hierarchischen Ansatz mit dem von Kanatani, der in der Literatur bisher nicht vorgestellt worden ist. Um wie beim theoretischen Vergleich in Kapitel 3.2.7 die Verfahren direkt zueinander in Bezug setzen zu können, wird auch hier wieder angenommen, daß das Rauschen in der zeitlichen sowie in der örtlichen Komponente durch einen Gaußschen Fehler gegeben ist, keine a priori Varianz vorliegt und kein weiterer Fehler auftritt: $\lambda_t = \lambda_s = 1, \lambda_p = \lambda_{\boldsymbol{v}} = 0$. In Tabelle 3.2 findet man für jede Testsequenz den bei einer vorgegebenen Dichte erzielten Winkelfehler, die zugehörige Standardabweichung und die mittlere Länge des Geschwindigkeitsvektors.

Simoncelli-Verfahren $\longleftrightarrow$ Kanatani-Verfahren

Yosemite-Fly-Through-Sequenz					
Verfahren	*zeitl. Abtastrate*	*Dichte*	*Winkelfehler*	*Stdabw.*	*mittlere Länge*
Simoncelli	1 : 1	40	4.3	10.9	1.0
Kanatani	1 : 1	40	5.9	12.5	1.4
Simoncelli	1 : 2	40	9.2	17.6	2.0
Kanatani	1 : 2	40	9.4	20.7	2.3
Simoncelli	1 : 3	30	17.0	23.4	3.0
Kanatani	1 : 3	30	12.6	24.6	2.7
Insight-Sequenz					
Verfahren	*zeitl. Abtastrate*	*Dichte*	*Winkelfehler*	*Stdabw.*	*mittlere Länge*
Simoncelli	1 : 1	60	4.0	3.3	1.3
Kanatani	1 : 1	60	8.4	13.3	1.5
Simoncelli	1 : 2	50	5.6	7.6	2.4
Kanatani	1 : 2	50	6.4	10.5	2.4
Simoncelli	1 : 3	40	19.1	20.8	3.8
Kanatani	1 : 3	40	14.3	24.4	3.5
Diverging-Tree-Sequenz					
Verfahren	*zeitl. Abtastrate*	*Dichte*	*Winkelfehler*	*Stdabw.*	*mittlere Länge*
Simoncelli	1 : 1	50	4.7	2.2	0.80
Kanatani	1 : 1	50	6.7	5.1	0.9
Simoncelli	1 : 2	50	2.1	1.6	1.6
Kanatani	1 : 2	50	2.7	2.3	1.7
Simoncelli	1 : 3	50	3.2	2.5	2.4
Kanatani	1 : 3	50	3.2	4.1	2.3

Tabelle 3.2: Vergleich des Verfahrens von Simoncelli ohne Pyramidenverarbeitung ($\lambda_t = 1, \lambda_{\boldsymbol{v}} = \lambda_p = 0$) mit dem von Kanatani ($\lambda_s = \lambda_t = 1, \lambda_p = 0$) bei vorgegebener Dichte

Wie aus Tabelle 3.2 ersichtlich ist, erzielt das Verfahren von Simoncelli et al. bei den Originalsequenzen die besseren Resultate, während das Kalman-Verfahren bei den im Verhältnis 1 : 3 zeitlich unterabgetasteten Sequenzen bessere Ergebnisse liefert. Bei den im Verhältnis 1 : 2 zeitlich unterabgetasteten Sequenzen ist der Fehler und die Standardabweichung der mit dem Verfahren von Simoncelli et. al. Ergebnisse geringer, aber die Resultate des Kanatani-Verfahrens weichen davon weniger stark ab als bei den Originalsequenzen. Ohne Pyramidenverarbeitung und ohne spezielle Rauschbetrachtung liefert also das Verfahren von Simoncelli et al. für kleine Geschwindigkeiten und das von Kanatani für große Geschwindigkeiten die besseren Ergebnisse.

Das Resultat des theoretischen Vergleichs in Kapitel 3.2.7 war, daß die Verfahren von Simoncelli et al. und von Kanatani sich dadurch unterscheiden, daß bei ersterem die a priori Varianz auf die Kovarianzmatrix addiert wird und bei letzterem ein positiver Faktor subtrahiert wird. Man erwartet also, daß die Geschwindigkeit, die mit dem Simoncelli-Verfahren berechnet wird, kleiner ist als die mit dem Verfahren von Kanatani erzielte. Zwar ist bei beiden Verfahren die a priori Varianz auf null gesetzt, doch auch dann muß das Verhältnis gelten oder zumindest Gleichheit vorliegen. Um dies zu überprüfen, ist in Tabelle 3.2.7 die mittlere Länge des Geschwindigkeitsvektors angegeben. Wie man sieht, tritt die erwartete Beziehung für die Originalsequenzen und die im Verhältnis 1 : 2 zeitlich unterabgetasteten ein, aber für die im Verhältnis 1 : 3 zeitlich unterabgetasteten gilt dies nicht, was auf numerische Fehler hindeutet.

Test 5: Gegenüberstellung der hierarchischen Verfahren: Simoncelli $\leftrightarrow$ Eigenwertverfahren

Es folgt nun der Vergleich der beiden Berechnungsansätze bei hierarchischer Verarbeitung. Dazu betrachtet man das hierarchische Verfahren von Simoncelli et al. gegenüber der Erweiterung des Verfahrens von Kanatani, dem hierarchischen Eigenwertverfahren. Da die Annahme eines fehlerbehafteten Warpvorgang, wie schon gezeigt, bessere Ergebnisse liefert als die fehlerfreie Variante, wird die fehlerbehaftete Strategie zur Gegenüberstellung benutzt. Wiederum wird das Rauschen in der örtlichen bzw. zeitlichen Komponente durch einen Gaußschen Fehler modelliert und der Modellfehler vernachlässigt: $\lambda_s = \lambda_t = 1,\ \lambda_v = 0$. Außerdem wird keine a priori Varianz benutzt ($\lambda_p = 0$).

In Tabelle 3.3 sind die Ergebnisse der verschiedenen hierarchischen Algorithmen gegenübergestellt. Dabei ist jeweils das Ergebnis derjenigen Berechnungen mit der verwendeten Stufenzahl der Pyramide angegeben, welche den kleineren Fehler liefert. Dargestellt werden der Winkelfehler und die zugehörige Standardabweichung bei vorgegebener Dichte des Flußfeldes.

Wie der Tabelle zu entnehmen ist, liefert das hierarchische Simoncelli-Verfahren mit verschiedenen Pyramidenstufen bei jeder Testsequenz bessere Schätzungen des optischen Flusses als das hierarchische Eigenwertverfahren. Auch die Standardabweichung ist kleiner. Der Fehler der mit dem Eigenwertwertverfahren erzielten Ergebnisse bewegt sich jedoch in einer ähnlichen Größenordnung wie der mit dem Simoncelli-Verfahren erzielte. Die schlechtere Leistung des Eigenwertverfahrens ist in Anbetracht der Ergebnisse der nicht hierarchischen Verarbeitung dadurch zu erklären, daß zwar die Schätzung auf

Hierarchisches Simoncelli-Verfahren $\longleftrightarrow$ Hierarchisches Eigenwertverfahren

Yosemite-Fly-Through-Sequenz					
Verfahren	*zeitl. Abtastrate*	*Hierarchiestufen*	*Dichte*	*Winkelfehler*	*Stdabw.*
Simoncelli	1 : 1	1	40	3.6	8.5
Eigenwert	1 : 1	1	40	4.5	9.6
Simoncelli	1 : 2	1	40	3.4	8.7
Eigenwert	1 : 2	1	40	3.8	9.1
Simoncelli	1 : 3	2	30	2.6	5.5
Eigenwert	1 : 3	2	30	3.2	6.7
Insight-Sequenz					
Verfahren	*zeitl. Abtastrate*	*Hierarchiestufen*	*Dichte*	*Winkelfehler*	*Stdabw.*
Simoncelli	1 : 1	0	60	4.0	3.3
Eigenwert	1 : 1	0	60	8.4	13.3
Simoncelli	1 : 1	1	60	4.2	3.2
Eigenwert	1 : 1	1	60	4.6	4.3
Simoncelli	1 : 2	1	50	3.0	2.9
Eigenwert	1 : 2	1	50	3.4	3.8
Simoncelli	1 : 3	1	40	2.1	1.9
Eigenwert	1 : 3	1	40	2.9	3.6
Simoncelli	1 : 3	2	40	2.0	1.9
Eigenwert	1 : 3	2	40	3.7	5.6
Diverging-Tree-Sequenz					
Verfahren	*zeitl. Abtastrate*	*Hierarchiestufen*	*Dichte*	*Winkelfehler*	*Stdabw.*
Simoncelli	1 : 1	0	50	4.7	2.2
Eigenwert	1 : 1	0	50	6.7	5.1
Simoncelli	1 : 1	1	50	5.9	1.8
Eigenwert	1 : 1	1	50	6.1	2.2
Simoncelli	1 : 2	1	50	2.2	1.8
Eigenwert	1 : 2	1	50	2.5	2.1
Simoncelli	1 : 2	3	50	2.0	1.6
Eigenwert	1 : 2	3	50	4.3	3.7
Simoncelli	1 : 3	1	50	2.1	1.7
Eigenwert	1 : 3	1	50	2.4	1.9
Simoncelli	1 : 3	4	50	1.9	1.4
Eigenwert	1 : 3	4	50	2.7	1.7

Tabelle 3.3: Vergleich des hierarchischen Verfahrens von Simoncelli ($\lambda_t = 1, \lambda_{\boldsymbol{v}} = \lambda_p = 0$) mit dem von uns entwickelten hierarchischen Eigenwertverfahren ($\lambda_s = \lambda_t = 1, \lambda_p = 0$) , bei dem das Warpen als fehlerbehaftet angenommen wird.

der gröbsten Stufe besser ist, jedoch wegen der schlechteren Resultate auf kleineren Geschwindigkeiten die Verfeinerung des Ergebnisses auf feineren Auflösungsstufen nach dem Warpen das Gesamtergebnis verschlechtert.

Test 6: Variation der a priori Varianz bei den hierarchischen Verfahren

Bisher sind für den direkten Vergleich der Verfahren untereinander die Rauschparameter nicht variiert worden. Im folgenden werden daher die Auswirkungen der Veränderung einzelner Parameter auf die Berechnung getestet. Dazu wird zunächst die frei wählbare inverse a priori Varianz λ_p variiert.

Um zu testen, wie sich die Ergebnisse bei Variation der a priori Varianz verändern, werden die hierarchischen Verfahren auf eine Pyramide mit der Stufenzahl, die in den vorherigen Versuchen zum kleinsten Fehler geführt hat, angewendet. Das heißt, daß zur Berechnung des Fehlers auf der Yosemite-Fly-Through-Sequenz das zweistufige Simoncelli-Verfahren und das zweistufige Eigenwertverfahren benutzt wird, zur Berechnung des Fehlers auf der Insight-Sequenz das zweistufige Simoncelli- und das einstufige Eigenwertverfahren und auf der Diverging-Tree-Sequenz das vierstufige Simoncelli- und das einstufige Eigenwertverfahren.

Bei dem hierarchischen Eigenwertverfahren ist bisher für die erste Berechnungsstufe keine a priori Varianz benutzt worden. Man kann jedoch auch hier eine a priori Varianz einführen, indem man zur Berechnung die Vorschrift (3.80) benutzt und auch hier die a priori Kovarianzmatrix $\mathbf{\Lambda}_p$ durch eine Diagonalmatrix mit den Diagonalelementen $\frac{1}{\lambda_p}$ ersetzt. Die inverse a priori Varianz durchläuft bei beiden Verfahren die Werte $0.005, 0.05, 0.5, 0.75, 1.0, 1.25, 1.5, 2.0, 3.0, 5.0$. Die Rauschparameter bleiben unverändert: $\lambda_s = \lambda_t = 1, \lambda_v = 0$. Der Winkelfehler ist in Abb. 3.17 für die Yosemite-Fly-Through-Sequenz bei einer Dichte von 30%, für die Insight-Sequenz bei einer Dichte von 40% und für die Diverging-Tree-Sequenz bei einer Dichte von 50% dargestellt.

Variation der a priori Varianz

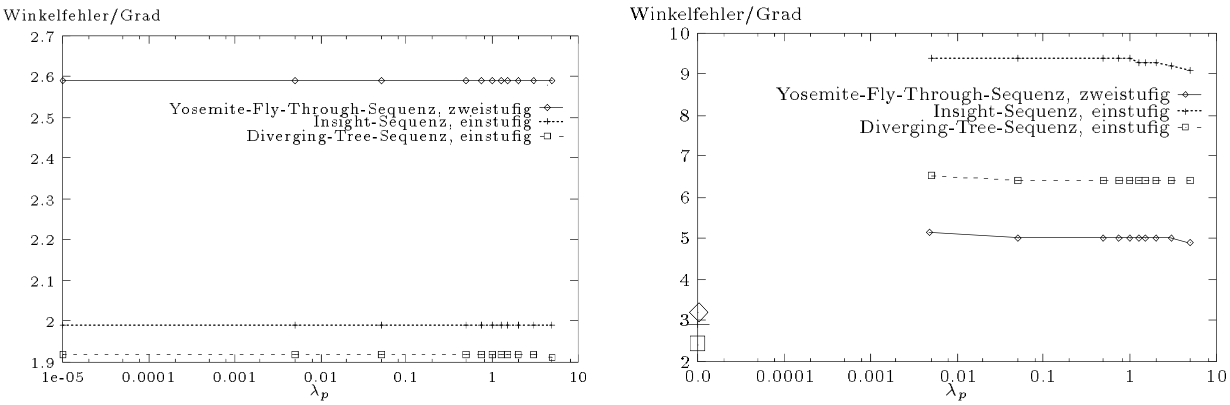


Abbildung 3.17: Darstellung des Winkelfehler bei Variation der inversen a priori Varianz λ_p; links: hierarchisches Simoncelli-Verfahren ($\lambda_t = 1, \lambda_v = 0$), rechts: hierarchisches Eigenwertverfahren ($\lambda_s = \lambda_t = 1$)

Aus Abb. 3.17 geht direkt hervor, daß die Variation der inversen a priori Varianz beim hierarchischen Simoncelli-Verfahren keine deutlichen Auswirkungen auf den Fehler des berechneten optischen Flusses hat. Man kann also zur Vermeidung von Problemen, die dadurch entstehen, daß die Kovarianzmatrix bei der Berechnung des optischen Flusses auf der gröbsten Stufe nicht invertierbar ist, eine beliebige a priori Varianz wählen.

Anders ist das Verhalten hingegen bei dem hierarchischen Eigenwertverfahren. Hier ist von Bedeutung, ob die inverse a priori Varianz ignoriert wird ($\lambda_p = 0$), oder ob sie benutzt wird ($\lambda_p > 0$), was daran liegt, daß sowohl die inverse a priori Kovarianzmatrix als auch die a priori Kovarianzmatrix selbst in die Berechnung einfließen. Daher ist auch keine lineare Fortsetzung des Fehlers zwischen den Berechnungen mit $\lambda_p > 0$ und $\lambda_p = 0$ möglich. Wie man sieht, sind die Ergebnisse ohne Berücksichtigung einer a priori Varianz deutlich besser. Berücksichtigt man jedoch die a priori Varianz, so ändert eine Variation des Wertes von λ_p den Fehler nicht wesentlich.

Test 7: Variation der Rauschparameter bei den nicht hierarchischen Verfahren

Zuletzt werden die Auswirkungen der Variation der Rauschparameter auf das Ergebnis untersucht. Dazu wird zunächst das nicht hierarchische Simoncelli-Verfahren auf die Originalsequenzen angewendet. Die a priori Varianz wird bei diesem Test wiederum ignoriert ($\lambda_p = 0$). Das Rauschen in der zeitlichen Komponente wird nun durch Normalverteilungen mit dem Mittelwert null und der Varianz λ_t modelliert, wobei λ_t die Werte $0.1, 0.25, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$ durchläuft. Der Modellfehler wird ebenfalls durch eine Normalverteilung gegeben, die den Mittelwert null und die Varianz $\lambda_{\boldsymbol{v}}$ hat.

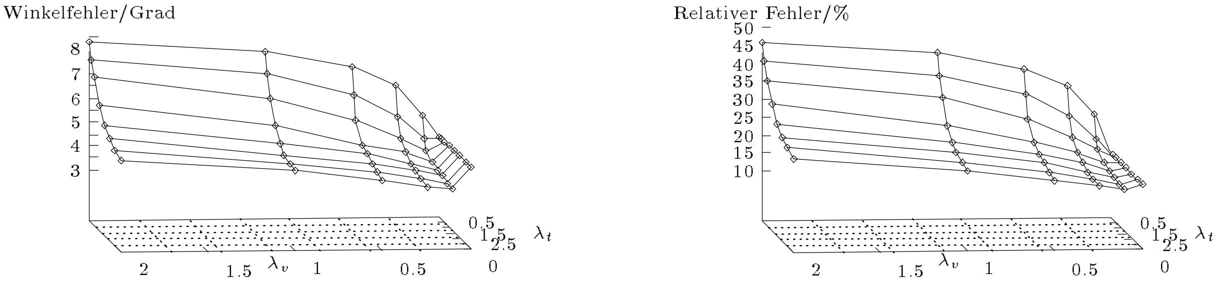
Der Parameter $\lambda_{\boldsymbol{v}}$ durchläuft bei diesem Versuch die Werte $0.0, 0.1, 0.25, 0.5, 1.0, 2.0$. In Abb. 3.18 sind der Winkelfehler und der relativer Fehler des damit berechneten Flusses zu sehen, wobei die Dichte des Flußfeldes bei der Yosemite-Fly-Through-Sequenz und bei der Insight-Sequenz 40% beträgt und bei der Diverging-Tree-Sequenz 50%.

