

Universität Bielefeld
Technische Fakultät
Arbeitsgruppe Angewandte Informatik

Diplomarbeit

Detektion von Bewegung bei der Berechnung
von Mosaikbildern

von

Birgit Möller

Betreuer: Prof. Dr.-Ing. Stefan Posch
Prof. Dr.-Ing. Gerhard Sagerer

Bielefeld, 30. März 2001

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbständig bearbeitet und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Bielefeld, den 30. März 2001

Birgit Möller

Vorwort

Ich möchte mich an dieser Stelle bei allen bedanken, die mir bei der Bearbeitung dieser Diplomarbeit zur Seite gestanden haben. Vorrangig mein Betreuer Stefan Posch hat mir durch seine Diskussionsbereitschaft und die daraus resultierenden Ratschläge in vielen Situationen weitergeholfen. Darüberhinaus ist meinen Kommilitonen Thomas Plötz, Marc Hanheide und Frank Seifert zu danken, die mich durch ihre konstruktive Kritik bei der schriftlichen Ausarbeitung der vorliegenden Arbeit unterstützt haben. Auch die Mitglieder der Arbeitsgruppe Angewandte Informatik um Gerhard Sagerer waren in allen Belangen hilfsbereit.

Insbesondere Daniel Schlüter gilt mein Dank. Durch meine zweijährige Tätigkeit als studentische Hilfskraft bei ihm konnte ich wertvolle Programmiererfahrung sammeln und einen Einblick in das Gebiet der Bildanalyse gewinnen. Die Entscheidung für dieses Diplomarbeitsthema wurde nicht zuletzt dadurch maßgeblich beeinflusst.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Aufgabenstellung	3
1.2	Gliederung	3
2	Mosaikbilder	5
2.1	Generierung von Mosaikbildern	6
2.1.1	Transformation	7
2.1.2	Pixelintegration	11
2.2	Bewegte Szenen	13
3	Bewegungsdetektion	15
3.1	Intensitätsdifferenz	17
3.2	Normal Flow	18
3.2.1	Grundlagen des optischen Flusses	18
3.2.2	Theorie des Normal Flow	20
3.2.3	Gewichteter mittlerer Normal Flow	22
3.2.4	Rekonstruktion der Bewegungsrichtung	23
3.3	Konvergenzanalyse	27

3.4	Evaluation	28
3.4.1	Konfigurationsparameter	28
3.4.2	Direkter Vergleich der Verfahren	33
3.4.3	Fazit	37
4	Regionentracking	39
4.1	Regionensegmentierung	40
4.1.1	Morphologische Operationen	41
4.2	Zeitliche Korrespondenzanalyse	44
4.2.1	Merkmale	45
4.2.2	Zuordnung detektierter Regionen	46
4.2.3	Zusammenhangskomponenten	51
4.2.4	Der Korrespondenzgraph	57
4.3	Trajektorienextraktion	58
5	Hintergrundmosaikbilder	61
5.1	Statische Integration	62
5.2	Wahl des Referenzbildes	64
5.3	Zusammenfassung	66
6	Realisierung	67
6.1	Datenstrukturen	68
6.2	Algorithmen	70
7	Auswertung	73
7.1	Ergebnisse	74
7.2	Komplexitätsanalyse	83

8 Zusammenfassung & Ausblick	87
Literaturverzeichnis	91
A Theorie des χ^2-Tests	95
B Bildfolgen	97
C Hinweise zum Mosaikprogramm	103

INHALTSVERZEICHNIS

Abbildungsverzeichnis

2.1	Beispiel für ein Mosaikbild	5
2.2	Zweistufiges Verfahren zur Mosaikbildgenerierung	7
2.3	Korrelation im <i>Point-To-Line-Verfahren</i>	10
2.4	Regionenbasierte Integrationsverfahren	12
2.5	Integrationsfehler bei bewegten Szenen	13
3.1	Visualisierung der Einschränkungsgleichung	20
3.2	Gewinnung des Normal Flow durch Projektion	20
3.3	Visualisierung des Normal Flow am bewegten Objekt	21
3.4	Erfolgreiche Rekonstruktion der Bewegungsrichtung	25
3.5	Instabile Rekonstruktion der Bewegungsrichtung	26
3.6	Nachbarschaftsgrößen in der Berechnung der Intensitätsdifferenz .	29
3.7	Vergleich der Verfahren zur Ableitungsberechnung	31
3.8	Nachbarschaftsgrößen und Binarisierung in der Konvergenzanalyse	33
3.9	Robustheit der Detektionsverfahren	34
3.10	Definitionslücken in der Normal Flow Berechnung	35
3.11	Unvollständige Objektdetektion bei fehlenden Grauwertdifferenzen	35
3.12	Direkter Vergleich der drei Detektionsverfahren	37

4.1	"Ausgefrante" und lückenhafte Gebiete in einer Bewegungskarte .	41
4.2	Morphologische Grundoperationen	42
4.3	Opening und Closing	43
4.4	Auswirkungen der Dilatation	44
4.5	Ausreißer im χ^2 -Test	48
4.6	Schnittfläche von zwei Histogrammen	49
4.7	Varianz in der Regionensegmentierung	51
4.8	Kürzeste Distanz zwischen zwei Regionen	53
4.9	Skizze einer Zusammenhangskomponente	53
4.10	Mehrdeutigkeiten in der Zuordnung von Komponenten	54
4.11	Varianz in der Regionengruppierung	55
4.12	Verbleibende Probleme in der Korrespondenzanalyse	56
4.13	Zeitliche Korrespondenz bewegter Regionen	57
4.14	Die drei möglichen Konstellationen bei der Trajektoriengenerierung	60
5.1	Statische Integration in Mosaikbildern bewegter Szenen	63
5.2	Auswirkungen der Referenzbildwahl	65
5.3	Mosaikbildentwicklung bei einer Szene mit Verdeckungen	66
6.1	Einbettung der Bewegungsanalyse in die Mosaikbildgenerierung .	72
7.1	Ergebnisse der Bewegungsanalyse in der Bildfolge "Baufix4" . . .	75
7.2	Ergebnisse der Bildfolge "Baufix2"	76
7.3	Trajektorie einer unstetigen Bewegung in der Bildfolge "Baufix5"	78
7.4	Mosaikbild und resultierende Trajektorie der Bildfolge "Office1" .	79
7.5	Mosaikbild der Bildfolge "Desk2" mit Integrationsfehlern	79

7.6	Mosaikbild und Trajektorien der Bildfolge "Desk1"	80
7.7	Ähnliche Trajektorien sich parallel bewegender Komponenten . .	82
A.1	χ^2 -Verteilungen verschiedener Freiheitsgrade	96
A.2	Signifikanzniveau beim χ^2 -Test	96
B.1	Bildfolge Baufix1	97
B.2	Bildfolge Baufix2	98
B.3	Bildfolge Baufix3	98
B.4	Bildfolge Baufix4	99
B.5	Bildfolge Baufix5	100
B.6	Bildfolge Desk1	101
B.7	Bildfolge Desk2	101
B.8	Bildfolge Office1	102
C.1	Screenshot der graphischen Oberfläche	104

ABBILDUNGSVERZEICHNIS

Kapitel 1

Einleitung

Aus der natürlichen Integration des Menschen in seine Umwelt resultiert eine stetige Interaktion. Auf der einen Seite nimmt der Mensch durch Handlungen Einfluss. Andererseits registriert er über seine Sinnesorgane die Eigenschaften und aktuellen Zustände der Umgebung und reagiert auf deren Veränderungen. Wie in [MF92] beschrieben, stellt die Aufnahme visueller Informationen dabei eine wichtige Informationsquelle dar. Sie ermöglicht ihm die Gewinnung statischer Informationen von Objekten, wie Farben, Formen oder Größen, und dynamischer Daten über ihre Bewegungen und Veränderungen im Verlauf der Zeit. Die visuelle Wahrnehmung ist in einen aktiven Explorationsprozess eingebettet. Durch eine kontinuierliche Fokussierung neuer Szenenbereiche und Objekte werden aktuell relevante Informationen der Umgebung akquiriert.

Die Speicherung der extrahierten Daten findet unter anderem im visuellen Kurzzeitgedächtnis statt [Sch98]. Es stellt jederzeit eine Repräsentation der Umwelt zur Verfügung, in der auch Informationen über aktuell nicht mehr im Blickfeld liegende Objekte enthalten sind. Durch Rückgriff auf diese Informationen kann der Mensch sich schnell neuen Situationen anpassen, ohne die Szene erneut explorieren zu müssen.

Bei der Entwicklung flexibler interaktiver Systeme besteht auf Grund der oben erläuterten Eigenschaften menschlicher Wahrnehmung eine der Hauptaufgaben in der Realisierung von Komponenten zur Verarbeitung visueller Informationen (*Computer Vision*). In [FA94] wird dazu ein an den Aufbau biologischer Systeme

angelehnter Ansatz vorgeschlagen, der den Wechselwirkungen zwischen den Systemen und ihren Umgebungen Rechnung trägt. Bildanalyssysteme sollen demnach unter anderem aus einzelnen Bildverarbeitungsmodulen zusammengesetzt werden, die der Gewinnung der visuellen Informationen aus mit Kameras aufgenommenen Bildern dienen, sowie ergänzend geeignete Aktoren für auszuführende Handlungen und ein Gedächtnis besitzen.

Um die in biologischen Systemen vorzufindende Leistungsfähigkeit zu erzielen, ist die Analyse von *Bildfolgen* unerlässlich. Insbesondere Bewegungsinformationen, die nach [MF92] in der visuellen Wahrnehmung eine Schlüsselrolle einnehmen, lassen sich aus Einzelbildern nicht gewinnen. Durch den Einsatz aktiver Kameras bei der Aufnahme der Folgen wird darüberhinaus die Umsetzung einer aufmerksamkeitsgesteuerten Szenenexploration möglich. Mit dem aus der Analyse der Bildfolgen resultierenden Informationsgewinn steigt allerdings gleichermaßen die zu verarbeitende Datenmenge für die Bildverarbeitungsmodule. Dies erschwert eine Echtzeitverarbeitung der Daten, die für eine hohe Interaktivität der Systeme zwingend notwendig ist.

Einen Lösungsansatz, um das Datenvolumen zu reduzieren, stellen Mosaikbilder dar. Ein Mosaikbild ist eine kompakte Zusammenfassung der in einer Sequenz enthaltenen Bilder, unter weitgehender Eliminierung der Redundanz. Da alle im Verlauf der bisherigen Bildfolge fokussierten Szenenbereiche repräsentiert sind, können Bildverarbeitungsmodule auf Basis des Mosaikbildes eine Untersuchung der aktuellen Szene durchführen, ohne die Einzelbilder der Sequenz analysieren zu müssen. Daher kann ein Mosaikbild als einfache Realisierung eines visuellen Kurzzeitgedächtnisses angesehen werden. Die fortschreitende Integration neuer Informationen mit jedem Bild führt dabei zu einer sukzessiven Verdrängung alter Informationen. Dies gleicht dem Prozess des Vergessens beim Menschen.

Gängige Verfahren zur Generierung von Mosaikbildern setzen statische Szenen voraus. Ihre Anwendung auf Bildsequenzen bewegter Szenen führt zu fehlerhaften Mosaikbildern und einem Verlust der, erst durch die Verwendung der Bildfolgen zugänglich gewordenen, dynamischen Informationen. Daraus folgt die Notwendigkeit, geeignete Verfahren zur Mosaikbildgenerierung für die Anwendung auf Sequenzen bewegter Szenen zu entwerfen. Sie sollen die angesprochenen Integrationsfehler vermeiden und die dynamischen Informationen erhalten.

1.1 Aufgabenstellung

Im Rahmen einer Diplomarbeit an der Universität Bielefeld wurden Verfahren zur Generierung von Mosaikbildern evaluiert und implementiert [ST99]. Zwingende Voraussetzung für fehlerfreie Mosaikbilder ist dabei die Verwendung von Bildfolgen statischer Szenen. Um den Einsatzbereich des entstandenen Programms auszudehnen und bewegte Szenen zuzulassen, besteht die Zielsetzung der vorliegenden Arbeit in der Entwicklung geeigneter Erweiterungen. Durch sie soll es möglich werden, nachfolgenden Bildanalysekomponenten fehlerfreie Mosaikbilder bewegter Szenen und die in den Bildsequenzen enthaltenen Bewegungsinformationen zur Verfügung zu stellen.

Ausgangsbasis ist zunächst die Detektion bewegter Bereiche einer Szene. Im Anschluss daran sollen die resultierenden Bewegungsinformationen bei der Mosaikbilderstellung geeignet berücksichtigt werden, um die im vorangegangenen Abschnitt angesprochenen Fehler im Mosaikbild zu vermeiden.

Die Extraktion der dynamischen Informationen soll durch eine parallel zur Mosaikbildgenerierung ablaufende Verfolgung der zuvor als bewegt detektierten Bereiche über die Zeit (*Tracking*) realisiert werden. Dazu sind Merkmale auszuwählen, die eine Zuordnung korrespondierender Regionen in aufeinanderfolgenden Bildern der Sequenz (*Matching*) erlauben. Ergänzend sollen eine geeignete Strategie bei der Suche nach Korrespondenzen entwickelt, sowie Datenstrukturen zu ihrer Verwaltung entworfen werden. Letztere müssen einen effizienten Zugriff auf die Bewegungsdaten detektierter Gebiete gewährleisten, die neben dem Mosaikbild den Analysemodulen als Basis dienen sollen.

1.2 Gliederung

Einleitend wird in Kapitel 2 die Grundidee von Mosaikbildern vorgestellt. Daran schließt sich eine Erläuterung der üblichen zweistufigen Vorgehensweise zu ihrer Generierung an, bevor detailliert die Probleme bei der Verarbeitung von Szenen mit bewegten Objekten dargestellt werden.

Zu ihrer Lösung enthält Kapitel 3 zunächst eine Diskussion und Evaluation verschiedener Möglichkeiten zur Detektion bewegter Bereiche in Bildsequenzen. Die

dabei betrachteten Verfahren liefern pixelbasierte Bewegungsinformationen für die Bilder einer Sequenz. Um eine Verfolgung der Bewegungen über die Zeit zu realisieren, findet daher eine Gewinnung regionenhafter Informationen aus den Bewegungsdaten statt, die zu Beginn von Kapitel 4 erläutert wird. Des Weiteren beinhaltet das Kapitel eine Beschreibung der Zuordnung von bewegten Bereichen in aufeinanderfolgenden Bildern. Neben der bei der Suche nach Zuordnungen benutzten Strategie, umfasst diese insbesondere eine Erläuterung der beim Matching verwandten Merkmale für die Regionen, sowie der zur Verwaltung der Korrespondenzen eingesetzten Datenstrukturen. Den Abschluss bildet die Darstellung der Extraktion von Bewegungsdaten, d.h. insbesondere Richtungsinformationen, aus den gewonnenen Zuordnungen. In Kapitel 5 wird die Generierung von Mosaikbildern des statischen Szenenhintergrundes vorgestellt. Durch eine geeignete Berücksichtigung der zuvor als bewegt detektierten Pixel werden dabei Integrationsfehler vermieden und konsistente Szenenrepräsentationen erstellt.

Die Realisierung der präsentierten Verfahren erfolgt auf Basis des von Dirk Sypli und Heiko Tappe im Rahmen ihrer Diplomarbeit [ST99] angefertigten Mosaikprogramms. Dazu wird dieses um geeignete Datenstrukturen und Algorithmen erweitert, die im sechsten Kapitel skizziert werden. Darüberhinaus findet sich dort eine Erläuterung der Einbettung der Bewegungsanalyse in den Prozess der Mosaikbildgenerierung. Die auf diese Weise erzielbare Leistungsfähigkeit bei der Verarbeitung von Bildfolgen bewegter Szenen wird anschließend in Kapitel 7 anhand von Beispielen illustriert. Für einzelne Bildfolgen werden dazu jeweils Mosaikbilder des Szenenhintergrundes berechnet und die extrahierten dynamischen Informationen geeignet visualisiert. Die dabei zu Grunde gelegten Bildfolgen, denen auch die in den einzelnen Kapiteln zu findenden Beispiele entnommen wurden, sind im Anhang B aufgeführt. Kapitel 8 fasst abschließend die Resultate der Diplomarbeit zusammen und gibt einen Ausblick auf weitere interessante Fragestellungen.

Kapitel 2

Mosaikbilder

Bildsequenzen, die aus der Aufnahme von Szenen mit aktiven Kameras resultieren, stellen ein großes Datenvolumen dar. In Abhängigkeit von der verwendeten Bildrate und der gegebenen Kameradynamik überlappen dabei aufeinanderfolgende Bilder mehr oder minder stark. Die dadurch in der Sequenz enthaltene Redundanz dient als Grundlage zur Erstellung von Mosaikbildern.

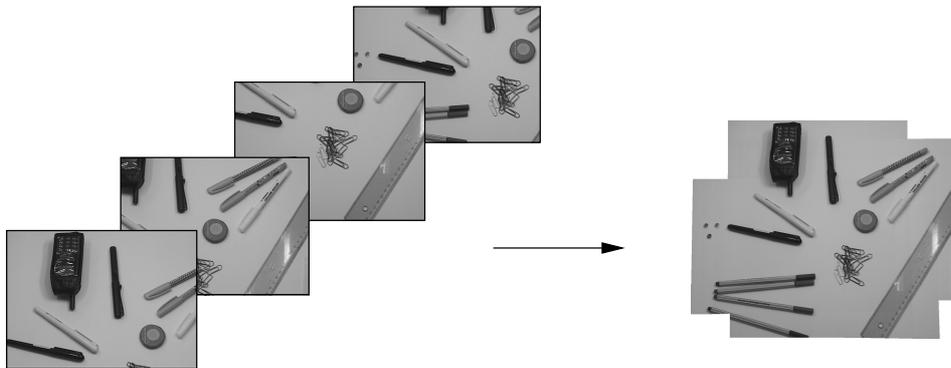


Abbildung 2.1: Beispiel für ein Mosaikbild

Ein Mosaikbild entsteht durch Fusion der einzelnen Bilder einer Sequenz zu einem gemeinsamen Bild. Wie in Abb. 2.1 dargestellt, sind dabei Szenenbereiche, die in mehreren Bildern auftreten, nur einmal im Mosaik enthalten, was zu einer Datenreduktion führt. Des Weiteren sind alle innerhalb der bisherigen Bildfolge explorierten Teile der Szene im Mosaikbild repräsentiert, so dass auch Informationen

über aktuell nicht mehr im Fokus der Kamera platzierte Objekte zur Verfügung stehen.

Bei der Erstellung von Mosaikbildern findet im Allgemeinen ein zweistufiges Verfahren Anwendung, bestehend aus einem Transformations- und einem Integrations-schritt. In Kapitel 2.1 wird diese Vorgehensweise genauer beschrieben, wobei sich die Darstellung im Wesentlichen an der Arbeit von Dirk Sypli und Heiko Tappe [ST99] orientiert, der weitere Details zu entnehmen sind. Die in der Einleitung bereits angesprochene Problematik von Bewegungen innerhalb der Szene während der Aufnahme der Bildsequenz wird in Kapitel 2.2 aufgegriffen, wobei die in dieser Arbeit verfolgten Lösungsansätze präsentiert werden.

2.1 Generierung von Mosaikbildern

Die Bilder einer Sequenz liegen, bedingt durch eine bei der Aufnahme stattfindende Kamerabewegung, in unterschiedlichen Kamerakoordinatensystemen vor. Identische Objektpunkte der Szene sind auf unterschiedliche Bildpunkte in den Einzelbildern abgebildet worden. Der erste Schritt zur Erstellung eines Mosaikbildes besteht in der Umrechnung der einzelnen Bildkoordinatensysteme in ein gemeinsames Referenzsystem. Dazu findet unter Annahme eines Bewegungsmodells für die Kamera eine Schätzung von Transformationsparametern zwischen den Einzelbildern und dem ausgewählten Referenzbild statt. In Abschnitt 2.1.1 werden, nach einer Beschreibung des verwendeten Modells, drei Verfahren hierzu vorgestellt.

Sind die Parameter bekannt, so können die Informationen aus den transformierten Einzelbildern im Mosaikbild zusammengefügt werden. Die unterschiedlichen Projektionen desselben Szenenpunktes in den Bildern werden dabei unter Anwendung geeigneter Integrationsheuristiken, wie sie exemplarisch in Abschnitt 2.1.2 vorgestellt werden, zu einem gemeinsamen Bildpunkt im Mosaik fusioniert. Steht bei der Erstellung des Mosaikbildes bereits die komplette Sequenz zur Verfügung, so können dabei alle Bilder gleichermaßen berücksichtigt werden. Falls eine mit der Bildfolge schritthaltende Mosaikbilderstellung erfolgen soll, werden lediglich die Informationen des aktuellen Bildes und des bisher erzeugten Mosaikbildes fusioniert. Abb. 2.2 zeigt eine schematische Übersicht des Verfahrens.

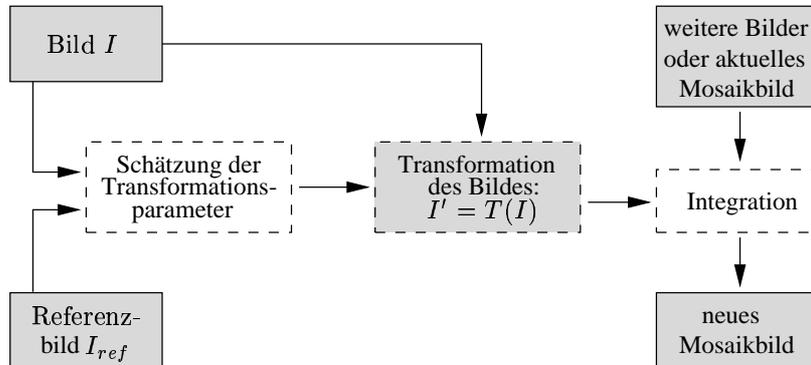


Abbildung 2.2: Zweistufiges Verfahren zur Mosaikbildgenerierung

2.1.1 Transformation

Aktive Kameras können bei der Aufnahme einer Bildfolge Rotationen um alle drei optischen Achsen, Translationen in allen drei Raumrichtungen und Zoom ausführen. Zur Beschreibung dieser Bewegungen existieren verschiedene Modelle, die sich im Wesentlichen hinsichtlich ihrer Flexibilität, d.h. der Vielfalt zugelassener Kamerabewegungen, und der Anzahl notwendiger Parameter unterscheiden. Zu berücksichtigen ist dabei, dass keines der Modelle eine Translation des optischen Zentrums der Kamera in 3D-Szenen erfassen kann. Die Kamerabewegung führt in diesen Fällen zu einer weiteren, unabhängigen Bewegung im Bild, die durch die scheinbare Verlagerung von Objekten auf Grund des Standortwechsels der Kamera verursacht wird (Parallaxe). Zu ihrer Modellierung wäre neben dem Modell für die Kamerabewegung ein weiteres Modell notwendig.

In [ST99] wurde die projektive Abbildung als Bewegungsmodell ausgewählt. Sie lässt sich durch acht Parameter beschreiben (Gleichung 2.1) und umfasst alle modellierbaren Bewegungen der Kamera in 2D- und 3D-Szenen.

$$\vec{x}' = \frac{A\vec{x} + \vec{b}}{\vec{c}^T \vec{x} + 1} = \frac{\begin{pmatrix} a_{11}x + a_{12}y + b_1 \\ a_{21}x + a_{22}y + b_2 \end{pmatrix}}{c_1x + c_2y + 1} \quad \text{mit} \quad \vec{x} = \begin{pmatrix} x \\ y \end{pmatrix} \quad (2.1)$$

Da viele Verfahren zur Parameterschätzung Linearität voraussetzen, und diese bei der projektiven Abbildung nicht gegeben ist, sind die Parameter nur unter

großem Aufwand zu ermitteln. Daher werden sie in der Regel zunächst durch ein lineares Modell iterativ approximiert und erst dann exakt berechnet. Das in diesem Fall gewählte Modell ist die pseudo-perspektivische Abbildung (Gleichung 2.2), die mit ebenfalls acht Parametern in der Lage ist, die von der projektiven Abbildung erfassten Bewegungen zu approximieren.

$$\vec{x}' = \begin{pmatrix} a + bx + cy + gx^2 + hxy \\ d + ex + fy + gxy + hy^2 \end{pmatrix} \quad (2.2)$$

Im Folgenden werden drei Verfahren vorgestellt, die auf der Basis zweier gegebener Bilder eine Schätzung der acht Transformationsparameter des pseudo-perspektivischen Modells ermöglichen. Sie lassen sich in merkmalslose und merkmalsbasierte Verfahren unterteilen. Die erste Klasse arbeitet direkt auf den Grauwertinformationen der kompletten Bilder. Ihr gehören das *Korrelationsverfahren* sowie der *Perspective Flow* an. Verfahren der zweiten Kategorie ermitteln die Transformationsparameter zwischen den Bildern auf Basis ausgewählter Merkmale. Hierzu zählt das Verfahren *Point-To-Line*.

Korrelationsverfahren

Ausgangspunkt der Parameterschätzung in diesem Verfahren ist die quadratische Intensitätsdifferenz zwischen beiden Bildern I_{ref} und I , die durch iterative Verbesserung des Parametervektors $\vec{\alpha} = (a, b, c, d, e, f, g, h)^T$ minimiert werden soll:

$$o(\vec{\alpha}) = \sum_{\vec{x}} (I_{ref}(\vec{x}) - I(\vec{x} - \vec{v}_{\alpha}))^2 \quad (2.3)$$

\vec{v}_{α} entspricht der Verschiebung der Bilder gegeneinander, in Abhängigkeit vom aktuellen Vektor $\vec{\alpha}$, \vec{x} bezeichnet jeweils die einzelnen Bildpunkte.

Die obige Zielfunktion o beschreibt ein stark überbestimmtes, nicht-lineares Gleichungssystem. Durch Anwendung geeigneter Linearisierungsverfahren kann das Minimum auf Basis linearer Gleichungssysteme angenähert werden. In diesem Fall wurde dazu die House-Holder-Transformation eingesetzt, bei der die zu optimierende Funktion zunächst durch eine Taylorreihen-Entwicklung linearisiert und das resultierende lineare Gleichungssystem dann mit bekannten Verfahren gelöst wird. Der errechnete Wert für $\vec{\alpha}$ dient jeweils als Startwert der nächsten Iteration, bis ein Abbruchkriterium erfüllt ist.

Perspective Flow

In diesem Verfahren wird zur Parameterschätzung eine Berechnung des optischen Flusses durchgeführt. Er beschreibt die beobachteten Grauwertveränderungen zwischen zwei Bildern. Für jeden Bildpunkt wird dabei ein Verschiebungsvektor \vec{v} bestimmt, der die Bewegung des Punktes charakterisiert. Als Basis der Berechnung dieser Vektoren dient die bekannte Einschränkungsgleichung (siehe auch Kap. 3.2.1):

$$0 = v_x I_x + v_y I_y + I_t \quad (2.4)$$

v_x und v_y entsprechen den pixelweisen Verschiebungen entlang der Achsen, I_x , I_y und I_t den räumlichen und zeitlichen Ableitungen der Bildfunktion I . Das Gleichungssystem ist unterbestimmt, so dass weitere Bedingungen festgelegt werden müssen. Dies geschieht durch direkte Einbeziehung des zu Grunde gelegten Bewegungsmodells, für das Parameter zu schätzen sind. Es folgt

$$\begin{aligned} v_x = x' - x &= (a + bx + cy + gx^2 + hxy) - x \\ v_y = y' - y &= (d + ex + fy + gxy + hy^2) - y \end{aligned} \quad (2.5)$$

Eingesetzt in Gleichung 2.4 resultiert ein Gleichungssystem, dessen gesuchtes Minimum mit der Methode der kleinsten Quadrate bestimmt werden kann. Die errechneten Parameter werden auch hier bis zum Erreichen eines Abbruchkriteriums iterativ verbessert.