Aus Abb. 3.18 wird deutlich, daß die kleinsten Fehler bei einem angenommenen Modellrauschen mit der Varianz $\lambda_{\boldsymbol{v}} = 0$ bzw. $\lambda_{\boldsymbol{v}} = 0.1$ (Yosemite-Fly-Through-Sequenz) erzielt wurden, wobei der Unterschied bei den Sequenzen mit der unterschiedlichen lokalen, örtlichen Konstanz des Flusses zu begründen ist, was einen unterschiedlichen Modellfehler bedingt. Wird für das Rauschen in der zeitlichen Komponente eine Varianz von ca. $\lambda_t \approx 2.5$ oder größer gewählt, so ist eine gute Schätzung des optischen Flusses möglich. Kombiniert man diese Parameterwerte, so erhält man bei jeder der drei Testsequenzen einen kleineren Fehler als bei der rauschfreien Betrachtung ($\lambda_t = 1, \lambda_{\boldsymbol{v}} = 0$).

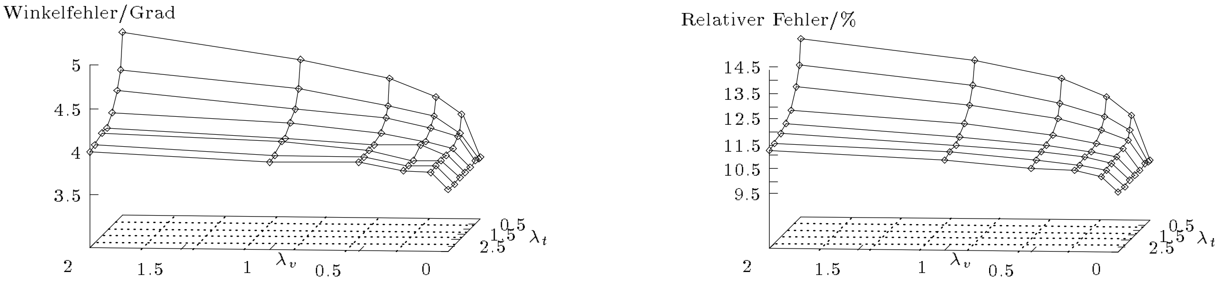
Die Variation der Rauschparameter bei dem hierarchischen Eigenwertverfahren erfolgt in ähnlicher Weise. Hier werden das Rauschen in der örtlichen und zeitlichen Komponente durch Normalverteilungen mit Mittelwert null und Varianz λ_s bzw. λ_t modelliert. Die Rauschparameter nehmen dabei folgende Werte an: $0.5, 1.0, 1.5, 2.0, 2.5, 3.0$. Der zeitliche Rauschparameter nimmt zusätzlich die Werte $3.5, 4.0, 5.0$ an. Der Winkelfehler und der relative Fehler des optischen Flußfeldes, welches mit diesen Parametern berechnet wird, ist in Abb. 3.19 dargestellt. Die Dichte der Flußfelder beträgt dabei wieder 40% bzw. 50%.

In Abb. 3.19 ist zu erkennen, daß sowohl der Winkel- als auch der relative Fehler bei einem Wert größer gleich 4.5 für den örtlichen Rauschparameter λ_s die kleinsten Werte annimmt.

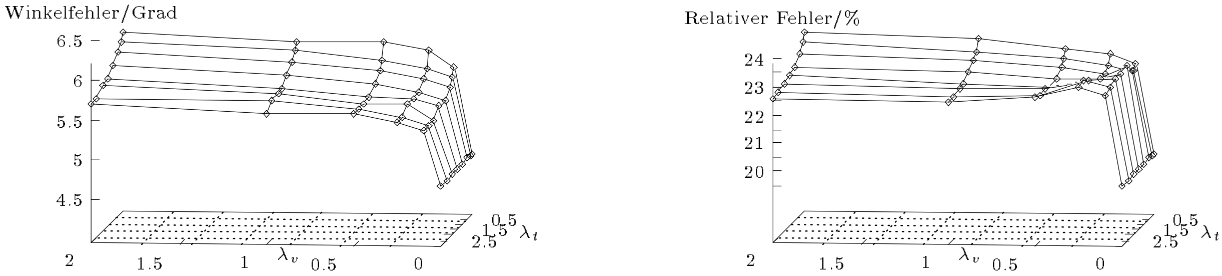
Simoncelli-Verfahren - Variation der Rauschparameter



Yosemite-Fly-Through-Sequenz



Insight-Sequenz



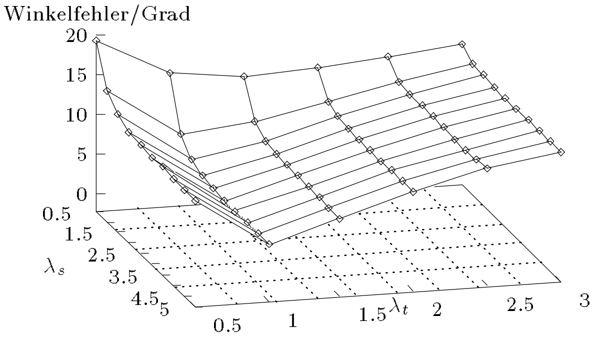
Diverging-Tree-Sequenz

Abbildung 3.18: Winkelfehler und relativer Fehler der Ergebnisse, die beim Anwenden des Simoncelli-Verfahrens ($\lambda_p = 0$) auf die zeitlich nicht unterabgetasteten Testsequenzen unter Variation der Rauschparameter $\lambda_{\boldsymbol{v}}, \lambda_t$ erzielt wurden.

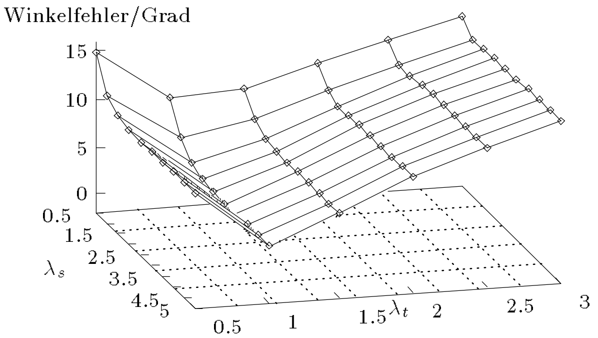
Die Wahl von 1.0 bzw. 1.5 (Diverging-Tree-Sequenz) für den zeitlichen Rauschparameter λ_t stellt gleichermaßen eine gute Berechnungsgrundlage dar. Kombiniert man diese beiden Parameterwerte, so erhält man einen minimalen Fehler der Berechnung. Die Berechnung, die das Rauschen berücksichtigt, liefert auch hier bessere Ergebnisse als die konventionelle Methode ($\lambda_s = \lambda_t = 1$).

Insgesamt ist daraus ersichtlich, daß eine explizite Berücksichtigung des Rauschens zu besseren Resultaten führt als die übliche Methode, die das Rauschen durch das verwendete Schätzungsverfahren allein behandelt. Bei dem Verfahren von Simoncelli et al. ist eine Parameterwahl von $\lambda_t \approx 2.5$ und $\lambda_{\boldsymbol{v}} \approx 0$ ratsam, während das Verfahren von Kanatani bei der Kombination $\lambda_s \approx 4.5$ und $\lambda_t = 1.0$ die beste Leistung erbringt.

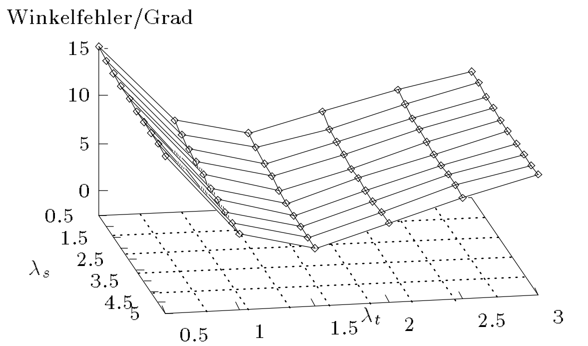
Kanatani-Verfahren - Variation der Rauschparameter



Yosemite-Fly-Through-Sequenz



Insight-Sequenz



Diverging-Tree-Sequenz

Abbildung 3.19: Winkelfehler und Relativer Fehler der Ergebnisse, die beim Anwenden des Kanatani-Verfahrens ($\lambda_p = 0$) auf die zeitlich nicht unterabgetasteten Testsequenzen unter Variation der Rauschparameter λ_s, λ_t für die örtliche und zeitliche Komponente erzielt wurden.

Fazit

Abschließend sind die besten Resultate, die durch das Rauschen berücksichtigende Berechnungen des optischen Flusses mittels dem Simoncelli-Verfahren und dem Eigenwertverfahren erzielt worden sind, dem konventionellen Verfahren von Lukas & Kanade gegenübergestellt. Die in Tabelle 3.4 dargestellten Fehler gelten für die Bestimmung des optischen Flusses auf den Originalsequenzen.

Zusammenfassend ist zu sagen, daß die hierarchische Verarbeitung schneller Bildbewegungen bei beiden Verfahren zu exakteren Schätzungen des optischen Flusses führt. Dabei stellt die Kompositionsmethode, die von einem fehlerbehafteten Warpvorgang ausgeht,

Gegenüberstellung der Verfahren von Lukas & Kanade, Simoncelli und Kanatani

Yosemite-Fly-Through-Sequenz			
Verfahren	*Dichte*	*Winkelfehler*	*Stdabw.*
Simoncelli	35	3.1	7.0
Kanatani	35	3.6	9.5
Lukas & Kanade	35.1	4.1	9.6
Insight-Sequenz			
Verfahren	*Dichte*	*Winkelfehler*	*Stdabw.*
Simoncelli	56	3.9	3.2
Kanatani	56	4.2	3.7
Lukas & Kanade	56	4.2	3.7
Diverging-Tree-Sequenz			
Verfahren	*Dichte*	*Winkelfehler*	*Stdabw.*
Simoncelli	50	4.7	2.2
Eigenwert	50	3.3	2.5
Lukas & Kanade	50	5.2	2.7

Tabelle 3.4: Vergleich der nicht hierarchischen Verfahren von Simoncelli ($\lambda_t = 3, \lambda_{\boldsymbol{v}} = 0, \lambda_p = 0.005$), von Kanatani ($\lambda_s = 4.5, \lambda_t = 1, \lambda_p = 0$) und dem von Lukas & Kanade

für das hierarchische Eigenwertverfahren die bessere Variante dar. Die Berücksichtigung des Rauschens ermöglicht in beiden Fällen eine Optimierung der Ergebnisse.

Bei der Anwendung der nicht hierarchischen Verfahren auf schnelle Bewegungen liefert das Verfahren von Kanatani die besseren Ergebnisse, während das Verfahren von Simoncelli et al. bei kleineren Bewegungen überragt. Auch die erzielten Resultate des von uns entwickelten hierarchischen Verfahrens werden von den Ergebnissen des Verfahrens von Simoncelli übertroffen. Sie liegen jedoch in einem ähnlichen Bereich.

Die Berechnung des optischen Flusses unter Verwendung einer frei wählbaren a priori Varianz stellt für das Verfahren von Simoncelli eine Garantie der Invertierbarkeit der Kovarianzmatrix dar und hat nur minimale Auswirkungen auf das Ergebnis. Die Berücksichtigung einer a priori Varianz verschlechtert die Ergebnisse des hierarchischen Eigenwertverfahrens jedoch erheblich.

Kapitel 4

Beitrag der Flußberechnung zur Gelenkverfolgung

In den vorigen Kapiteln sind zum einen die Grundlagen der Berechnung des optischen Flusses gegeben worden, und zum anderen sind spezielle hierarchische Algorithmen vorgestellt worden, die sowohl das Problem der Fusion von Sensordaten durch die Formulierung eines Gütemaßes lösen als auch das Rauschen der Messungen berücksichtigen. Bisher ist der Beitrag der Flußberechnung zur Gelenkverfolgung nur kurz dargestellt worden. Welche Rolle die Flußberechnung bei der Gelenkverfolgung genau spielt, wird im nun folgenden Kapitel ausführlich geklärt.

Nach einem Überblick über die bisherigen Arbeiten zur Realisierung der Verfolgung von gegliederten Objekten, folgt die Beschreibung der Implementierung des Verfahrens zur Bewegungsanalyse und die nähere Erklärung einiger Teilschritten. Im Anschluß daran folgt die Darstellung der in Experimenten erzielten Leistungen des Verfahrens.

4.1 Bisherige Arbeiten zur Bewegungsverfolgung artikulierter Objekte

Die Bewegungslehre in der Biomechanik kennt mehrere Techniken zur direkten Messung von menschlichen Bewegungen. Eine Möglichkeit ist die Messung mittels eines Teilspannungsmessers, eines sogenannten Goniometers [35], welches an den Gelenken befestigt wird. Die gemessene Spannung hängt dabei direkt von dem Gelenkwinkel ab und stellt somit ein Maß der Bewegung dar. Ein anderes Gerät zur Bewegungsmessung ist das Accelerometer [35], welches basierend auf einem Kraftumwandler die Reaktionskraft zu einer Beschleunigung mißt. Über die Beschleunigung können dann Schlüsse auf die Bewegung gezogen werden. Die Vorteile dieser Meßapparate liegen in der direkten Verfügbarkeit der Bewegungsdaten. Nachteilig ist jedoch, daß die Apparate nur relative Daten liefern und zudem die Bewegungsfreiheit einschränken.

Doch auch die Bewegungsanalyse aus Bilddaten wird in der Biomechanik verwandt. Zum einen gibt es dort die Methode der Mehrfachbelichtung [35]. Dabei wird mit einem Stroboskop in einem abgedunkelten Raum die sich bewegende Person, die dunkle Kleidung und reflektierende Streifen oder Marker trägt, in regelmäßigen Abständen angestrahlt. Der Film wird bei jedem Blitz erneut belichtet, so daß man den Ablauf der Bewegung vom Bild ablesen kann. Zum anderen wird auch die visuelle Bewegungsanalyse mittels Bildsequenzen benutzt, bei der die Person spezielle Marker an den charakteristischen Körperstellen tragen muß [35]. Die Analyse erfolgt über Zuordnung der einander entsprechenden Positionen der Marker von Bild zu Bild. Gegenüber den oben genannten Meßapparaten liegt der Vorteil dieser Techniken darin, daß die Bilddaten eine repräsentative (d.h. anschauliche) Darstellung der Analyse sind, die Analyse an beliebigen Stellen des Körpers erfolgen kann und die Bewegungsfreiheit der Person weniger stark eingeschränkt ist. Doch ist es nicht möglich, zu anderen Zwecken aufgenommene Bildsequenzen zu analysieren. Außerdem ist eine Zuordnung der korrespondierenden Marker in zwei aufeinander folgenden Bildern nicht immer durchführbar. Psychologische Untersuchungen belegen zudem, daß eine Bewegungsanalyse, die nur auf Gelenkmustern beruht, schon beim menschlichen Gehirn zu Schwierigkeiten führt [10].

Mit dem Ziel, die Bewegungsverfolgung als Teil der visuellen Bewegungsanalyse artikulierter Objekte von Markern unabhängig zu machen, werden in der Literatur unterschiedliche Ansätze verfolgt. Man unterscheidet zum einen zwischen 2D-Verfahren, die in der Bildebene operieren, und 3D-Verfahren, die mit räumlichen Strukturen arbeiten. Ein weiterer Unterschied liegt in der Art und Weise, wie die Freiheitsgrade der Bewegung, die mit der Struktur und der Bewegung des menschlichen Körpers assoziiert sind, eingeschränkt werden. Die meisten Verfahren erreichen dies entweder durch modellbasierte Ansätze, die vordefinierte Körpermodelle benutzen, oder durch eine Einschränkung der Bewegungsmöglichkeiten. Ein vollständiger Verzicht auf vordefinierte Modelle oder Einschränkungen der Bewegungsfreiheit ist wegen der großen Zahl an Freiheitsgraden unpraktikabel.

Die meisten 3D-Verfahren konzentrieren sich auf die Verfolgung der Konturkanten, welche durch ein Anpassen der Projektion eines detailierten, artikulierten, dreidimensionalen Modells an die aus der Grauwertsequenz extrahierten Kantendaten realisiert wird. Dazu gehören die im folgenden kurz vorgestellten Verfahren von Gavrila und Davis, von Goncalves et al. und von Rohr.

Gavrila und Davis benutzen in ihrem Verfahren [12], [13] die Daten von mehreren zueinander orthogonalen Kameras, was auch als Mehransichten-Ansatz bezeichnet wird. Die Verfolgung einer sich bewegenden Person wird hier durch die Bestimmung der Körperhaltung in jedem einzelnen Bild realisiert. Dazu generieren Gavrila und Davis mögliche Körperhaltungen aus einem dreidimensionalen Modell und projizieren diese in Bildebenen, die zu den optischen Achsen der benutzten Kameras orthogonal sind. Die aktuelle Haltung der Person wird durch eine Kantenanpassung zwischen der synthetisierten Modellansicht und der aktuellen Szene bestimmt. Das Ähnlichkeitsmaß ist die Summe der Abstände von jedem Punkt aus der einen Kantenmenge zu dem nächsten Punkt aus der anderen Menge. Das erste gefundene Minimum zeigt die größte Ähnlichkeit an. Das benutzte

Körpermodell besteht aus zwei Komponenten: einer Strichfigur, die die Skelettstruktur wiedergibt, und einer Darstellung des umgebenden Weichteilgewebes. Die Skelettstruktur ist eine Anordnung einzelner Segmente und räumlicher Gelenkwinkel, denen insgesamt 22 jeweils unterschiedliche Freiheitsgrade zugeordnet sind. Die einzelnen Segmente werden durch Volumenprimitiva, nämlich zugespitzte Super-Quadriken (Zylinder, Sphaeren, Ellipsoide, Hyperrechtecke ohne Deformation), plastisch gemacht.

Durch den Mehransichten-Ansatz können Doppeldeutigkeiten, die bei nur einer Ansicht durch Selbstverdeckung auftreten können, aufgehoben werden. Außerdem sind die Eingabebildfolgen nicht auf eine bestimmte Ansicht beschränkt. Gavrila und Davis setzen die Verwendung mehrerer starrer, kalibrierter Kameras und eng anliegender Kleidung für die Anwendung ihres Verfahrens voraus. Zur Vorverarbeitung der Bilder gehört die Hintergrundsubtraktion, was zusätzlich die Kenntnis des Hintergrundes verlangt.

Goncalves et al. beschränken sich in ihrem Verfahren auf die Verfolgung eines Armes [15]. Grundlage ihres Verfahrens ist ein dreidimensionales Modell des Armes, welches aus zwei Kegeln für Unter- und Oberarm und zwei Raumwinkeln mit insgesamt vier Freiheitsgraden für Schulter- und Ellenbogengelenk besteht. Die Hand wird einfach als Fortsetzung des Unterarms modelliert, so daß keine weiteren Freiheitsgrade durch das Handgelenk hinzukommen. Zur Anpassung des Modells werden sieben absolute Größen benutzt, was die Verwendung einer kalibrierten Kamera voraussetzt. Als zusätzliche Voraussetzung muß die Position des Schultergelenks in jedem Bild bekannt sein. Die vier Gelenkwinkel dienen als Parameter der Armhaltung und bilden somit den gesuchten Zustand. Als rekursiven Schätzer für diesen Zustand benutzen Goncalves et al. ein erweitertes Kalman-Filter. Die zur Korrektur des prädizierten Zustandes notwendige Messung erfolgt über ein Anpassen der Projektion der prädizierten Instanz des Armmodells an ein aus dem Grauwertbild berechnetes Binärbild. Durch die Erweiterung dieses Verfahrens auf die Verfolgung des ganzen Körpers erhält man wesentlich mehr Freiheitsgrade, was aber die Messung der Position komplizierter macht. Zudem wird vorausgesetzt, daß der Hintergrund dunkel ist, wodurch die Anwendbarkeit stark einschränkt wird.

Sowohl bei dem Verfahren von Gavrila und Davis als auch bei dem von Goncalves et al. ist die Menge der möglichen Modellinstanzen nur durch die festgelegten Freiheitsgrade eingeschränkt, weswegen der Parameterraum sehr groß ist und die Messung der Position, also die Bestimmung der passenden Modellparameter viel Rechenzeit kostet.

Auch das Verfahren von Rohr [29] zur Verfolgung von Gehbewegungen beruht auf einem Zuordnungsverfahren zwischen der Projektion möglicher Modellinstanzen und dem extrahierten Kantenbild der Person. Das dabei benutzte Körpermodell ist nach drei von Marr [24] erarbeiteten Kriterien entworfen:

- ▹ Das Koordinatensystem ist objektorientiert.
- ▹ Die einzelnen Körperteile sind durch Volumenprimitiva gegeben.
- ▹ Die Modellbeschreibung unterliegt einer modularen, hierarchischen Organisation.

Dementsprechend ist das Modell aus 14 Zylindern mit elliptischer Grundfläche aufgebaut, die an den Gelenkpunkten verbunden sind. Bestimmt wird jeder Zylinder durch drei Parameter: die jeweilige Länge des Zylinders und der beiden Hauptachsen. Daher ist auch bei diesem Verfahren das Kalibrieren der Kamera notwendig. Der Ursprung des Körperkoordinatensystems liegt im Zentrum des Rumpfes. Um die Menge der möglichen Instanzen des Körpermodells einzuschränken, benutzt Rohr zusätzlich ein Bewegungsmodell, welches die Gehbewegung eines Menschen durch eine Bewegungskurve, die von einem Posenparameter abhängt, modelliert. Da sich die Bewegungsmuster der Körperteile individuell nur wenig voneinander unterscheiden, ist das Modell aus Daten von Bewegungsstudien des Menschen aufgebaut. Als Parameter dient der Gelenkwinkel zwischen zwei Körperteilen, dessen gemessenen, diskreten Werte durch periodische, kubische Splines interpoliert werden.

Zunächst wird nun die die Person enthaltende Region des Bildes per Änderungsdetektion bestimmt, um daraus mit Hilfe der Kamera- und Projektionsparameter die 3D-Position der Person zu finden. Zur Bestimmung der Körperhaltung in einem Bild wird dann diejenige Instanz des Körpermodells gesucht, deren Projektion in die Bildebene dem extrahierten Grauwertkantenbild am ähnlichsten ist. Dazu wird ein auf dem Winkel zwischen Grauwertkante und Modellkante und auf den Längen dieser Kanten definiertes Ähnlichkeitsmaß in Abhängigkeit von dem durch das Bewegungsmodell gegebenen Parameter minimiert, wobei die Gewichtung von längeren Modellkanten stärker ist und verdeckte Modellkanten nicht berücksichtigt werden. Die Kosten dieser Minimierung steigen exponentiell mit der Zahl der Parameter. Die rekursive Schätzung des Ortes, der Pose und deren Änderungen erfolgt nach der Initialisierungsphase mit einem Kalman-Filter. Nachteilig an diesem Verfahren ist, daß es nur zur Verfolgung von Bewegungen benutzt werden kann, dessen Bewegungsmuster gegeben sind.

Eine weitere Methode der modellbasierten Verfolgung ist die Betrachtung dynamischer Konturmodelle, wozu der folgende 3D-Ansatz von Kakadiaris und Metaxas sowie die Verfahren von Baumberg und Hogg und von Murase und Sakai zu zählen sind, die zu den 2D-Verfahren gehören und mit Eigenraumrepräsentation [33] arbeiten.

Das von Kakadiaris und Metaxas vorgestellte, modellbasierte Mehransichten-Verfahren zur Verfolgung der Bewegung von Körperteilen [19] arbeitet mit einem dreidimensionalen, deformierbaren Körpermodell, das vor dem eigentlichen Verfolgungsprozeß konstruiert wird. Zur Vermeidung von Fehlern, die durch Verdeckungen entstehen, benutzen sie die Aufnahmen von bis zu drei zueinander orthogonal ausgerichteten Kameras, wobei interaktiv anhand von zwei Kriterien entschieden wird, wieviele und welche Ansichten benutzt werden. Um eine Relation zwischen Punkten der verdeckenden Kontur zu Punkten auf dem Modell zu finden, verwenden sie Konzepte der projektiven Geometrie.

Vor der eigentlichen Verfolgung wird für jede der drei Ansichten zunächst per Konturanalyse ein 2D-Modell erarbeitet, woraus anschließend durch Integration das 3D-Modell gebildet wird ([20],[18]). Die zweidimensionale Körperdarstellung basiert auf einer parametrischen Komposition von deformierbaren, geometrischen Primitiva. Über ein sequentielles Abarbeiten der Bilder einer Bildfolge wird ein Graph erstellt, dessen Knoten die Körperteile und dessen Kanten die Gelenke darstellen. Damit die Struktur des Körpers an der Kontur erkennbar wird, muß die Person eine bestimmte Abfolge von Bewegungen durchführen, die zum einen die Segmentierung aller Körperteile garantiert und zum anderen die Reihenfolge der Segmentierung festlegt. Ausgehend von einem einzigen Teil für die ganze Person als Wurzel des Graphen werden nach der *HBPIS* (Human Body Part Identification Strategy) durch Kompositions- und Dekompositionsprozesse neue Primitiva einer Liste hinzugefügt und der Graph erweitert. Eine Konkavität der Kontur wird durch eine Differenz und eine Vorwölbung durch eine Vereinigung von Primitiva modelliert.

Die Verfolgung der Bewegung beruht auf einem dynamischen Modell, das aus den vereinfachten Gleichungen der Lagrange'schen Mechanik für jedes Körperteil besteht. Die Zustände dieses Modells sind durch den Ursprung und die Freiheitsgrade der Rotation eines Körperteils gegeben, deren Änderung durch externe, von den Messungen abhängige Kräfte und allgemeine Restriktionen bestimmt wird, und aus denen die Geschwindigkeit eines Punktes mit Hilfe der Jacobi-Matrix berechnet werden kann. Aus der Geschwindigkeit läßt sich dann die neue Position berechnen. Um dabei falschen Berechnungen durch Zuordnungsfehler zwischen Kontur- und Modellpunkten vorzubeugen, entscheidet ein interaktiver Prozeß, welche Kombination der drei Ansichten für die Berechnung der Kräfte ausreicht. Zu welchem Körperteil die Messung an einem Bildpunkt den Kräftebeitrag liefert, wird über einen Vergleich der Normalen auf die Kontur und der auf das Modell des entsprechenden Körperteils entschieden, wozu auf der Grundlage der projektiven Geometrie Konturpunkte zu Modellpunkten in Beziehung gesetzt werden. Zur Stabilisierung der Schätzungen wird dieses dynamische Modell in ein erweitertes Kalman-Filter-Verfahren eingebettet. Da für dieses Verfahren Bildsequenzen der Bewegung aus drei verschiedenen Ansichten benötigt werden und die Sequenzen anfangs den für die Modellinitialisierung festgelegten Bewegungsablauf zeigen müssen, kann man das Verfahren nicht auf bereits existierende Sequenzen anwenden.

Baumberg und Hogg betrachten in ihrer Arbeit [4] die Gehbewegung eines Fußgängers. Das von ihnen benutzte, dynamische Konturmodell [5] gibt die Variationen der menschlichen Silhouette während einer Gehbewegung wieder. Es wird aus Trainingsdaten generiert, die einen gehenden Menschen aus verschiedenen Blickwinkeln zeigen, und besteht aus einer Sammlung von B-Splines, die die jeweilige Kontur der Person annähern. Da nur die Variationen der Kontur modelliert werden sollen, benutzen Baumberg und Hogg für die Modellvektoren die Eigenraumrepräsentation der B-Splines. Die Projektion einer Modellkontur in ein Eingabebild ist durch die Konturparameter, den Ursprung des Objekts und die Anordnungskoeffizienten, welche eine skalierte Rotation beschreiben, bestimmt. Durch einen Anpassungsprozeß der Modellkontur an die beobachtete Kontur durch einen Kontrastvergleich werden die Anordnungskoeffizienten bestimmt und somit eine Verfolgung realisiert. Ein Kalman-Filter wird benutzt, um den Bereich für die Merkmalssuche im Laufe der Sequenz einzuschränken, was über die Größe der Varianz der Position der Stützpunkte gesteuert wird, und um die Kontur in einem Bild besser anzupassen.

Das von Murase und Sakai vorgeschlagene ortszeitliche Korrelationsverfahren auf Konturdaten zur *Erkennung* von Bewegungen [25] kann abgewandelt auch zur Verfolgung genutzt werden. Um den hohen Rechenaufwand einer Korrelation zu reduzieren, benutzen sie eine Eigenraumrepräsentation der Objektsilhouette, wobei durch eine Hauptachsentransformation in den Raum der Eigenvektoren der Kovarianzmatrix der Silhouettenbilder die Zahl der zur Repräsentation eines Bildes notwendigen Basisvektoren reduziert wird. Der Eigenraum wird dabei über eine Hauptachsentransformation von den dynamischen Konturdaten, die bei einer Bewegung beobachtet werden, aufgespannt. Nach Hintergrundsubtraktion und Grauwertnormierung wird durch Subtraktion von zwei zeitlich aufeinander folgenden Eingabebildern die Silhouette des Objekts bestimmt und bzgl. Größe und Position normalisiert in den durch die Trainingsdaten konstruierten Eigenraum projiziert. Da die Repräsentation der Objektsilhouette im Eigenraum in einer erheblich kleineren Dimension möglich ist, wird so der Rechenaufwand der Korrelation reduziert. Vorausgesetzt wird dazu ein konstanter Hintergrund und frontoparallele Bewegung des Objekts. Man braucht auch hier eine Trainingssequenz zur Erstellung des Eigenraums. Die Probleme der Verdeckung und der schnellen Bewegungen werden mit diesem Ansatz nicht gelöst.

Bei beiden Verfahren arbeitet die Verfolgung nur für Posen und Ansichten einer Bewegung zuverlässig, die durch die Trainingsmenge ausreichend repräsentiert werden, so daß auch hier die Anwendung auf beliebige Sequenzen problematisch ist.

Die Eigenraumtechnik zur Realisierung der Bewegungsverfolgung ist auch direkt auf die Grauwertbilder einer Sequenz anwendbar wie das folgende Verfahren von Black und Jepson zeigt.

Statt alle möglichen Orientierungen des Objekts zu repräsentieren, beschränken sich Black und Jepson [7] auf eine Auswahl von Ansichten des Objekts und erlauben zusätzlich eine parametrische Transformation zwischen dem Eingabebild und dem Eigenraum, z.B. eine affine Bewegung, welche sie auch mit *Warpen* bezeichnen. Um für diese Zuordnung die Eigenraumtechnik benutzen zu können, formulieren Black und Jepson eine der BCCE ähnliche Grauwerterhaltung, die sogenannte Unterraumerhaltung, welche besagt, daß für eine Ansicht des Objekts deren Grauwerte bei der Eigenraumrepräsentation erhalten bleiben. Zur Bestimmung der aktuellen Orientierung muß man nun gleichzeitig die Parameter der Bewegung und den zur Projektion des Bildes in den Eigenraum nächsten Punkt finden. Dazu wird die Unterraumerhaltung in eine der Kontinuitätsgleichung des optischen Flusses entsprechende Bedingung umgeformt, welche bzgl. der Bewegungsparameter und Projektionskoeffizienten schrittweise minimiert wird. Bei einer gegebenen groben Anfangsschätzung kann so mit dem Gradientenabstiegsverfahren die Orientierungsschätzung optimiert werden. Schnelle Bewegungen werden bei diesem Ansatz über eine Pyramidenverarbeitung berücksichtigt. Durch den Eigenraumansatz ist bei einer alle Ansichten abdeckenden Trainingsmenge die Verfolgung der Bewegung vom Blickpunkt der Kamera unabhängig. Black und Jepson haben diese Technik bisher an artikulierten Objekten nur zur Erkennung einer bewegten Hand getestet. Eine Erweiterung der Anwendung des Verfahrens zur Verfolgung der gesamten Bewegung einer Person ist mit einer Vervielfachung der Freiheitsgrade der Bewegung und somit der nötigen Trainingsmenge gekoppelt, die die Bestimmung der optimalen Parameter verlangsamt.

Bei den 2D-Verfahren zur Verfolgung artikulierter Bewegungen gibt es ferner modellbasierte Ansätze, die ebenfalls auf einem Abgleich eines zweidimensionalen Modells mit dem aktuellen Grauwertbild beruhen wie das im folgenden beschriebene Verfahren von Leung und Yang.

Bei dem System 'First Sight', das Leung und Yang entwickelt haben [23], handelt es sich um ein Verfahren, das in zwei Schritten die verschiedenen Körperteile (Kopf, Rumpf, Arme und Beine) einer sich bewegenden Person kennzeichnet. Im ersten Schritt wird aus der Grauwertsequenz für jedes Bild die Silhouette der Person bestimmt und im zweiten aus der Silhouette eine Bandfigur erarbeitet. Die Extraktion der Silhouette erfolgt über sogenannte Koinzidenzkanten ([22], [23]), welche aus der ortszeitlichen Analyse von Grauwertkantenbildern gewonnen werden. Durch einen ortszeitlichen Prozeß wird das Problem gelöst, welche Seite der Kante zum Körper gehört, so daß anschließend Bänder angepaßt werden können und man neben der Silhouette eine Sammlung verbundener Bänder untersucht. Der Kennzeichnungsprozeß basiert auf einem zweiteiligen Modell, das aus dem *Basismodell* und dem *Stützmodell* besteht.