Beide vorgestellten Verfahren nutzen zur Schätzung der Parameter Ableitungswerte der lokalen Bildfunktion (beim Korrelationsverfahren bedingt durch die Taylorreihen-Entwicklung). Diese werden in der Regel aus einer Nachbarschaft um das betrachtete Pixel berechnet und sind damit lokal begrenzt. Um Bewegungen zu erfassen, die den betrachteten Bereich überschreiten, werden die Verfahren innerhalb einer AuflösungsPyramide organisiert. Dabei wird die Auflösung der Bilder zunächst schrittweise verringert. Angefangen bei der kleinsten Auflösung erfolgt dann stufenweise eine Bestimmung von Transformationsparametern, die jeweils initial auf die Bilder der nächsthöheren Stufe angewendet werden und die dortigen Bewegungen näherungsweise kompensieren. Auf diese Weise wird auf jeder Stufe die maximal zulässige Größe der Bewegungen eingehalten, und auch bei großen Verschiebungen zwischen den Bildern können adäquate Transformationsparameter berechnet werden.

Point-To-Line

Die merkmalslosen Verfahren weisen eine hohe Komplexität auf, da sie alle Bildpunkte berücksichtigen. Das merkmalsbasierte *Point-To-Line*-Verfahren beschränkt sich auf ca. 20 – 30 markante Punkte, die zur Parameterschätzung herangezogen werden. Im zu transformierenden Bild I werden diese Punkte festgelegt, wobei Konturinformation als Auswahlkriterium dient. Anschließend wird für jeden Punkt eine Korrelationsmatrix berechnet, die beschreibt, wie gut die Nachbarschaft des Punktes im aktuellen Bild I mit Bildbereichen im Referenzbild I_{ref} übereinstimmt. Falls der betrachtete Punkt auf einer Kante liegt, so wird die Korrelation an allen Stellen der Kante im Referenzbild hoch sein (siehe Abb. 2.3).

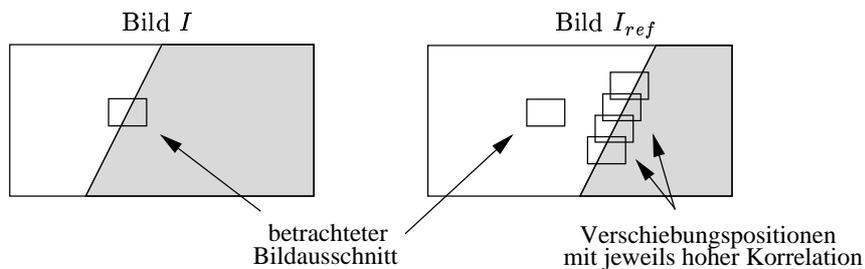


Abbildung 2.3: Korrelation im *Point-To-Line-Verfahren*

Durch eine Hough-Transformation der Korrelationsmatrix erfolgt die Berechnung einer Geraden, die diese Kante, auf die der Punkt im Referenzbild verschoben wurde, beschreibt. Die resultierenden Paare aus Punkten und ihnen zugeordneten Geraden dienen als Grundlage für die Parameterschätzung, bei der unter Annahme des gewählten Bewegungsmodells der Abstand zwischen den ermittelten Geraden und den ins Referenzbild transformierten Punkten minimiert wird.

In allen drei präsentierten Verfahren entspricht das bei der Transformationsparameterschätzung zu Grunde gelegte Referenzkoordinatensystem des Mosaikbildes im Allgemeinen dem Koordinatensystem des ersten Bildes. Insbesondere in langen Sequenzen ist der Überlapp einzelner Bilder mit dem ersten Bild jedoch oft zu gering, um eine direkte Schätzung der Parameter zuzulassen. Aus diesem Grund werden sie in der Regel differenziell bestimmt. Dabei wird das Vorgängerbild $I^{(t-1)}$

der Sequenz als Referenzbild für die Schätzung der Parameter des aktuellen Bildes $I^{(t)}$ benutzt. Der resultierende Parametersatz $P_{t-1,t}$, der die Transformation des Bildes $I^{(t)}$ in die Koordinaten des Bildes $I^{(t-1)}$ beschreibt, wird dann mit den bereits bekannten Transformationsparametern $P_{0,t-1}$ zwischen Vorgängerbild und Mosaikbild konkateniert und auf das aktuelle Bild $I^{(t)}$ angewendet.

Da hierbei nur jeweils zwei aufeinanderfolgende Bilder in Beziehung zueinander gesetzt werden, können leicht inkonsistente Parametersätze die Folge sein, die zu Fehlern bei der Erstellung des Mosaikbildes führen (Beispiele in [ST99]). Um dem entgegenzuwirken, wurde eine direkte Schätzung der Parameter relativ zum Mosaikbild realisiert (*Frame-To-Mosaic*). Dabei wird unter Verwendung der inversen Parameter $P_{0,t-1}^{-1}$ des Vorgängerbildes ein geeigneter Ausschnitt aus dem aktuellen Mosaikbild herausgeschnitten, der als Referenz dient. In ihm sind Informationen aller vorherigen Bilder der Sequenz enthalten, so dass eine robuste Schätzung der Parameter unter Berücksichtigung der bisherigen Bilder der Folge möglich wird und Fehler vermieden werden.

2.1.2 Pixelintegration

Nach Anwendung der ermittelten Transformationsparameter liegen alle verfügbaren Bilder im gegebenen Referenzkoordinatensystem vor. Der nächste Schritt besteht nun in der Fusion der Grauwertinformationen aus den einzelnen Bildern und eines eventuell schon bestehenden Mosaikbildes. Nach [MS99] sind dabei grundsätzlich pixel- und regionenbasierte Vorgehensweisen zu unterscheiden.

Im ersten Fall werden die für jeden Mosaikbildpunkt $M(\vec{x})$ vorhandenen Informationen in den gegebenen Bildern $I^{(t)}$, $t = 1 \dots n$, unter Anwendung einer Kombinationsfunktion f zusammengefasst:

$$M(\vec{x}) = f(I^{(1)}(\vec{x}'_1), I^{(2)}(\vec{x}'_2), \dots, I^{(N)}(\vec{x}'_n)) \quad (2.6)$$

Die \vec{x}'_i entsprechen dabei jeweils den Koordinaten der zu \vec{x} korrespondierenden Bildpunkte in den einzelnen Bildern $I^{(t)}$ der Folge. Übliche Kombinationsfunktionen sind die Bildung des arithmetischen Mittels oder die Berechnung des Medians. Die Berücksichtigung mehrerer Bilder bei der Fusion führt unter Umständen zu Unschärfe bzw. Verwischungen im Mosaikbild. Dies passiert insbesondere dann,

wenn die aufeinander abgebildeten Pixelwerte der Einzelbilder durch Szenenveränderungen im Verlauf der Aufnahme (z.B. Varianz in der Beleuchtung) stark voneinander abweichen.

Regionenbasierte Integrationsverfahren, die vornehmlich bei Vorliegen der kompletten Bildsequenz eingesetzt werden, wirken dem entgegen. Für das zu erstellende Mosaikbild wird eine Zerlegung in disjunkte Zellen ermittelt, in denen nur jeweils ein Bild $I^{(s(\vec{x}))}$, festgelegt durch den Index $s(\vec{x})$, als Informationsquelle dient:

$$M(\vec{x}) = p_{s(\vec{x})}(I^{(1)}(\vec{x}'_1), I^{(2)}(\vec{x}'_2), \dots, I^{(n)}(\vec{x}'_n)) = I^{(s(\vec{x}))}(\vec{x}'_{s(\vec{x})}) \quad (2.7)$$

$p_{s(\vec{x})}$ ist eine Projektionsfunktion, $p_s(y_1, y_2, \dots, y_n) = y_s$, die die Partitionierung des Mosaiks in Abhängigkeit von der gegebenen Indexfunktion s durchführt. Eine solche Partitionierung stellt beispielsweise die Voronoi-Zerlegung dar. Dabei wird jedem Bildpunkt im Mosaik die Bildinformation desjenigen Bildes zugewiesen, dessen Zentrum dem betrachteten Punkt am nächsten liegt. Falls eine zeitliche Ordnung der Bilder gegeben ist, so kann auch diese zur Zerlegung verwendet werden, indem für die Mosaikbildpunkte jeweils das Bild als Basis dient, das die neueste Information enthält. Abb. 2.4 illustriert beide Vorgehensweisen. Durch die Zerlegung des Bildes können an den Grenzen harte Übergänge entstehen, die eine nachträgliche Glättung erfordern.

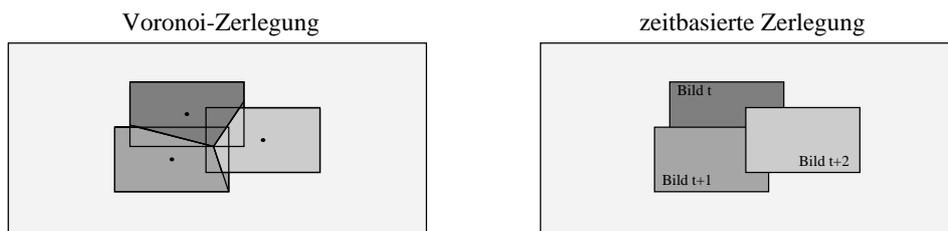


Abbildung 2.4: Regionenbasierte Integrationsverfahren

In [ST99] sind als Methoden zur Integration eines Bildes in ein bestehendes Mosaikbild die Mittelwertbildung und zwei regionenbasierte Verfahren implementiert worden. Im einen Fall wird jeweils die neue Information aus dem Bild übernommen, im anderen die im Mosaikbild enthaltene alte Information beibehalten und nur in bislang nicht repräsentierten Bereichen neue Information eingefügt.

2.2 Bewegte Szenen

Das in den vorangegangenen Abschnitten vorgestellte zweistufige Verfahren zur Erzeugung von Mosaikbildern setzt statische Szenen voraus. Falls die Kamerabewegung nicht der dominierenden Bewegung im Bild entspricht, weil etwa zusätzlich große bewegte Einzelobjekte innerhalb der Szene vorhanden sind, kann die Schätzung der Transformationsparameter fehlschlagen. Falls trotz der Objektbewegungen Parameter ermittelt werden können, so treten in solchen Fällen Integrationsfehler im Mosaikbild auf. In der nachfolgenden Abbildung ist ein Beispiel dafür zu sehen. Links sind zwei Bilder einer Sequenz abgebildet, die eine typische

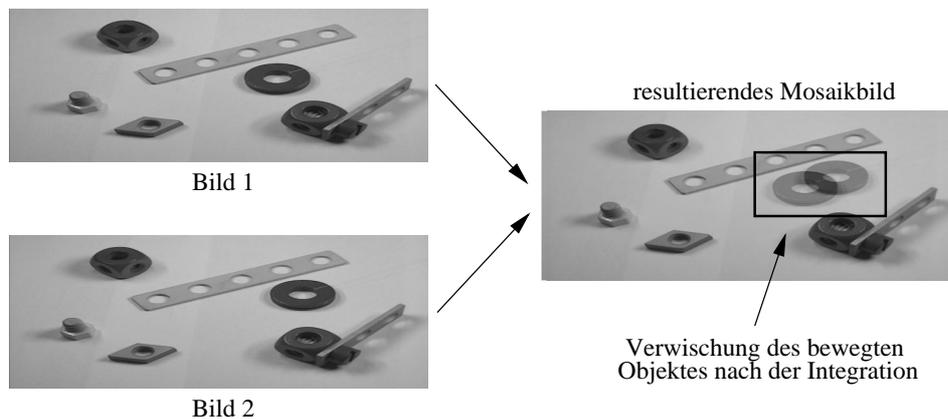


Abbildung 2.5: Integrationsfehler bei bewegten Szenen

Szene aus dem Konstruktionszenario des SFB 360¹ zeigt. Neben einer geringen Kamerabewegung wurde der Ring in der Mitte des Bildes parallel zur Leiste nach rechts verschoben. Das rechts gezeigte Mosaikbild ist aus den beiden Bildern unter Mittelung der Grauwertinformationen entstanden. Es ist deutlich zu erkennen, dass der Ring durch die Transformation der Bilder ins Referenzkoordinatensystem auf zwei unterschiedliche Positionen abgebildet worden ist, da seine Bewegung von dem globalen Bewegungsmodell für die Kamera nicht erfasst wird. Die Konturen des Ringes verschwimmen, so dass beispielsweise eine Objekterkennung kaum mehr möglich ist. Im Nachhinein können ferner allein auf Basis des Mosaikbildes nähere Informationen zur Bewegung des Ringes, wie Richtungs-

¹Situierte Künstliche Kommunikatoren, Universität Bielefeld

oder Geschwindigkeitsdaten, nicht mehr ermittelt werden.

Da in der Regel bei der Verarbeitung von Bildfolgen nicht von statischen Szenen ausgegangen werden kann, sind Verfahren notwendig, die die Verwendung von Mosaikbildern trotz Bewegungen in der Szene zulassen. Im Rahmen des in dieser Arbeit vorgestellten Ansatzes werden die statischen und dynamischen Informationen der Szene während der Generierung der Mosaikbilder voneinander getrennt. Die statischen Informationen werden zu einem Mosaikbild des Szenenhintergrundes zusammengefügt, die bewegten Objekte und ihre Bewegungscharakteristik parallel dazu repräsentiert.

Als Basis für die Trennung dieser Informationen findet zunächst eine Einteilung der Pixel eines Bildes in statische und bewegte Bildpunkte statt. Die bewegten Pixel werden dann zu Regionen zusammengefasst, die zum einen bei der Erstellung des statischen Mosaikbildes ausgeklammert, sowie zum anderen über die Zeit verfolgt werden, um ihre Bewegungsdaten zu erfassen.

Anzumerken ist, dass die Verarbeitung der Bildsequenzen in dem hier vorgestellten Ansatz schritthaltend erfolgen soll. Die Bewegungsdetektion ist grundsätzlich robuster, wenn mehrere Bilder, im Optimalfall die gesamte Bildfolge, zur Verfügung stehen (vergleiche [MS99]). In einem solchen Fall sind gleichzeitig mehrere Abbildungen eines Szenenpunktes in den Bildern gegeben und die seiner Klassifikation zu Grunde liegende Datenbasis ist größer. Die damit verbundene Bedingung, dass alle Bilder zum Zeitpunkt der Bewegungsdetektion vorliegen müssen, stellt allerdings eine große Einschränkung für den Einsatz der Mosaikbilder dar. Eine schritthaltende Verarbeitung lässt sich nicht durchführen. Das schließt interaktive Systeme aus dem Einsatzbereich von Mosaikbildern aus und wird deshalb in der vorliegenden Arbeit nicht näher betrachtet.

Kapitel 3

Bewegungsdetektion

Bewegungen stellen eine wichtige Informationsquelle für die Interpretation von Szenen dar. In der maschinellen Bildverarbeitung lassen sie sich über die Analyse von Bildfolgen detektieren. Aus Grauwertveränderungen zwischen Einzelbildern sind Rückschlüsse auf bewegte Objekte innerhalb einer Szene und die Charakteristiken ihrer Bewegungen möglich. Beim Einsatz aktiver Kameras zur Aufnahme von Bildsequenzen werden in den Bildern Veränderungen auf Grund der Kamerabewegung hervorgerufen, die keiner Objektbewegung entsprechen. Bei der Detektion bewegter Objekte in derartigen Sequenzen ist eine explizite Trennung dieser Veränderungen von den durch reale Bewegungen in der Szene verursachten Veränderungen notwendig.

Die der vorliegenden Arbeit zu Grunde liegende Vorgehensweise ist an die Mosaikbildgenerierung in Kap. 2.1 angelehnt. Der eigentlichen Detektion geht dabei die Schätzung von Parametern für ein Bewegungsmodell zur Umrechnung der Bilder in ein gemeinsames Koordinatensystem voraus. Unter der Annahme, dass die Kamerabewegung innerhalb der Sequenz dominiert, erfasst das Modell ihre Bewegung. Sie wird damit durch Anwendung der Transformationsparameter auf die Bilder kompensiert. Anschließend lassen sich die bewegten Objekte der Szene anhand der in den Bildern verbliebenen Unterschiede detektieren. Falls die Kamerabewegung nicht dominiert, beschreibt das Modell weitere Bewegungen innerhalb der Szene, und eine Extraktion der reinen Kameradynamik ist nicht möglich. In den im Rahmen dieser Diplomarbeit verwendeten Bildfolgen sind die

bewegten Objekte jeweils klein im Verhältnis zum aufgenommenen Ausschnitt, so dass die Voraussetzungen für den gewählten Ansatz erfüllt sind.

Die notwendigen Transformationsparameter werden auf Grund der Parallelen des vorgestellten Ansatzes zur Mosaikbildgenerierung direkt daraus übernommen. Zur anschließenden Detektion der bewegten Bereiche in den Bildern werden im Folgenden drei Ansätze präsentiert und evaluiert. Das in Kap. 3.1 vorgestellte, einfachste Verfahren besteht in der Auswertung der *Intensitätsdifferenzen* zwischen den transformierten Bildern. Hierbei werden nur die pixelweisen Grauwerte betrachtet, ergänzende Informationen über den Verlauf der Bildfunktion bleiben unberücksichtigt. In die Berechnung des aus dem optischen Fluss abgeleiteten *Normal Flow* zur Ermittlung von Bewegungen (Kap. 3.2) geht demgegenüber der lokale Gradient ein. Das dritte Verfahren in Kap. 3.3 setzt im Unterschied zu den ersten beiden eine Kamerabewegung zwischen den Aufnahmen der beiden Bilder zwingend voraus. Die Detektion basiert auf der Auswertung des *Konvergenzverhaltens* von Intensitätsdifferenzen unter Anwendung der zuvor ermittelten Transformationsparameter auf die zu untersuchenden Bilder [BEPR94]. Dazu muss sich die Abbildung deutlich von der Identitätsabbildung unterscheiden, was bei Aufnahmen mit statischer Kamera nicht gegeben ist. In Kapitel 3.4 werden die Verfahren evaluiert und verglichen. Im Vordergrund der Ausführungen stehen dabei zum einen die Wahl geeigneter Werte für die jeweiligen Konfigurationsparameter, sowie zum anderen die Eigenschaften und Unterschiede in der praktischen Anwendung.

Der gewählte Ansatz beschränkt sich auf die ausschließliche Detektion der bewegten Bereiche, eine exakte mathematische Modellierung ihrer Bewegungscharakteristiken, wie beispielsweise in [SA96] vorgeschlagen, wird auf Grund des damit verbundenen Aufwandes nicht angestrebt. Die Extraktion der dynamischen Informationen umfasst lediglich die Berechnung einer Translationsrichtung, wobei angenommen wird, dass sie die Bewegung des Objektes hinreichend genau approximiert (vgl. Kap. 4.3). In Kapitel 3.2.4 wird ein Verfahren beschrieben, das eine Rekonstruktion der Translationsrichtung aus dem Normal Flow ermöglicht. In der vorliegenden Arbeit ergeben sich aus der verwendeten Vorgehensweise zur Bestimmung des Normal Flow allerdings Einschränkungen für seine Anwendbarkeit, die detailliert analysiert werden.

3.1 Intensitätsdifferenz

Die Transformation der zu untersuchenden Bilder unter Anwendung eines Bewegungsmodells, das die Kamerabewegung beschreibt, kompensiert die durch diese in den Bildern verursachten Grauwertveränderungen. Korrespondierende Bildpunkte statischer Szenenbereiche weisen anschließend identische Koordinaten in den Bildern auf, an denen die Intensitätswerte bis auf kleine Abweichungen (Bildrauschen) gleich sind. Für bewegte Regionen trifft dies nicht zu. Die Eigenbewegung der einzelnen Objektpunkte zwischen den Aufnahmen der Bilder führt zu einer Projektion dieser Punkte auf unterschiedliche Bildkoordinaten. Daraus resultieren in der Regel größere Intensitätsdifferenzen in den betroffenen Bereichen, die als Basis für eine Detektion dieser Gebiete dienen.

Zwischen dem Referenzbild I_{ref} und dem neuen Bild I wird gemäß der folgenden Gleichung pixelweise der Absolutwert der Grauwertdifferenz d zwischen allen Bildpunkten (x, y) im Überlappungsbereich der beiden Bilder bestimmt:

$$d(x, y) = |I_{ref}(x, y) - I(x, y)| \quad (3.1)$$

Anhand eines geeigneten Schwellwertes θ_{diff} für die maximal zulässige Intensitätsdifferenz zwischen Punkten, die als statisch eingeordnet werden sollen, erfolgt anschließend die binäre Klassifikation K_I der Pixel des Bildes I als unbewegt oder bewegt relativ zum Referenzbild:

$$K_I(x, y) = \begin{cases} 0, & \text{falls } d(x, y) \leq \theta_{diff}, \\ 1, & \text{sonst.} \end{cases} \quad (3.2)$$

Da hierfür nur die absolute Differenz an einer Stelle im Bild herangezogen wird, ist die Klassifikation sehr sensitiv gegenüber Rauschen. Um dem entgegenzuwirken, werden neben der Differenz des betrachteten Punktes auch die Werte in einer Nachbarschaft N berücksichtigt. Dabei ergibt sich der Detektionswert für jeden Pixel aus dem arithmetischen Mittel aller Differenzen innerhalb der Umgebung:

$$d_N(x, y) = \frac{1}{|N|} \cdot \sum_{(x,y) \in N} d(x, y) = \frac{1}{|N|} \cdot \sum_{(x,y) \in N} |I_{ref}(x, y) - I(x, y)| \quad (3.3)$$

Auf diese Weise wird die fehlerhafte Klassifikation einzelner statischer Pixel vermieden, bei denen auf Grund von Rauschen hohe Intensitätsdifferenzen zwischen den Bildern aufgetreten sind.

3.2 Normal Flow

Grauwertveränderungen zwischen zwei Bildern lassen sich mit Hilfe des *optischen Flusses* quantitativ erfassen. Er ordnet jedem Bildpunkt einen Verschiebungsvektor zu, der dessen Bewegung relativ zum zweiten Bild beschreibt. Das Konzept entstammt nach [Jäh97] der Hydrodynamik, in der Kontinuitätsgleichungen zur Modellierung von Strömungsbewegungen in Flüssigkeiten eingesetzt werden. Sie spezifizieren die Tatsache, dass Teilchen innerhalb einer Flüssigkeit unter Bewegung erhalten bleiben. In analoger Weise lassen sich ähnliche Bedingungen auch für die Grauwerte zwischen Bildern einer Sequenz formulieren. Durch Bewegungen in der Szene während der Aufnahme der Sequenz werden die Projektionen von Objektpunkten in den Bildern verschoben, ihre Grauwerte bleiben aber in der Regel konstant. Die mathematische Ausarbeitung dieser Idee, die im folgenden Abschnitt vorgestellt wird (nach [HS80]), führt auf eine Einschränkungsgleichung, die als Grundlage für eine Berechnung des optischen Flusses dient. Bildbereiche, in denen sein Betrag einen Schwellwert überschreitet, lassen auf Bewegungen innerhalb der Szene schließen, in statischen Bildbereichen geht er gegen Null.

Zur eindeutigen Berechnung des optischen Flusses sind neben der Einschränkungsgleichung weitergehende Bedingungen notwendig und die Verfahren sind aufwändig. Demgegenüber lässt sich die Komponente des optischen Flusses, die parallel zum lokalen Gradienten verläuft, der *Normal Flow*, direkt aus der Bildfunktion ermitteln (Kap. 3.2.2). Die vollständige Richtungsinformation der Bewegung geht zwar verloren, in bewegten Bereichen jedoch wird der Betrag des Normal Flow in der Regel höher sein als in statischen, so dass eine schwellwertbasierte Detektion von Bewegungen möglich wird. Beispiele hierfür finden sich unter anderem in [CM98], [MS99] und [PBA99].

3.2.1 Grundlagen des optischen Flusses

Gegeben sei zunächst eine Bildfunktion $I(x, y, t)$, die jedem Bildpunkt (x, y) zum Zeitpunkt t einen Grauwert zuweist. Durch eine Bewegung innerhalb der Szene wird der Bildpunkt zum nachfolgenden Zeitpunkt $t + \delta t$ unter Erhalt seines

Intensitätswertes auf andere Bildkoordinaten abgebildet:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \quad (3.4)$$

Zu jedem Punkt im Bild I_t ist demnach an der durch den Vektor $(\delta x, \delta y)^T$ spezifizierten Position im Bild $I_{t+\delta t}$ ein korrespondierender Punkt vorhanden. Eine Taylorreihen-Entwicklung der rechten Seite der Gleichung führt, unter Vernachlässigung der Terme höherer Ordnung, im Grenzwert für $\delta t \rightarrow 0$ zur Einschränkungsgleichung:

$$\begin{aligned} I(x, y, t) &= I(x, y, t) + \delta x \frac{\partial I}{\partial x} + \delta y \frac{\partial I}{\partial y} + \delta t \frac{\partial I}{\partial t} + \epsilon \\ \Rightarrow 0 &= \frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = I_x v_x + I_y v_y + I_t \end{aligned} \quad (3.5)$$

Sie bildet die Basis für eine Berechnung des optischen Flusses $\vec{v} = (v_x, v_y)^T$ für jeden Bildpunkt. Im mathematischen Sinn entspricht \vec{v} einem Geschwindigkeitsvektor. Unter der Annahme von einem Zeitschritt zwischen den Bildern lässt er sich jedoch als Verschiebungsvektor interpretieren.

Die durch Gleichung 3.5 gegebene Bedingung legt lediglich eine zum lokalen Gradienten senkrechte Gerade fest (siehe Abb. 3.1), auf der \vec{v} liegen soll. Deshalb ist allein auf dieser Grundlage die eindeutige Berechnung der Verschiebung nicht durchführbar. Eine Möglichkeit zur Formulierung ergänzender Nebenbedingungen sind Glattheitsforderungen, die für benachbarte Punkte ähnliche Verschiebungsvektoren propagieren. Dazu kann beispielsweise die quadratische Summe S^2 über die Abweichungen der Verschiebung $\vec{v} = (v_x, v_y)^T$ eines Punktes gegenüber der mittleren Verschiebung $\bar{v} = (\bar{v}_x, \bar{v}_y)^T$ in seiner Nachbarschaft minimiert werden:

$$S^2 = \|\bar{v} - \vec{v}\|^2 = (\bar{v}_x - v_x)^2 + (\bar{v}_y - v_y)^2 \quad (3.6)$$

Diese Nebenbedingungen werden über einen Gewichtungsfaktor β^2 additiv in die Einschränkungsgleichung einbezogen:

$$0 = \beta^2 S^2 + I_x v_x + I_y v_y + I_t \quad (3.7)$$

Die Bestimmung der pixelweisen Verschiebungen v_x und v_y ist nun durch eine Minimierung der rechten Seite der Gleichung 3.7 möglich. Dies geschieht unter Anwendung iterativer Optimierungsverfahren (Details in [HS80]).

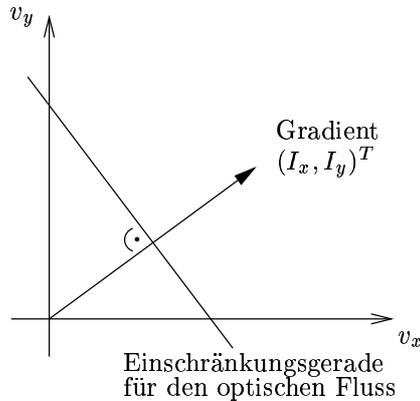


Abbildung 3.1: Visualisierung der Einschränkungsgleichung nach [HS80]

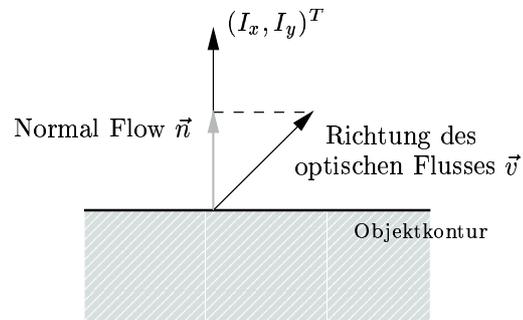


Abbildung 3.2: Gewinnung des Normal Flow durch Projektion

Alternativ zu der vorgestellten Vorgehensweise lassen sich Nebenbedingungen auch in Form eines Bewegungsmodells (vgl. Berechnung des *Perspective Flow* in Kapitel 2.1.1) oder durch Glattheitsforderungen für das aus der Ableitung des Verschiebungsvektorfeldes des optischen Flusses resultierende Geschwindigkeitsfeld festlegen [Sin91].