Das Basismodell ist ein Baum bestehend aus fünf U-förmigen Bändern und einem Rumpfband, die jeweils an Gelenken zusammenhängen. Jedes U-förmige Band verfügt über eine Mittellinie (Wirbelsäule); zusätzlich gibt es eine Schulterlinie. Außerdem ist das Modell durch sieben Hauptpunkte festgelegt, die die Verbindungen zwischen Kopf, Armen, Beinen und Rumpf und die Endpunkte der Wirbelsäule kennzeichnen. Dieser äußeren Struktur werden Auflagen gemacht: Die maximale räumliche Abweichung zwischen bestimmten Hauptpunkten ist eingeschränkt, nur gewisse Bänder dürfen mit anderen Bändern kombiniert werden, und das Verhältnis der Ausdehnung von Kopf zu Armen zu Beinen zu Rumpf ist festgelegt. Das Stützmodell enthält zudem Informationen über die das Körpergewicht unterstützenden Glieder (z.B. die Füße bei einer Gehbewegung). Zu den zu verschiedenen Bewegungsarten gehörenden Umrißsequenzen werden dazu die Glieder, die den Körper stützen, angegeben.

Die Bänder werden nun in drei Gruppen eingeteilt: solche, die in ihrer Orientierung vom Zentrum weg strahlen, die bewegungslosen Teile und der Rumpf. Mittels einer genauen Analyse der Endungen der Bänder, ihrer Ausmaße und relativen Positionen zueinander wird die gemäß den obigen Auflagen beste Kombination der Körperteile gewählt, wobei das Stützmodell durch die Bestimmung der stützenden Körperteile den Suchraum einschränkt.

Bedingt durch das notwendige Stützmodell ist die Anwendung des Verfahrens entweder auf bestimmte Bewegungen eingeschränkt oder der Suchraum bei der Verwendung eines allgemeinen Stützmodells sehr groß.

Die bisher genannten Verfahren nutzen die Bildbewegung nicht explizit aus, sondern erreichen die Verfolgung über Zuordnungsprozesse. Im Gegensatz dazu benutzen die folgenden Verfahren von Pentland und Horowitz und von Ju, Black und Yacoob zur Verfolgung den optischen Fluß.

In ihrem Algorithmus [28] passen Pentland und Horowitz ein dreidimensionales, physiologisch-basiertes, gegliedertes Modell den zweidimensionalen optischen Flußdaten an, um so die 3D-Bewegung zu bestimmen. Die Freiheitsgrade der Bewegung werden dabei in starre und nicht starre Komponenten zerlegt. Das 3D-Modell beruht auf der Methode finiter Elemente, die in der Technik für die Simulation der Dynamik eines Objekts benutzt wird und die Eigenschaften von Variationen der Position, der Form und der Orientierung beschreibt. Das Problem der Anpassung eines 3D-Modells an 2D-Daten ist i.a. unterbestimmt. Gelöst wird es durch die modale Repräsentation, die man durch eine Diagonalisierung des Gleichungssystems erhält und wobei wegen der intrinsischen, elastischen Eigenschaften des Objekts nur die modalen Komponenten mit niedrigen Frequenzen berücksichtigt werden. Man kann daher die Zahl der Modi so wählen, daß das Anpassungsproblem überbestimmt ist. Position und Geschwindigkeit als dynamische Eigenschaften der modalen Parameter der finiten Elemente werden daraus mit Hilfe eines Kalman-Filters berechnet.

Pentland und Horowitz setzen für ihren Ansatz orthographische Projektion voraus, die man aber nicht bei jeder beobachteten Bewegung erreichen kann (z.B. nicht bei Translation entlang der optischen Achse). Daraus entstehen unvermeidliche Fehler bei der Schätzung der 3D-Geschwindigkeit. Ferner treten Zuordnungsprobleme wegen der Verwendung des optischen Flusses beim Vorliegen schneller Bewegungen auf.

Am engsten mit der vorliegenden Arbeit verwandt ist das Verfahren von Ju et al. [17], welches die artikulierte Bewegung durch die Anpassung eines parametrischen Modells des optischen Flusses an Teilgebiete des Objekts darstellt. Dazu benutzen sie für die Begrenzung der Gebiete, in denen ein Flußmodell angepaßt werden soll, ein zweidimensionales Modell des Menschen, welches die Glieder durch über je ein Gelenk verbundene, viereckige Flächen modelliert. Die Aktivität kann daher durch die Bewegungsparameter jeder Fläche beschrieben werden. Zur Bestimmung der Parameter benutzen Ju et al. die BCCE für jede Fläche und als zusätzliche Bedingung, daß die berechnete Bewegung in den verbundenen Flächen an den Gelenkpunkten gleich ist. Insgesamt erhalten Ju et al. daraus die sogenannte Federkraft-Energie zwischen zwei Punkten, die zur Bestimmung der Bewegungsparameter mit dem Gradientenabstiegsverfahren minimiert wird. Die Verfolgung wird erreicht, indem die berechnete Bewegung zur Prädiktion des Ortes benutzt wird. Zur Behandlung schneller Bewegung wird eine hierarchische Verarbeitung erwähnt, aber nicht näher ausgeführt. Die Selbstverdeckung stellt für dieses Verfahren ein noch nicht gelöstes Problem dar.

Zu erwähnen ist hier ferner das von Wren et al. entwickelte System zur Verfolgung der Bewegungen von Menschen, *Pfinder* [36], welches auf im voraus definierte Modelle verzichtet. Im Gegensatz zu den anderen Verfahren arbeitet es auf Farbbildern. Erst während des Verarbeitungsprozesses werden zweidimensionale Modelle für die beobachtete Person und die Szene gebildet, die auf stochastischen, örtlich gebundenen Merkmalen basieren, nämlich je einer Gaußverteilung für Orts- und Farbinformation. Das Körpermodell besteht aus verbundenen Flächen, *"blobs"*, deren Ausdehnung jeweils durch eine Kombination der beiden Gaußverteilungen bestimmt wird. Jeder "blob" ist durch die Mittelwerte von Orts- und Farbwert und die entsprechende Kovarianzmatrix definiert und verfügt zusätzlich über eine Liste, die für jeden Bildpunkt die Zugehörigkeit zu dem "blob" angibt. Außerdem wird für jeden "blob" ein dynamisches Modell generiert, das bei der Verfolgung zur Prädiktion der örtlichen Komponenten des Mittelwerts dient. Die Szene wird als texturierte Oberfläche betrachtet und an jedem Szenepunkt durch den mittleren Farbwert und die zugehörige Verteilung modelliert. Zur Initialisierung der Modelle ist es daher notwendig, daß die ersten Bilder der Bildfolge nur die Szene zeigen, damit zunächst das Szenenmodell aufgebaut werden kann. Die Anwendung des Verfahrens auf bereits existierende Bildfolgen ist deshalb i.a. nicht möglich. Durch die Detektion der starken Veränderungen der

Farbwerte beim Eintreten einer Person in die Szene erfolgt die Detektion der Person. Gesteuert von einer Konturanalyse, die bei bestimmten Körperhaltungen zur Generierung neuer "blobs" für bestimmte Körperteile dient, werden per Berechnung der Farbverteilungen die Parameter des menschlichen Modells ermittelt.

Aus den Farbinformationen der folgenden Bilder werden die statistischen Werte zur Aktualisierung der Modelle gewonnen. Zunächst werden die Mittelpunkte der "blobs" anhand der a priori Information prädiziert. Anschließend wird für jeden Bildpunkt mit der MAP-Strategie bestimmt, zu welchem der "blobs" er gehört, wobei a priori Wissen über räumliche Zusammenhangsbedingungen einfließt. Dementsprechend werden die Zugehörigkeitslisten korrigiert. Als letzter Verarbeitungsschritt wird das dynamische Modell aktualisiert.

4.2 Implementierung

Das Verfahren zur Detektion und Verfolgung von Gelenkbewegungen, welches hier vorgestellt werden soll, gehört zu den modellfreien Verfahren. Die Bestimmung der aktuellen Gelenkposition erfolgt über die direkte Ausnutzung der Bilddaten, indem der optische Fluß zur Verschiebungsbestimmung verwendet wird. Es wird also auf eine Modellanpassung an jedes der Bilder verzichtet.

Ferner gehört das Verfahren zu den passiven Verfahren, was bedeutet, daß es auf bereits existierendem Bildmaterial arbeitet. Daher wird keine spezielle Aufnahmetechnik gefordert und auch die Art der Bewegung ist nicht beschränkt. Das heißt ebenfalls, daß das Verfahren keine Kennzeichnung der Gelenke durch Marker o.ä. verlangt.

Durch die Applikation des optischen Flusses in der Bestimmung der Verschiebung der Gelenke von Bild zu Bild stellen schnelle Bewegungen wegen des zeitlichen Aliasingproblems eine Schwierigkeit dar. Das Problem der schnellen Bewegungen wird durch eine hierarchische Verarbeitung der Bildsequenzen bei der Berechnung des optischen Flusses gelöst.

Bei der visuellen Bewegungsanalyse unterscheidet man zwei Teilbereiche: die Detektion des Objekts, in unserem Falle der Person, und die Detektion und Verfolgung der Gelenke. In beiden Fällen werden wegen des Auftretens von schnellen Bewegungen hierarchische Algorithmen verwandt.

Dieses Kapitel beschränkt sich auf eine Darstellung der Implementierung der Gelenkverfolgung. Die Detektion der Person und die Zuordnung der Gelenke wird im folgenden Kapitel näher erläutert. Bei der Verfolgung der Gelenke sei vorausgesetzt, daß die initialen Gelenkpositionen bekannt sind. Der Hauptbestandteil des Verfahrens ist die hierarchische Berechnung des affinen optischen Flusses für das jeweilige Körperteil. Aus der Geschwindigkeit der einzelnen Glieder wird die Verschiebung der Gelenke durch Kombination der Daten derjenigen Glieder, die durch das Gelenk verbunden werden, bestimmt. Einen Überblick über die Struktur des Verfahrens ist in Abb. 4.1 gegeben.

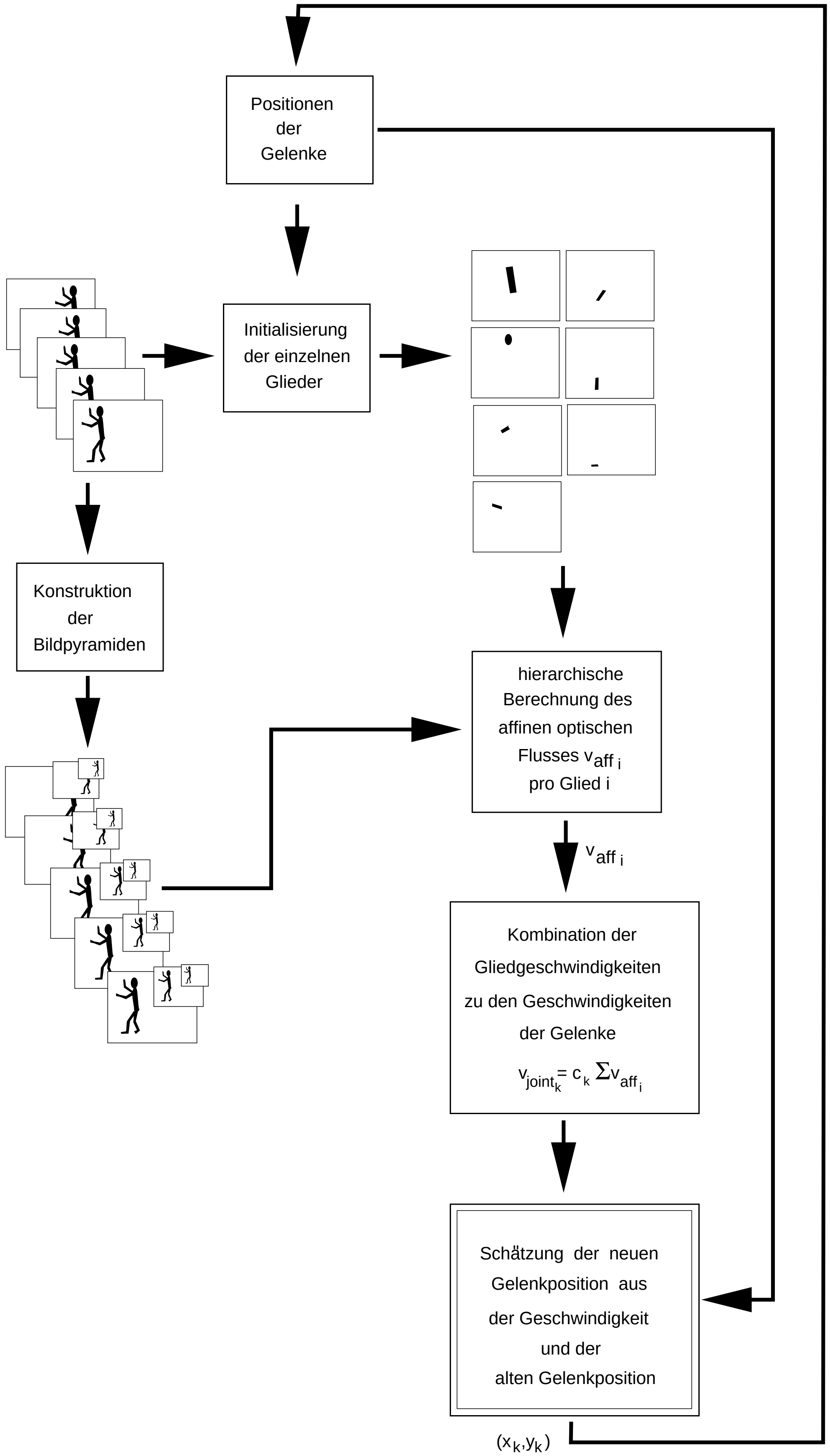


Abbildung 4.1: Flußdiagramm des Systems zur Verfolgung der Gelenke

Im groben besteht das Verfahren aus vier Schritten:

1. Generierung der Schablonen für die einzelnen Körperteile
2. Hierarchische Berechnung des affinen optischen Flusses für jedes durch die Schablone festgelegte Körperteil
3. Bestimmung der Gelenkgeschwindigkeit
4. Schätzung der neuen Gelenkposition

Generierung der Körperteilschablonen

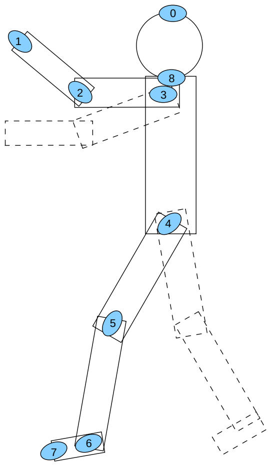


Abbildung 4.2: Zweidimensionales Modell des Menschen für die Initialisierung der Gliedregionen und Kennzeichnung der zu verfolgenden Gelenke

Als Eingabe für die Erzeugung der Schablonen der verschiedenen Körperteile benötigt man nur die aktuellen Gelenkpositionen. Wie aus Abb. 4.2 abzulesen ist, ist jedes Körperteil, welches für die Gelenkverfolgung von Interesse ist, durch zwei Gelenkpunkte festgelegt. Der Kopf ist durch Scheitel und Hals definiert, der Rumpf durch Hals und Hüfte und der Oberarm durch Schulter und Ellenbogen. Der Unterarm wird durch Ellenbogen und Handgelenk gegeben, der Oberschenkel durch Hüft- und Kniegelenk, der Unterschenkel durch Knie- und Fußgelenk und der Fuß durch Fußgelenk und Fußspitze.

Körperteil	Breite	Länge
Oberarm	0.7w	0.7/o
Unterarm	0.6w	0.6/o
Schulter	1.3w	1.3/o
Hüfte	1.6w	1.6/o
Oberschenkel	1.6w	1.6/o
Unterschenkel	1.0w	1.0/o
Fuß	0.8w	0.8/o

Tabelle 4.1: Ausdehnung der einzelnen Körperteile in Abhängigkeit von den Parametern w (width) und o (offset)

Die beiden ein Körperteil begrenzenden Gelenkpunkte werden durch einen Vektor verbunden und ihr Abstand zueinander berechnet. Als nächstes wird der Verbindungsvektor um ein vom Körperteil abhängigen Bruchteil (s. Tabelle 4.1) in beide Richtungen verlängert, wobei die Größenordnung vom Benutzer gesteuert werden kann. Orthogonal zu diesem Vektor werden in einem ebenfalls vom Körperteil abhängigen Abstand die Eckpunkte für das entsprechende Körperteil gesetzt (s. Abb. 4.3). Die einzige Ausnahme bildet die Generierung des Kopfes. Hier wird ein Kreis generiert, dessen Durchmesser dem Abstand zwischen Scheitel und Hals zuzüglich eines festen Offsets entspricht.

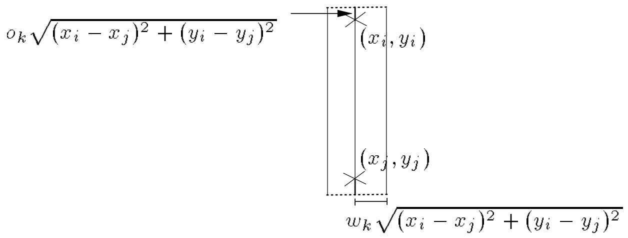


Abbildung 4.3: Schema zur Erzeugung einer Körperteilschablone aus den zwei angrenzenden Gelenken

Hierarchische Berechnung des affinen optischen Flusses

Sind die Schablonen erzeugt worden, folgt die Berechnung des affinen optischen Flusses für jedes einzelne Körperteil. Dazu wird aus den Eingabebildern zunächst durch Glätten und Unterabtasten jeweils eine Bildpyramide gebildet und anschließend der optische Fluß nach einem der in Kapitel 3.2 erklärten hierarchischen Verfahren berechnet. Da die einzelnen Körperteile als starre Teile angesehen werden können, ist das affine Modell für die Berechnung der Geschwindigkeit eines Körperteils geeignet.

Die für die Berechnung notwendigen Grauwertableitungen werden mittels Faltung mit einer Binomialmaske bestimmt. Als lokaler Ursprung jedes Körperteils dient für das affine Modell der jeweilige Mittelpunkt. Das in die Berechnung integrierte Warpen geschieht jeweils nur auf dem durch die Schablone begrenzten Bereich des Bildes. Als Ergebnis erhält man die Modellparameter $u_{0_i}, u_{x_i}, u_{y_i}, v_{0_i}, v_{x_i}, v_{y_i}$ des affinen optischen Flusses für jedes Körperteil i.

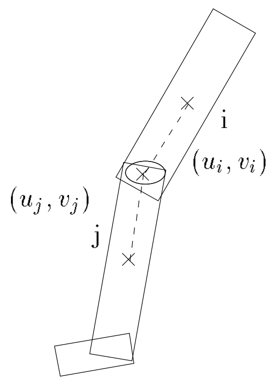


Abbildung 4.4: Berechnung von zwei Geschwindigkeiten des Kniegelenks aus den affinen Modellen für Ober- und Unterschenkel

Bestimmung der Gelenkgeschwindigkeit

Die Geschwindigkeit an einem bestimmten Punkt des Körperteils läßt sich aus den Parametern gemäß (2.16) über die relative Lage des Punktes berechnen. Dies geschieht so bei der Berechnung der Geschwindigkeit der Gelenke. Als Eingabe wird daher die aktuelle

Position des Gelenks im Bild benötigt. Es wird daraus pro Körperteil die Geschwindigkeit der begrenzenden Gelenke berechnet, wie in Abb 4.4 am Beispiel des Kniegelenks zu sehen ist.

Durch Kombination der verschiedenen Geschwindigkeiten eines Gelenks für die einzelnen Körperteile wird nun die mittlere Geschwindigkeit des Gelenks berechnet. Dabei ist zu beachten, daß nicht jeder der betrachteten Punkte ein Gelenk zwischen den benutzten Körperteilen ist und somit dessen Geschwindigkeit nur von einem Körperteil bestimmt wird. Die jeweils zur Gelenkgeschwindigkeit beitragenden Körperteile sind Tabelle 4.2 zu entnehmen.

Gelenk	Körperteile
Scheitel	Kopf
Handgelenk	Unterarm
Ellenbogen	Unterarm + Oberarm
Schulter	Oberarm
Hüfte	Rumpf + Oberschenkel
Knie	Oberschenkel + Unterschenkel
Fußgelenk	Unterschenkel + Fuß
Fußspitze	Fuß
Hals	Kopf + Rumpf

Tabelle 4.2: Schema über die zur Berechnung einer Gelenkgeschwindigkeit betrachteten Körperteile

Wie man aus den so gewonnenen Geschwindigkeiten der Gelenke die neue Position im Bild schätzt, wird in Kapitel 4.4 erklärt.

4.3 Detektion und Zuordnung der Gelenke

Das Ziel ist, aus rein visuellen Daten auf die Position verschiedener Gelenke eines Menschen zu schließen. Dazu braucht man zum einen eine Methode, um das von der Person eingenommene Gebiet aus einer Folge von Grauwertbildern zu extrahieren, und zum anderen eine Technik, um aus den extrahierten Daten die Gelenkposition zu finden.

Die Lokalisierung der Person im Bild erfolgt durch Änderungsdetektion. Dabei wird vorausgesetzt, daß die größte Bewegung im Bild von der Person ausgeführt wird und außerdem nur eine Person betrachtet wird. Weitere restriktive Bedingungen an die Szene werden nicht gestellt.

Die konventionelle Methode der Änderungsdetektion besteht in der Bestimmung der Region, in der die zeitliche Ableitung der Grauwertfunktion als Maß für die zeitliche Änderung einen Schwellwert übersteigt. Im Falle von langsamen Bewegungen ist, wie das Beispiel in Abb. 4.5 (links) zeigt, diese Art der Detektion zuverlässig. Falls die betrachteten Bewegungen jedoch sehr schnell sind, kommt es dabei wegen des zeitlichen Aliasings zu einer falschen Begrenzung der Region, wie in Abb. 4.5 (rechts) zu sehen ist. Statt der aktuellen Armregion werden die Bereiche, die die Arme in den zur Bildung der zeitlichen Ableitung notwendigen Bildern einnehmen, detektiert. Um diesen Effekt zu vermeiden, ist auch schon im Detektionsschritt ein hierarchischer Algorithmus notwendig.

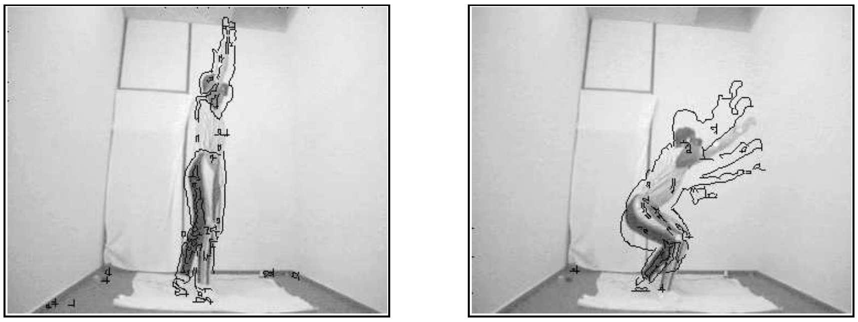


Abbildung 4.5: Ergebnis der Detektion der Person über Änderungsdetektion mittels der zeitlichen Ableitung bei langsamer Bewegung (links) und bei schneller Bewegung (rechts) am Beispiel der Sequenz aus Abb. 4.8

Die in Kapitel 3.2 erläuterten hierarchischen Verfahren zur Berechnung des optischen Flusses erweisen sich dafür als geeignet. Zunächst wird hierarchisch der optische Fluß bestimmt und anschließend ein Schwellwertverfahren auf das Flußfeld angewandt.

Der Ablauf der hierarchischen Detektion ist in Abb. 4.6 dargestellt. Als Eingabe benötigt das Verfahren der Größe des Ableitungsfilters entsprechend viele Bilder der Sequenz. Für jedes Bild wird eine Pyramide gebildet, woraus mit einem der hierarchischen Algorithmen, die in Kapitel 3.2 vorgestellt worden sind, der optische Fluß des mittleren Bildes bestimmt wird. Durch ein anschließendes Schwellwertverfahren werden die Regionen bestimmt, in denen der optische Fluß einen Grenzwert übersteigt. Diese Region ist gerade der Bereich des Bildes, in der sich die Person aufhält. Durch eine Bestimmung der Kontur dieses

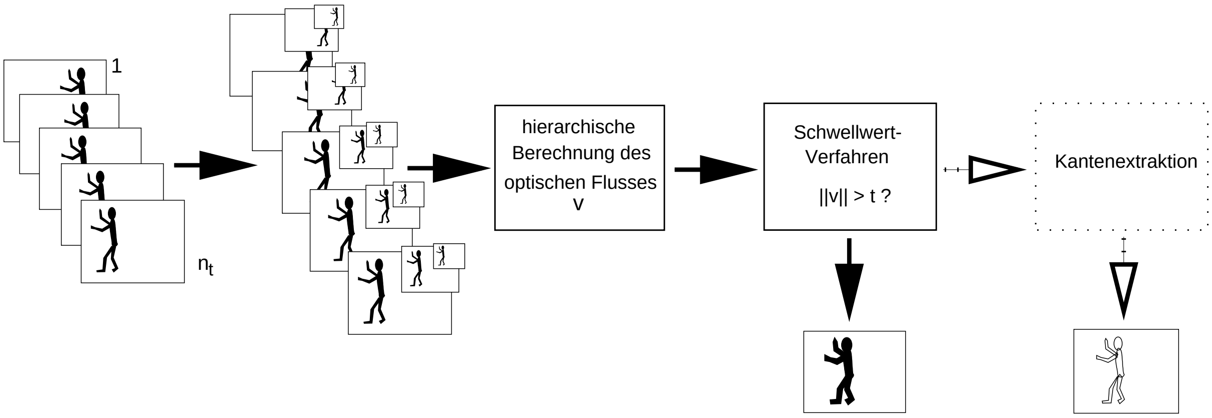


Abbildung 4.6: Verfahrensablauf der auf die hierarchische Berechnung des optischen Flusses gestützte Detektion der Person

Gebiets über eine Kantenextraktion kann aus diesem Binärbild die grobe Silhouette der Person gewonnen werden.

Wenn die Bildregion bestimmt worden ist, in der sich die Person befindet, dann muß noch die Position der Gelenke spezifiziert werden. Soll dies automatisch erfolgen, so braucht man ein Modell des Menschen. Zur Definition eines Modells des menschlichen Körpers benutzt man das Wissen über die Anatomie, Umrißeigenschaften und seine Beweglichkeit. Je nach Dimension des Raumes, in dem das Modell definiert wird, kann man Strichfigurmodelle oder Volumenmodelle benutzen. Die dazu bekannten Arbeiten sind bereits in Kapitel 4.1 kurz vorgestellt worden.

Die Gelenke haben keine besonderen Struktureigenschaften. Sie sind im allgemeinen lediglich über ihre Funktion als Verbindung zwischen starren Körperteilen oder aber über ihre relative Lage definiert. Um ihre Position über eine Modellanpassung bestimmen zu können wird nach dem heutigen Wissensstand die Kenntnis von a priori Informationen über die absolute Größe der Person, das initiale Bewegungsmuster oder ähnliches vorausgesetzt.

Soll das Verfahren auf bereits vorliegendem Bildmaterial arbeiten können, ist dieses Wissen nicht gegeben. Bei dem hier vorgestellten System wird daher die Lokalisation der Gelenke manuell durchgeführt. Da eine Markierung der Gelenke per Hand nur in den ersten Bildern notwendig ist, wobei die Menge von der Größe der zur Bestimmung des affinen Flusses benutzten Ableitungsfilter abhängt, stellt dieser Eingriff des Benutzers einen vertretbaren Aufwand dar.

Bei der Markierung der Gelenke ist eine bestimmte Reihenfolge einzuhalten, die Abb. 4.2 zu entnehmen ist. Die Positionen der Gelenke sollten nach folgendem Schema abgelegt werden:

1. Scheitel
2. Handgelenk
3. Ellenbogen
4. Schulter
5. Hüfte
6. Knie
7. Fußgelenk
8. Fußspitze
9. Hals

Es wird bei dem Verfahren auf eine Verfolgung beider Gelenke bei Gelenkpaaren (Handgelenke, Ellenbogen, Schultern, Knie, Fußgelenke, Zehen) verzichtet. Bei der Markierung ist daher zu beachten, daß die Gelenke derselben Körperhälfte markiert werden.

4.4 Schätzung der Position

Aus den vorausgegangenen Berechnungsschritten hat man eine Messung für die Geschwindigkeit jedes Gelenks erhalten. Die weitere Verarbeitung ist in Abb. 4.7 dargestellt. Per

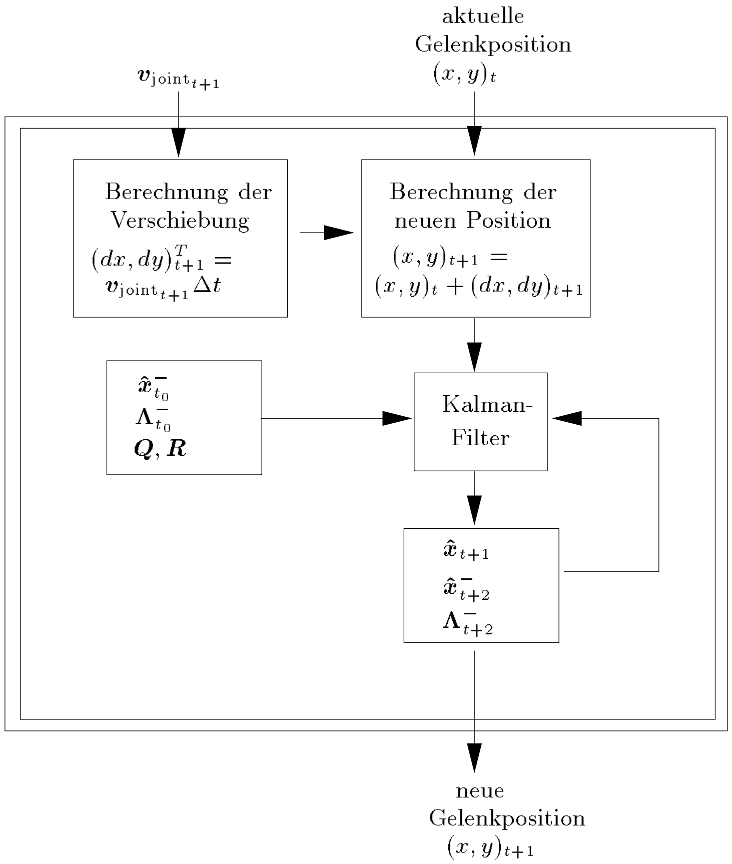


Abbildung 4.7: Ablauf der Schätzung der Position des Gelenks

Integration über die Zeit läßt sich aus der Geschwindigkeit die Verschiebung eines Gelenks berechnen. Unter der Voraussetzung, daß die betrachteten Zeitintervalle so klein sind, daß die Geschwindigkeit während dieser Zeit konstant ist, kann man die Integration durch eine einfache Summation vereinfachen:

$$\begin{pmatrix} x \\ y \end{pmatrix}_{t+1} = \begin{pmatrix} x \\ y \end{pmatrix}_t + \begin{pmatrix} u \\ v \end{pmatrix} \Delta t.$$

Die gesuchte Verschiebung der Gelenke soll in Pixel/Bild angegeben werden. Folglich ist das Zeitintervall gleich eins zu wählen. Die Werte der gemessenen Geschwindigkeiten stellen somit die Werte der gesuchten Verschiebungen dar.

Die so berechneten Verschiebungen resultieren direkt aus Messungen der Grauwerte. Um diese Werte zu stabilisieren, folgt die Anwendung eines Kalman-Filters über die Zeit. Das zugrunde liegende Modell für die Verwendung des Kalman-Filters zur Schätzung der Bewegungszustände ist bereits in Kapitel 3.1.4 vorgestellt worden. Der benutzte Bewegungszustand umfaßt die Position, die Geschwindigkeit und die Beschleunigung, welche als konstant angenommen wird. Die Transitionsvorschrift (3.31) ist durch die Bewegungsgleichungen einer gleichförmig beschleunigten Bewegung gegeben:

$$\boldsymbol{x} = \begin{pmatrix} x \\ y \\ u \\ v \\ a_x \\ a_y \end{pmatrix}, \qquad \boldsymbol{A} = \begin{pmatrix} 1 & 0 & t & 0 & \frac{1}{2}t^2 & 0 \\ 0 & 1 & 0 & t & 0 & \frac{1}{2}t^2 \\ 0 & 0 & 1 & 0 & t & 0 \\ 0 & 0 & 0 & 1 & 0 & t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Als Messung wird die aktuelle Gelenkposition verwendet, d.h. die Messungsmatrix ist durch (3.33) gegeben. Die das Systemrauschen charakterisierende Varianzmatrix $\boldsymbol{Q}$ aus (3.32) ist durch nur zwei Werte für die Beschleunigung q_{a_x}, q_{a_y} bestimmt, während die das Meßrauschen charakterisierende Matrix $\boldsymbol{R}$ durch die Skalare r_x, r_y definiert ist. Diese Parameter und die Anfangsschätzungen des Zustands und der Varianz sind als Eingabeparameter des Verfahrens frei wählbar und ermöglichen ein Tuning der Schätzungen.

Die aus der Kalman-Filterung resultierenden Gelenkpositionen, die man durch Projektion des Bewegungszustands auf die ersten beiden Komponenten erhält, sind die gesuchten Schätzungen der Position der Gelenke im nächsten Bild.

4.5 Experimente

Um die Leistung des vorgestellten Verfahrens zu veranschaulichen, wird es auf verschiedene Testsequenzen angewendet, welche jeweils einen sich bewegenden Menschen in einer beliebigen Szene zeigen. In jeder der Bildfolgen kommen schnelle und langsame Bewegungen vor. Besonders die Bewegung der Arme ist in allen Bildfolgen sehr schnell. Damit die in den Testsequenzen gezeigten Bewegungsabläufe deutlich werden, sind Ausschnitte der Sequenzen in den folgenden Abbildungen gezeigt.

In der ersten Bildfolge (s. Abb. 4.8) ist ein Sportler zu sehen, der eine Turnübung, nämlich die große "Körperwelle" ausführt. Die Bewegung der Beine ist dabei sehr gering, wohingegen die Geschwindigkeit der Arme besonders schnell ist. Der Sportler ist von der Seite gezeigt und trägt Marker, die von dem Verfahren jedoch nicht benutzt werden. Während des Bewegungsablaufs kommt es zu Verdeckung von Teilen des Körpers durch die Arme.