3.2.2 Theorie des Normal Flow

Im Gegensatz zu den aufwändigen, zur Berechnung des vollständigen optischen Flusses notwendigen Verfahren, ist die Bestimmung der Komponente, die parallel zum lokalen Gradienten verläuft, direkt aus der Bildfunktion möglich. Der *Normal Flow* \vec{n} ergibt sich aus der senkrechten Projektion des lokalen Verschiebungsvektors \vec{v} auf den Gradienten $\nabla I = (I_x, I_y)^T$ (siehe Abb. 3.2):

$$\vec{n} = \frac{\nabla I \cdot \vec{v}}{\|\nabla I\|} \frac{\nabla I}{\|\nabla I\|} \quad (3.8)$$

Der Zähler des obigen Bruches lässt sich durch Einsetzen der zu

$$-I_t = (I_x, I_y) \cdot (v_x, v_y) = \nabla I \cdot \vec{v} \quad (3.9)$$

umgestellten Einschränkungsgleichung 3.5 weiter vereinfachen:

$$\vec{n} = \frac{-I_t}{\|\nabla I\|} \frac{\nabla I}{\|\nabla I\|} \quad (3.10)$$

Eine Detektion von Bewegungen ist nun durch die Auswertung der Norm des Normal Flow

$$\|\vec{n}\| = \frac{|I_t|}{\|\nabla I\|} \quad (3.11)$$

an jedem Bildpunkt möglich. Sie ist umso größer, je besser die durch den optischen Fluss vorgegebene Bewegungsrichtung mit der lokalen Gradientenrichtung übereinstimmt. In Abbildung 3.3 ist der Normal Flow eines sich nach rechts bewegendes, homogenen Objektes skizziert. Durch Wahl eines geeigneten Schwellwertes lassen sich, analog zur Klassifikation in Kapitel 3.1, bewegte Pixel detektieren. Die Berücksichtigung von Bewegungsinformationen aus der Nachbarschaft N eines Pixels erhöht die Robustheit des Verfahrens gegenüber Rauschen. Wie durch die gemittelte Intensitätsdifferenz nach Gleichung 3.3 ist damit die Eliminierung einzelner statischer Pixel mit verrauschten Bewegungsdaten möglich. Die Berechnung des gemittelten Normal Flow Betrags $n(x, y)$ erfolgt anhand der nachstehenden Gleichung, wobei in Abschnitt 3.2.3 ergänzend eine alternative Berechnungsvorschrift vorgestellt wird.

$$n(x, y) = \frac{1}{|N|} \cdot \sum_{(x,y) \in N} \|\vec{n}(x, y)\| \quad (3.12)$$

Grundsätzlich muss bei der Bestimmung des Normal Flow Betrags gemäß Gleichung 3.11 beachtet werden, dass sich die Berechnung an Bildpunkten mit sehr kleinem Gradientenbetrag numerisch instabil verhält. Bei verschwindendem Gradienten ist sie nicht möglich. Aus diesen Gründen wird hier zur Berechnung ein Mindestgradientenbetrag vorausgesetzt. An Pixeln, an denen die festgelegte Schwelle unterschritten wird, resultieren Definitionslücken in der Detektionsfunktion, die geeignet zu behandeln sind. In weitgehend inhomogenen Bildbereichen

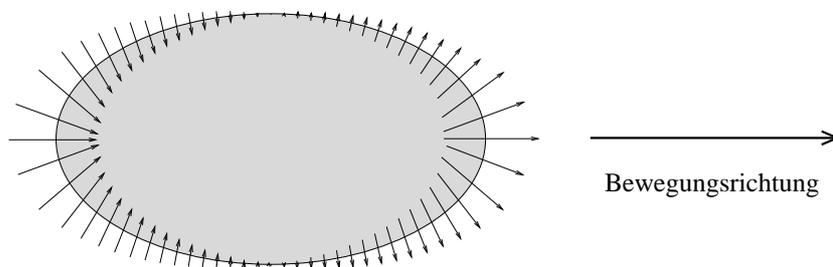


Abbildung 3.3: Visualisierung des Normal Flow am bewegten Objekt

genügt bereits die Mittelung der Bewegungsinformationen in einer Umgebung des jeweils betrachteten Pixels nach Gleichung 3.12, um an einzelnen Pixeln mit einem Gradientenbetrag gleich Null die Singularität auszugleichen. Die Bewegungsinformationen der Nachbarschaft werden dabei auf das aktuelle Pixel übertragen. Größere homogene Bereiche, die die Nachbarschaft überschreiten, erfordern demgegenüber allerdings eine explizite Behandlung in der weiteren Verarbeitung, auf die in Kapitel 4.1 näher eingegangen wird.

Abschließend ist zu berücksichtigen, dass die in die Berechnung der Normal Flow Beträge eingehenden Ableitungen der Bildfunktion lediglich diskret approximiert werden können. Daher stimmen die errechneten Werte für den Normal Flow unter Umständen mit den tatsächlichen Beträgen der Projektionen des optischen Flusses nicht überein. Die Bewegungsdetektion wird dadurch nicht beeinträchtigt (vgl. auch [CM98]), da sich in bewegten Bereichen auf Grund der Intensitätsdifferenzen weiterhin höhere Werte ergeben als in statischen, die in Abschnitt 3.2.4 behandelte Rekonstruktion der Bewegungsrichtung wird allerdings erschwert.

3.2.3 Gewichteter mittlerer Normal Flow

In [MS99] und [IRP94] wird für die Bewegungsdetektion eine alternative Formel zur Berechnung des mittleren Normal Flow innerhalb einer Nachbarschaft N um das jeweils betrachtete Pixel vorgeschlagen. Im Unterschied zu Gleichung 3.12 werden die einzelnen Beträge dabei mit dem quadratischen Betrag des lokalen Gradienten gewichtet:

$$\tilde{n}(x, y) = \frac{1}{\sum_{(x,y) \in N} \|\nabla I(x, y)\|^2 + C} \sum_{(x,y) \in N} \frac{|I_t(x, y)|}{\|\nabla I(x, y)\|} \|\nabla I(x, y)\|^2 \quad (3.13)$$

C ist eine Konstante, die an Stellen mit einem Gradientenbetrag gleich Null numerische Instabilitäten ausgleichen soll.

Das Berechnungsverfahren wird in der vorliegenden Arbeit alternativ implementiert und der ungewichteten Mittelung des Normal Flow gegenübergestellt (Kapitel 3.4). Abweichend von der obigen Formel erfolgt die Behandlung der Singularitäten allerdings, analog zu der im vorangegangenen Abschnitt vorgestellten Heuristik, durch Festlegung eines Mindestbetrags für den Gradienten.

3.2.4 Rekonstruktion der Bewegungsrichtung

Aus der Berechnung des optischen Flusses folgt für jeden Bildpunkt ein Verschiebungsvektor, der die im Rahmen der Extraktion dynamischer Informationen gesuchte Translationsrichtung des Punktes beschreibt. Der Normal Flow spezifiziert lediglich die zum lokalen Gradienten parallele Komponente dieses Vektors. Im Folgenden wird ein Verfahren vorgestellt, das eine Rekonstruktion der vollständigen Translationsrichtung aus dem Normal Flow mehrerer zu einer Region¹ zusammengefasster Pixel ermöglicht. Dabei werden starre Regionen vorausgesetzt, die einer rein translatorischen Bewegung unterliegen, so dass alle Pixel einen identischen Verschiebungsvektor \vec{v} aufweisen. Ein vergleichbarer Ansatz findet sich in [WW88]. Dort wird der Normal Flow allerdings aus der Verschiebung korrespondierender Konturen in den Bildern geschätzt, da den Autoren eine, wie in der vorliegenden Arbeit realisierte, Berechnung auf Basis der Intensitätsdifferenzen zu unsicher erschien. Durch die hierbei auftretenden, in Abschnitt 3.2.2 bereits erwähnten Abweichungen zwischen dem berechneten Normal Flow und den tatsächlichen Projektionsbeträgen sind unter Umständen Fehler in der rekonstruierten Richtung zu erwarten. An die Erläuterung des Verfahrens schließt sich daher eine Analyse der in der Praxis tatsächlich auftretenden Differenzen und ihrer Signifikanzen an, verbunden mit einer Darstellung der daraus folgenden Einschränkungen für die Anwendbarkeit im gegebenen Kontext.

Minimierung der Projektionsdifferenzen

Jedem Punkt (x, y) einer Region mit signifikanten Bewegungsinformationen in seiner Nachbarschaft wird bei der Bewegungsdetektion ein Betrag $n(x, y)$ für den Normal Flow zugeordnet. Dieser geht jeweils aus der senkrechten Projektion des Translationsvektors $\vec{v} = (v_x, v_y)^T$ der Region hervor. \vec{v} lässt sich daher durch Minimierung der pixelweisen quadratischen Differenzen zwischen dem Betrag seiner Projektion $\vec{v} \cdot \nabla \hat{I}$ auf den auf Einheitslänge normierten lokalen Gradienten

¹Die Regionensegmentierung wird in Kapitel 4.1 beschrieben.

$\nabla \hat{I} = (\hat{I}_x, \hat{I}_y)^T$ und dem berechneten Betrag $n(x, y)$ rekonstruieren:

$$\begin{aligned} \vec{v}^* &= \arg \min_{\vec{v}} \sum_{(x,y)} (\vec{v} \cdot \nabla \hat{I}(x, y) - n(x, y))^2 \\ &= \arg \min_{\vec{v}} \sum_{(x,y)} (v_x \cdot \hat{I}_x(x, y) + v_y \cdot \hat{I}_y(x, y) - n(x, y))^2 \end{aligned} \quad (3.14)$$

Zur Lösung des Optimierungsproblems findet die *Methode der kleinsten Quadrate* Anwendung (vgl. [ST99], Berechnung des Perspective Flow). Dabei werden zunächst die partiellen Ableitungen der zu minimierenden Zielfunktion

$$z(v_x, v_y) = \sum_{(x,y)} (v_x \cdot \hat{I}_x(x, y) + v_y \cdot \hat{I}_y(x, y) - n(x, y))^2 \quad (3.15)$$

bezüglich der gesuchten Vektorkomponenten v_x und v_y berechnet:

$$\begin{aligned} \frac{\partial z(x, y)}{\partial v_x} &= \sum_{(x,y)} 2 \cdot (v_x \cdot \hat{I}_x(x, y) + v_y \cdot \hat{I}_y(x, y) - n(x, y)) \cdot \hat{I}_x(x, y) \\ \frac{\partial z(x, y)}{\partial v_y} &= \sum_{(x,y)} 2 \cdot (v_x \cdot \hat{I}_x(x, y) + v_y \cdot \hat{I}_y(x, y) - n(x, y)) \cdot \hat{I}_y(x, y) \end{aligned} \quad (3.16)$$

Aus dem Gleichsetzen mit Null und anschließenden Umformungen resultiert das nachfolgende lineare Gleichungssystem:

$$\begin{pmatrix} \sum_{(x,y)} \hat{I}_x^2 & \sum_{(x,y)} \hat{I}_x \hat{I}_y \\ \sum_{(x,y)} \hat{I}_x \hat{I}_y & \sum_{(x,y)} \hat{I}_y^2 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} \sum_{(x,y)} n \hat{I}_x \\ \sum_{(x,y)} n \hat{I}_y \end{pmatrix} \quad (3.17)$$

Der gesuchte Translationsvektor \vec{v} ergibt sich aus der Lösung des Systems, beispielsweise mit dem Eliminationsverfahren nach Gauß.

Auswertung

In die Berechnung des Normal Flow gehen, wie aus den Gleichungen in Abschnitt 3.2.2 ersichtlich ist, die räumlichen und zeitlichen Ableitungen der Bildfunktion ein. Diese lassen sich in der Praxis lediglich diskret lokal approximieren. Die zeitliche Ableitung I_t wird üblicherweise durch die Intensitätsdifferenz zwischen den betrachteten Bildern I und I_{ref} angenähert (Gleichung 3.18), während für die

räumlichen Ableitungen I_x und I_y beispielsweise der symmetrische Grauwertgradient (Gleichungen 3.19 und 3.20) zu Grunde gelegt wird.

$$I_t(x, y) = I(x, y) - I_{ref} \quad (3.18)$$

$$I_x(x, y) = \frac{1}{2} \cdot (I(x + 1, y) - I(x - 1, y)) \quad (3.19)$$

$$I_y(x, y) = \frac{1}{2} \cdot (I(x, y + 1) - I(x, y - 1)) \quad (3.20)$$

Falls die aus diesen Gleichungen folgenden Ableitungswerte die tatsächlichen lokalen Charakteristiken der untersuchten Bildfunktion nur unzureichend modellieren, sind Fehler in den ermittelten Normal Flow Beträgen die Folge. Insbesondere bei größeren Verschiebungen der Bildfunktion zwischen zwei Bildern und einem nicht-linearen lokalen Verlauf, resultieren unter Umständen *signifikante* Abweichungen. Sie wirken sich auch auf die Richtungsrekonstruktion aus. Die unter Verwendung des im vorangegangenen Abschnitt vorgestellten Verfahrens berechneten Translationsrichtungen weichen in solchen Fällen deutlich von der vorliegenden Bewegungsrichtung ab.

Dies wird auch in den nachfolgend gezeigten Ergebnissen aus der praktischen Anwendung des Verfahrens auf zwei Bildfolgen deutlich. In Abbildung 3.4 ist die rekonstruierte Richtung für das bewegte Objekt aus der Bildfolge "Baufix5" dargestellt. Links ist das Objekt gezeigt, sowie die weiß eingezeichnete, tatsächliche Bewegungsrichtung. In den übrigen drei Bildausschnitten sind die jeweils als bewegt detektierten Bereiche aus aufeinanderfolgenden Bildern visualisiert, sowie die rekonstruierten Translationsrichtungen (graue Pfeile). Das Objekt hat sich zwischen den Bildern nur minimal bewegt. Die approximierten Ableitungswerte scheinen die Funktion zutreffend zu beschreiben, so dass die ermittelte Bewegungsrichtung mit der gegebenen Richtung weitgehend übereinstimmt.

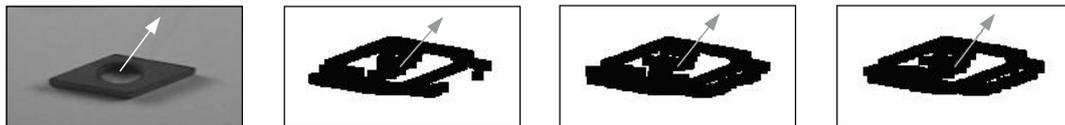


Abbildung 3.4: Erfolgreiche Rekonstruktionen der Bewegungsrichtung

Demgegenüber zeigt Abbildung 3.5 das Objekt aus der Bildfolge "Desk2", bei dem insbesondere im ersten und dritten Detektionsbild deutliche Differenzen zwischen der tatsächlichen und der rekonstruierten Bewegungsrichtung für die segmentierten Regionen zu erkennen sind. Die Bewegungen zwischen den Bildern sind hier relativ zum Objekt größer. Offensichtlich lassen sich die Ableitungen der Funktion nur noch unzureichend diskret approximieren, so dass die resultierenden Normal Flow Beträge signifikant von den Projektionsbeträgen des Verschiebungsvektors \vec{v} abweichen. Ein Rückschluss auf die Bewegungsrichtung des Objektes ist nicht mehr möglich.

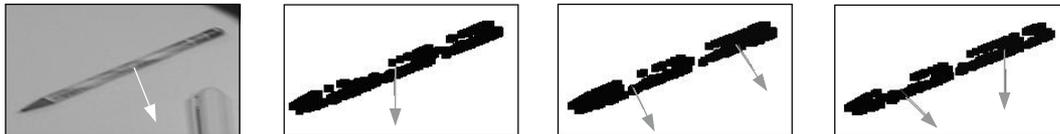


Abbildung 3.5: Instabile Rekonstruktion der Bewegungsrichtung

Zusammenfassend lässt sich daraus folgern, dass das vorgestellte Verfahren für eine zuverlässige Richtungsrekonstruktion korrekte Beträge für den Normal Flow eines bewegten Objektes benötigt. Das hier zur Berechnung eingesetzte, vom lokalen Gradienten und den Intensitätsdifferenzen abhängige Verfahren, gewährleistet dies nicht immer. Insbesondere bei großen Verschiebungen, die innerhalb der zu verarbeitenden Bildfolgen nicht ausgeschlossen werden können, treten Fehler auf und die Rekonstruktion misslingt. Durch den Einsatz des Verfahrens in der vorliegenden Arbeit ist ohne eine Modifikation der Normal Flow Berechnung damit eine robuste Richtungsrekonstruktion im Allgemeinen nicht zu erzielen. Eine Bestimmung des Normal Flow innerhalb einer Auflösungspyramide, in der zu große Bewegungen zunächst kompensiert werden (vgl. Kap. 2.1.1), garantiert weitgehend, dass Bewegungen eine vorgegebene Größenordnung nicht übersteigen und die diskreten Ableitungen die lokalen Charakteristiken der Funktion hinreichend genau beschreiben. Auf Grund des damit jedoch verbundenen, erhöhten Aufwandes, wird stattdessen eine Bestimmung der Translationsrichtung durch die zeitliche Zuordnung und Verfolgung der detektierten Gebiete realisiert, die in Kapitel 4 ausführlich dargestellt wird.

3.3 Konvergenzanalyse

In den beiden bislang vorgestellten Verfahren zur Detektion von bewegten Einzelobjekten innerhalb von Bildfolgen werden die untersuchten Bilder zunächst transformiert und dann anhand der verbleibenden Differenzen bewegte Bereiche detektiert. Ausschlaggebend sind dabei im Wesentlichen die absoluten Intensitätsdifferenzen zwischen den Bildern. Das im Folgenden präsentierte Verfahren nach [BEPR94] abstrahiert von den konkreten Differenzen und untersucht demgegenüber ihr Konvergenzverhalten unter Anwendung der gegebenen Transformationsparameter². An Bildpunkten, die dem globalen Bewegungsmodell unterliegen, d.h. dem statischen Hintergrund zuzurechnen sind, wird sich die Intensitätsdifferenz zwischen den Bildern durch die Transformation verringern, während sie sich an Punkten, die zu bewegten Einzelobjekten gehören, in der Regel konstant verhalten oder vergrößern wird.

Die Analyse des Konvergenzverhaltens m erfolgt durch Berechnung des nachstehenden Ausdrucks für jeden Bildpunkt:

$$m = \frac{|I_t| - |I_t^\omega|}{|I_t| + |I_t^\omega|} \quad \text{mit} \quad I_t^\omega = I_{ref} - I' \quad (3.21)$$

I_t entspricht der zeitlichen Ableitung von I bezüglich I_{ref} , I_t^ω erfasst die Differenzen zwischen dem transformierten Bild I' und dem Referenzbild I_{ref} . Der Wertebereich von m liegt im Intervall $[-1, 1]$, wobei die Werte aus den im Folgenden aufgeschlüsselten Konstellationen resultieren:

$ I_t \gg I_t^\omega $	$m \rightarrow 1$	Differenz nimmt ab, Pixel erscheint statisch
$ I_t \ll I_t^\omega $	$m \rightarrow -1$	Differenz nimmt zu, Pixel ist bewegt
	$m = 0$	Transformation hat keinen Einfluss

An Bildpunkten, an denen zu keinem Zeitpunkt Intensitätsdifferenzen auftreten, d.h. $|I_t| = |I_t^\omega| = 0$, wird m auf 1 gesetzt. Eine Mittelung der Informationen in einer Nachbarschaft führt auch in diesem Verfahren wiederum zu einer größeren

²Diese müssen sich deutlich von der Identität unterscheiden, da sich unter einer Identitätsabbildung keine Veränderungen in den Grauwerten der Bilder ergeben, die zur Detektion in diesem Verfahren zwingend notwendig sind.

Robustheit gegenüber Ausreißern.

Durch Festsetzung eines geeigneten Schwellwertes, der zwischen -1 und 0 liegen sollte, erfolgt abschließend eine binäre Bewegungsklassifikation der Pixel, die eine Trennung bewegter Bereiche vom statischen Hintergrund zulässt.

3.4 Evaluation

Die aus der Anwendung der drei Verfahren resultierenden Bewegungsinformationen weisen unterschiedliche Charakteristiken auf. Diese hängen von den zur Detektion herangezogenen Daten (Gradienteninformationen, Intensitätsdifferenzen, Transformationsparametern, usw.) und den Werten der jeweils gegebenen Konfigurationsparameter ab. Im nachfolgenden Abschnitt werden zunächst die Parameter für jedes Verfahren analysiert. Dazu findet eine Diskussion der Wahl geeigneter Werte statt, sowie die Darstellung der jeweiligen Auswirkungen auf die Detektionsergebnisse. Anschließend erfolgt in Abschnitt 3.4.2 ein direkter, qualitativer Vergleich der Verfahren anhand ausgewählter Beispiele. Schwerpunktmäßig werden dabei die Intensitätsdifferenz und der Normal Flow einander gegenübergestellt, da sie in jedem Fall, d.h. auch bei statischer Kamera, anwendbar sind. Der Vergleich beider Verfahren mit der Konvergenzanalyse erfolgt abschließend durch ein Beispiel, in dem eine ausreichende Kamerabewegung zwischen den Bildern gegeben ist.

3.4.1 Konfigurationsparameter

Jedes der hier behandelten Verfahren lässt sich durch spezifische Konfigurationsparameter gezielt beeinflussen. Die Größe der Nachbarschaft, die als Grundlage zur Berechnung des jeweiligen Bewegungsmaßes dient, und ein Schwellwert für die binäre Klassifikation der Pixel, sind in allen Verfahren festzulegen. Ergänzend erfordert der Normal Flow die Wahl eines Mindestgradientenbetrags, und es besteht die Möglichkeit, zwischen verschiedenen Algorithmen zur Berechnung der lokalen Ableitungen und des mittleren Normal Flow zu wählen (vgl. Kap. 3.2.3).

Intensitätsdifferenz

Die Größe μ der Nachbarschaft, innerhalb der die pixelweisen Differenzen gemittelt werden, beeinflusst die Robustheit des Verfahrens gegenüber einzelnen vertauschten Bildpunkten. Je größer der gewählte Bereich ist, desto besser können Störungen eliminiert werden. Andererseits werden mit wachsender Nachbarschaft die Bewegungsinformationen auch in tatsächlich bewegten Bereichen zunehmend verwischt und die Schärfe der Objektkonturen nimmt ab (Abbildung 3.6 zeigt dies am Beispiel einer sich bewegenden Holzleiste).

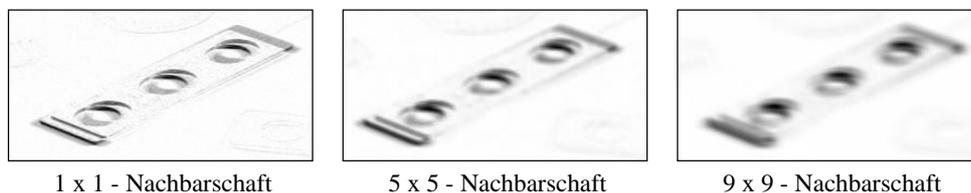


Abbildung 3.6: Einfluss der Nachbarschaftsgröße in der Berechnung der Intensitätsdifferenz (dunkle Bereiche deuten Bewegung an)

Daraus folgt die Notwendigkeit, bei der Wahl der Nachbarschaftsgröße einen Kompromiss zwischen Robustheit auf der einen Seite und Informationsverlust auf der anderen Seite einzugehen. Versuche haben gezeigt, dass für $\mu = 3$ bzw. $\mu = 5$ beiden Aspekten in ausreichendem Maße Rechnung getragen wird. Diese Wahl wird zusätzlich dadurch unterstützt, dass der zur Berechnung notwendige Aufwand quadratisch mit der Nachbarschaftsgröße ansteigt und von einer zu großen Nachbarschaft daher abzuraten ist.

Die Wahl eines geeigneten Schwellwertes für die binäre Klassifikation der Pixel hat entscheidenden Einfluss auf den weiteren Verlauf der Bewegungsanalyse. Insbesondere die Regionenanzahl, die im Rahmen der zeitlichen Zuordnung verarbeitet werden muss, ist ausschlaggebend für den Aufwand. Je niedriger die Klassifikationsschwelle liegt, desto mehr Pixel werden als bewegt klassifiziert und desto größer ist in der Regel die Anzahl der Regionen. Der Schwellwert sollte auf Grund dessen nicht zu niedrig angesetzt werden. Dem steht entgegen, dass Bewegungsinformationen, die unterhalb der Schwelle liegen und damit bei der Klassifikation verloren gehen, nicht rekonstruiert werden können. Bei geringen

Bewegungen kann es dadurch beispielsweise an einzelnen Objektkanten zu einem Verlust wichtiger Informationen für die Erkennung des bewegten Objektes kommen. Aus diesem Grund ist trotz der obigen Gegenargumente ein eher niedriger Schwellwert zu empfehlen, der den Informationsverlust weitgehend begrenzt. Die daraus unter Umständen resultierende, höhere Regionenanzahl wird im Nachhinein durch Anwendung geeigneter Heuristiken ausgeglichen (Kapitel 4.2). Der Wert der Klassifikationsschwelle muss allerdings in jedem Fall über dem in den Bildern gegebenen Rauschniveau liegen. In der Praxis hat sich im Allgemeinen ein Wert im Bereich von 20 bis 40 bewährt.

Normal Flow

Die zur Wahl geeigneter Werte für die Nachbarschaftsgröße in der Normal Flow Berechnung anzuführenden Argumente entsprechen im Wesentlichen denen aus dem vorangegangenen Abschnitt. Für die ungewichtete Mittelung nach Gl. 3.12 haben sich Werte von $\mu = 3$ bzw. $\mu = 5$ als sinnvoll erwiesen. Aus der Anwendung der gewichteten Mittelwertberechnung (Gl. 3.13) resultieren in homogenen Bereichen, in denen die Varianz der innerhalb der Nachbarschaft auftretenden Gradientenbeträge klein ist, keine Unterschiede. Selbst an Objektkanten sind keine Auswirkungen bei den angesprochenen Nachbarschaftsgrößen erkennbar. Erst bei sehr großen Werten ($\mu \geq 9$), wenn der betrachtete Bereich die Kante weiträumig erfasst, ist eine ausreichende Varianz innerhalb der Gradientenbeträge gegeben. In diesem Fall werden die Konturpixel bei der Berechnung höher gewichtet. Die Unterschiede im Ergebnis sind dennoch nicht signifikant, so dass der aus der großen Nachbarschaft folgende, höhere Aufwand eine Anwendung der gewichteten Mittelung im Allgemeinen nicht rechtfertigt.

Auch die Schwellwertfestlegung erfolgt nach denselben Kriterien, die bei der Intensitätsdifferenz zu Grunde gelegt werden. Ein hoher Wert verhindert weitgehend Fehlklassifikationen, führt aber unter Umständen zu einem Informationsverlust, ein niedriger Wert verursacht einen erhöhten Aufwand bei der Korrespondenzanalyse. Werte, die einen geeigneten Kompromiss realisieren, liegen in der Regel zwischen 2 und 4.

Die Behandlung der numerischen Instabilitäten in der Berechnung erfolgt über einen Mindestgradientenbetrag. Bildpunkte, an denen dieser Betrag unterschrit-

ten wird, werden bei der Berechnung ausgeklammert. Je höher die zu setzende Schwelle gewählt wird, desto mehr Definitionslücken sind zu erwarten (wie z.B. auf dem Normal Flow Detektionsbild in Abb. 3.10 zu sehen). Ein niedriger Schwellwert wirkt dem entgegen, wobei sich allerdings der Wertebereich der Funktion vergrößert. An Bildpunkten mit kleinem Gradienten resultieren unverhältnismäßig hohe Detektionswerte. Geeignete Schwellen, die beide Aspekte berücksichtigen, liegen zwischen 2 und 5.

Zur Berechnung des Gradienten werden üblicherweise die Gleichungen 3.19 und 3.20 (symmetrischer Gradient) verwendet. Dabei gehen nur Informationen *eines* Bildes ein. In großen homogenen Bereichen dieses Bildes wird der Mindestgradientenbetrag unter Umständen großflächig unterschritten. Trotz signifikanter Intensitätsdifferenzen zwischen den Bildern, die auf Bewegungen hindeuten, können diese durch den Normal Flow nicht erkannt werden. In Abbildung 3.7 sind links zwei Beispielbilder gezeigt, in denen dieser Fall eintritt. Die Detektion der zweiten Objektposition schlägt auf Grund der verschwindenden Gradientenbeträge im Bild I fehl.