Als weitere Testsequenz betrachten wir die "Hampelmann"-Bewegung (s. Abb. 4.9), bei der zunächst das linke Bein und beide Arme seitlich ausgestreckt und anschließend abgewinkelt werden, bevor die Person in die Ausgangsstellung zurückkehrt. In dieser Bildfolge kommt es nicht zu Selbstverdeckung. Die Besonderheit dieser Sequenz ist die Frontalansicht der Person, in der alle Körperglieder zu sehen sind.

Die Testsequenz "Beuge" (s. Abb. 4.10) zeigt die Seitenansicht einer Person, die sich beugt und dabei eine Hand zum Boden führt, während die andere ausgestreckt bleibt. Anschließend kehrt die Person in die Ausgangsstellung zurück. Bei dieser Bewegung stellt die Verfolgung des Armes eine Schwierigkeit dar, weil er anfangs den anderen Arm verdeckt

Testsequenz: "Körperwelle"

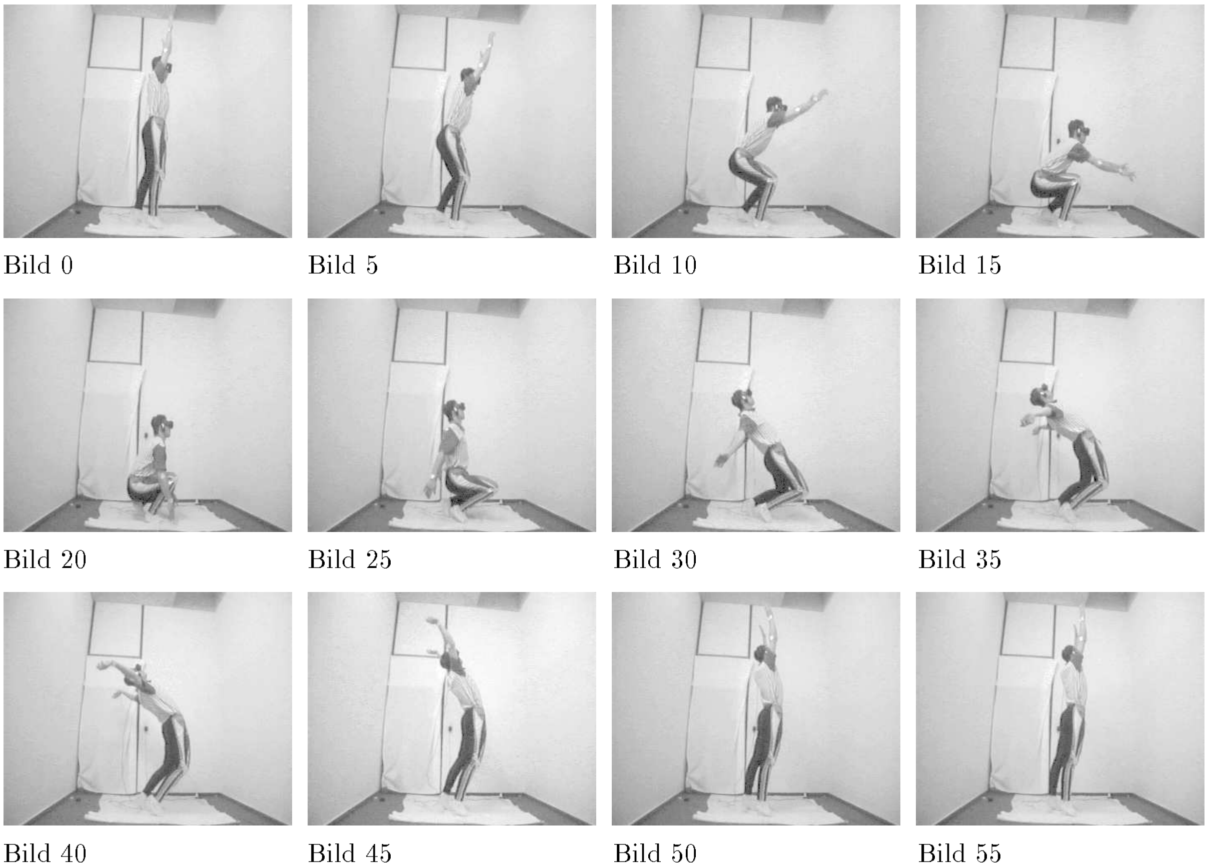


Abbildung 4.8: Ausschnitte aus der Testsequenz "Körperwelle" (jedes 5. Bild)

und daher eine Trennung der Arme erfolgen muß. Davon abgesehen kommt es nicht zu weiteren Verdeckungen. Die betrachteten Bewegungen sind von mittlerer Geschwindigkeit.

Zuletzt betrachten wir die Bildfolge "Kniebeuge" (Abb. 4.11). Sie zeigt die Seitenansicht einer Kniebeuge bei gleichzeitigem Schwingen der Arme. Durch die synchrone Bewegung der Arme überdecken diese sich größtenteils während der ganzen Bewegung. Außerdem verdecken sie zeitweise Beine und Rumpf, was eine Schwierigkeit darstellt. Die Bewegung der Arme ist während der ganzen Übung sehr schnell.

Für die Bewegungen der Testsequenzen gibt es keine Grundwahrheiten, so daß eine Darstellung der Leistung des Verfahrens durch eine Fehlerabschätzung nicht möglich ist. Statt dessen werden im Folgenden die Resultate des Verfahrens direkt dargestellt, wodurch man sich einen Eindruck der Leistung verschaffen kann.

Testsequenz: "Hampelmann"


Bild 0


Bild 10


Bild 20


Bild 30


Bild 40


Bild 50


Bild 60


Bild 70


Bild 80


Bild 90


Bild 100


Bild 110

Abbildung 4.9: Ausschnitte aus der Testsequenz "Hampelmann" (jedes 10. Bild)

Testsequenz: "Beuge"


Bild 0


Bild 10


Bild 20


Bild 30


Bild 40


Bild 50


Bild 60


Bild 70


Bild 80


Bild 90

Abbildung 4.10: Ausschnitte aus der Testsequenz "Beuge" (jedes 10. Bild)

Testsequenz: "Kniebeuge"

Bild 0

Bild 4

Bild 8

Bild 12

Bild 16

Bild 20

Bild 24

Bild 28

Abbildung 4.11: Ausschnitte aus der Testsequenz "Kniebeuge" (jedes 4. Bild)

4.5.1 Detektion der Person

Sollen die Gelenke automatisch lokalisiert werden, ist ein vorausgehendes Einschränken der Suchregion auf das von der Person eingenommene Gebiet unerläßlich. Der in dieser Arbeit vorgeschlagene Detektions-Algorithmus besteht aus der hierarchischen Berechnung des optischen Flusses und einer anschließenden Schwellwertbehandlung.

Wie gut die mit diesem Algorithmus erzielten Resultate im Vergleich zu der üblichen auf der zeitlichen Ableitung beruhenden Änderungsdetektion sind, wird im folgenden veranschaulicht. Zur hierarchischen Berechnung des lokal konstanten optischen Flusses wird dabei das Verfahren von Simoncelli benutzt. Anschließend wird der Betrag des optischen Flusses berechnet, wobei nur ein über das Gütemaß bestimmter Teil des optischen Flußfeldes betrachtet wird. Ein Schwellwertverfahren liefert dann die gesuchte Region.

Die Größe der im Verfahren von Simoncelli benutzten Ableitungsfilter beträgt bei den folgenden Experimenten jedesmal 5×5. Das Fenster, in dem der optische Fluß als konstant angenommen wird, hat die Abmessung 7×7, und die Rauschparameter haben die Werte $\lambda_v = 0.0, \lambda_t = 1.0$. Bei der konventionellen Änderungsdetektion wird ein Ableitungsfilter der Größe 5×5 benutzt.

Da für die Detektion genügend Bewegung im Bild vorhanden sein muß, ist es bei den gegebenen Testsequenzen sinnvoll, nicht das erste berechenbare Bild, welches bei der angegebenen Größe der Ableitungsfilter das dritte ist, sondern ein späteres zu benutzen. Deshalb wird bei der "Körperwelle"-Sequenz und der "Hampelmann"-Sequenz das fünfte und bei der "Beuge"-Sequenz und der "Kniebeuge"-Sequenz jeweils das zehnte Bild betrachtet.

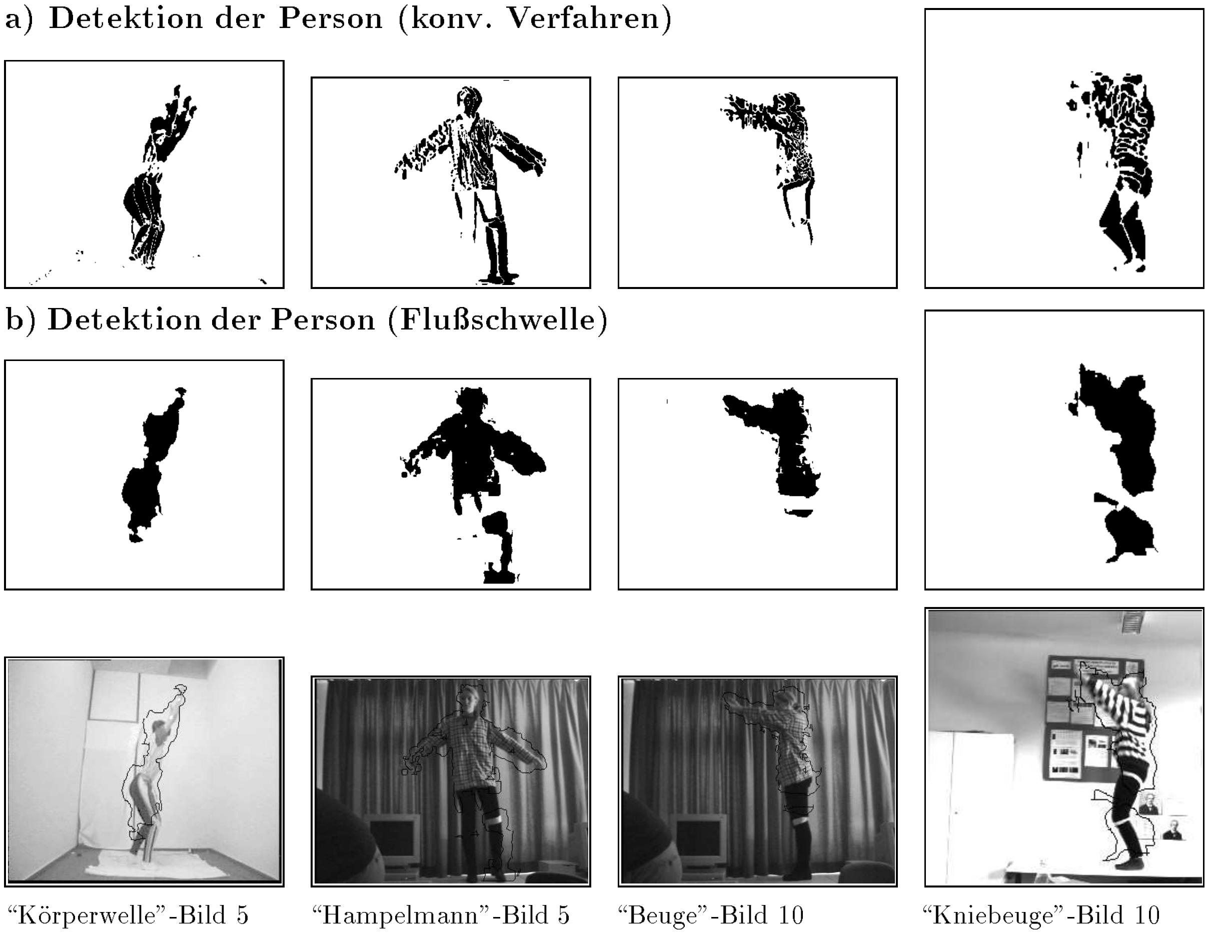


Abbildung 4.12: Detektion der Person a) mit der konventionellen auf der zeitlichen Ableitung basierenden Änderungsdetektion, b) mit dem auf der hierarchischen Berechnung des optischen Flusses beruhenden Detektionsalgorithmus. Unten: Kennzeichnung der gefundenen Silhouette im Bild.

Zum Vergleich der Leistung der Verfahren sind in Abb. 4.12 a) die Resultate der konventionellen Änderungsdetektion und b) der auf dem optischen Fluß basierenden Detektion dargestellt. Unten in der Abbildung ist die Einordnung der gefundenen Silhouette in das entsprechende Bild zu sehen. Der Betrag der zeitlichen Ableitung wird bei der konventionellen Detektion im Falle der "Körperwelle"-Sequenz durch 2.0 Pixel, im Falle der "Hampelmann"-Sequenz und der "Beuge"-Sequenz durch 3.0 Pixel und im Falle der "Kniebeuge"-Sequenz durch 14 Pixel beschränkt. Die bei der so bezeichneten Flußschwellendetektion zwischen den einzelnen Sequenzen variierenden Parameter, nämlich Anzahl der Pyramidenstufen, Dichte des Flußfeldes und Schwellwert, sind der folgenden Tabelle zu entnehmen:

Sequenz	*Stufenzahl*	*Dichte/%*	*Schwellwert*
“Körperwelle”	2	20	2.0
“Hampelmann”	1	35	0.5
“Beuge”	2	30	0.5
“Kniebeuge”	3	25	2.0

.

Man sieht, daß gerade bei der “Körperwelle”-Sequenz und bei der “Kniebeuge”-Sequenz die Detektion über den optischen Fluß eine genauere Bestimmung des von der Person eingenommenen Bereichs liefert. Bei der “Hampelmann”- und bei der “Beuge”-Sequenz ist die Bewegung so klein, daß man auch mit der konventionellen Änderungsdetektion gute Ergebnisse erhält.

Es zeigt sich, daß die konventionelle Methode im Beinbereich bessere Ergebnisse liefert, was an der mangelnden Textur der Beine liegt, wodurch es bei der Flußberechnung zum Aperturproblem kommt. Insgesamt zeigt sich, daß speziell dann, wenn schnelle Bewegungen vorliegen, die Detektion über den optischen Fluß der konventionellen Änderungsdetektion vorzuziehen ist. Auch bei langsamen Bewegungen liefert sie eine vergleichbar gute Detektion der Person.

4.5.2 Verfolgung der Gelenke

Nach der Detektion der Person und der Detektion der betreffenden Gelenke erfolgt zur Verfolgung für jedes der jeweils durch zwei Gelenke definierten Körperteile eine hierarchische Berechnung des affinen optischen Flusses, aus dem dann die Verschiebung der Gelenke ermittelt wird. Das Korrigieren der aktuellen Gelenkposition mit der geschätzten Verschiebung bildet letztendlich die Verfolgung.

Um die Leistung des im Rahmen dieser Arbeit entwickelten Verfahrens zur Gelenkverfolgung zu testen, wird es auf verschiedene in den Abb. 4.8 bis 4.11 dargestellte Testsequenzen angewendet, in denen unterschiedliche Schwierigkeiten auftreten. Da es für die unter Realbedingungen aufgenommenen Bildfolgen keine Grundwahrheiten gibt, ist es nicht möglich, den Fehler der Verfolgung quantitativ anzugeben. Statt dessen werden die Ergebnisse der Verfolgung durch Markierung der geschätzten Gelenkpositionen in den Bildern der Sequenz veranschaulicht. Zusätzlich wird eine Strichfigur generiert, die die Bewegung der Person nachzeichnet.

In einer den Tests vorausgehenden “Off line”-Initialisierung werden die Schablonen für die Körperteile mit den Größen $w = o = 10$ generiert. Die Verfolgung der Gelenke wird daher im folgenden von einem Bild zum nächsten durchgeführt. Verwendet man die Schätzung als neuen Ausgangswert, so kann es durch Abweichungen der geschätzten Position von der zur Generierung der Schablonen benutzten Position zu Fehlern kommen. Im Extremfall liegt die geschätzte Position außerhalb der generierten Schablone, wodurch eine gute Schätzung der neuen Gelenkposition unmöglich wird. Aus diesem Grund werden in den Tests als aktuelle Gelenkpositionen die zur Generierung der Schablonen verwendeten Positionen benutzt.

Bei der Berechnung des optischen Flusses innerhalb des Verfolgungsalgorithmus' haben die Binomialfilter zur Messung des Grauwertgradienten die Größe 5×5. Die frei wählbare a priori Varianz λ_p wird bei der Berechnung in allen Tests ignoriert. Zusätzlich werden die Schätzungen durch ein Kalman-Filter stabilisiert. Dabei wird als a priori Schätzung der Nullvektor und als a priori Varianz die Einheitsmatrix verwendet. Ferner wird die Systemvarianz durch $q_{a_x} = q_{a_y} = 5.0$ und die Meßvarianz durch $r_x = r_y = 0.5$ modelliert.