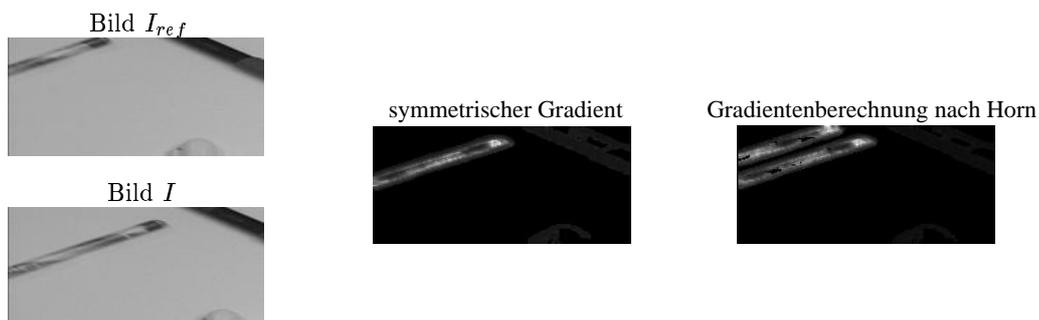


Abbildung 3.7: Vergleich der beiden Verfahren zur Ableitungsberechnung: helle Werte in den Bewegungskarten deuten auf Bewegungen zwischen den Bildern hin, die nur rechts korrekt detektiert werden.

Eine alternative Vorgehensweise zur Bestimmung der räumlichen und zeitlichen Ableitungen nach Horn [HS80] bietet einen Ausweg für eine robustere Erkennung:

$$\begin{aligned}
 I_x(x, y) &= \frac{1}{4}(I_{ref}(x, y+1) - I_{ref}(x, y) + I_{ref}(x+1, y+1) - I_{ref}(x+1, y) \\
 &\quad + I(x, y+1) - I(x, y) + I(x+1, y+1) - I(x+1, y)) \\
 I_y(x, y) &= \frac{1}{4}(I_{ref}(x+1, y) - I_{ref}(x, y) + I_{ref}(x+1, y+1) - I_{ref}(x, y+1) \\
 &\quad + I(x+1, y) - I(x, y) + I(x+1, y+1) - I(x, y+1)) \\
 I_t(x, y) &= \frac{1}{4}(I(x, y) - I_{ref}(x, y) + I(x+1, y) - I_{ref}(x+1, y) \\
 &\quad + I(x, y+1) - I_{ref}(x, y+1) + I(x+1, y+1) - I_{ref}(x+1, y+1))
 \end{aligned}$$

Hier werden *beide* Bilder berücksichtigt, wodurch die Detektion insbesondere in homogenen Bereichen sicherer wird. Das Ergebnis unter Verwendung der oben aufgeführten Gleichungen ist rechts in der Abbildung gezeigt.

Konvergenzanalyse

Die Resultate der Konvergenzanalyse hängen zum einen von der gewählten Nachbarschaftsgröße ab, sowie zum anderen vom festgelegten Klassifikationsschwellwert. Analog zu den beiden vorher beschriebenen Verfahren sollte die Nachbarschaft nicht zu groß sein. In diesem Verfahren treten die Auswirkungen der Glättung jedoch deutlicher hervor. Die pixelweise Berechnung des Konvergenzmaßes ergibt im gesamten Bild sehr verrauschte Werte, während sich in den anderen Verfahren signifikante Detektionswerte hauptsächlich auf Objektkanten und bewegte Bereiche in den Bildern konzentrieren (siehe nachfolgender Abschnitt). Eine Vergrößerung der Nachbarschaft erhöht damit die Robustheit der Detektion erheblich, woraus $\mu = 5$ oder $\mu = 7$ als geeignete Größen folgen. In Abbildung 3.8 sind in der oberen Zeile exemplarisch Detektionsergebnisse gezeigt, die jeweils mit variierenden Nachbarschaften erzielt wurden.

Ein geeigneter Schwellwert für die binäre Klassifikation der Pixel sollte, auf Grund des in Abschnitt 3.3 skizzierten Wertebereichs der Funktion, zwischen -1 und 0 liegen. Dadurch ist in der Regel gewährleistet, dass überwiegend Pixel als bewegt erkannt werden, die tatsächlich einer Eigenbewegung unterliegen. In der praktischen Anwendung haben sich Werte zwischen -0.5 und -0.3 als geeignet erwiesen, bei denen Fehldetektionen vermieden werden und trotzdem kaum Bewegungsinformation verloren geht. Die drei unteren Bilder in Abbildung 3.8 zeigen die

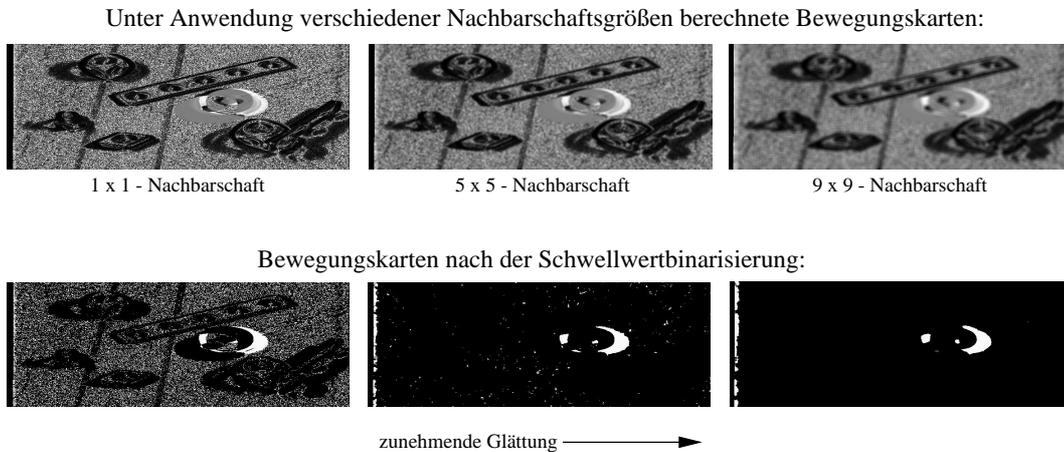


Abbildung 3.8: Einfluss der Nachbarschaftsgröße in der Konvergenzanalyse: Helle Bereiche deuten auf Bewegungen hin, graue und dunkle Bereiche charakterisieren vorrangig den Szenenhintergrund.

Ergebnisse einer Binarisierung der Bewegungsbilder mit einem Schwellwert von -0.3 . Mit zunehmender Nachbarschaftsgröße sind deutlich die Auswirkungen der Glättung auf die Detektionsresultate zu erkennen. Die als bewegt klassifizierten Pixel beschränken sich auf das bewegte Objekt, wobei im rechten Bild auch der bei großen Nachbarschaften auftretende Informationsverlust zu beobachten ist.

3.4.2 Direkter Vergleich der Verfahren

Nachfolgend werden die Detektionsergebnisse der unterschiedlichen Verfahren direkt miteinander verglichen. Dabei stehen zunächst die Intensitätsdifferenz und der Normal Flow im Vordergrund, ein Vergleich mit der Konvergenzanalyse findet am Ende des Abschnitts statt.

Die Intensitätsdifferenz legt zur Detektion bewegter Bereiche in den Bildern lediglich die absolute Grauwertdifferenz zu Grunde. Der Normal Flow setzt diese hingegen zum lokalen Gradienten in Beziehung. Dadurch wird die Tatsache berücksichtigt, dass in Bereichen mit großem Gradientenbetrag hohe Grauwertdifferenzen nicht notwendigerweise auf große Bewegungen schließen lassen. Der Normal Flow reagiert daher weniger sensitiv auf Differenzen an Bildpunkten mit

großen Gradientenbeträgen, die hauptsächlich an Objektkonturen zu finden sind. Auf der einen Seite führt dies zu einer größeren Robustheit, andererseits ist damit ein Informationsverlust verbunden. Abbildung 3.9 zeigt ein entsprechendes Beispiel.

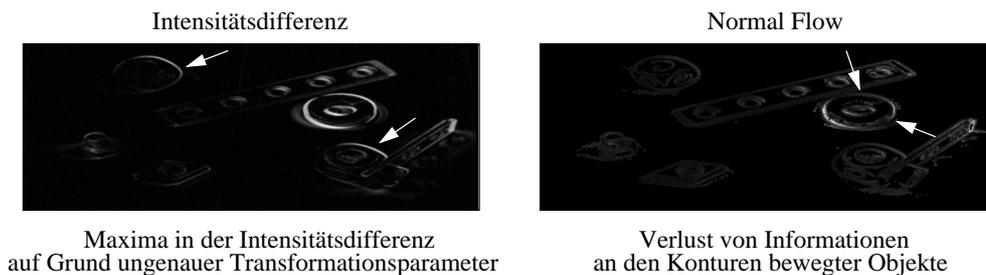


Abbildung 3.9: Robustheit der Detektionsverfahren: Links sind auch an statischen Objekten helle Maximalwerte sichtbar (Pfeile), rechts resultieren an Objektkanten grundsätzlich kleinere Werte.

Als Grundlage dienen zwei Bilder der Folge "Baufix2", zwischen denen Transformationsparameter geschätzt werden. Sie sind in diesem Fall ungenau, so dass sich auch an den Kanten statischer Objekte hohe Intensitätsdifferenzen ergeben. Im linken Bild ist das Resultat der Berechnung der Intensitätsdifferenz gezeigt, wobei die hohen Detektionswerte an den Kanten der statischen Objekte deutlich hervortreten. Sie ziehen Fehlklassifikationen nach sich. Der Normal Flow, dessen Ergebnis rechts abgebildet ist, reagiert demgegenüber weniger sensitiv auf diese Störungen. Die Detektionsmaxima konzentrieren sich hier im Wesentlichen auf den tatsächlich bewegten Ring. Zu beachten ist allerdings, dass auch an bewegten Objekten die Kanten nur unvollständig detektiert werden, wodurch wichtige Informationen für eine Rekonstruktion verloren gehen.

Ein weiterer Unterschied in den Detektionsergebnissen der beiden Verfahren ist durch die Definitionslücken innerhalb der Normal Flow Berechnung gegeben. Die Bewegungsinformation des Normal Flow ist weitaus weniger flächenhaft als die Information aus der Berechnung der Intensitätsdifferenz. Das Beispiel in Abbildung 3.10 illustriert dies. Während zwischen den beiden Bildern (Bildfolge "Office1") keine Kamerabewegung stattgefunden hat, variiert die Position der abgebildeten

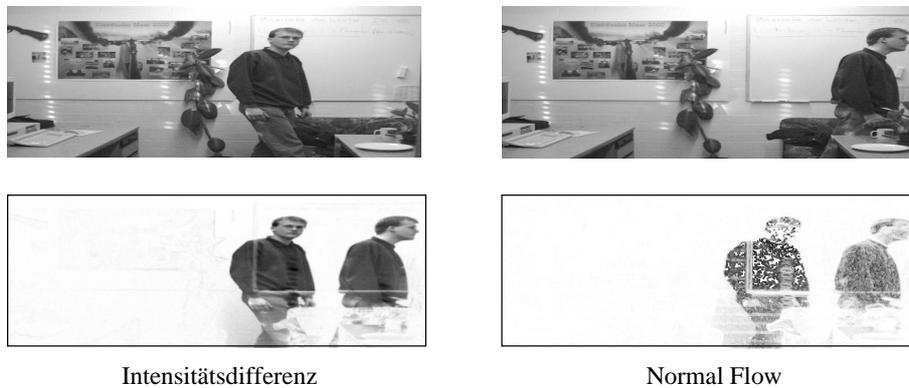


Abbildung 3.10: Definitionslücken in den Bewegungskarten des Normal Flow
(dunkle Bereiche zeigen Bewegungen zwischen den Bildern an)

Person deutlich. Links ist das Ergebnis der Intensitätsdifferenz zu sehen. Die bewegten Bereiche sind jeweils vollständig erkannt worden. Auf der rechten Seite sind demgegenüber die Auswirkungen der angesprochenen Definitionslücken des Normal Flow zu sehen, die keine flächenhafte Detektion der Bereiche zulassen. Grundsätzlich sind beide Verfahren zur Erkennung von Bewegungen auf Grauwertdifferenzen zwischen den Bildern angewiesen. In Bereichen, in denen diese trotz Bewegung nicht gegeben sind, ist eine Detektion nicht möglich. Verschwindende Differenzen sind entweder auf zu geringe Kontraste zwischen dem bewegten Objekt und dem jeweiligen Hintergrund zurückzuführen, oder sie resultieren aus einer geringen Verschiebung eines weitgehend homogenen Objektes, bei der sich die Objektpositionen in den Bildern überlagern (Abb. 3.11). Beide Ursachen verhindern eine vollständige Rekonstruktion des bewegten Objektes und erfordern weitergehende Maßnahmen.



Abbildung 3.11: Unvollständige Detektion bei fehlenden Grauwertdifferenzen

Die in Abschnitt 3.3 präsentierte Konvergenzanalyse setzt zur Detektion von Bewegungen eine deutliche Kamerabewegung zwischen den Bildern voraus. In interaktiven Systemen, in denen eine schritthaltende Verarbeitung stattfinden soll und die Kameradynamik variiert, ist diese Bedingung nicht immer gegeben. Sie kommen daher als Einsatzbereich für das Verfahren nicht in Frage. Denkbar ist allerdings, das Verfahren ergänzend zu den beiden anderen einzusetzen, sobald eine ausreichend große Kamerabewegung festgestellt worden ist (beispielsweise anhand der Transformationsparameter). Dadurch könnten robustere Detektionsergebnisse erzielt werden. Dieser Ansatz wird hier nicht weiter verfolgt, die Konvergenzanalyse soll aber dennoch mit den beiden im vorangegangenen Abschnitt diskutierten Verfahren verglichen werden. Die Ausführungen können als Ausgangspunkt für weitere Untersuchungen in der aufgezeigten Richtung dienen.

Der Vergleich beruht auf dem in Abbildung 3.12 gezeigten Detektionsbeispiel. Die zwei betrachteten Bilder I und I_{ref} entstammen der Bildfolge "Baufix4", wobei eine hinreichende Kamerabewegung zwischen den Bildern vorhanden ist, um eine Anwendung der Konvergenzanalyse zu rechtfertigen. Das Bild I' ist durch Anwendung der errechneten Transformationsparameter auf das Bild I entstanden und bildet zusammen mit den Originalbildern die Basis für die Bewegungsdetektion, deren Ergebnisse in der rechten Spalte der Abbildung gezeigt sind.

Der Normal Flow und die Intensitätsdifferenz detektieren die sich ins Bild bewegende Leiste im Rahmen ihrer jeweiligen Möglichkeiten vollständig. Im übrigen Bild treten darüberhinaus keine weiteren signifikanten Detektionswerte auf. Die Konvergenzanalyse liefert demgegenüber im gesamten Bild Bewegungsinformationen. Dunkle Bereiche entsprechen Gebieten, in denen die Intensitätsdifferenz unter Anwendung der Transformationsparameter abgenommen hat. Insbesondere an den unbewegten Objekten der Szene, deren Positionen in den Bildern durch die Kamerabewegung zuvor verschieden waren, und die erst durch die Transformation deckungsgleich verschoben worden sind, ist dies zu sehen. Helle Bereiche zeigen einen Anstieg der Differenzen unter der Transformation an, wie sie im Bereich der Leiste am unteren Bildrand auftreten. Die Bewegungsdaten sind insgesamt mit der flächigen Information der Intensitätsdifferenz zu vergleichen.

Im Gegensatz zu den beiden anderen Verfahren lässt die Konvergenzanalyse in diesem Fall keinen Rückschluss auf das vollständige Objekt zu. Im mittleren Be-

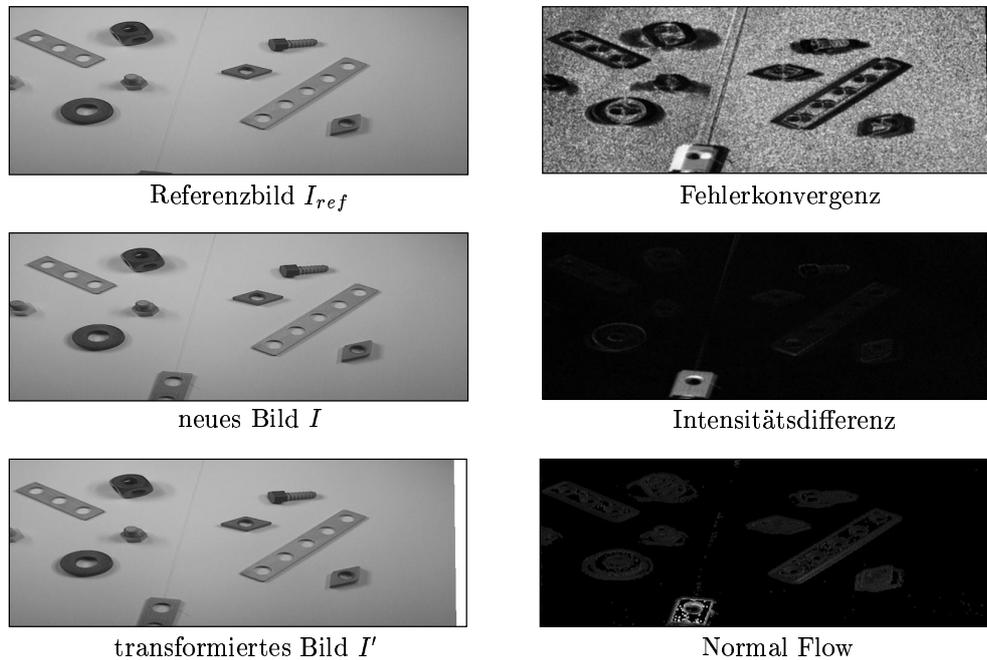


Abbildung 3.12: Direkter Vergleich der drei Detektionsverfahren

reich der Leiste hat sich die Grauwertdifferenz zwischen den Bildern unter Anwendung der Transformationsparameter nicht verändert, so dass keine Aussagen über die Bewegungscharakteristiken der Bildpunkte getroffen werden können. Eine Erkennung des kompletten Objektes ist nur möglich, falls die aus den Kamerabewegungen resultierenden Objektpositionen in den Einzelbildern disjunkt sind, die Größenordnung der Kamerabewegung also die Objektgröße übersteigt. Diese Voraussetzung ist hier nicht erfüllt.

3.4.3 Fazit

Alle drei in den vorangegangenen Abschnitten diskutierten Verfahren eignen sich für eine Detektion von Bewegungen zwischen Bildern, sofern die jeweils notwendigen Voraussetzungen gegeben sind. Aus den Grauwertveränderungen sind Rückschlüsse auf Bewegungen innerhalb der Szene möglich. Die Erkennung homogener Objekte, deren Bewegungen in Teilbereichen unter Umständen keine Intensitätsveränderungen bewirken, stellt dabei ein Problem dar. In den meisten Fällen

ist mit den hier vorgestellten Verfahren eine direkte vollständige Detektion der Objekte nicht realisierbar. Über die Gruppierung sich parallel bewegender Regionen, die unter Umständen auf ein gemeinsames Bewegungsereignis zurückzuführen sind, oder durch Nutzung ergänzender Kontur- oder Segmentierungsinformationen, kann aber im weiteren Verlauf der Analyse gegebenenfalls eine Rekonstruktion des bewegten Objektes gelingen. Dieser Aspekt wird in der vorliegenden Arbeit jedoch nicht weiter betrachtet.

Die von den Verfahren gelieferten Daten sind pixelorientiert. Um die in den folgenden Verarbeitungsschritten vorausgesetzten, regionenhaften Bewegungsdaten zu generieren, ist eine Nachbearbeitung der Bewegungsbilder notwendig, die durch Anwendung morphologischer Operatoren umgesetzt wird (Kapitel 4.1).

Der Einsatzbereich der Konvergenzanalyse ist durch die zwingend erforderliche Kamerabewegung am weitesten eingeschränkt. Sofern diese vorhanden ist, sind die Ergebnisse jedoch mit denen der anderen Verfahren vergleichbar. Ohne vorherige Kenntnis der Kameradynamik ist nur der Einsatz eines der anderen Verfahren sinnvoll. Der Normal Flow zeichnet sich dabei durch eine erhöhte Robustheit gegenüber ungenauen Transformationsparametern aus. Allerdings ist der gleichermaßen auftretende Informationsverlust an den Konturen bewegter Objekte durch eine geeignete Nachbearbeitung auszugleichen, ebenso wie die unvermeidlichen Detektionslücken auf Grund verschwindender Gradientenbeträge geeignet zu schließen sind. Die Intensitätsdifferenz liefert von vornherein flächigere Bewegungsinformationen für die Bilder.

Die Detektionsergebnisse aller Verfahren sind abhängig von den Konfigurationsparametern. Zu niedrig oder zu hoch eingestellte Werte können zu unvollständigen oder fehlerhaften Bewegungsdaten führen. Insbesondere aus zu niedrig gesetzten Klassifikationsschwellwerten folgen unter Umständen sehr viele, zum Teil fehlklassifizierte Regionen, die im Rahmen der nachfolgenden Korrespondenzanalyse identifiziert und ausgeschlossen werden müssen. Bei der schritthaltenden Verarbeitung von Bildfolgen, deren Charakteristik sich über die Zeit verändert, ist eine kontinuierliche Adaption der Werte erforderlich. Dazu sind Heuristiken notwendig, die eine automatische Schwellwertfestlegung auf Basis der gegebenen Bilder ermöglichen. Ihre Diskussion übersteigt allerdings den Rahmen dieser Diplomarbeit.

Kapitel 4

Regionentracking

Die vollständige Repräsentation einer gegebenen Szene umfasst statische und dynamische Informationen über die enthaltenen Objekte. Während zur Analyse der statischen Daten einzelne Bilder eine hinreichende Grundlage darstellen, sind Bewegungen nur durch die Verarbeitung von Bildfolgen zu erfassen. Im vorangegangenen Kapitel wurden Verfahren vorgestellt, die eine Detektion bewegter Pixel innerhalb von Bildern einer gegebenen Sequenz ermöglichen. Die Bewegungscharakteristiken der Punkte werden dabei nicht erfasst, so dass zu deren Extraktion weitere Schritte notwendig sind. In der vorliegenden Arbeit erfolgt die Gewinnung der vollständigen dynamischen Daten durch die Beobachtung der Bewegungen über die Zeit. Das hierzu eingesetzte Verfahren wird in diesem Kapitel vorgestellt.

Den Ausgangspunkt bilden bewegte Regionen in den einzelnen Bildern. Sie resultieren aus morphologischen Operationen auf den von den Detektionsverfahren gelieferten, pixelorientierten Bewegungskarten (Kapitel 4.1). Im Anschluss daran wird zwischen aufeinanderfolgenden Bildern der betrachteten Sequenz nach korrespondierenden Regionen gesucht. Die Zuordnung erfolgt dabei über den Vergleich der Regionen auf Basis ausgewählter Merkmale. Diese werden in Abschnitt 4.2 beschrieben, gemeinsam mit zwei alternativen Ähnlichkeitsmaßen.

Für eine optimale Zuordnung der korrespondierenden Regionen aus zwei gegebenen Bildern ist der paarweise Vergleich aller detektierten Regionen notwendig. Der damit verbundene Aufwand wächst jedoch quadratisch mit der Anzahl der Regio-

nen, so dass diese Vorgehensweise in der Praxis im Allgemeinen nicht realisierbar ist. Aus diesem Grund werden die Regionen zunächst in *Zusammenhangskomponenten* gruppiert (Kap. 4.2.3), die sich darüberhinaus auch zum Ausgleich von Varianzen in den Segmentierungsergebnissen eignen. Die Anzahl resultierender Komponenten für ein Bild liegt in der Regel deutlich unter der Anzahl enthaltener Regionen, so dass sie mit weniger Aufwand paarweise miteinander verglichen werden können. Die Komplexität wird dabei durch die Festlegung einer Maximaldistanz für den jeweiligen Abstand der Schwerpunkte zusätzlich reduziert. Komponenten, deren Schwerpunkte zu weit voneinander entfernt sind, werden a-priori als nicht korrespondierend angenommen und nicht verglichen.

In Einzelfällen, die in Abschnitt 4.2.3 detailliert beschrieben werden, ist allein mit der skizzierten Vorgehensweise eine robuste Verfolgung bewegter Bereiche nicht möglich. Komponenten, die als Ganzes nicht zugeordnet werden können, werden daher in einem nachfolgenden Schritt in *Teilzusammenhangskomponenten* zerlegt. Zwischen diesen wird anschließend nach weiteren Korrespondenzen gesucht. Alle gefundenen Zuordnungen werden in einem *Korrespondenzgraphen* verwaltet, der in Abschnitt 4.2.4 erläutert wird. Parallel zum Aufbau dieses Graphen werden für die verfolgten Komponenten Trajektorien generiert (Kapitel 4.3). Sie beschreiben ihre Bewegungscharakteristiken, spezifiziert durch die Verschiebungen der Schwerpunkte, und bilden damit die Ausgangsbasis für eine Analyse der dynamischen Informationen der verarbeiteten Bildfolge.

4.1 Regionensegmentierung

Die aus der Bewegungsdetektion resultierenden Daten sind pixelorientiert. In den durch die Schwellwertbinarisierung erzeugten, binären Bewegungskarten sind daher im Allgemeinen viele "zerfranste" und lückenhafte Gebiete (Abb. 4.1) enthalten. Da in der weiteren Verarbeitung kompaktere Regionen vorteilhaft sind, werden zunächst in einem Vorverarbeitungsschritt *morphologische Operationen* auf die Karten angewendet.



Abbildung 4.1: "Ausgefrante" und lückenhafte Gebiete in einer Bewegungskarte

4.1.1 Morphologische Operationen

Morphologische Operationen verändern die Form von Objekten. In der vorliegenden Arbeit werden lediglich die zwei Grundoperationen der *Erosion* und *Dilatation* betrachtet, so dass an dieser Stelle auf eine fundierte theoretische Herleitung verzichtet werden soll. Stattdessen beschränkt sich die Darstellung auf die Erläuterung der beiden angesprochenen Operationen, sowie die hier gewählte Realisierung durch *Rangordnungsoperatoren*. Weitere Details sind [Nie81] oder [Jäh97] zu entnehmen.

Ausgangspunkt für die Definition von Rangordnungsoperatoren, die pixelweise auf ein Bild angewendet werden, ist die Rangordnung $R(x, y)$ der Grauwerte r_i in einer festgelegten Nachbarschaft $N(x, y)$ des aktuellen Pixels (x, y) :

$$R(x, y) = \{r_1 \leq r_2 \leq \dots \leq r_{M-1} \leq r_M \mid r_i \in V(x, y)\} \quad (4.1)$$

$V(x, y)$ entspricht der M -elementigen Menge der Grauwerte in N :

$$V(x, y) = \{I(i, j) \mid (i, j) \in N(x, y)\}$$

Eine Rangordnungsoperation stellt eine Funktion f dar, durch die jeweils ein Wert $h(x, y)$ aus R als neuer Wert des betrachteten Pixels ausgewählt wird:

$$h(x, y) = f(R(x, y)) \quad (4.2)$$

Die Erosion und die Dilatation werden auf dieser Basis wie folgt definiert:

$$h(x, y) = r_1 \quad \text{Erosion} \quad (4.3)$$

$$h(x, y) = r_M \quad \text{Dilatation} \quad (4.4)$$

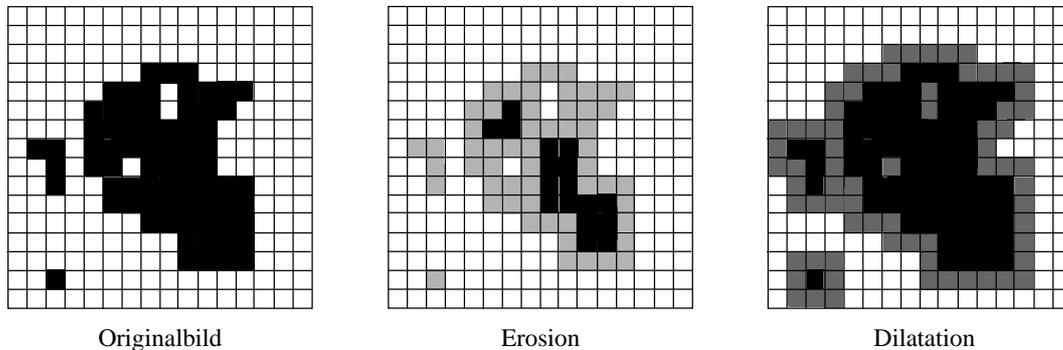


Abbildung 4.2: Morphologische Grundoperationen (angelehnt an [Jäh97])

Bei der Erosion wird jeweils der minimale Grauwert innerhalb der betrachteten Nachbarschaft als neuer Wert übernommen, bei der Dilatation der maximale Wert. In Binärbildern treten als mögliche Werte für die r_i lediglich 0 und 1 auf. Im Rahmen einer Erosion wird damit jeder Objektpunkt, in dessen Nachbarschaft mindestens ein Pixel den Wert 0 aufweist, eliminiert. Nur Bildpunkte, die vollständig im Inneren des Objektes liegen, bleiben erhalten. In Abbildung 4.2 sind in der mittleren Skizze die Auswirkungen einer Erosion (Nachbarschaftsgröße $\mu = 3$) auf das links gezeigte Originalobjekt dargestellt. Die hell gezeichneten Pixel werden durch die Erosion entfernt. Damit können Störungen, wie beispielsweise der einzelne Pixel unten links, beseitigt werden. Gleichzeitig werden Löcher innerhalb von Objekten vergrößert und die Objektfläche insgesamt nimmt ab. Eine Dilatation erzeugt gegenteilige Effekte. Hierbei werden alle Bildpunkte dem Objekt zugefügt, in deren Nachbarschaft mindestens ein Pixel den Wert 1 aufweist. Dadurch werden Lücken geschlossen, eng benachbarte Regionen zusammengefasst und Objektkonturen geglättet. Rechts in Abbildung 4.2 ist dies zu sehen (zugefügte Punkte sind dunkelgrau gezeichnet). Allerdings ist mit der Dilatation auch ein Wachstum der Objekte verbunden, durch das unter anderem Störungen verstärkt werden.

In der Praxis werden die vorgestellten Operationen in der Regel miteinander kombiniert, um die angesprochenen Größenveränderungen der bearbeiteten Objekte zu vermeiden. Durch die Anwendung einer Dilatation auf ein erodiertes Bild (*Opening*) werden den Objekten in kompakten Teilbereichen zuvor entfernte Randpixel wieder zugefügt. Die Objektgrößen bleiben dort nahezu konstant,

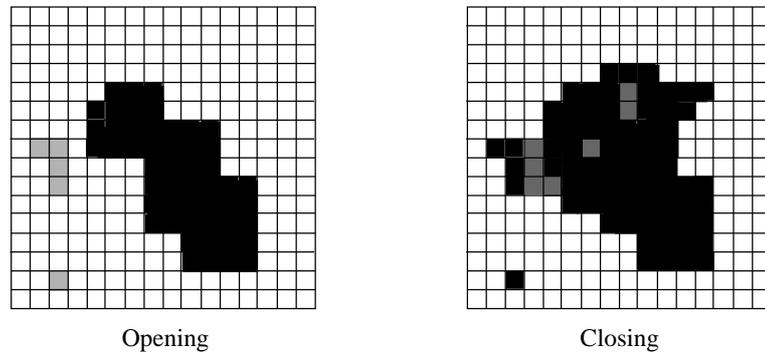


Abbildung 4.3: Opening und Closing

während in den übrigen Bereichen eine Ausdünnung des Objektes stattfindet. Die Erosion eines Bildes, das zuvor durch eine Dilatation bearbeitet wurde (*Closing*), führt zu einer rückwirkenden Verkleinerung der durch die Dilatation vergrößerten Objekte, wobei Löcher im Objektinneren geschlossen werden. Im ersten Schritt zusammengefasste Regionen bleiben dabei als Einheit erhalten. In Abbildung 4.3 sind die Ergebnisse der Opening- und Closing-Operationen für das Objekt aus Abbildung 4.2 gezeigt.

Anhand der vorstehenden Ausführungen erscheint ein Closing-Operator für die Bearbeitung der binären Bewegungskarten am besten geeignet. Durch seine Anwendung gehen keine Informationen verloren und die in den Regionen vorhandenen Löcher können geschlossen werden. Im Hinblick auf die in Kapitel 3.4 diskutierten Probleme bei der Detektion von Objektkonturen unter Verwendung des Normal Flow, und auf Grund der allgemeinen Zielsetzung, bewegte Objekte möglichst vollständig zu detektieren, wird dennoch die Dilatation zur Bearbeitung der Bewegungskarten ausgewählt. Das damit verbundene Wachstum der bewegten Regionen ermöglicht, auch bei Einsatz der Intensitätsdifferenz oder Konvergenzanalyse, eine robustere Konturdetektion und bietet damit eine bessere Grundlage für eine nachfolgende Objekterkennung. Die gleichzeitige Verstärkung von Störungen, die sich bei der Dilatation nicht vermeiden lässt, wird durch die Festlegung einer Mindestgröße für signifikante Regionen ausgeglichen. Unter der Annahme, dass aus verrauschten Daten resultierende Regionen in der Regel weitaus kleiner sind als tatsächlich bewegte Regionen, ist damit eine Begrenzung der Anzahl zu

untersuchender Regionen möglich. Es muss jedoch beachtet werden, dass die skizzierte Vorgehensweise die Verarbeitung sehr kleiner Objekte nicht gestattet, da sie nur schwer von Detektionsstörungen zu unterscheiden sind. In der nachstehenden Abbildung sind abschließend die Ergebnisse der Dilatation für zwei Bewegungskarten gezeigt. Als geeignete Nachbarschaftsgrößen haben sich im vorliegenden Kontext $\mu = 5$ bzw. $\mu = 7$ erwiesen.

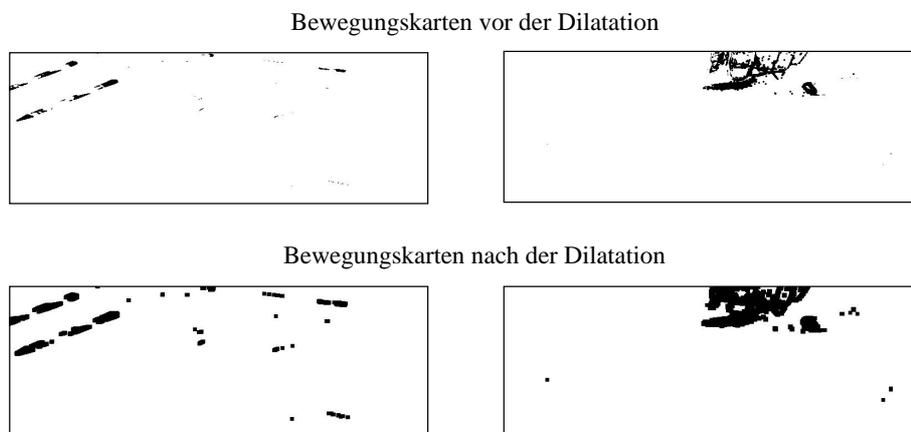


Abbildung 4.4: Auswirkungen der Dilatation

4.2 Zeitliche Korrespondenzanalyse

Aus der Anwendung der Dilatation auf die Bewegungskarten resultieren weitgehend kompakte Regionen, die bewegten Objekten oder Teilen dieser entsprechen. Für die angestrebte Zuordnung korrespondierender Regionen bzw. Zusammenhangskomponenten in aufeinanderfolgenden Bildern der Bildfolge werden diese miteinander verglichen. Die Grundlage hierfür bilden die im Folgenden vorgestellten Merkmale, die zunächst für jede Region berechnet werden. Anhand eines geeigneten Abstandsmaßes wird anschließend über die Ähnlichkeit zweier Regionen entschieden. Die Merkmale sind dabei so gewählt, dass sie sich leicht auf die angesprochenen Zusammenhangskomponenten übertragen lassen und damit für deren Vergleich dieselben Heuristiken angewendet werden können (Kap. 4.2.3).

4.2.1 Merkmale

Zur Charakterisierung von Regionen sind in der Literatur ([Jäh97], [Nie81]) eine Vielzahl von Merkmalen zu finden. Häufig kommen Texturmaße oder Kennzahlen, die die Regionenform spezifizieren, zum Einsatz. Die Eigenschaften der Regionen im vorliegenden Kontext lassen sich durch diese Merkmale nur unzureichend erfassen. Da sie nicht durch Anwendung festgelegter Segmentierungskriterien aus Einzelbildern entstanden sind, sondern auf der Basis von Grauwertdifferenzen zwischen zwei Bildern generiert wurden, weisen sie im Allgemeinen weder eine einheitliche Textur auf noch sind sie homogen. Innerhalb ihrer Intensitätswerte liegt vielmehr häufig eine große Varianz vor. Diese hängt zum einen von den spezifischen Objekteigenschaften ab. Zum anderen resultiert sie in vielen Fällen aus der Tatsache, dass die Regionen oft an den Rändern bewegter Objekte zu finden sind und somit sowohl Objekt- als auch Hintergrundpixel in ihnen zusammengefasst werden. Die Segmentierungsergebnisse in den Bildern können weiterhin signifikant variieren, und die eingesetzte Dilatation wirkt bei den zu Grunde gelegten Nachbarschaftsgrößen stark formverändernd, so dass sich auch Formfaktoren als Ausgangsbasis für die Suche nach Korrespondenzen nicht eignen.

Die Regionen werden auf Grund der vorstehenden Ausführungen durch die drei folgenden Merkmale beschrieben, die den Varianzen innerhalb der Grauwerte Rechnung tragen und gegenüber Formveränderungen weitgehend invariant sind. Im Einzelnen trifft dies auf das *Histogramm* der Grauwerte, den *Regionenschwerpunkt* und die *Pixelanzahl* zu. Der Schwerpunkt $s = (s_x, s_y)$ einer Region wird auf Basis von Momenten m_{pq} berechnet [Nie81]:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (4.5)$$

Sie spezifizieren die Verteilung der Funktionswerte der untersuchten Funktion f . In der Analyse diskreter Bildfunktionen I werden die aus Gleichung 4.5 abgeleiteten, diskreten Momente verwendet:

$$m_{pq} = \sum_x \sum_y I(x, y) x^p y^q \quad (4.6)$$

Aus ihnen lässt sich der Schwerpunkt einer Region unter Berücksichtigung aller zugehörigen Pixel ermitteln:

$$s_x = \frac{m_{10}}{m_{00}} \quad , \quad s_y = \frac{m_{01}}{m_{00}} \quad (4.7)$$

Das Grauwerthistogramm dient als Grundlage für den Vergleich zweier Regionen. Anhand der im nächsten Abschnitt vorgestellten Abstandsmaße für Histogramme lassen sich korrespondierende Regionen mit ähnlichen Intensitätsverteilungen ermitteln. Im Vorfeld der Analyse werden dabei zur Aufwandsreduktion zu weit voneinander entfernte Regionen identifiziert und ausgeschlossen, wobei die Schwerpunktdistanz der Regionen als Kriterium zu Grunde gelegt wird.

Darüberhinaus findet eine Überprüfung der Größendifferenzen zwischen den zu vergleichenden Regionen statt. Wird eine zu spezifizierende Maximaldifferenz überschritten, so werden die Regionen als nicht korrespondierend betrachtet. Auf diese Weise resultiert eine größere Robustheit der Zuordnungsergebnisse. Dabei ist allerdings zu berücksichtigen, dass der Schwellwert nicht zu niedrig gewählt werden darf, da beispielsweise bei einer Bewegungsdetektion im *Frame-To-Mosaic*-Modus relativ zum Mosaikbild des statischen Szenenhintergrundes (vgl. Kapitel 5) ein Regionenwachstum durch das Detektionsverfahren gegeben ist und bei der Korrespondenzanalyse in angemessener Weise beachtet werden muss.

4.2.2 Zuordnung detektierter Regionen

Zwei Regionen, die auf Grund ihres Schwerpunktabstandes und der vorliegenden Größendifferenzen potenziell korrespondierend sind, werden auf der Basis ihrer Grauwerthistogramme miteinander verglichen. Im Rahmen dieser Diplomarbeit wurden dazu zwei verschiedene Heuristiken untersucht, die im Folgenden vorgestellt werden. Beim ersten Verfahren handelt es sich um den χ^2 -*Test*, ein statistisches Standardverfahren für den Vergleich diskreter Verteilungen [PTVF92]. Er wird insbesondere im Hinblick auf seine Eignung im gegebenen Kontext analysiert. Als alternative Vorgehensweise wird anschließend die Berechnung der *Schnittfläche* zweier Histogramme präsentiert. Den Abschluss des Abschnitts bilden einige Anmerkungen zur Bewertung von Größendifferenzen zwischen Regionen, durch die eine erhöhte Robustheit der Detektionsergebnisse angestrebt wird.

χ^2 -Test

Ein bei der Analyse statistischer Daten häufig auftretendes Problem stellt die Überprüfung von Verteilungsannahmen dar. In der Regel werden dabei aus empirischen Untersuchungen erhaltene Verteilungen mit angenommenen Standardverteilungen verglichen, um im Vorfeld formulierte Hypothesen zu verifizieren oder zu widerlegen. Für die Analyse diskreter Daten, die in Form von Histogrammen gegeben sind, eignet sich dazu der χ^2 -Test. Unter Vorgabe eines Signifikanzniveaus lässt sich durch die Untersuchung der Differenzen zwischen den Histogrammen die angenommene Ähnlichkeit überprüfen. Details zu den theoretischen Grundlagen finden sich im Anhang A. Durch eine Abwandlung der dort eingeführten Berechnungsformeln kann der Test auch für den Vergleich zweier, aus empirischen Untersuchungen resultierender Verteilungen herangezogen werden. Die modifizierte Formel zur Berechnung der gewichteten quadratischen Differenzen zwischen den Histogrammen (vgl. mit Gleichung A.3) ergibt sich nach [PTVF92] zu:

$$\chi^2 = \sum_i \frac{(\sqrt{A/B} \cdot b_i - \sqrt{B/A} \cdot a_i)^2}{b_i + a_i} \quad (4.8)$$

$$A = \sum_i a_i \quad , \quad B = \sum_i b_i \quad (4.9)$$

a_i und b_i entsprechen dabei den Treffern innerhalb der einzelnen Diskretisierungszellen i , A und B jeweils der Gesamtanzahl Treffer in beiden Histogrammen.

Die hier zu untersuchenden Grauerthistogramme zweier Regionen lassen sich auf dieser Basis vergleichen, wenn sie als Verteilungen einer diskreten Zufallsvariablen interpretiert werden. Durch Festlegung eines geeigneten Signifikanzniveaus für eine erfolgreiche Zuordnung von Histogrammen können Korrespondenzen zwischen Regionen etabliert bzw. zurückgewiesen werden.

In der praktischen Anwendung hat sich der Test zum Vergleich von Regionen allerdings als zu restriktiv erwiesen. Dies liegt hauptsächlich in der Sensitivität gegenüber großen Abweichungen innerhalb einzelner Diskretisierungszellen begründet. Da die Differenzen, wie aus Gleichung 4.8 ersichtlich, quadratisch gewichtet bei der Berechnung eingehen, sind bereits wenige Ausreißer ausschlaggebend für die Gesamtdifferenz. Regionen, deren Histogramme im gegebenen Kontext als ähnlich einzustufen sind, werden auf Grund dessen als nicht korrespondierend zurückgewiesen. Abbildung 4.5 zeigt ein entsprechendes Beispiel. Für χ^2

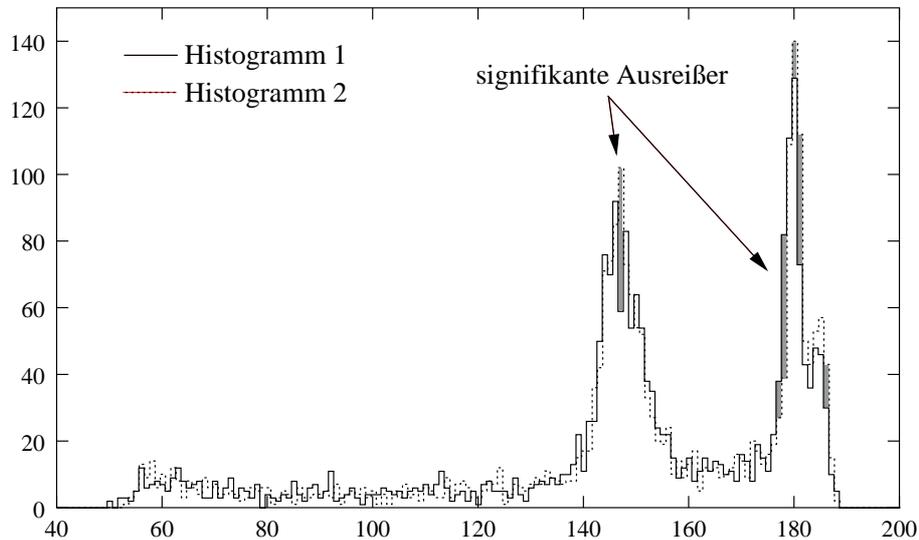


Abbildung 4.5: Ausreißer im χ^2 -Test

resultiert in diesem Fall, trotz sehr weniger großer Differenzen (beispielhaft grau markiert), ein Wert, der unter der gegebenen Anzahl von Freiheitsgraden eine Rückweisung der Nullhypothese nach sich zieht. Durch die im vorangegangenen Abschnitt skizzierten Eigenschaften der Regionen, insbesondere dem durch die Detektionsverfahren hervorgerufenen Wachstum, lassen sich große Differenzen in einzelnen Zellen der Histogramme jedoch nicht vermeiden. Daraus folgt, dass für einen Vergleich der hier gegebenen Histogramme Verfahren Anwendung finden müssen, die weniger sensitiv auf Abweichungen reagieren. Die dem χ^2 -Test zu Grunde liegenden, sehr restriktiven statistischen Annahmen über die Ähnlichkeit von Verteilungen schließen seinen Einsatz aus.

Schnittfläche

Ein einfaches Verfahren zur Abschätzung der Ähnlichkeit zweier Histogramme besteht in der Berechnung der Schnittfläche F , die aus einer Überlagerung der Histogramme resultiert. Unter der Voraussetzung, dass die Diskretisierungsskalen der betrachteten Histogramme identisch sind, lässt sie sich durch Gleichung 4.10 ermitteln, wonach über alle Diskretisierungszellen i jeweils das Minimum der

Einträge a_i und b_i aus beiden Histogrammen aufsummiert wird:

$$F = \sum_i \min(a_i, b_i) \quad (4.10)$$

In Abbildung 4.6 ist die Schnittfläche für zwei Histogramme visualisiert.

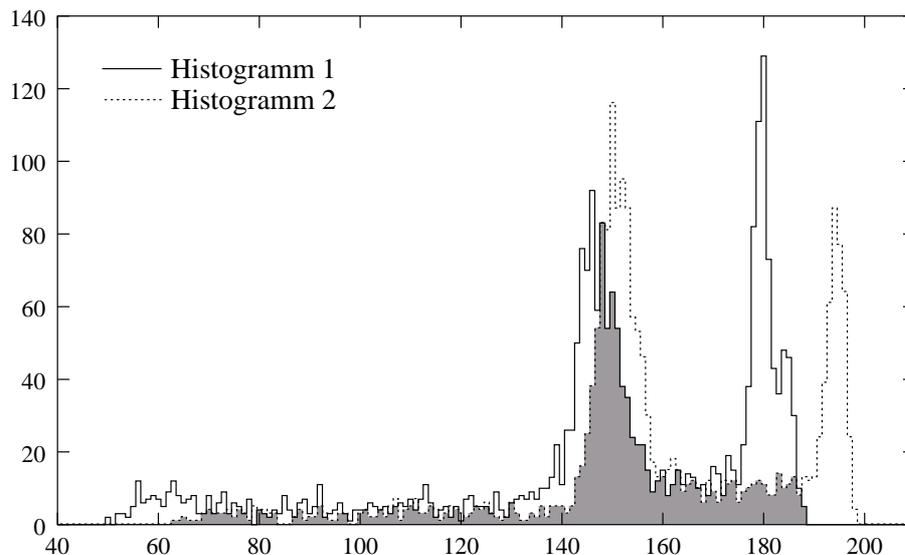


Abbildung 4.6: Schnittfläche von zwei Histogrammen

Als Maß für die Ähnlichkeit der Regionen wird die errechnete Fläche jeweils ins Verhältnis gesetzt zu den Gesamtflächen A und B (Gleichung 4.9) der beiden Histogramme:

$$q_1 = F/A \quad \text{und} \quad q_2 = F/B \quad (4.11)$$

Wenn die resultierenden Quotienten q_1 und q_2 ein festgelegtes Signifikanzniveau überschreiten, d.h. die Schnittfläche mehr als einen bestimmten Prozentsatz der einzelnen Histogrammflächen ausmacht, werden die Regionen auf Grund ihrer Intensitätshistogramme als korrespondierend angenommen. In der Praxis hat sich ein Signifikanzniveau von 0.75 bis 0.85 als geeignet erwiesen, um den Ähnlichkeitsanforderungen im gegebenen Kontext gerecht zu werden.

Auch in diesem Fall stellt das Wachstum der Regionen jedoch ein Problem dar, das berücksichtigt werden muß. Je größer die Differenzen zwischen den Flächen

A und B der beiden Histogramme sind, desto signifikantere Abweichungen resultieren in der Regel zwischen den Einträgen a_i und b_i der einzelnen Diskretisierungszellen. Im Allgemeinen sind sehr kleine Werte für q_1 und q_2 die Folge, die zu Rückweisungen von Zuordnungen zwischen Regionen führen, deren Grauwertverteilungen im gegebenen Kontext ähnlich sind. Da Zuordnungen bei nur kleinen Größendifferenzen nicht ausgeschlossen werden sollen, bietet sich eine Normierung der Histogramme an. Dazu werden die Trefferzahlen beider Histogramme anhand der nachfolgenden Gleichungen auf die Gesamtsumme 1 normiert:

$$\tilde{a}_i = \frac{a_i}{A} \quad \text{und} \quad \tilde{b}_i = \frac{b_i}{B} \quad (4.12)$$

Auf diesem Weg lassen sich Varianzen innerhalb der Einträge in beiden Histogrammen ausgleichen und korrespondierende Regionen sicher zuordnen.

Die in den vorangegangenen Absätzen mehrfach angesprochenen Differenzen innerhalb der Größen detektierter Regionen führen beim χ^2 -Test zu den dargestellten Problemen, die seinem Einsatz im gegebenen Kontext entgegenstehen. Obwohl bei der Berechnung der Schnittfläche Größendifferenzen durch eine Normierung der Histogramme ausgeglichen werden können, ist dennoch eine explizite Überprüfung der Regionengrößen sinnvoll. Zum einen wird durch den frühzeitigen Ausschluss von Regionen, deren Pixelzahlen sich signifikant unterscheiden, der bei der Korrespondenzanalyse anfallende Aufwand reduziert, zum anderen sind auf diese Weise robustere Ergebnisse zu erzielen. Fehlerhafte Zuordnungen von Regionen, die sehr große Differenzen aufweisen, werden vermieden. Die im Vorfeld der Vergleiche stattfindende Überprüfung wird durch die Berechnung des nachstehenden Quotienten realisiert:

$$\frac{|A - B|}{A + B} \quad (4.13)$$

Durch ihn werden die Trefferzahlen A und B der beiden Histogramme zueinander in Beziehung gesetzt. Nur wenn die Differenz zwischen ihnen relativ zur Summe der Histogrammflächen unterhalb einer festgesetzten Schwelle bleibt, werden die Regionen miteinander verglichen.

4.2.3 Zusammenhangskomponenten

Aus der Bewegungsdetektion und der anschließenden Regionensegmentierung resultieren unter Umständen eine große Anzahl bewegter Regionen. Der paarweise Vergleich dieser ist mit erheblichem Aufwand verbunden. Weiterhin treten in den Segmentierungsergebnissen mitunter deutliche Varianzen auf. Selbst bei geringen Objekt- und Kamerabewegungen zerfallen oder verschmelzen Regionen im Verlauf der Segmentierung (Abbildung 4.7). Allein durch einen Vergleich einzelner Regionen lassen sich in solchen Fällen die vorhandenen Korrespondenzen nicht detektieren.

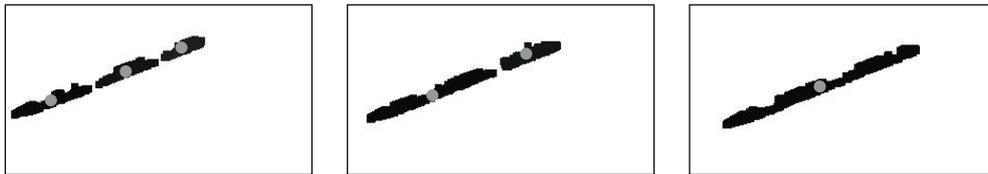


Abbildung 4.7: Varianz in der Regionensegmentierung: Die einem Objekt zuzurechnenden, bewegten Regionen (schwarz) verändern sich von Bild zu Bild (die Schwerpunkte sind grau markiert).

Auf Grund der beiden vorstehenden Aspekte, und im Hinblick darauf, dass eng benachbarte Regionen häufig demselben Objekt zuzurechnen sind, werden die Regionen vor der Suche nach Korrespondenzen innerhalb von *Zusammenhangskomponenten* gruppiert (vgl. auch [Reh98]). Die Suche beschränkt sich anschließend zunächst auf die Komponenten als Ganzes, die mit einem geringeren Gesamtaufwand paarweise miteinander verglichen werden können. Dieser Ansatz ermöglicht die Zuordnung von *Regionengruppen* zueinander. Innerhalb der betrachteten Gruppen muß dabei nicht notwendigerweise die gleiche Anzahl von Regionen enthalten sein, wodurch die Flexibilität bei der Suche zusätzlich erhöht wird. Durch die Etablierung zeitlicher Korrespondenzen auf dieser Basis wird eine größere Robustheit erzielt, die unter anderem den Ausgleich von Varianzen innerhalb der Segmentierungsergebnisse zulässt.

Unter einer Zusammenhangskomponente ist eine Menge von Regionen (oder genau eine Region) eines Bildes zu verstehen, innerhalb der zu jeder Region mindestens eine weitere, gemäß eines entsprechenden Abstandskriteriums, benachbarte Region vorhanden ist. Im vorliegenden Kontext sollen lediglich räumliche Nachbarschaftskriterien, d.h. die Abstände der Regionen, für die Gruppierung herangezogen werden. Weitere Merkmale, wie ähnliche Farben oder Formen, scheiden aus den in Abschnitt 4.2.1 aufgeführten Gründen aus.

Zur Berechnung des Abstandes zwischen zwei Regionen gibt es verschiedene Distanzmaße. In [Lan00] findet sich dazu eine Übersicht. Die einfachste Abschätzung der Distanz zwischen zwei Regionen R_1 und R_2 besteht in der Berechnung des Schwerpunktabstandes:

$$d_{com} = \| s(R_1) - s(R_2) \| \quad (4.14)$$

Da für jede Region der Schwerpunkt initial als charakterisierendes Merkmal berechnet wird, ist auf dieser Basis eine effiziente Gruppierung realisierbar. Der Schwerpunktabstand ist im vorliegenden Anwendungsgebiet allerdings in vielen Fällen nicht sehr aussagekräftig für den tatsächlichen Abstand zu untersuchender Regionen. In Abbildung 4.7 sind die Schwerpunkte der Regionen grau markiert worden. Sie weisen einen großen Abstand auf und beschreiben damit die offensichtlich vorhandene räumliche Nähe der Regionen nur unzureichend.