In den folgenden Tests wird stets das auf ein affines Modell erweiterte hierarchische Verfahren von Simoncelli et al. zur Bestimmung der Bildgeschwindigkeit benutzt. Die Anwendung der affinen Variante des hierarchischen Eigenwertverfahrens ist prinzipiell auch möglich. Wegen der komplexeren Rauschparameterabhängigkeit ist es jedoch schwieriger, für eine Verfolgung über mehrere Bilder günstige Parameter zu bestimmen. Außerdem ist bei der hierarchischen Berechnung die Leistung des Verfahrens von Simoncelli et al. besser.

Zum Vergleich der Leistung des Mehrauflösungsverfahrens mit der eines nicht hierarchischen Ansatzes zur Gelenkverfolgung ist in Abb. 4.13 das Ergebnis der Gelenkverfolgung unter Verwendung der Erweiterung des Simoncelli-Algorithmus' zur auf der Testsequenz "Körperwelle" in Originalauflösung gezeigt. Es ist zu erkennen, daß die Prädiktion der Position der eher langsam bewegten Gelenke z.B. im Beinbereich auch ohne hierarchische Verarbeitung möglich ist. Die schnelle Bewegung im Arm wird jedoch stark unterschätzt.

Nicht hierarchische Gelenkverfolgung: "Körperwelle"

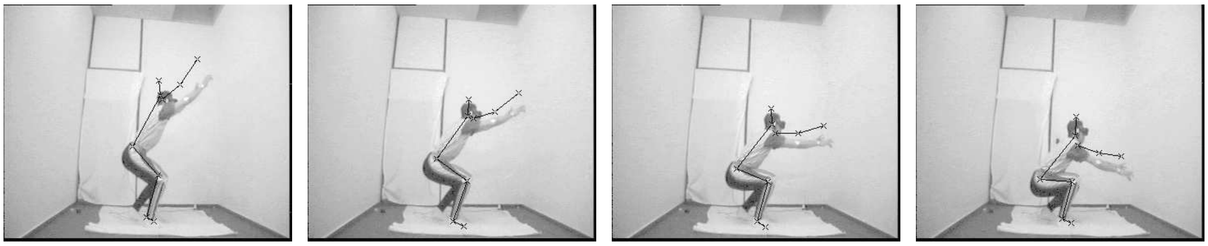


Bild 29 Bild 31 Bild 33 Bild 35

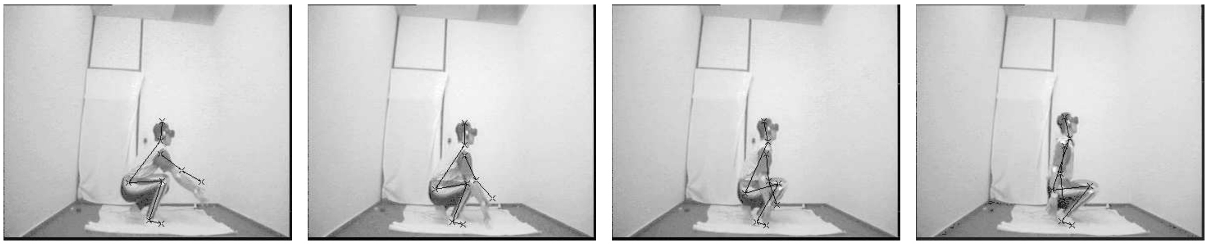


Bild 37 Bild 39 Bild 41 Bild 43

Abbildung 4.13: Beispiel für die Gelenkverfolgung ohne hierarchische Verarbeitung bei der Testsequenz "Körperwelle", wobei das auf ein affines Flußmodell erweiterte Simoncelli-Verfahren mit den Parametern $\lambda_{\boldsymbol{v}} = 0.0, \lambda_t = 1.0$ verwendet worden ist.

Benutzt man nun die hierarchische Variante des Verfahrens zur Gelenkverfolgung, so ist die Verfolgung der langsamen Bewegungen noch immer gewährleistet, wie in Abb. 4.14 zu sehen ist.

Gelenkverfolgung: “Körperwelle”

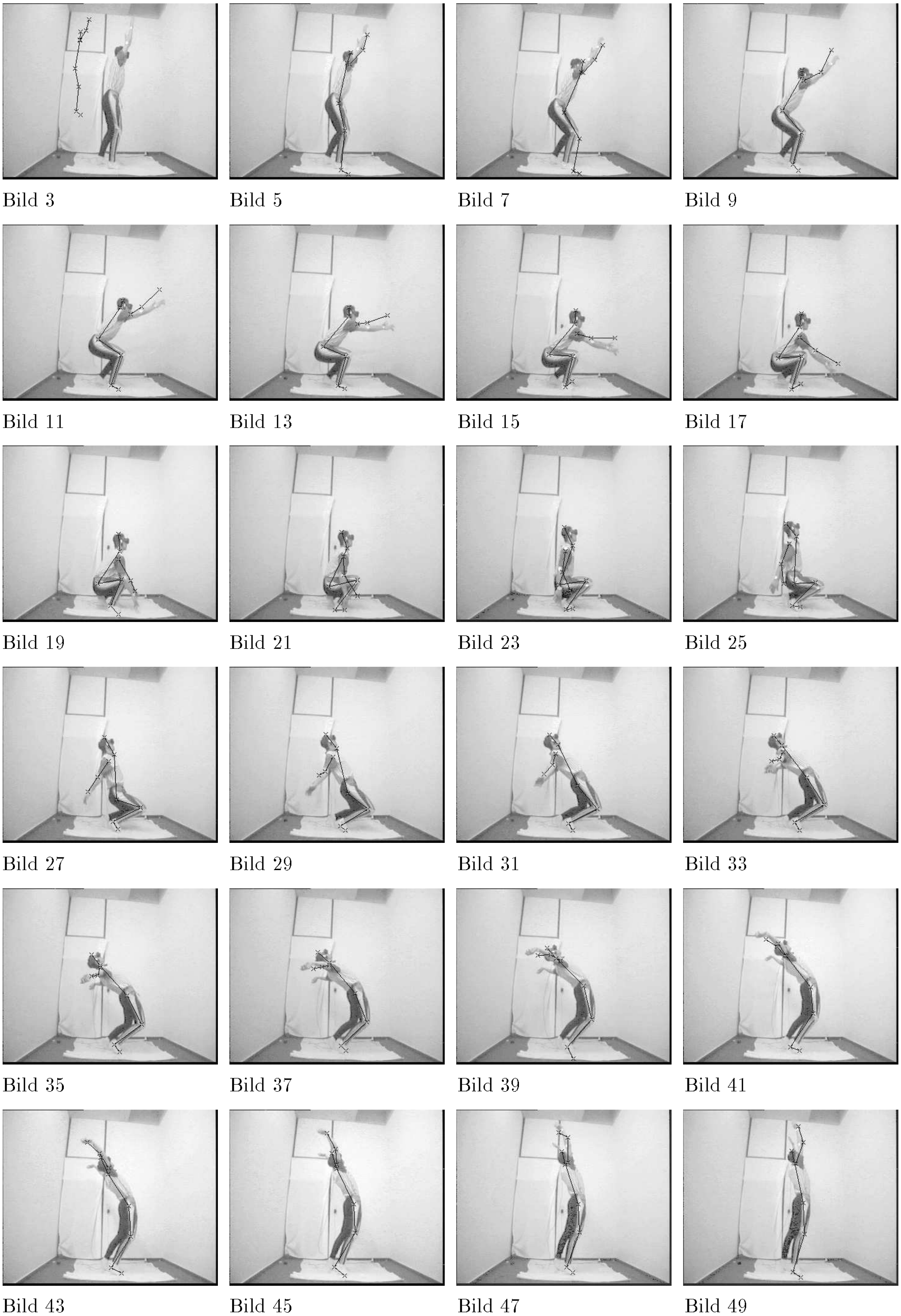


Abbildung 4.14: Veranschaulichung der Gelenkverfolgung bei der Testsequenz “Körperwelle” (jedes 2. Bild), wobei das auf ein affines Flußmodell erweiterte Simoncelli-Verfahren mit den Parametern $\lambda_v = 0.1, \lambda_t = 1.0$ auf einer zweistufigen Bildpyramide verwendet worden ist.

Eine deutlich bessere Leistung zeigt das Verfahren bei der Verfolgung der schnellen Armbewegung. Das Ellenbogengelenk wird zuverlässig lokalisiert und auch bei der Prädiktion der Position des Handgelenks sind bessere Resultate vorzuweisen. Das Handgelenk wird zwar bei der Verwendung einer zweistufigen Pyramide nicht immer exakt lokalisiert, aber die Prädiktion ist schon besser als beim nicht hierarchischen Ansatz. Man könnte theoretisch eine weitere Pyramidenstufe verwenden, was jedoch praktisch durch die geringe räumliche Ausdehnung des Unterarms i.a. keine besseren Ergebnisse liefert.

Die Geschwindigkeit des Handgelenks beträgt in der Testsequenz "Körperwelle" bis zu ca. 18 Pixeln pro Bild, was die Ursache für die verbleibenden Probleme bei der Verfolgung mit Hilfe eines zweistufigen Verfahrens ist. Die anderen Gelenke, deren maximale Geschwindigkeit unter 12 Pixeln pro Bild liegt, lassen sich mit dem zweistufigen Algorithmus gut verfolgen. Auch die teilweise Verdeckung des restlichen Körpers durch die Arme stellen bei der schrittweisen Verfolgung von Bild zu Bild keine Schwierigkeit dar.

Im den ersten Bild der Abb. 4.14 bis 4.17 ist deutlich die Initialisierung des Kalman-Filters zu erkennen. Erst nach der Verarbeitung der ersten Bilder sind die Schätzungen des Kalman-Filters zuverlässig. Charakteristisch für die Kalman-Filterung ist eine Verzögerung der Aktualisierung, d.h. die Bewegung wird bei großer Beschleunigung im allgemeinen unterschätzt und auf eine abrupte Änderung der Bewegungsrichtung reagiert das Filter erst ein paar Schritte später.

In den Abb. 4.15 und 4.16 sieht man wieder die Anwendung des zweistufigen Verfolgungsalgorithmus' auf die Testsequenzen "Hampelmann" und "Beuge". Hier werden alle Gelenke mit akzeptabler Abweichung verfolgt. In Bild 73 aus Abb. 4.15 sieht man die durch das Kalman-Filter verspätete Reaktionszeit. Die maximale Armbewegung in der Sequenz "Hampelmann" bzw. die maximale Beinbewegung in der Sequenz "Beuge" liegt bei ca. 12 Pixeln pro Bild. die Geschwindigkeit der anderen Gelenke ist geringer. Die Resultate in Abb. 4.15 und 4.16 zeigen, daß das Verfahren unabhängig von der Ansicht der Person ist.

Zuletzt ist in Abb. 4.17 die Anwendung des zweistufigen Verfahrens auf die Testsequenz "Kniebeuge" zu sehen. Hier ist die Armbewegung noch schneller als in der Sequenz "Körperwelle". Sie liegt bei über 20 Pixeln pro Bild. Mit einem zweistufigen Ansatz kann man auch diese Bewegung nicht mehr zuverlässig verfolgen. Die anderen Gelenke, deren Geschwindigkeit unter 10 Pixeln pro Bild liegt, lassen sich von Bild zu Bild in guter Näherung prädizieren.

Insgesamt lassen diese Beobachtungen den Schluß zu, daß der vorgestellte hierarchische Verfolgungsalgorithmus die Schwierigkeiten der schnellen Bewegungen bewältigt und auch kurzzeitige Verdeckungen die Verfolgung nicht beeinträchtigen. Außerdem ist das Verfahren von der Ansicht der Person unabhängig. Dabei ist zu bemerken, daß zum einen die Anzahl der Pyramidenstufen durch die Ausdehnung der einzelnen Körperglieder begrenzt wird und zum anderen die maximale Geschwindigkeit eines zu verfolgenden Gelenks nicht mehr als ungefähr das vierfache der Zahl der zu verarbeitenden Auflösungsebenen (Pyramidenstufenzahl+1) betragen sollte. Unter Beachtung dieser Bedingungen liefert das Verfahren eine zuverlässige Verfolgung der Gelenke.

Gelenkverfolgung: "Hampelmann"


Bild 13

Bild 17

Bild 21

Bild 25

Bild 29

Bild 33

Bild 37

Bild 41

Bild 45

Bild 49

Bild 53

Bild 57

Bild 61

Bild 65

Bild 69

Bild 73

Bild 77

Bild 81

Bild 85

Bild 89

Bild 93

Bild 97

Abbildung 4.15: Veranschaulichung der Gelenkverfolgung bei der Testsequenz "Hampelmann" (jedes 4. Bild), wobei das auf ein affines Flußmodell erweiterte Simoncelli-Verfahren mit den Parametern $\lambda_v = 0.1, \lambda_t = 1.0$ auf einer zweistufigen Bildpyramide verwendet worden ist.

Gelenkverfolgung: "Beuge"

Abbildung 4.16: Veranschaulichung der Gelenkverfolgung bei der Testsequenz "Beuge" (jedes 3. Bild), wobei das auf ein affines Flußmodell erweiterte Simoncelli-Verfahren mit den Parametern $\lambda_v = 0.1, \lambda_t = 1.0$ auf einer zweistufigen Bildpyramide verwendet worden ist.

Gelenkverfolgung: "Kniebeuge"

Abbildung 4.17: Veranschaulichung der Gelenkverfolgung bei der Testsequenz "Kniebeuge", wobei das auf ein affines Flußmodell erweiterte Simoncelli-Verfahren mit den Parametern $\lambda_{\boldsymbol{v}} = 0.1, \lambda_t = 1.0$ auf einer zweistufigen Bildpyramide verwendet worden ist.

Kapitel 5

Probleme und Ausblick

Das vorgestellte Mehrauflösungsverfahren zur Detektion und Verfolgung der Bewegung von Menschen stellt eine Möglichkeit zur Lösung des Problems schneller Bewegungen dar. Es ist jedoch zu bemerken, daß dadurch im allgemeinen noch kein vollständig robustes System garantiert ist.

Neben den schnellen Bewegungen stellt das Auftreten von Selbstverdeckungen eine Schwierigkeit bei der Verfolgung artikulierter Objekte dar. Die Einbettung des Mehrauflösungsverfahren in ein Kalman-Filter ermöglicht zwar eine Stabilisierung der geschätzten Gelenkpositionen, bei längeren Verdeckungsphasen reicht dies jedoch nicht mehr aus. Durch die Kenntnis der Position und der aktuellen Bewegung ist die Vorhersage einer Verdeckung innerhalb des beschriebenen Verfahrens möglich. Erweitert man das Verfahren in ein dreidimensionales Verfahren, indem man statt einer Kamera mehrere Kameras benutzt und die Verfolgung parallel ausführt, kann man bei der Vorhersage einer Verdeckung aus den Informationen einer anderen Ansicht die Bewegung des verdeckten Körperteils ermitteln. Außerdem ist es bei einer solchen Erweiterung möglich, dreidimensionale Bewegungen zu rekonstruieren.

Treten besonders schnelle Bewegungen auf, die wegen der Begrenzung der Zahl von zu verwendenden Pyramidenstufen durch die räumliche Ausdehnung der Körperteile nicht mehr mit einem hierarchischen Ansatz des Verfahrens zu verfolgen sind, ist der Einsatz von Hochfrequenzkameras notwendig.

Ein weiteres Problem liegt in der Generierung der Schablonen für die einzelnen Körperteile. Bei der hierarchischen Berechnung des optischen Flusses wird von der zweitobersten, also zweitgröbsten, Stufe an bis zur untersten die zeitliche Ableitung der gewarpten Sequenz benötigt. Durch die affine Modellierung des optischen Flusses jedes einzelnen Körperteils gelten für verschiedene Bildbereiche unterschiedliche Flußparameter. Das heißt, für das Warpen ist die Begrenzung der einzelnen Geltungsbereiche der Flußparameter durch Schablonen notwendig. Werden zur Berechnung der zeitlichen Ableitung wie bei der vorgestellten Implementierung FIR-Filter verwendet, so bedeutet dies, daß diese Schablonen schon für Bilder bekannt sein müssen, in denen die aktuelle Position der Gelenke erst noch bestimmt werden muß. Für die Experimente sind daher die Schablonen im voraus generiert worden. Dieses Problem kann durch eine Ersetzung der FIR-Filter durch zeitlich rekursive IIR-Filter [9] zur Berechnung der zeitlichen Ableitung gelöst werden. Die zeitlich rekursiven IIR-Filter sind asymmetrisch und benötigen nur das aktuelle und vorherige Bilder zur Berechnung der zeitlichen Ableitung. Von diesen ist die Gelenkposition bekannt, so daß die Schablonen innerhalb des Verfahrens generiert werden können. In diesem Fall ist es möglich, das Verfahren so zu implementieren, daß es fähig ist, in Echtzeit zu arbeiten. Die berechneten Bewegungsinformationen können dann von dem System für eine Rückkopplung benutzt werden.

Das Problem der abrupten Richtungsänderung kann durch die komplexere Formulierung eines dynamischen Modells für das Kalman-Filter angegangen werden. Berücksichtigt man auch die Verschiebung als Maß für die Geschwindigkeit, so ist die Annahme der Konstanz der Beschleunigung nicht mehr notwendig.