Als Alternative bietet sich die Berechnung des kürzesten Abstandes zwischen zwei Regionen an. Er ergibt sich aus der Distanz der zwei nächstgelegenen Punkte p_1 und p_2 der beiden betrachteten Regionen (Abbildung 4.8):

$$d_{short} = \min_{p_1 \in R_1, p_2 \in R_2} \| p_1 - p_2 \| \quad (4.15)$$

Die Regionen werden hierfür zunächst durch umschreibende Polygone approximiert, bevor auf dieser Basis das Punktepaar mit minimalem Abstand gesucht wird. Der damit verbundene Aufwand hängt von der gewählten Genauigkeit für die Approximation ab. In jedem Fall ist er jedoch höher als bei der Distanzberechnung über Schwerpunkte. Da im Allgemeinen jedoch auf der Grundlage der aus dieser Vorgehensweise folgenden Komponenten (Beispielskizze in Abbildung 4.9) robustere Ergebnisse zu erzielen sind, lässt sich der erhöhte Aufwand rechtfertigen.

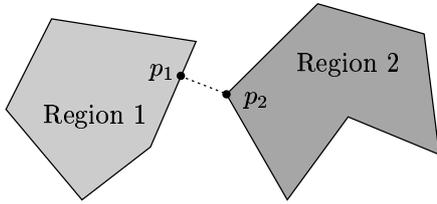


Abbildung 4.8: Kürzeste Distanz zwischen zwei Regionen

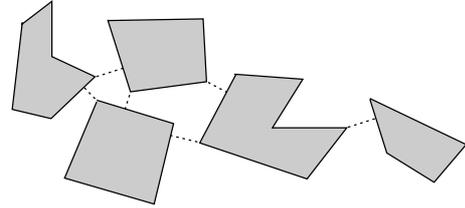


Abbildung 4.9: Skizze einer Zusammenhangskomponente

Nachdem für die beiden betrachteten Bilder jeweils die Zusammenhangskomponenten ermittelt wurden, werden diese paarweise miteinander verglichen. Dabei finden dieselben Merkmale Anwendung, die in Abschnitt 4.2.1 für die einzelnen Regionen vorgestellt wurden. Der Schwerpunkt einer Zusammenhangskomponente s_c resultiert aus dem gewichteten Mittel der Schwerpunkte s aller Regionen R , indiziert durch j . Als Gewichte werden die Pixelanzahlen A_j der Regionen verwendet:

$$s_c = \frac{1}{\sum_j A_j} \sum_j A_j \cdot s(R_j) \quad (4.16)$$

Die Größe A_c der Komponente ergibt sich aus der Gesamtsumme der Pixel in den beteiligten Regionen (Gl. 4.17), und das Intensitätshistogramm stellt die Summe über alle Regionenhistogramme dar (Gl. 4.18). Der Eintrag a_i^c im i -ten Diskretisierungsintervall des Komponentenhistogramms resultiert dabei aus der Summe der Einträge a_i^j in den einzelnen Histogrammen der Regionen R_j .

$$A_c = \sum_j A_j \quad (4.17)$$

$$a_i^c = \sum_j a_i^j \quad (4.18)$$

Wenn die Schwerpunktdistanz zweier Zusammenhangskomponenten in aufeinanderfolgenden Bildern die festgelegte Maximaldistanz nicht überschreitet und sich ihre Größen entsprechend des in Abschnitt 4.2.2 vorgestellten Kriteriums nicht zu sehr voneinander unterscheiden, werden die Komponenten auf Basis ihrer Histogramme miteinander verglichen. Falls eine hinreichende Ähnlichkeit besteht, wird der *Match* in den Korrespondenzgraphen eingetragen (Kap. 4.2.4).

Für eine Komponente sind dabei mehrere Zuordnungen zugelassen. Auf diese Weise werden im Fall von Mehrdeutigkeiten, die beispielsweise aus Objektüberdeckungen resultieren können (Abbildung 4.10), zunächst alle potenziellen Korrespondenzen in den Graphen aufgenommen. Sie lassen sich, falls notwendig, zu einem späteren Zeitpunkt anhand ergänzender Informationen wieder eliminieren. Entscheidend ist allerdings, dass dennoch jede detektierte Region nur als Teil genau *einer* Komponente in den Graphen eingefügt wird. Dadurch werden konkurrierende Hypothesen über die korrekte Gruppierung von Regionen vermieden und die nachträgliche, aufwändige Auflösung dieser Konflikte umgangen.

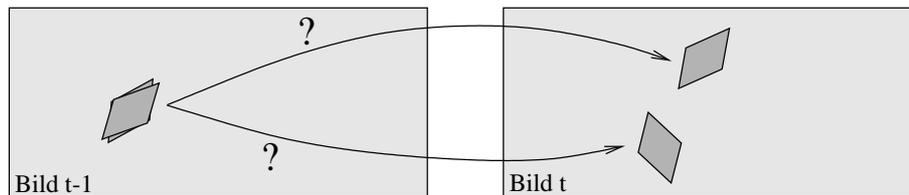


Abbildung 4.10: Mehrdeutigkeiten in der Zuordnung von Komponenten:

Für die beiden im linken Bild einander überlagerten Objekte resultiert nur eine Region, die zu den beiden Regionen im nächsten Bild nicht eindeutig zugeordnet werden kann.

In vielen Fällen ist durch die vorgestellte Vorgehensweise eine robuste Verfolgung bewegter Bereiche zu gewährleisten. Allerdings ergeben sich Probleme, wenn Gruppen korrespondierender Regionen in aufeinanderfolgenden Bildern jeweils in unterschiedliche Zusammenhangskomponenten zerlegt werden. In Abbildung 4.11 ist ein solcher Fall skizziert. Während im linken Bild die Regionen R_1 und R_2 in einer Komponente zusammengefasst werden, trifft dies im rechten Bild auf die Regionen R'_2 und R'_3 zu. Durch den reinen Vergleich der Zusammenhangskomponenten sind die vorhandenen Korrespondenzen zwischen den Regionen nicht zu detektieren.

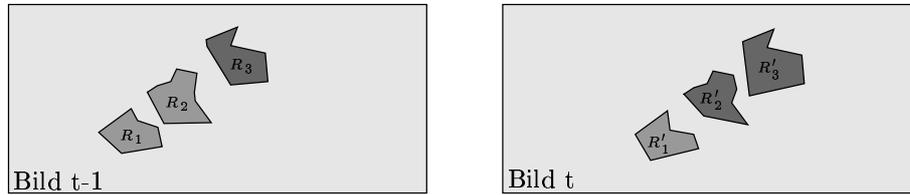


Abbildung 4.11: Varianz in der Regionengruppierung

Eine Behandlung derartiger Problemfälle wird durch die Zerlegung von Komponenten, die als Ganzes nicht zugeordnet werden können, in *Teilzusammenhangskomponenten* realisiert. Eine Teilzusammenhangskomponente enthält eine Teilmenge der zu einer Zusammenhangskomponente gehörenden Regionen, wobei die zu Grunde gelegten Nachbarschaftskriterien erfüllt sein müssen. Anschließend werden auf Basis der zwei Mengen aller Teilzusammenhangskomponenten der beiden Bilder weitere Korrespondenzen gesucht. Dabei erfolgt iterativ eine Auswahl der jeweils besten Zuordnung zwischen den Mengen, bis keine weiteren Korrespondenzen mehr gefunden werden. Die zugeordneten Komponenten, sowie alle weiteren, in denen Regionen aus ihnen enthalten sind, werden jeweils aus den Mengen entfernt. Auf diese Weise wird auch hier garantiert, dass jede Region nur einmal in den Graphen eingefügt wird. Für das vorstehende Beispiel aus Abbildung 4.11 wird durch den Vergleich der Teilzusammenhangskomponenten eine Detektion aller Korrespondenzen möglich.

Die Bewertung der Matches erfolgt auf Basis der aus den Vergleichsverfahren resultierenden Ähnlichkeitswerten für die Histogramme. Bei der Schnittflächenberechnung dienen die errechneten Flächenanteile (Gleichung 4.11) als Maßstab. In den untersuchten Bildfolgen können damit in der Regel gute Ergebnisse erzielt werden (vgl. Kapitel 7). Dennoch ist eine Erweiterung denkbar, die beispielsweise die Größen der Komponenten explizit berücksichtigt und der Zuordnung größerer Komponenten zunächst den Vorzug gibt.

Trotz der Zerlegung in Teilzusammenhangskomponenten sind in Einzelfällen Korrespondenzen nicht zu erkennen. Falls Regionen bei der Segmentierung zerfallen oder verschmelzen und die Teile gleichzeitig jeweils in unterschiedliche Komponenten gruppiert werden (Abb. 4.12), so ist dies durch die vorgestellte Vorgehensweise nicht erfasst. An dieser Stelle sind Erweiterungen notwendig, die bei-

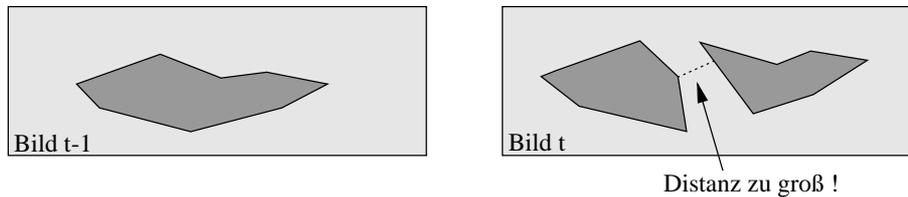


Abbildung 4.12: Verbleibende Probleme in der Korrespondenzanalyse:

Die Gruppierung der beiden Regionen aus dem rechten Bild in verschiedene Zusammenhangskomponenten verhindert eine Zuordnung zu der Region im linken Bild.

spielsweise auch Mengen von Zusammenhangskomponenten bei den Vergleichen berücksichtigen. Ein solcher Fall tritt auch im nachfolgenden Abschlussbeispiel auf, das die mit der vorgestellten Vorgehensweise erzielbaren Resultate illustriert. Im linken Teil der Abbildung 4.13 sind die Bewegungskarten von fünf aufeinanderfolgenden Bildern der Bildfolge "Baufix1" gezeigt. Die resultierenden Zusammenhangskomponenten sind jeweils nummeriert und die zugehörigen Regionen gekennzeichnet. Auf der rechten Seite sind die Zuordnungen skizziert, die sich aus der Anwendung des vorgestellten Verfahrens ergeben. Es wird deutlich, dass die vorhandenen Korrespondenzen im Allgemeinen robust detektiert werden, trotz des sichtbaren Regionen- und Komponentenwachstums. Zwischen dem dritten und vierten Bild resultiert die Zuordnung der drei einzelnen Komponenten des dritten Bildes zu der einen Komponente des vierten Bildes aus dem Vergleich der Teilzusammenhangskomponenten. Die Größenunterschiede stehen einem *direkten* Vergleich entgegen. Zwischen dem ersten und dem zweiten Bild lassen sich nur zwei der drei vorhandenen Korrespondenzen detektieren. Die Regionen innerhalb der Komponenten zwei und drei des ersten Bildes verschmelzen im zweiten Bild zu einer Region und bilden *eine* Komponente. Aus den oben aufgeführten Gründen ist eine Zuordnung zu den beiden korrespondierenden Komponenten des ersten Bildes nicht möglich.

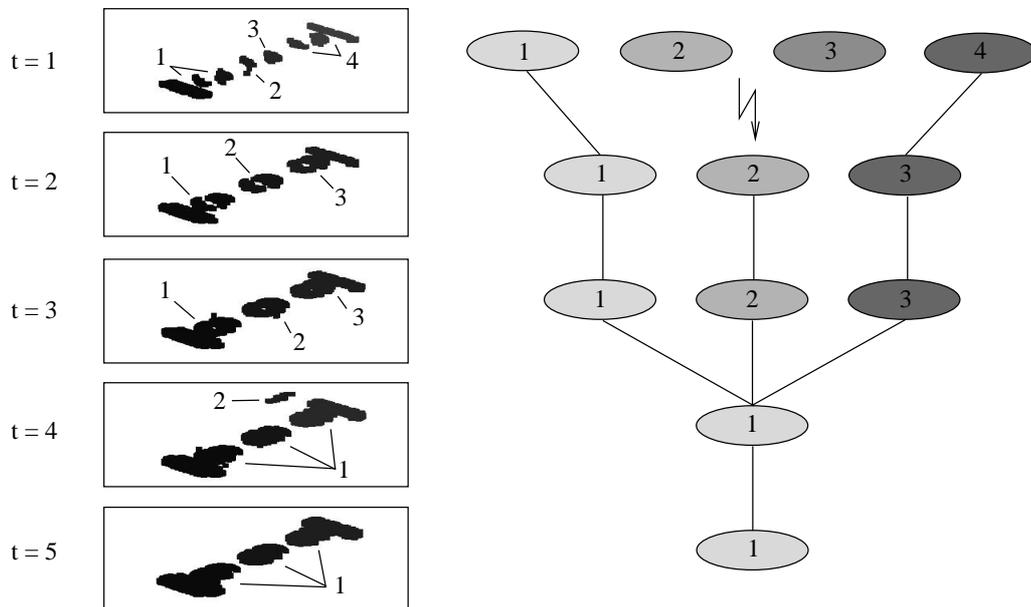


Abbildung 4.13: Zeitliche Korrespondenz bewegter Regionen

4.2.4 Der Korrespondenzgraph

Die im vorangegangenen Abschnitt bestimmten Zuordnungen zwischen (Teil-)Zusammenhangskomponenten aufeinanderfolgender Bilder kodieren die dynamischen Informationen innerhalb der Bildsequenz. Aus ihnen lassen sich die Bewegungsinformationen der Regionen gewinnen, die bei der Generierung von Mosaikbildern ohne eine explizite Bewegungsdetektion verloren gehen würden. Um diese Informationen nachfolgenden Bildanalysemodulen zugänglich zu machen, ist eine geeignete Datenstruktur notwendig, aus der bei Bedarf die jeweils interessierenden Informationen entnommen werden können. Die Charakteristik der Korrespondenzen, die über die Zeit zwischen jeweils zwei Zusammenhangskomponenten etabliert werden, legt die Verwendung einer Graphdatenstruktur nahe, wie sie auch in [CM98] Anwendung findet.

Der hier eingesetzte Graph ist gerichtet. Seine Knoten resultieren aus den Komponenten, die zu einer Zuordnung gehören. Die Korrespondenzen zwischen ihnen werden durch zeitlich gerichtete Kanten repräsentiert. Jeder Kante wird ein Ge-

wicht zugeordnet, das aus der Bewertung der Korrespondenz resultiert. Für einen Vergleich der Komponenten auf Basis der Histogrammschnittflächen wird demnach der prozentuale Überlapp der Flächen als Bewertung übernommen.

Auf Basis dieses Graphen lassen sich für jede eingefügte Komponente durch eine Rückverfolgung der bei ihr endenden Pfade alle in der Vergangenheit korrespondierenden Komponenten ermitteln. Die Bewertungen der Kanten, die sich unter Anwendung einer geeigneten Heuristik auch als Bewertungen der einzelnen Pfade interpretieren lassen, können dabei als Anhaltspunkt für die Signifikanz der Korrespondenzen dienen. Durch diese lassen sich unter Umständen auch Mehrdeutigkeiten in den Zuordnungen auflösen, sofern die Bewertungen der resultierenden Pfade hinreichend deutliche Unterschiede aufweisen.

Darüberhinaus kann der Graph als Grundlage für eine nachträgliche Verifikation oder Ergänzung der gefundenen Korrespondenzen dienen. Zuordnungen oder Pfade mit sehr kleinen Bewertungen deuten auf Fehldetektionen hin. Andererseits besteht durch eine gezielte Nachbearbeitung der Pfade mitunter die Möglichkeit, Lücken in den Zuordnungen zu erkennen und vorhandene Korrespondenzen im Nachhinein einzufügen. Insgesamt sollten auf diese Weise robustere Resultate in der zeitlichen Korrespondenzanalyse zu erzielen sein.

4.3 Trajektorienextraktion

Bewegungen von Objekten lassen sich anhand von Bewegungsmodellen mathematisch erfassen. Je komplexer diese Modelle sind, desto exaktere Beschreibungen der Objektdynamik resultieren aus ihrer Anwendung. Wie in Kapitel 3 einleitend erwähnt wurde, besteht die Zielsetzung der vorliegenden Arbeit vorrangig in der *Bewegungsdetektion* und eine detaillierte mathematische Modellierung ist nicht vorgesehen. Die Bewegungen der Objekte sollen lediglich durch Translationsvektoren approximiert werden. Diese ermöglichen in vielen Fällen eine hinreichend genaue Beschreibung der Bewegungscharakteristiken. Insbesondere bei hohen Bildraten und nur kleinen Bewegungen der Objekte lassen sich auf diese Weise nahezu beliebige Bewegungsbahnen annähern.

Zur Extraktion der Translationsvektoren im gegebenen Kontext werden die Verlagerungen der Komponentenschwerpunkte zwischen zwei aufeinanderfolgenden Bildern zu Grunde gelegt. Lässt sich eine Komponente dabei über mehrere Bilder hinweg verfolgen, so resultiert aus der Aneinanderreihung der einzelnen Schwerpunktpositionen eine *Trajektorie*. Sie kann auf der einen Seite als Ausgangsbasis für die Analyse der Bewegungscharakteristiken dienen, sowie auf der anderen Seite als Grundlage zur Anwendung komplexerer Bewegungsmodelle genutzt werden. Wenn das Objekt sich entlang einer Geraden bewegt, so genügt bereits eine lineare Regression der Trajektorienpunkte, um Parameter für ein lineares Modell zu bestimmen, bei komplizierteren Bewegungsmustern sind entsprechend flexiblere Interpolationsverfahren anzuwenden.

Die einzelnen Trajektorien bestehen aus einer Liste von Punkten, mit denen jeweils eine Zusammenhangskomponente bzw. der zugehörige Knoten des Korrespondenzgraphen assoziiert werden. Ihre Generierung erfolgt parallel zum Aufbau des Graphen. Wenn eine gefundene Korrespondenz eingefügt wird, so findet eine Untersuchung der enthaltenen Komponente aus dem Vorgängerbild dahingehend statt, ob sie bereits durch frühere Zuordnungen im Graphen repräsentiert ist. Dabei sind drei mögliche Fälle zu unterscheiden (Abbildung 4.14). Falls die Komponente bereits enthalten ist, so existieren Trajektorien, die ihre Schwerpunktverlagerung bis zum vorherigen Zeitpunkt beschreiben. Sie müssen lediglich um den aus der neuen Zuordnung hinzukommenden Punkt erweitert werden. Ist die Komponente noch nicht im Graphen vorhanden, so wird überprüft, ob eine oder mehrere ihrer Regionen als Bestandteil anderer Komponenten eingefügt wurden. Dies trifft beispielsweise zu, wenn zum aktuellen Zeitpunkt nur Teilzusammenhangskomponenten einer Komponente zugeordnet werden können, die im vorausgegangenen Schritt als Ganzes gematcht wurde. Dann existiert für die aktuelle Komponente keine direkte Zuordnung, obwohl durch die beteiligten Regionen Korrespondenzen gegeben sind. Solche Konstellationen sind häufig bei einem Zerfall oder einer Verschmelzung bewegter Bereiche zu beobachten. Sind entsprechende Regionen vorhanden, so findet eine Erweiterung der mit den zugehörigen Komponenten assoziierten Trajektorien statt. Falls weder die Komponente selbst noch eine ihrer Regionen innerhalb des Graphen repräsentiert sind, wird eine neue Trajektorie initialisiert und gegebenenfalls im weiteren Verlauf expandiert.

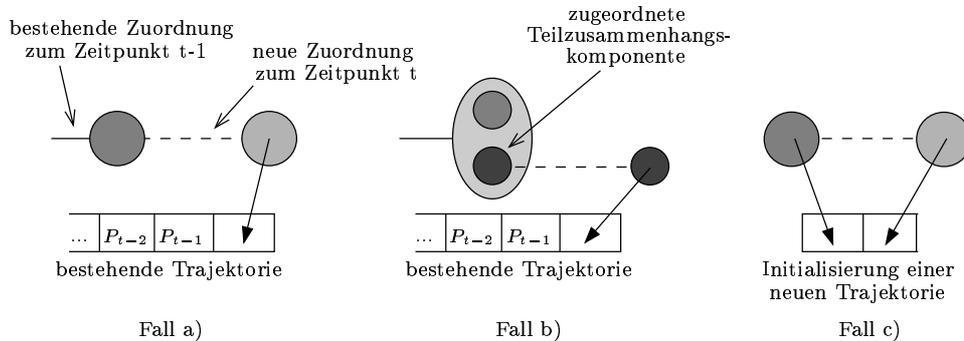


Abbildung 4.14: Die drei möglichen Konstellationen bei der Trajektoriengenerierung: a) die Komponente oder b) einzelne Regionen sind bereits im Graphen enthalten, d.h. bestehende Trajektorien werden expandiert, oder c) eine neue Trajektorie wird initialisiert.

Die gewonnenen Trajektorien eignen sich unter Umständen zur Lösung verschiedener Probleme, die in vorangegangenen Abschnitten aufgezeigt wurden. Zunächst könnten sich durch ihre Analyse Fehlklassifikationen ermitteln lassen. Die Schwerpunktpositionen eines als bewegt erkannten, tatsächlich aber statischen Bereichs unterliegen im Allgemeinen keiner Verschiebung. Die Varianz der zugehörigen Trajektorienpunkte sollte somit gering, und der Bereich durch eine Varianzanalyse von real bewegten Bereichen zu unterscheiden sein. Insbesondere die Initialpositionen bewegter Objekte, die bei einer Detektion im *Frame-To-Mosaic*-Modus im Mosaikbild verbleiben (Kapitel 5), wären auf diese Weise zu lokalisieren und nachträglich zu entfernen. Des Weiteren ist denkbar, dass die Trajektorien auch als Grundlage zur Rekonstruktion unvollständig detektierter Objekte dienen können. Wenn sich zwischen den einzelnen, zu einem Objekt gehörenden Regionen auf Basis der Bewegungsdetektion keine Verbindung herstellen lässt, so sollten sie dennoch eine ähnliche Bewegungscharakteristik aufweisen. Durch einen Vergleich ihrer Trajektorien ist gegebenenfalls eine Gruppierung sich parallel bewegendender Regionen möglich, auf deren Basis eine Rekonstruktion des Objektes erfolgen kann. Ein analoger Ansatz wird auch in [Bom97] vorgestellt, dessen Übertragbarkeit auf den vorliegenden Anwendungsbereich zu prüfen ist.

Kapitel 5

Hintergrundmosaikbilder

Die Zusammenfassung von Bildern einer Sequenz in einem einzigen Mosaikbild führt zu einer erheblichen Datenreduktion und erleichtert für Bildverarbeitungssysteme die Analyse der aufgenommenen Szene. Durch die gezielte Detektion von Objektbewegungen und ihrer Beobachtung über die Zeit ist dabei ein weitgehender Erhalt der dynamischen Informationen gewährleistet. Die in den vorangegangenen Kapiteln diskutierten Verfahren bilden damit eine wichtige Basis für die effiziente Verarbeitung von Bildfolgen mit Hilfe von Mosaikbildern.

Aus der Bewegungsdetektion und dem Regionentracking resultieren statische und dynamische Daten, die die Szene vollständig beschreiben. Sie werden hier nachfolgenden Bildanalysemodulen in Form des angesprochenen Korrespondenzgraphen und eines Mosaikbildes des statischen Szenenhintergrundes zur Verfügung gestellt. Aus dem Graphen lassen sich die dynamischen Informationen der Szene in Form von Trajektorien- und Regionendaten extrahieren. Das Mosaikbild repräsentiert demgegenüber den statischen, innerhalb der Bildfolge unveränderten Anteil der aufgenommenen Szene. In diesem Kapitel wird die Generierung des Hintergrundmosaikbildes behandelt. Durch die explizite Berücksichtigung der zuvor gewonnenen Bewegungsdaten werden dabei die in dieser Arbeit einleitend aufgezeigten Integrationsfehler bei der Mosaikbildgenerierung von Bildfolgen bewegter Szenen weitgehend vermieden.

5.1 Statische Integration

In Kapitel 2.1.2 wurden verschiedene Heuristiken zur Integration neuer Bilder in ein Mosaikbild vorgestellt. Die in der Regel verwendete Berechnung des arithmetischen Mittels aus den Grauwerten der gegebenen Bilder verursacht in bewegten Bereichen Integrationsfehler im Mosaikbild. Um diese zu vermeiden, ist im Rahmen der vorliegenden Arbeit ergänzend ein statischer Integrationsmodus realisiert worden, in dem für jedes Pixel das Resultat der Klassifikation als bewegt oder unbewegt berücksichtigt wird. An Pixeln, die als unbewegt eingeordnet wurden, resultiert der neue Intensitätswert wie bisher aus dem arithmetischen Mittelwert der Grauwerte. Bildpunkte, die einer Eigenbewegung unterliegen, werden bei der Integration ausgespart. An den entsprechenden Bereichen im Mosaikbild bleibt die vorhandene Information erhalten.

Abbildung 5.1 zeigt einen Vergleich der aus der Anwendung der beiden Vorgehensweisen resultierenden Mosaikbilder. Die obere Zeile enthält die beiden zu Grunde gelegten Bilder aus der Bildfolge "Office1". Darunter sind das aus der Berechnung der Intensitätsdifferenz erhaltene Bewegungsbild (inverse Darstellung) und die binäre Bewegungskarte abgebildet. Dunkle Bereiche kennzeichnen Bewegungen. Im Mosaikbild unten links, das aus der einfachen Mittelwertbildung hervorgegangen ist, sind die Integrationsfehler deutlich zu erkennen. Die sich bewegende Person wird mehrfach integriert, wobei durch die arithmetische Mittelung jeweils eine Verwischung des Objektes hervorgerufen wird. Im Gegensatz dazu sind im rechten Mosaikbild die bewegten Bereiche bei der Integration berücksichtigt worden. Die aktuelle Position der Person wird nicht integriert, so dass Fehler vermieden werden. Auffallend ist allerdings, dass die alte Position weitgehend erhalten bleibt. Der Bereich wird zwar als bewegt klassifiziert, die in diesem Fall korrekte, neue Information jedoch zur Integration nicht herangezogen. Dies zeigt, dass die Unterscheidung zwischen Bereichen, die ein aktuell bewegtes Objekt beschreiben, und Bereichen, die aus früheren Positionen der Objekte in den Bildern resultieren, notwendig ist. Allein durch die hier verwendeten Verfahren zur Bewegungsdetektion ist das allerdings nicht realisierbar. Eine ergänzende Überprüfung, in deren Rahmen die Ursachen für die Klassifikation einzelner Pixel festgestellt werden können, lässt sich beispielsweise durch die Auswertung von

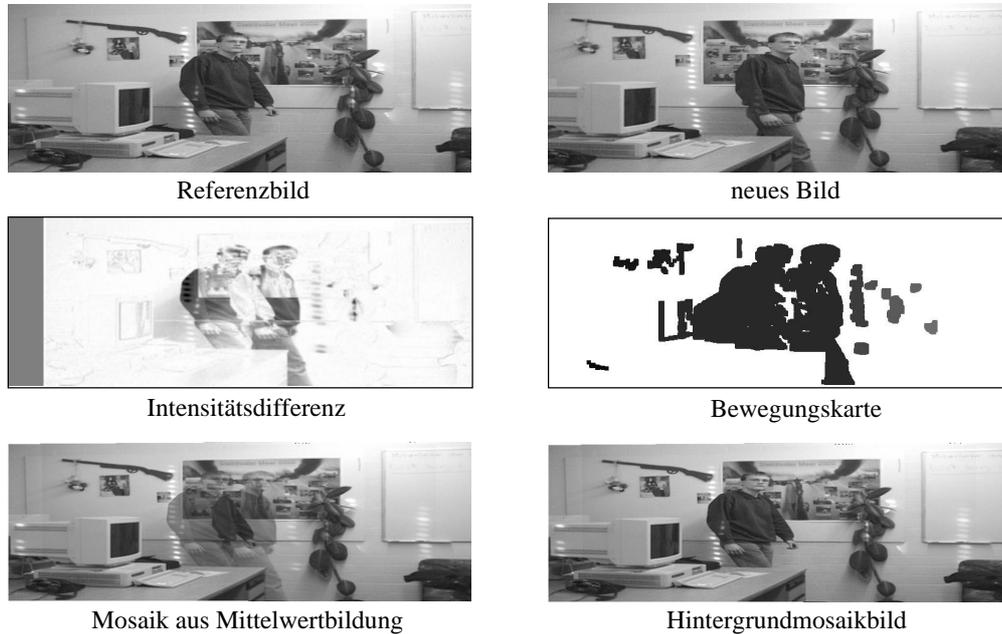


Abbildung 5.1: Statische Integration in Mosaikbildern bewegter Szenen

lokalen Konturinformationen oder durch die Analyse von Trajektorien- und Bewegungsdaten der betroffenen Gebiete (vgl. Kapitel 4.3) umsetzen.