Wie man bei dem experimentellen Vergleich der hierarchischen Berechnungsverfahren des optischen Flusses gesehen hat, liefert beim hierarchischen Eigenwertverfahren die Annahme, daß der Warpvorgang fehlerbehaftet ist, bessere Ergebnisse als die fehlerfreie Variante. Es liegt nah, die Fusion der Daten verschiedener Hierarchiestufen auch bei dem Verfahren von Simoncelli unter der Annahme eines fehlerbehafteten Warpvorgangs zu testen.

Kapitel 6

Zusammenfassung

Das im Rahmen dieser Arbeit implementierte System dient der visuellen Bewegungsanalyse. Betrachtet werden dabei die Bewegungen eines Menschen, die sich durch ihre nur partiell starre Struktur als besonderes Gebiet der Bewegungsanalyse auszeichnen. Als Schwerpunkt der Bewegungsanalyse werden die durch schnelle Bewegungen auftretenden Schwierigkeiten des zeitlichen Aliasings behandelt.

Das entwickelte, passive Verfahren arbeitet auf bereits existierenden Bildfolgen und verzichtet im Gegensatz zu einer Vielzahl der konventionellen Verfahren auf Kennzeichnungen durch Marker. Des weiteren werden keine restriktiven Bedingungen an den Hintergrund gestellt. Es wird lediglich vorausgesetzt, daß nur eine Person betrachtet wird und die Lichtverhältnisse sich nur wenig ändern. Das Verfahren gehört außerdem zu den modellfreien Verfahren, die in der zweidimensionalen Bildebene arbeiten und direkt die Bildbewegung zur Analyse ausnutzen.

Da es sowohl im ersten Teilbereich der Bewegungsanalyse, nämlich der Detektion der Person, als auch im zweiten Teilbereich, der Verfolgung der Bewegung, zu Aliasingproblemen durch schnelle Bewegungen kommen kann, sind für beide Bereiche hierarchische Verfahren entwickelt worden.

Die Hauptkomponente beider Teilbereiche ist der optische Fluß. Bei der Detektion tritt das optische Flußfeld an die Stelle der bei der üblichen Änderungsdetektion verwendeten zeitlichen Ableitung des Grauwertbildes. Mittels eines Schwellwertverfahrens werden die Bereiche des Bildes detektiert, in denen der Betrag des optischen Flusses über einer vorgegebenen Schranke liegt. Bei der Verfolgung der Bewegung liefert der optische Fluß die Schätzung der Geschwindigkeit an einer Stelle des Körpers, aus der sich die Verschiebung von Bild zu Bild, und damit die neue Position berechnen läßt.

In beiden Fällen ist also die Berechnung des optischen Flusses die Basis der Algorithmen. Die dabei angenommenen Modelle des optischen Flusses unterscheiden sich jedoch. Im Detektionsschritt geht man von einem lokal konstanten optischen Flußfeld aus, während wegen der partiellen Starrheit der Bewegung eines artikulierten Objekts bei der Verfolgung ein affines Modell vorausgesetzt wird. Das erfordert zwei verschiedene hierarchische Berechnungsverfahren für den optischen Fluß, die in dieser Arbeit vorgestellt werden.

Zur Lösung des Problems der Fusion von Sensordaten, welches bei einer hierarchischen Verarbeitung in Form der Kombination von Ergebnissen aus verschiedenen Auflösungsstufen auftritt, sind für jedes Modell des optischen Flusses zwei alternative Berechnungsverfahren implementiert worden. Diese Verfahren liefern ein für die Fusion notwendiges Gütemaß der Berechnung. Außerdem benutzen sie die Technik des Warpens, wobei die Berechnung des optischen Flusses auf einer Stufe dazu verwendet wird, die Bildbewegung auf der nächst feineren Stufe durch ein entsprechendes Verschieben der Grauwerte zu reduzieren. So tragen die feineren Auflösungsstufen der Bildpyramide zu einer Korrektur der Berechnung auf der gröberen Stufe bei.

Zusätzlich wird in den Verfahren berücksichtigt, daß die zur Berechnung verwendeten Messungen verrauscht sind. Der Unterschied der beiden Verfahren liegt in den zugrunde gelegten Modellen, die das Rauschen darstellen. Das von Simoncelli et al. entwickelte Verfahren geht nur auf das Rauschen in der zeitlichen Komponente des Grauwertgradienten ein, während das Verfahren von Kanatani, das zu einem hierarchischen Verfahren erweitert worden ist, außerdem das Rauschen in den örtlichen Komponenten berücksichtigt. Basierend auf dem Verfahren von Simoncelli et al. und dem von Kanatani sind des weiteren hierarchische Algorithmen für die Berechnung des affinen optischen Flusses entwickelt worden.

In Experimenten auf synthetischen Testsequenzen ist die Leistung der Verfahren zur Berechnung des optischen Flusses sowohl bei hierarchischer als auch bei nicht hierarchischer Verarbeitung ermittelt worden. Dabei hat sich herausgestellt, daß beim Vorliegen von schnellen Bewegungen durch eine hierarchische Berechnung genauere Ergebnisse erzielt werden können. Es ist dabei jedoch zu beachten, daß der Zusammenhang, je höher die verwendete Bildpyramide, desto besser die Resultate, nicht gilt. Die sinnvolle Wahl der Pyramidenhöhe hängt von der Größe des Bildes und der Geschwindigkeit der Bewegung ab.

Außerdem sind das Verfahren von Simoncelli et al. und das auf Kanatani basierende Eigenwertverfahren quantitativ gegenübergestellt worden, was in der Literatur bisher nicht erfolgt ist. Das Fazit dieses Vergleichs ist, daß das von Kanatani vorgeschlagene Verfahren ohne hierarchische Verarbeitung bei schnellen Bewegungen bessere Ergebnisse liefert als das nicht hierarchische Simoncelli-Verfahren. Bei der Anwendung der hierarchischen Versionen der Algorithmen überragt jedoch das Simoncelli-Verfahren. Dies ist darauf zurückzuführen, daß die auf den einzelnen Auflösungsebenen zu bestimmenden Geschwindigkeiten kleiner sind und das Simoncelli-Verfahren bei kleinen Geschwindigkeiten leistungsstärker ist.

Es zeigt sich in den Experimenten ferner, daß die Berücksichtigung des Rauschens eine Optimierung der Berechnung des optischen Flusses ermöglicht. Dies wird durch die Ergebnisse der Flußberechnung unter Variation der Rauschparameter und die Gegenüberstellung der Verfahren zu dem konventionellen Verfahren von Lukas & Kanade belegt.

In einer weiteren Versuchsreihe ist die Leistung des Verfahrens zur Bewegungsanalyse auf realen Bildfolgen, die sportliche Übungen zeigen, getestet worden. Da für diese Versuche keine Grundwahrheiten zur Verfügung stehen, konnten die Ergebnisse nur qualitativ dargestellt und beurteilt werden.

Der erste Teil der Analyse, nämlich die Detektion der Person, ist in den Tests durch den vorgestellten hierarchischen Algorithmus erfolgt. Durch den Mehrauflösungsansatz des Verfahrens sind die Probleme des zeitlichen Aliasings gelöst worden, die durch schnelle Bewegungen hervorgerufen werden. Die Detektion der Person mit Hilfe des hierarchisch berechneten optischen Flusses stellt also eine leistungsfähige Alternative zu den üblichen Verfahren dar.

Die Verfolgung der Gelenke ist ebenfalls in Experimenten getestet worden. Dabei hat sich herausgestellt, daß das Verfahren eine zuverlässige Verfolgung der Gelenke erreicht, wenn die Bewegungen nicht eine durch die noch sinnvolle Pyramidenstufenzahl festgelegte Schranke übersteigen. Die noch sinnvolle Anzahl an Pyramidenstufen hängt dabei von der Größe des Bildes bzw. der räumlichen Ausdehnung der Körperteile ab. Das vierfache der noch sinnvollen Pyramidenstufenzahl beschränkt die Höhe der Geschwindigkeit der noch verfolgbaren Gelenkbewegung.

Abschließend sind die Probleme des Verfahrens diskutiert worden und mögliche Lösungsansätze vorgeschlagen worden. Das Problem längerer Selbstverdeckungen kann durch die Erweiterung des Verfahrens auf ein Mehransichten-Mehrauflösungsverfahren bewältigt werden. Ferner gibt es die Möglichkeit, bei der Bestimmung der Bildgeschwindigkeit zur Berechnung der zeitlichen Ableitung statt FIR-Filtern zeitlich rekursive IIR-Filter zu benutzen. Dadurch kann erreicht werden, daß die Generierung der Körperteilschablonen für die Berechnung des affinen optischen Flusses on line ohne a priori Wissen erfolgen kann. Kombiniert man das Verfahren mit einem Algorithmus, der aus der Umrißinformation die Gelenke lokalisiert, so hat man ein vollständiges System zur Gelenkverfolgung.

Literaturverzeichnis

[1] Anandan, P.: *A computational framework and an algorithm for the measurement of visual motion.* International Journal of Computer Vision, 2:283–310, 1989.

[2] Barron, J.L., D.J. Fleet, S.S. Beauchemin und T.A. Burkitt: *Performance of optical flow techniques.* In: *IEEE Conf. Computer Vision and Pattern Recognition*, Seiten 236–242, Champaign, IL, June 15-18, 1992.

[3] Bar-Shalom, Y. und T.E. Fortmann: *Tracking and Data Association.* Academic Press, New York, NY, 1988.

[4] Baumberg, A. und D.C. Hogg: *An Efficient Method for Contour Tracking Using Active Shape Models.* Technischer Bericht94.11, School of Computer Studies, University of Leeds, April 1994.

[5] Baumberg, A. und D.C. Hogg: *Learning Flexible Models from Image Sequences.* In: *Proc. Third European Conference on Computer Vision*, Band 1, Seiten 299–308. Stockholm, Sweden, May 2-6, J.O. Eklundh (Ed.), Springer LNCS 800, 1994.

[6] Beauchemin, S.S. und J.L. Barron: *The computation of optical flow.* ACM Computing Surveys, 27:433–467, 1995.

[7] Black, M.J. und A.D. Jepson: *EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View–Based Reprasentation.* Proc. Fourth European Conference on Computer Vision, Seiten 329–342, 1996.

[8] Daniilidis, K.: *Zur Fehlerempfindlichkeit in der Ermittlung von Objektbeschreibungen und relativen Bewegungen aus monokularen Bildfolgen.* Dissertation, Fakultät für Informatik, Universität Karlsruhe (TH), Juli 1992.

[9] Daniilidis, K., Ch. Krauss, M. Hansen und G. Sommer: *Real Time Tracking of Moving Objects with an Active Camera.* Technischer Bericht Nr. 9509, Inst. f. Inf. u. Prakt. Math., Oktober 1995. submitted for publication in the Real Time Imaging Journal.

[10] DiFranco, D.: *Visual Processing of Human Movement.* PhD Thesis, Dept. of Psychology, Queen's University, Kingston, Canada, 1980.

[11] Fischer, G.: *Lineare Algebra.* Vieweg Studium, Braunschweig/Wiesbaden, April 1987.

[12] Gavrila, D.M. und L.S. Davis: *Towards 3-D model-based tracking and recognition of human movement: a multi-view approach.* In: *International Workshop on Automatic Face- and Gesture-Recognition*, Seiten 272–277, Zürich, 1995.

[13] Gavrila, D.M. und L.S. Davis: *3-D model-based tracking of humans in action: a multi-view approach.* San Fransisco, CA, June 18-20, 19:73–80, 1996.

[14] Golub, G.H. und C.F. van Loan: *Matrix computations.* Johns Hopkins University Press, 1996.

[15] Goncalves, L., E. Di Bernardo, E. Ursella und P. Perona: *Monocular tracking of the human arm in 3D.* Boston, MA, June 20-23, 8:764–770, 1995.

[16] Jähne, B.: *Digitale Bildverarbeitung.* Springer Verlag, Berlin et al., 1991.

[17] Ju, S.X., M.J. Black und Y. Yacoob: *Cardboard People: A Parameterized Model of Articulated Image Motion.* In: *IEEE Conf. Computer Vision and Pattern Recognition*, Band 9, Seiten 38–44, San Fransisco, CA, June 18-20, 1996.

[18] Kakadiaris, I.A. und D. Metaxas: *3D Human Body Model Acquisition from Multiple Views.* In: *Proc. Int. Conf. on Computer Vision*, Band 8, Seiten 618–623, Boston, MA, June 20-23, 1995.

[19] Kakadiaris, I.A. und D. Metaxas: *Model-Based Estimation of 3D Human Motion with Occlusion Based on Active Multi-Viewpoint Selection.* In: *IEEE Conf. Computer Vision and Pattern Recognition*, Band 19, Seiten 81–87, San Fransisco, CA, June 18-20, 1996.

[20] Kakadiaris, I.A., D. Metaxas und R. Bajcsy: *Active Motion-Based Segmentation of Human Body Outlines.* In: *Proc. Articulate and Nonrigid Motion Workshop*, 1994.

[21] Kanatani, K.: *Statistical Optimization for Geometric Computation: Theory and Practice.* Elsevier Science, Amsterdam, 1996.

[22] Leung, M.K. und Y.H. Yang: *Human Body Motion Segmentation in a Complex Scene.* Pattern Recognition, 20:55–64, 1987.

[23] Leung, M.K. und Y.H. Yang: *First Sight: A Human Body Outline Labeling System.* IEEE Trans. Pattern Analysis and Machine Intelligence, 17:359–377, 1995.

[24] Marr, D.: *Vision.* W.H. Freeman, San Fancisco, 1982.

[25] Murase, H. und R. Sakai: *Moving object recognition in eigenspace representation: gait analysis and lip reading.* Pattern Recognition Letters, 17:155–162, 1996.

[26] NAGEL, H.-H.: *On a constraint equation for the estimation of displacement rates in image sequences.* IEEE Trans. Pattern Analysis and Machine Intelligence, 11:13–30, 1989.

[27] NAGEL, H.H.: *Optical Flow Estimation and the Interaction Between Measurement Errors at Adjacent Pixel Positions.* International Journal of Computer Vision, 15:271–288, 1993.

[28] PENTLAND, A. und B. HOROWITZ: *Recovery of Nonrigid Motion and Structure.* IEEE Trans. Pattern Analysis and Machine Intelligence, 13:730–742, 1991.

[29] ROHR, K.: *Towards Model-Based Recognition of Human Movements in Image Sequences.* CVGIP: Image Understanding, 59:94–115, 1994.

[30] SIMONCELLI, E.P.: *Distributed Representation and Analysis of Visual Motion.* Technical Report 209, Vision and Modeling Group, M.I.T.Media Labroratory, 20 Ames Street, Cambridge, MA 02139, January 1993.

[31] SIMONCELLI, E.P., E.H. ADELSON und D.J. HEEGER: *Probability Distributions of Optical Flow.* In: *Proc. Conf. Comput. Vis. Recog.*, Seiten 310–315, Maui, Hawaii, 1991.

[32] SORENSON, H.W.: *Parameter Estimation, Principles and Problems.* Marcel Dekker, New York and Basel, 1980.

[33] TURK, M. und A. PENTLAND: *Face recognition using eigenfaces.* Maui, Hawaii, June 3-6, Seiten 586–591, 1991.

[34] WEBER, J. und J. MALIK: *Robust computation of optical flow in a multi-scale differential framework.* In: *ICCV*, Seiten 12–20, ICCV93, 1993.

[35] WINTER, D.A.: *Biomechanics and motor control of human movement.* A Wiley-Interscience Publication, John Wiley & Sons, Inc., Second Auflage, 1990.

[36] WREN, C., A. AZARBAYEGAN, T. DARRELL und A. PENTLAND: *Pfinder: Real Time Tracking of the Human Body.* Technical Report 353, M.I.T. Media Laboratory Perceptual Computing Section, 20 Ames St.,Cambridge,MA 02139 USA, 1995.

Danksagung

Diese Gelegenheit möchte ich dazu nutzen, mich bei Herrn Dr. Konstantinos Daniilidis zu bedanken, der, wann immer es ihm möglich war, durch Diskussionen und Denkanstöße dazu beigetragen hat, diese Arbeit zu verfassen.

Dank gilt auch Jens Ellerbrock, der mir bei der Aufnahme der Bildfolgen behilflich war, und Michael Benkwitz fürs Korrekturlesen.

Ein besonderes Dankeschön geht an die Leute in meiner unmittelbaren Arbeitsumgebung, die sich auch durch noch so verzwickte, LaTeX oder ähnliches betreffende Fragen nicht aus der Ruhe bringen ließen und für die Verbreitung einer angenehmen Arbeitsatmosphäre gesorgt haben.

Bedanken möchte ich mich außerdem bei Lars, dafür, daß er mir immer wieder den notwendigen Abstand verschafft hat, dafür, daß er mich durch konstruktive Kritik, insbesondere durch das Korrekturlesen, beim Gelingen meiner Arbeit unterstützt hat, und dafür, daß er meine schwankenden Launen ertragen hat.

Zuletzt möchte ich mich ganz besonders bei meinen Eltern bedanken, die mir dieses Studium ermöglicht haben und mir ermutigend zur Seite standen.

Erklärung:

Hiermit versichere ich, daß ich die vorliegende Arbeit selbständig verfaßt und ausschließlich die angegebenen Hilfsmittel und Quellen verwendet habe.

Kiel, den