Die Qualität des aus der statischen Integration resultierenden Mosaikbildes hängt davon ab, wie vollständig die bewegten Objekte detektiert werden können. Je besser dies gelingt, desto weniger Integrationsfehler treten auf und desto konsistenter und hochwertiger ist das Mosaikbild. Neben den, aus den Eigenschaften der Detektionsverfahren folgenden Einschränkungen, ist dafür auch die Wahl des Referenzbildes ein entscheidender Faktor. Seine Festlegung hat maßgeblichen Einfluss auf die detektierbaren Teilbereiche der Objekte. Der nachfolgende Abschnitt stellt die verschiedenen Alternativen vor.

5.2 Wahl des Referenzbildes

Analog zu den in Kapitel 2.1.1 skizzierten, verschiedenen Möglichkeiten zur Festlegung des Referenzbildes bei der Schätzung der Transformationsparameter, kann auch für die Bewegungsdetektion das vorangegangene Bild (*Frame-To-Frame*) oder ein geeigneter Ausschnitt des gegenwärtigen Mosaikbildes (*Frame-To-Mosaic*) als Referenz verwendet werden.

Im ersten Fall sind bei hohen Bildraten und entsprechend kleinen Verschiebungen der bewegten Objekte zwischen den Bildern, insbesondere bei hochgradig homogenen Objekten, in der Regel nur Teilstücke dieser zu detektieren (vgl. Abschnitt 3.4.2). Demgegenüber lassen sich durch die zweite Variante vollständigere Bewegungsdaten gewinnen. Die Ergebnisse sind allerdings von der jeweils verwendeten Integrationsheuristik für das Mosaikbild abhängig. Bei einer Integration neuer Bilder durch Mittelwertbildung unterscheiden sich die resultierenden Daten kaum von denen, die durch eine Detektion relativ zum Vorgängerbild erzielt werden. Wird hingegen die auf Basis der Bewegungsdaten realisierbare, statische Integration aus dem vorangegangenen Abschnitt eingesetzt, so kann das bewegte Objekt im Verlauf der schritthaltenden Verarbeitung der Bildfolge zunehmend vollständiger detektiert werden.

In Abbildung 5.2 sind die Detektionsergebnisse und die Hintergrundmosaikbilder für ausgewählte Bilder der Bildfolge "Baufix1" dargestellt, wobei die *Frame-To-Frame*-Detektion der Bewegungserkennung relativ zum Mosaikbild gegenübergestellt wird. Die obere Zeile zeigt die drei betrachteten Bilder der Sequenz, darunter sind jeweils die inversen Bewegungskarten beider Verfahren und die Mosaikbilder zu sehen. Zu Beginn der Verarbeitung weisen die Ergebnisse keine signifikanten Unterschiede auf. Während im weiteren Verlauf gegenüber dem Vorgängerbild jeweils nur minimale Veränderungen registriert werden können, die bei gleichbleibender Geschwindigkeit und Bildrate kaum variieren, wird das bewegte Objekt im *Frame-To-Mosaic*-Modus zunehmend vollständiger detektiert. In den Mosaikbildern der *Frame-To-Frame*-Detektion wird somit die aktuelle Position der Leiste jeweils weitgehend integriert, wobei nur an Objekträndern geringe Verwischungen auftreten. Im Hintergrundmosaikbild wird demgegenüber das bewegte Objekt ausgeklammert.

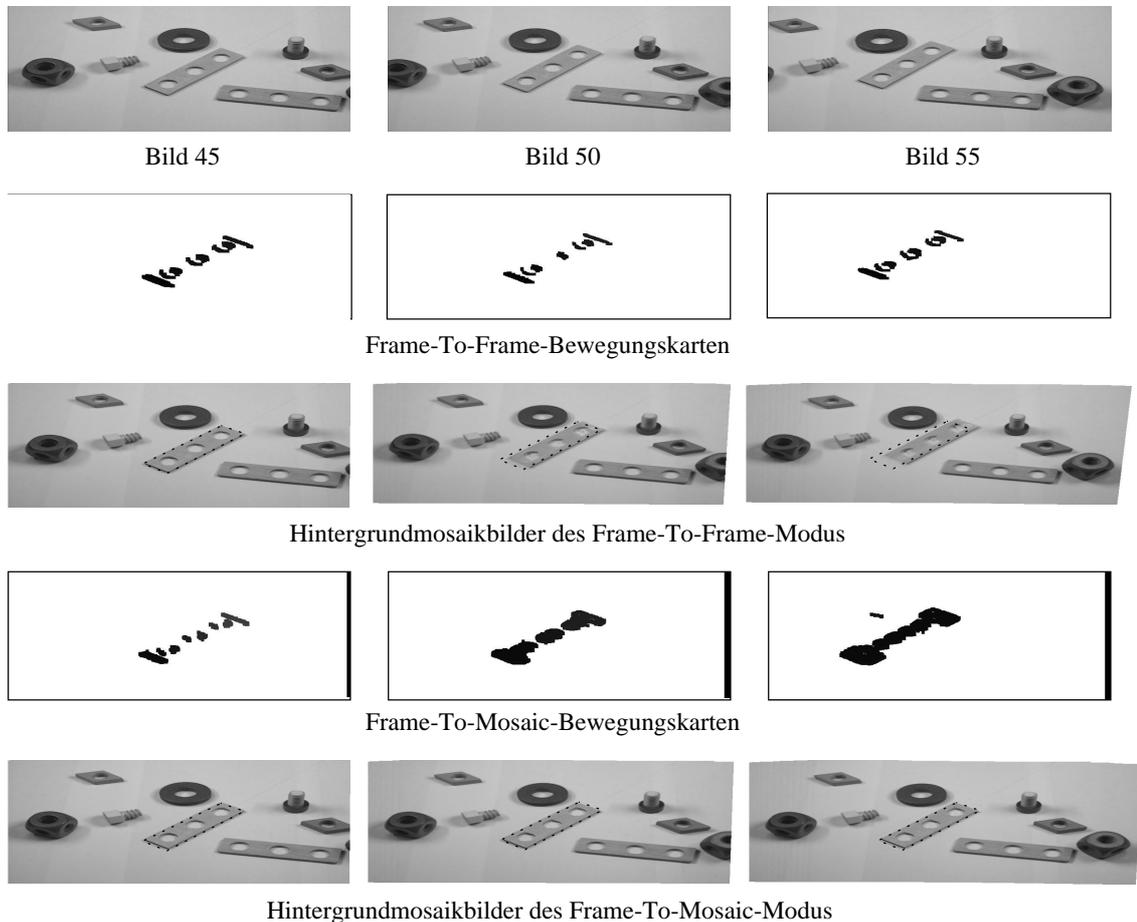


Abbildung 5.2: Auswirkungen der Referenzbildwahl

Abschließend ist anzumerken, dass die Initialposition des bewegten Objektes (in der Abbildung punktiert eingezeichnet), die bei der *Frame-To-Mosaic*-Detektion im Mosaikbild erhalten bleibt, im *Frame-To-Frame*-Modus durch die arithmetische Mittelwertbildung im weiteren Verlauf verschwindet. Da das Mosaikbild, in das das Objekt anfangs integriert wurde, in die Bewegungsdetektion nicht eingeht, werden die betroffenen Bereiche, nachdem das Objekt sich entfernt hat, als statisch klassifiziert. Die neuen Informationen gehen damit jeweils zur Hälfte bei der Integration ein und der Einfluss der initialen Grauwerte des Mosaikbildes nimmt exponentiell ab.

5.3 Zusammenfassung

Eine Bewegungsdetektion relativ zum aktuellen Mosaikbild, das durch Anwendung der statischen Integration entstanden ist, verbessert die Ausgangsbasis für die angestrebte Generierung weitgehend fehlerfreier Mosaikbilder. Mit zunehmender Detektionsdauer können die bewegten Objekte vollständiger detektiert werden und die Robustheit der Erkennung nimmt zu. Absolut fehlerfreie und zu jedem Zeitpunkt mit den statischen Anteilen der Szene übereinstimmende Mosaikbilder sind dennoch nicht zu garantieren, da Initialpositionen sich bewegender Objekte im Mosaikbild erhalten bleiben.

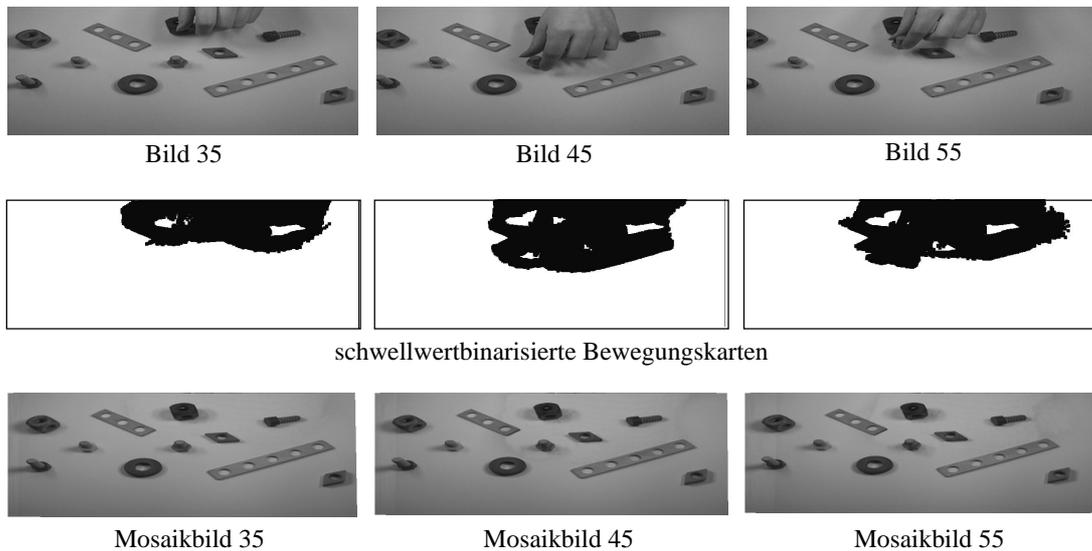


Abbildung 5.3: Mosaikbildentwicklung bei einer Szene mit Verdeckungen

Das Beispiel in Abb. 5.3 illustriert abschließend die Leistungsfähigkeit der Bewegungsdetektion im *Frame-To-Mosaic*-Modus in Kombination mit der statischen Integration. In den oben gezeigten Bildern der Folge "Baufix3" werden Objekte durch die ins Bild greifende Hand verdeckt und die Schraube in der Mitte des Bildes entfernt. Während die Hand nicht integriert wird, sind die partiell verdeckten Objekte Bestandteile des Hintergrundmosaiks. Die Schraube, die aus der Szene entfernt wird, bleibt jedoch im Mosaikbild erhalten, wodurch die Szenenrepräsentation in diesem Bereich mit der Realität nicht übereinstimmt.

Kapitel 6

Realisierung

Die in den vorangegangenen Kapiteln vorgestellten Verfahren zur Bewegungsanalyse in Bildfolgen ermöglichen die Erstellung konsistenter Mosaikbilder bewegter Szenen. Im Gegensatz zu konventionellen Vorgehensweisen werden vor der Mosaikbildberechnung zunächst Veränderungen innerhalb der Bilder registriert. Durch ihre anschließende Berücksichtigung bei der Integration der Bilder ins Mosaikbild und ihre Beobachtung über die Zeit werden Integrationsfehler vermieden, und die dynamischen Informationen der aufgenommenen Szene bleiben erhalten. Die resultierenden Mosaikbilder bilden somit in Kombination mit den extrahierten Bewegungsdaten eine gute Ausgangsbasis für die Analyse der betrachteten Szene durch nachfolgende Bildverarbeitungskomponenten.

Das im Rahmen ihrer Diplomarbeit von Dirk Sypli und Heiko Tappe [ST99] implementierte Mosaikprogramm zur Verarbeitung von Bildfolgen statischer Szenen diente bei der Realisierung der vorgestellten Verfahren als Ausgangspunkt. Es wurde um geeignete Datenstrukturen und Algorithmen zur Bewegungsdetektion und zum Regionentracking erweitert, so dass jetzt eine Verarbeitung von Bildfolgen bewegter Szenen möglich ist. In diesem Kapitel werden die vorgenommenen Ergänzungen erläutert, wobei für Details der bereits vorhandenen Implementierung auf die oben zitierte Diplomarbeit verwiesen wird. Konkrete Hinweise zur Bedienung des Programms finden sich im Anhang C.

6.1 Datenstrukturen

Neben den bereits realisierten Datenstrukturen für Bilder und Transformationsparameter werden zur Bewegungsanalyse Repräsentationsformalismen für Bewegungskarten, Regionen, Zusammenhangskomponenten, Trajektorien und den Korrespondenzgraphen benötigt. Während die Bewegungskarten in die bereits vorhandenen Datenstrukturen zur Verwaltung der Bilder eingegliedert wurden, erfolgte die Realisierung der übrigen Datentypen durch eigene Klassen. Zum Teil fanden dabei vordefinierte Datentypen der LEDA-Bibliothek¹ Verwendung.

Bewegungskarten

Die interne Verwaltung der zu verarbeitenden Bilder wird durch einen Datentyp `matrix` realisiert. Da für die Schätzung der Transformationsparameter die Bilder im Allgemeinen in mehreren Auflösungen vorliegen müssen (vgl. Kap. 2.1.1), existiert ergänzend ein Datentyp `matrixpyramid`. Durch ihn werden die verschiedenen Auflösungsstufen eines Bildes zusammengefasst und gemeinsam an die Verfahren zur Parameterschätzung übergeben. Dieser Datentyp wurde für die Bewegungsdetektion um Matrizen zur Aufnahme der Bewegungskarten erweitert. Sie sind wiederum in einer Auflösungspyramide strukturiert, wobei aktuell nur die unterste Ebene tatsächlich Verwendung findet. Die Pyramide schafft jedoch die Grundlage für eine stärkere Einbindung der Bewegungsinformationen bei der Parameterschätzung, die zu einem späteren Zeitpunkt realisiert werden kann. Der Datentyp `matrix` selbst, der neben den Funktionen zur Speicherverwaltung auch Routinen zur Manipulation von Matrizen zur Verfügung stellt, wurde um Methoden zur Anwendung morphologischer Operationen, sowie zur Binarisierung und Invertierung ergänzt.

Regionen

Aus der Binarisierung der Bewegungskarten und der anschließenden Anwendung eines Dilationsoperators resultieren Regionen, die als Grundlage zur Extraktion der dynamischen Informationen dienen. Sie werden durch den Datentyp `region`

¹<http://www.mpi-sb.mpg.de/LEDA/leda.html>

realisiert. Er beinhaltet neben Datenfeldern für die Merkmale (Schwerpunkt, Pixelanzahl, Grauwert histogramm, umschreibendes Rechteck bzw. Polygon) unter anderem auch zwei Listen. In der einen werden die Abstände der Region zu den anderen, im selben Bild detektierten Regionen abgelegt, die andere verwaltet Zeiger auf die gemäß des Nachbarschaftskriteriums benachbarten Regionen. Auf diese Weise wird ein effizienter Zugriff auf die zur Berechnung der Zusammenhangskomponenten notwendigen Informationen gewährleistet. Um darüberhinaus auch die aus der Verwendung der Zusammenhangskomponenten und des Korrespondenzgraphen resultierenden Verknüpfungen geeignet repräsentieren zu können, enthält jede Region Zeiger auf die Komponenten, denen sie zugeordnet ist, sowie den korrespondierenden Knoten des Graphen.

Neben Methoden zum Auslesen der erläuterten Datenfelder sind in der Region-Klasse weiterhin Funktionen zur Berechnung der Merkmale, zur Ermittlung der Translationsrichtung (Kapitel 3.2.4) und für das Zusammenspiel mit den Zusammenhangskomponenten und dem Graphen zu finden. Die Routinen zur Berechnung der Abstände zwischen einzelnen Regionen, die im Rahmen der Generierung der Zusammenhangskomponenten benötigt werden, wurden dabei im Wesentlichen aus der Diplomarbeit von Sebastian Lang [Lan00] übernommen.

(Teil-)Zusammenhangskomponenten

Zur Verwaltung der Zusammenhangskomponenten dienen zwei Klassen. Die Klasse `component` stellt eine Basisklasse für Zusammenhangskomponenten dar. Sie beinhaltet neben Datenfeldern für die Merkmale der Komponenten und die Bewegungsrichtung jeweils Listen zum Ablegen der zugehörigen Regionen und Teilzusammenhangskomponenten. Ferner sind in jedem Objekt Verweise auf die zugeordneten Trajektorien und den Knoten im Korrespondenzgraphen zu finden. Die implementierten Methoden umfassen im Wesentlichen die Berechnungen der Komponentenmerkmale und der Teilzusammenhangskomponenten, sowie Funktionen zur Interaktion mit dem Graphen und den Trajektorien.

Die Teilzusammenhangskomponenten selbst wurden in der abgeleiteten Klasse `partComponent` implementiert. Diese stellt ergänzende Datenfelder und Methoden bereit, die eine effiziente, rekursive Generierung der Teilzusammenhangskomponenten unterstützen.

Korrespondenzgraph

Im Korrespondenzgraphen werden die zwischen den Bildern der Sequenz gefundenen Zuordnungen abgelegt. Er wird durch den Datentyp `correspondenceGraph` realisiert, der von parametrisierten LEDA-Graphen abgeleitet ist. In den Knoten des Graphen sind jeweils Zusammenhangskomponenten enthalten, deren Korrespondenzen über Kanten zwischen den Knoten repräsentiert werden. Jeder Kante wird dabei eine Signifikanz zugeordnet, die aus dem Vergleich der Komponenten resultiert. Der implementierte Datentyp erweitert die Funktionalität des LEDA-Datentyps um Datenfelder und Funktionen zum Einfügen von Zuordnungen, zur Verwaltung der parallel generierten Trajektorien und zum Zugriff auf die enthaltenen Zusammenhangskomponenten.

Trajektorien

Die Bewegungscharakteristik der detektierten und verfolgten Zusammenhangskomponenten wird durch Trajektorien beschrieben. Sie werden dem Korrespondenzgraphen in Form von Objekten des Typs `trajectory` zugeordnet. Die Trajektorien werden dabei aus mehreren Teilstücken zusammengesetzt. Auf diese Weise können Aufspaltungen von Bewegungsbahnen, wie sie beispielsweise beim Zerfall einer Region auftreten, unter Vermeidung von Redundanzen in den Trajektorien Daten effizient repräsentiert werden. Dementsprechend enthält eine Trajektorie neben einer Liste der Punkte auch Listen mit Zeigern auf vorausgehende und nachfolgende Trajektorienstücke. Die Punkte sind als eigener Datentyp `trajectoryPoint` implementiert. Jedem Punkt wird neben seinen Koordinaten zusätzlich ein Zeiger auf die zugehörige Komponente des Graphen zugeordnet, durch den die effiziente Extraktion der dynamischen Informationen möglich wird.

6.2 Algorithmen

Die realisierten Verfahren zur Bewegungsanalyse innerhalb von Bildfolgen lassen sich in Anlehnung an das dritte und vierte Kapitel dieser Arbeit in zwei Kategorien einteilen. In die erste Kategorie fallen die Routinen, die zur Erkennung von Bewegungen innerhalb der Sequenz eingesetzt werden, die zweite Kategorie

umfasst Algorithmen zur Segmentierung der Regionen und zur Suche nach Korrespondenzen zwischen diesen.

Zur Detektion von Bewegungen zwischen den Bildern wurden die drei alternativen Verfahren aus Kapitel 3 implementiert. Ihnen werden als Argumente jeweils die zu verarbeitenden Bilder sowie gegebenenfalls benötigte Ableitungen und Konfigurationsdaten (Nachbarschaftsgröße, Mindestgradient, usw.) übergeben. Als Ergebnis liefern sie die Bewegungskarten in Form einer Bildmatrix zurück, die weiterverarbeitet werden kann. Die nachfolgende Dilatation und die Schwellwertbinarisierung wurden in einer Funktion zusammengefasst. Diese enthält darüberhinaus Routinen zur Vor- und Nachbearbeitung (beispielsweise Invertierung) der Bewegungskarten, die jeweils von dem verwendeten Detektionsverfahren abhängen.

Die zeitliche Korrespondenzanalyse beginnt mit der Regionensegmentierung. Dazu werden zunächst zusammenhängende Bereiche in den binarisierten Bewegungskarten mit Hilfe eines Verfahrens zum *Region Labeling* eindeutig bezeichnet, sowie ihre umschließenden Rechtecke bestimmt. Durch diese können die bei der anschließenden Merkmalsberechnung zu betrachtenden Bildbereiche auf ein Minimum beschränkt werden. Für die Generierung der Zusammenhangskomponenten werden unter Anwendung des gewählten Distanzmaßes die paarweisen Abstände zwischen allen detektierten Regionen des aktuellen Bildes berechnet. Parallel dazu wird bereits überprüft, ob die Abstände zwischen zwei Regionen unterhalb der zulässigen Maximaldistanz liegen. Wenn dies der Fall ist, so wird die bestehende Nachbarschaft in den beteiligten Regionen vermerkt. Damit ist die Gruppierung der Regionen durch eine Suche nach Nachbarn in den entsprechenden Listen möglich. Die aktuellen Komponenten werden zusammen mit dem Korrespondenzgraphen an eine Funktion zur Suche nach Korrespondenzen übergeben. Unter Anwendung der gewählten Heuristik findet ein paarweiser Vergleich mit den Komponenten des Vorgängerbildes statt. Detektierte Zuordnungen werden direkt in den Graphen eingetragen, Komponenten, die nicht zugeordnet werden konnten, durch die klasseneigenen Routinen in Teilzusammenhangskomponenten zerlegt und weiter untersucht. Der letzte Schritt in der Verarbeitung eines Bildes besteht schließlich in dessen Integration in das Mosaikbild, wobei die berechneten Bewegungskarten berücksichtigt werden. Hierzu wurde die entsprechende Funktion der `matrix`-Klasse, die zur Integration neuer Bilder verwendet wird, geeignet

modifiziert.

Die nachfolgende Skizze zeigt die Einbettung der Bewegungsanalyse in den Prozess der Mosaikbildgenerierung im Überblick, wobei das bestehende Mosaikbild als Referenz zu Grunde gelegt wird:

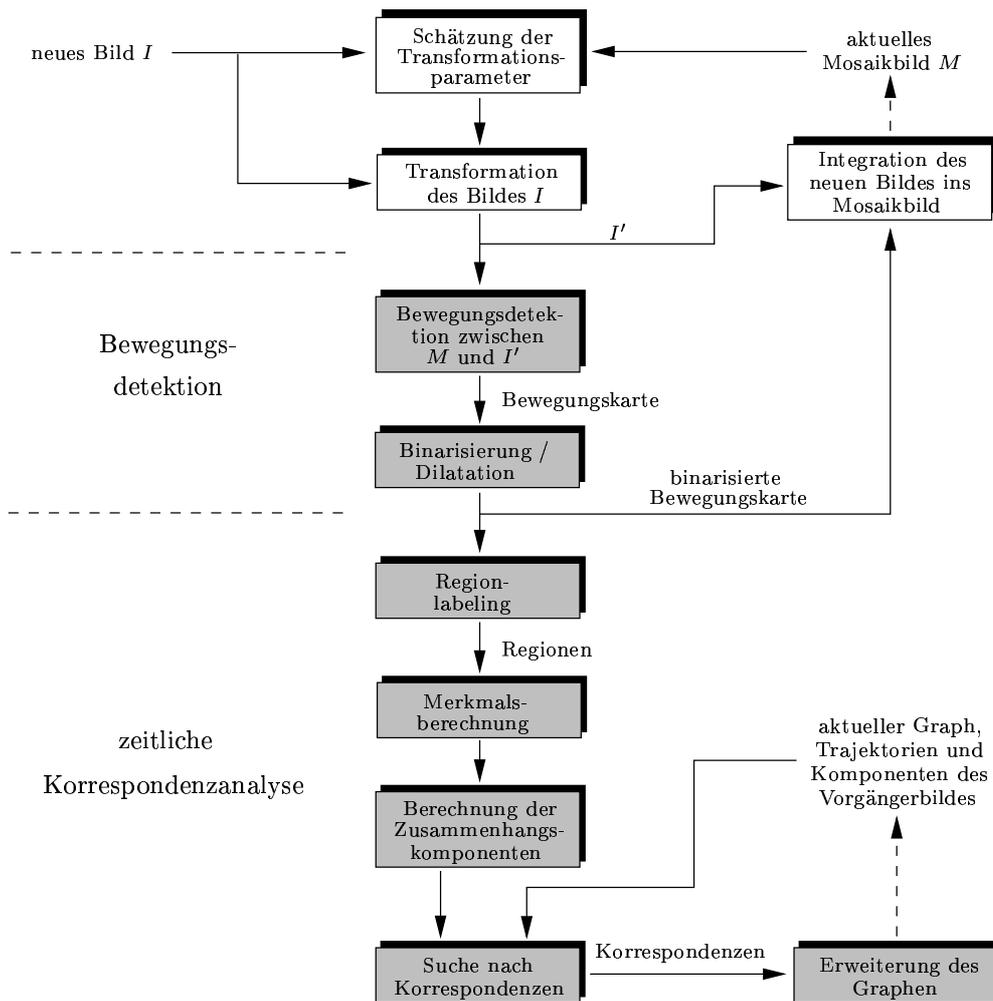


Abbildung 6.1: Einbettung der realisierten Bewegungsanalyse in die Mosaikbildgenerierung, dunkle Kästen markieren die neuen Routinen

Kapitel 7

Auswertung

Nachdem in den vorangegangenen Kapiteln die einzelnen Komponenten der Bewegungsanalyse im Wesentlichen unabhängig voneinander vorgestellt und untersucht wurden, soll an dieser Stelle abschließend eine integrierte Auswertung des realisierten Verfahrens erfolgen. Ausgehend von den Resultaten der Bewegungsdetektion, die durch Berechnung der Intensitätsdifferenz oder des Normal Flow erfolgt, werden für ausgewählte Bildfolgen aus dem Anhang B Mosaikbilder des statischen Szenenhintergrundes erzeugt. Parallel dazu findet im Rahmen der zeitlichen Korrespondenzanalyse die Extraktion der dynamischen Informationen der bewegten Objekte statt. Sie werden durch die resultierenden Trajektorien in den Bildern visualisiert. Die Diskussion der Ergebnisse umfasst insbesondere die Qualität und Konsistenz der Mosaikbilder mit der aktuellen Szene, die in der weiteren Verarbeitung eine entscheidende Rolle spielt. Außerdem wird die Eignung der gewonnenen Trajektorien für eine hinreichend exakte Beschreibung der Bewegungscharakteristiken der Objekte untersucht. Nur wenn sowohl das Mosaikbild und als auch die Trajektorien den Ansprüchen genügen, kann auf eine zusätzliche Analyse der Einzelbilder der verarbeiteten Bildfolge verzichtet werden.

Einen wichtigen Aspekt in der Auswertung der präsentierten Vorgehensweise stellt darüberhinaus die Analyse der Laufzeiten der implementierten Verfahren dar. Insbesondere im Hinblick auf eine Echtzeitverarbeitung von Bildfolgen in interaktiven Systemen sind effiziente Realisierungen der Algorithmen notwendig. Dirk Sypli und Heiko Tappe untersuchten im Rahmen ihrer Diplomarbeit

[ST99] die Komplexität der Transformationsparameterschätzung, die entscheidenden Einfluß auf den mit der Mosaikbildgenerierung verbundenen Gesamtaufwand hat. Aus den dort erzielten Resultaten geht hervor, dass derzeit eine schritthaltende Verarbeitung von Bildfolgen in Echtzeit nicht durchführbar ist. Die Laufzeiten der aufgesetzten Bewegungsanalyse werden daher vorrangig in Relation zu den Zeiten für die Parameterschätzung gesetzt, die als Maßstab für den Aufwand der zugefügten Routinen angesehen werden können (Abschnitt 7.2).

7.1 Ergebnisse

Zur Evaluation der Bewegungsdetektion und der zeitlichen Korrespondenzanalyse wurden die zu Grunde gelegten Bildfolgen aus dem Anhang im Allgemeinen auf Teilsequenzen beschränkt. In diesen ist jeweils eine signifikante Objektbewegung gegeben, während die nicht näher untersuchten Kamerabewegungen unter Umständen klein sind. Die Varianz innerhalb der Bilder einer Folge ist weiterhin gering, so dass geeignete Konfigurationsparameter für jede Folge gefunden werden konnten (siehe Tabelle 7.1). Bei der Interpretation der Ergebnisse ist dies zu berücksichtigen, da die Resultate in einer schritthaltenden Verarbeitung von Sequenzen mit größerer Varianz möglicherweise von den hier präsentierten abweichen (vergleiche auch Abschnitt 3.4). Als Referenzbild diente in Anlehnung an die Ausführungen im fünften Kapitel jeweils das aktuelle Mosaikbild.

Bei der Berechnung der Zusammenhangskomponenten aus den segmentierten Regionen wurde die Abstandsberechnung auf Basis des minimalen Punktabstandes verwendet, wobei sich als zulässige Maximaldistanz ein Wert von 10 als geeignet erwiesen hat. Um auch bei größeren Bewegungen und Veränderungen innerhalb der Schwerpunktpositionen sichere Zuordnungen vornehmen zu können, wurde die maximale Schwerpunktdistanz potenziell korrespondierender Komponenten in der Regel auf den hohen Wert von 60 gesetzt. In den hier untersuchten Bildfolgen ist die detektierte Komponentenanzahl im Allgemeinen klein, so dass der Aufwand für den paarweisen Vergleich einer Komponente mit allen potenziell korrespondierenden gering ist. In Folgen mit einer höheren Anzahl an Komponenten ist aber unter Umständen ein niedrigerer Wert zu wählen, der den Aufwand beschränkt, gleichzeitig aber auch nur kleinere Bewegungen erfasst.

Bildfolge	verarbeitete Bilder	Detektionsverfahren	minimaler Gradient	Schwellwert
Baufix1	30 – 69	Normal Flow ¹	3	2
Baufix2	15 – 49	Intensität	—	25
Baufix4	10 – 79	Normal Flow ¹	3	6
Baufix5	5 – 99	Intensität	—	35
Office1	0 – 50	Intensität	—	40
Desk1	36 – 98	Intensität	—	40
Desk2	20 – 79	Normal Flow	5	4

Tabelle 7.1: Konfigurationsparameter der Bewegungsdetektion in den Beispielen, die Größe der Nachbarschaft betrug in allen Fällen $\mu = 3$.

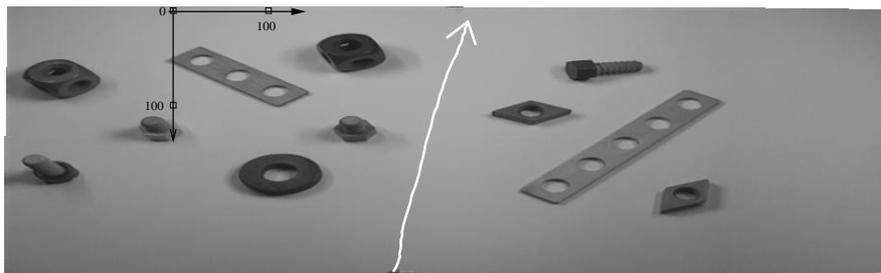


Abbildung 7.1: Mosaikbild des statischen Hintergrundes und rekonstruierte Trajektorie (weiß) des bewegten Objektes aus der Bildfolge "Baufix4"

Das erste Beispiel der Bildfolge "Baufix4" in Abbildung 7.1 zeigt zum Einstieg eine fehlerfreie Bewegungsdetektion, bei der die Trajektorie des bewegten Objektes exakt rekonstruiert werden kann. Als Orientierungspunkt ist hier, wie auch in den nachfolgenden Beispielen, das durch das erste verarbeitete Bild vorgegebene Referenzkoordinatensystem des Mosaikbildes eingezeichnet worden. Da das bewegte Objekt initial nicht in den Bildern der Folge enthalten ist und erst später in den Aufnahmebereich der Kamera gelangt, ist eine durchgehend vollständige Detektion möglich und im Mosaikbild sind keine Integrationsfehler zu erkennen.

¹Berechnung der Ableitungen nach Horn (Kap. 3.4)

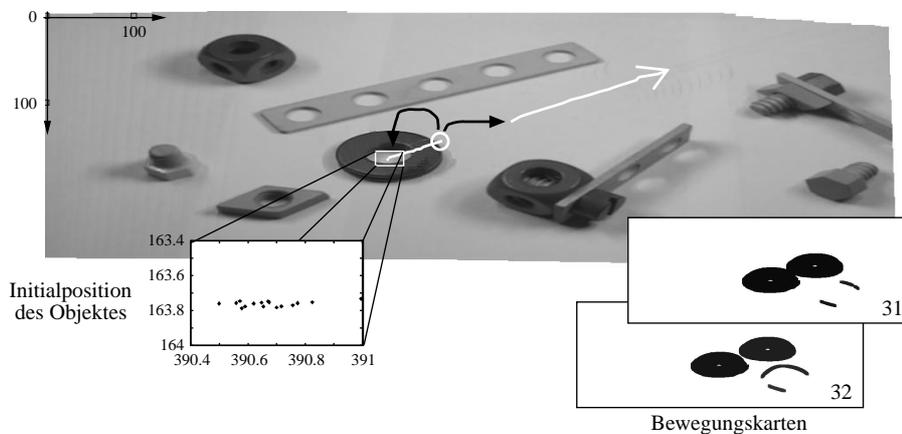


Abbildung 7.2: Die Ergebnisse für die Bildfolge "Baufix2": die Initialposition des Objektes verbleibt im Mosaikbild und die Trajektorie wird im Verlauf der Verfolgung aufgespalten (weißer Kreis, schwarze Pfeile).

Demgegenüber ist der bewegte Baufixring in Abbildung 7.2 zunächst Teil des statischen Szenenhintergrundes, bevor seine Bewegung beginnt. Wie in der Abbildung deutlich zu erkennen ist, verbleibt die Initialposition im Mosaikbild, das dadurch in dem betroffenen Bereich die tatsächlichen Gegebenheiten der Szene nur unzureichend repräsentiert. Alle weiteren Positionen des Ringes können allerdings auch in diesem Fall fehlerfrei und vollständig detektiert werden, so dass ein hochwertiges Mosaikbild generiert wird.

Die Trajektorie des Ringes wird im Wesentlichen korrekt extrahiert, wobei anzumerken ist, dass nach kurzer Verfolgungszeit eine Aufspaltung durch den Zerfall des bewegten Bereiches erfolgt. Zu Beginn der Bewegung überlagern sich die aktuellen Ringpositionen mit der Initialposition, so dass nur ein bewegter Bereich resultiert (vergleiche Bewegungskarte 31 in der Abbildung). Mit zunehmender Entfernung des Ringes von der Initialposition wird der Bereich auseinandergezogen, bis er schließlich entsprechend der nunmehr disjunkten Objektpositionen zerfällt (Bewegungskarte 32). Durch die zunächst noch vorhandene räumliche Nähe der beiden resultierenden Regionen werden sie in einer Zusammenhangskomponente gruppiert und dem zuvor detektierten Bereich zugeordnet. Auf diese Weise wird die vorhandene Trajektorie weiter expandiert. Erst wenn der wachsende Abstand eine Zusammenfassung innerhalb

einer Komponente nicht mehr zulässt, findet die, durch den weißen Kreis und die Pfeile in der Abbildung markierte Aufspaltung der Trajektorie statt. Die Schwerpunkte der beiden nunmehr detektierten Komponenten unterscheiden sich deutlich von dem Schwerpunkt der zuvor verfolgten Zusammenhangskomponente. Obwohl eine direkte Zuordnung der beiden neuen Komponenten zu der des Vorgängerbildes nicht möglich ist, lassen sich im Rahmen der Untersuchung der Teilzusammenhangskomponenten die vorhandenen Korrespondenzen ermitteln. Eine detailliertere Analyse der Trajektorienpunkte, die nach dem Zerfall für den der Initialposition zugeordneten bewegten Bereich resultieren (Diagramm unten links in der Abbildung 7.2), zeigt die in Kapitel 4.3 bereits angedeuteten Möglichkeiten zur nachträglichen Eliminierung der initialen Objektposition auf. Die geringe Varianz innerhalb der Punkte tritt offensichtlich hervor, so dass eine Unterscheidung dieses detektierten Bereichs von den tatsächlich bewegten, und eine nachträgliche, lokale Korrektur des Mosaikbildes realisierbar erscheint. Die Konstellation in der Bildfolge "Baufix5" (Abbildung 7.3) ist mit der des Vorgängerbeispiels vergleichbar. Das anfangs statische Baufixobjekt wird erfolgreich detektiert und verfolgt, wobei die Initialposition wiederum im Mosaikbild verbleibt. Auch hier weisen die Trajektorienpunkte jedoch eine sehr geringe Varianz auf, die eine Eliminierung der Initialposition ermöglichen sollte. Die Bewegung des Objektes in dieser Sequenz verläuft im Gegensatz zu den vorangegangenen Beispielen nicht stetig. Wie aus den Einzelbildern der Folge im Anhang ersichtlich wird, verändert sich die Objektposition nach einer Bewegung zu Beginn zwischenzeitlich nicht. Erst gegen Ende der Folge setzt sich die Bewegung fort. Das Objekt wird durchgehend als bewegt eingeordnet und auch während der Bewegungsunterbrechung nicht in das Mosaikbild aufgenommen, obwohl es zum betreffenden Zeitpunkt dem statischen Hintergrund zuzurechnen ist. Daraus folgt eine zwischenzeitlich inkonsistente Szenenrepräsentation durch das Mosaikbild. Aber auch in diesem Fall kann eine Analyse der Trajektorienpunkte nützlich sein. Die Unterbrechung tritt innerhalb der Trajektorie deutlich hervor (schwarzes Rechteck und Diagramm im oberen Teil der Abbildung), so dass ihre Analyse eine gute Ausgangsbasis bietet, um die Konsistenz des Mosaiks durch eine korrekte Behandlung der bewegten Bereiche wieder herzustellen.

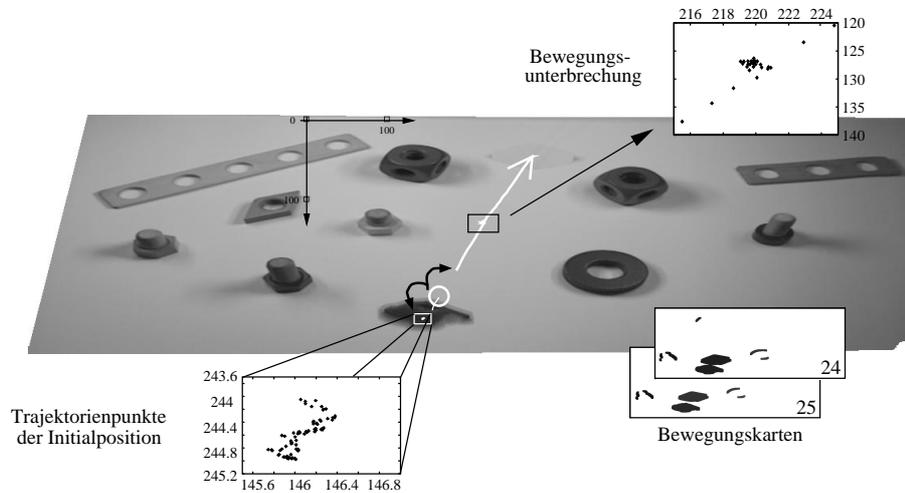


Abbildung 7.3: Das Resultat der Bewegungsanalyse in der Bildfolge "Baufix5": die Unterbrechung der Objektbewegung und die Initialposition lassen sich anhand der Trajektorienpunkte eindeutig identifizieren.

Die aus den Bildern der bislang untersuchten Folgen extrahierten Trajektorien repräsentieren die tatsächlichen Bewegungsrichtungen im Allgemeinen zutreffend. Die Segmentierungsergebnisse in den einzelnen Bildern variieren kaum. Dies liegt im Wesentlichen in der starren Form der beobachteten Objekte und ihrer ausreichenden Grauwertdifferenz gegenüber dem Hintergrund begründet. In der Bildfolge "Office1" wird die Bewegung einer Person verfolgt. Wie aus der Abbildung 7.4 deutlich wird, können die Positionen über die Folge hinweg erfolgreich lokalisiert werden und ein fehlerfreies Mosaikbild resultiert. Innerhalb der aus den Schwerpunktpositionen resultierenden Trajektorienpunkte liegen allerdings größere Varianzen vor. Das Objekt besitzt hier keine starre Form, so dass die Segmentierungsergebnisse zwischen den Bildern variieren und allein aus dieser Tatsache bereits Verschiebungen des Schwerpunktes folgen. Während in diesem Fall die Hauptbewegungsrichtung trotz des nicht geradlinigen Verlaufs der Trajektorie noch rekonstruierbar erscheint, müssen im Allgemeinen Veränderungen der Schwerpunktpositionen, die nicht allein aus der Bewegung des Objektes resultieren, explizit berücksichtigt werden. Sie können die tatsächliche Bewegungsrichtung unter Umständen signifikant verfälschen, wie das nachfolgende Beispiel aus der Bildfolge "Desk2" (Abb. 7.5) deutlich zeigt. Der hier bewegte Stift befindet



Abbildung 7.4: Mosaikbild der Bildfolge "Office1" und rekonstruierte Trajektorie, die auf Grund von Varianzen in den Segmentierungsergebnissen keinen geradlinigen Verlauf aufweist

sich anfangs nicht im Bild. Erst im weiteren Verlauf rollt er in den Aufnahmebereich der Kamera hinein, wobei allein durch die jeweils neu hinzukommenden Teile eine Verschiebung des Komponentenschwerpunktes verursacht wird. Aus diesem Grund weicht die berechnete Trajektorie zu Beginn deutlich von der tatsächlichen Bewegungsrichtung ab, die in der Abbildung durch den weißen Pfeil angedeutet wird. Eine exakte Rekonstruktion ist damit ohne weitergehende Maßnahmen zum Ausgleich der nicht durch die Bewegung verursachten Schwerpunktverschiebungen nicht möglich.

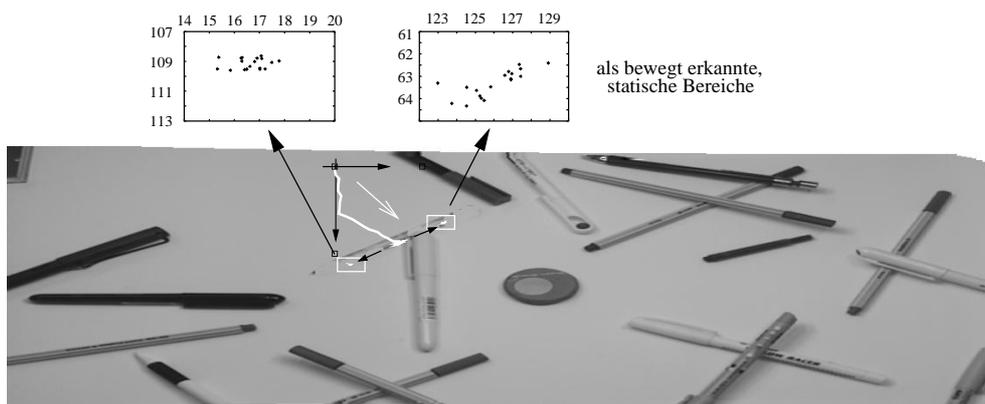


Abbildung 7.5: Resultierendes Mosaikbild der Bildfolge "Desk2" mit einer verfälschten Trajektorie und Integrationsfehlern, die aus einer unvollständigen Objektdetektion folgen

Neben der ungenauen Trajektorie innerhalb dieser Bildfolge sind im Mosaikbild auch Integrationsfehler zu erkennen. Die eingestellten Konfigurationsparameter der in diesem Fall zu Grunde gelegten Normal Flow Berechnung lassen offensichtlich eine durchgängige Detektion des bewegten Objektes nicht zu. Während der Stift anfangs noch vollständig erkannt wird, treten nach dem Ende seiner Bewegung Fehler in der Bewegungsdetektion auf. Insbesondere im mittleren Teil werden Objektpunkte im Mosaikbild integriert, obwohl zwischen den Bildern und dem Mosaikbild Unterschiede vorliegen. Der vormals weitgehend vollständig als bewegt klassifizierte Bereich zerfällt auf diese Weise, angedeutet durch die schwarzen Pfeile in der Abbildung. Während die zwei verbleibenden Regionen sich auf Grund ihrer relativ konstanten Schwerpunktpositionen als statisch identifizieren lassen, ist ein Ausgleich der Detektionsfehler im Zentrum des Objektes nur schwer möglich.

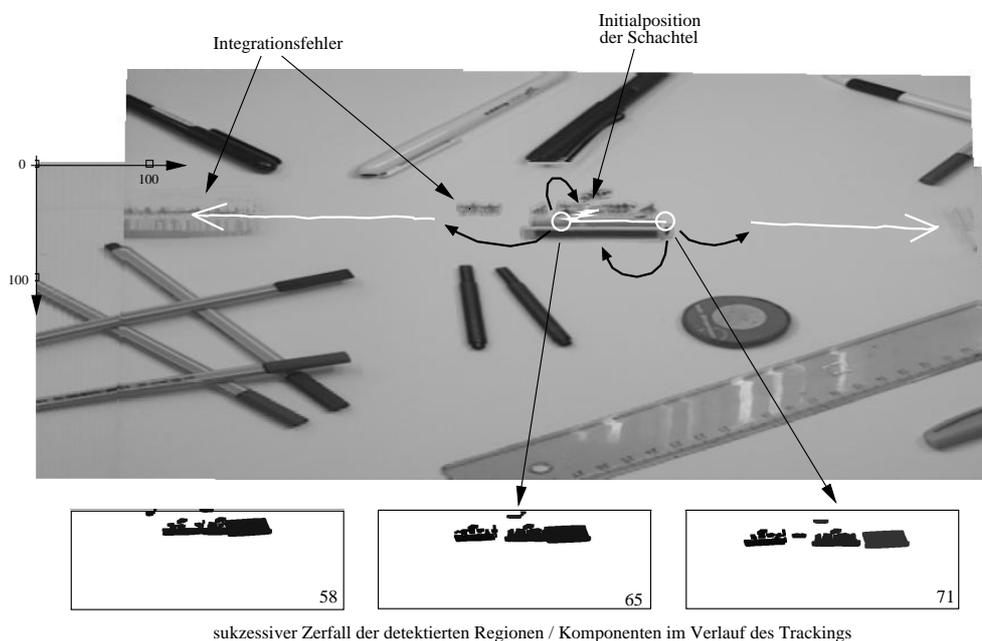


Abbildung 7.6: Mosaikbild und Trajektorien der Bildfolge "Desk1": eine unvollständige Detektion ist die Ursache für Integrationsfehler im Mosaikbild, während die Regionen trotz Zerfall verfolgt werden können.

Das aus einer Bewegungsdetektion unter Verwendung der Intensitätsdifferenz resultierende Hintergrundmosaikbild der Bildfolge "Desk1" in Abbildung 7.6 veranschaulicht noch einmal die bislang diskutierten Aspekte. In dieser Folge zerfällt das Objekt selbst in mehrere Einzelteile. Anfangs resultiert nur ein bewegter Bereich. Mit zunehmender Distanz zwischen den Teilstücken wird dieser Bereich dann stufenweise aufgespalten, wobei Sprünge der Schwerpunktpositionen in den Trajektorien (gekennzeichnet durch die weißen Kreise in der Abbildung) die Folge sind. Dennoch lassen sich für die einzelnen Teile die Bewegungsrichtungen korrekt rekonstruieren und auch von der Initialposition des Objektes in der Bildmitte unterscheiden. Aus der Beschaffenheit des beobachteten Objektes folgen aber auch in diesem Fall Probleme bei der Bewegungsdetektion und Integrationsfehler im Mosaikbild. Insbesondere das sich nach links bewegende Teilstück besitzt in einigen Objektbereichen nur eine sehr geringe Grauwertdifferenz gegenüber dem hellen Hintergrund. Eine vollständige Detektion ist nicht möglich, wodurch Fehler im Mosaik nicht ausgeschlossen werden können.

Alle betrachteten Beispiele illustrieren grundsätzlich die Leistungsfähigkeit des gewählten Ansatzes bei der Verarbeitung von Bildfolgen bewegter Szenen. Er ermöglicht die Generierung konsistenter Mosaikbilder und die Extraktion weitgehend zutreffender Trajektorien, sofern die Objekte sich vollständig detektieren lassen. Gleichzeitig werden allerdings auch Probleme in seiner Anwendung offenbar, die einen Ausblick auf verbleibende Fragestellungen und Verbesserungsmöglichkeiten für das Verfahren eröffnen. Einen vielversprechenden Ansatz scheint dabei insbesondere die erweiterte Analyse der Trajektorien Daten darzustellen, die im Rahmen dieser Arbeit noch nicht realisiert wurde. Durch eine Untersuchung der in den Trajektorienpunkten vorliegenden Varianz lassen sich vermutlich Initialpositionen von Objekten und Bewegungsunterbrechungen erkennen und bei der Erstellung konsistenter Mosaikbilder berücksichtigen.

Das abschließende Beispiel (Abbildung 7.7) demonstriert darüberhinaus auch die potenzielle Eignung der Trajektorienanalyse zur Rekonstruktion nur unvollständig detektierter Objekte. Diese Problematik wurde in Kapitel 3.4 bereits angesprochen, wobei als mögliche Lösung die Gruppierung sich parallel bewegender Komponenten zu einem Objekt vorgeschlagen wurde. Um die Trajektorien Daten in dieser Hinsicht zu evaluieren, wurden 20 aufeinanderfolgende Bilder

der Folge "Baufix1" ausgewählt und eine Bewegungsanalyse auf Basis des Normal Flow durchgeführt. Im Gegensatz zu den übrigen Beispielen wurde dabei zu Demonstrationszwecken der *Frame-To-Frame*-Modus verwendet, durch den eine Verschmelzung der bewegten Bereiche vermieden wird. Die Detektionsergebnisse bleiben weitgehend konstant und die Regionen werden, wie in den gezeigten Bewegungskarten durch die gestrichelten Linien angedeutet wird, jeweils in drei paarweise korrespondierende Zusammenhangskomponenten gruppiert. Zwischen diesen kann innerhalb eines Bildes keine Verbindung hergestellt werden, sie lassen sich aber im Rahmen der zeitlichen Korrespondenzanalyse verfolgen. Im Mosaikbild sind die drei resultierenden Trajektorien durch die weißen Pfeile markiert. Sie weisen deutliche Ähnlichkeiten auf, die als Ausgangspunkt für eine Zusammenfassung der Komponenten zu einem Objekt geeignet erscheinen und damit die grundsätzliche Eignung des Verfahrens zur Objektrekonstruktion bestätigen.

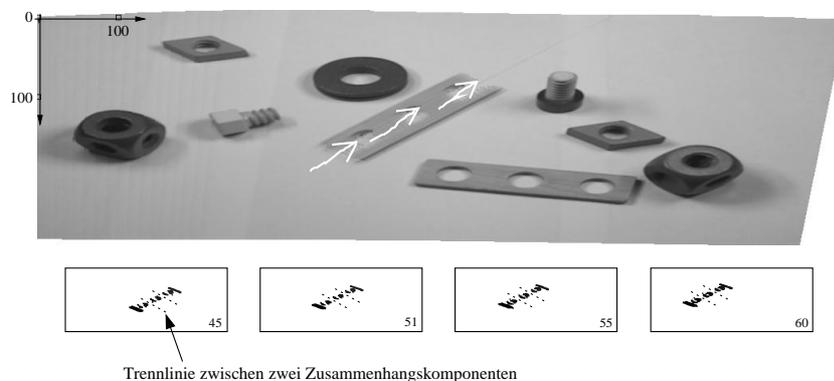


Abbildung 7.7: Demonstration der ähnlichen Trajektoriencharakteristiken sich parallel bewegnender Regionen als Ausgangspunkt für eine Objektrekonstruktion am Beispiel der Bildfolge "Baufix1"

7.2 Komplexitätsanalyse

Die Auswertung der im Rahmen dieser Diplomarbeit realisierten Erweiterungen des bestehenden Mosaikprogramms zeigt die Leistungsfähigkeit der präsentierten Vorgehensweise in der Verarbeitung von Bildfolgen bewegter Szenen. Die Qualität der Mosaikbilder und der Trajektorien­daten bietet eine solide Grundlage für eine detaillierte Szenenanalyse. Für die Erschließung breiter Einsatzfelder sollte die Verarbeitung der Bildfolgen allerdings auch effizient erfolgen, so dass eine weitgehend schritthaltende Integration der zu verarbeitenden Bilder erfolgen kann. Die Algorithmen zur Bewegungsanalyse werden im Folgenden in dieser Hinsicht untersucht. Dabei findet vorrangig ein Vergleich der Laufzeiten für die Bewegungsanalyse mit denen zur Schätzung der Transformationsparameter statt, die den größten Aufwand bei der Generierung von Mosaikbildern verursachen. Für die Komplexitätsanalyse wurden sieben Bildfolgen ausgewählt, auf denen jeweils sowohl mit der Intensitätsdifferenz als auch mit dem Normal Flow zufriedenstellende Ergebnisse erzielt werden konnten. Insgesamt standen dadurch 402 Bilder für die Evaluation zur Verfügung.

Auf jeder der Folgen wurde eine Bewegungsanalyse mit beiden Detektionsverfahren durchgeführt. Die Schätzung der Transformationsparameter erfolgte in allen Fällen unter Verwendung des *Perspective Flow*, da nur für dieses Verfahren der *Frame-To-Mosaic*-Modus implementiert ist. In Tabelle 7.2 sind die gemessenen mittleren Laufzeiten der einzelnen Schritte der Mosaikbildgenerierung und Bewegungsanalyse aufgelistet.

Es zeigt sich, dass der größte Aufwand in der Bewegungsanalyse durch die Bewegungsdetektion, die die Berechnung der Bewegungskarten umfasst, und die Regionensegmentierung gegeben ist. Die Berechnung des Normal Flow erweist sich als deutlich aufwändiger gegenüber der Intensitätsdifferenz, da für die zu untersuchenden Bilder zuvor die räumlichen Ableitungen und Gradientenbeträge zu bestimmen sind. Während die Regionensegmentierung (Schwellwertbinarisierung, Dilatation und Labeling) etwa eine halbe Sekunde benötigt, ist die Zeit für die Berechnung der Regionenmerkmale abhängig von den Größen der jeweils detektierten Regionen und ihrer Anzahl. Die angegebenen Werte sind daher nur als grobe Richtlinie anzusehen. In den untersuchten Folgen resultierten im Mittel

pro Bild etwa 50 ms für die Merkmalsberechnung, in den Folgen "Baufix3" und "Office1", die sehr große bewegte Objekte enthalten, allerdings schon zwischen 70 und 370 ms.

Diese beiden Bildfolgen verursachen auch bei der Berechnung der Zusammenhangskomponenten (polygonale Approximation und paarweise Abstandsberechnung zwischen den Regionen, sowie Generierung der Komponenten) sehr lange Laufzeiten. Unter Auslassung der beiden Folgen bei der Auswertung resultieren deutlich kleinere Zeiten, wie aus der Tabelle zu entnehmen ist. Insbesondere für den Normal Flow ist der Aufwand in den Folgen hoch, was auf ausgefranstere Regionen hindeutet, die sich nur unter größerem Aufwand durch umschreibende Polygone approximieren lassen.

	mittlere Laufzeiten²	
	Intensitätsdiff.	Normal Flow
Parameterschätzung	5177 ms (max. 15416 ms)	
Bewegungsdetektion	371 ms	1358 ms
Regionensegmentierung	540 ms	
Merkmalsberechnung	10 - 15 ms pro Region	
Zusammenhangskomponenten ³	54 ms (31 ms)	160 ms (38 ms)
Korrespondenzanalyse	≈ 1 ms	

Tabelle 7.2: Mittlere Laufzeiten der Parameterschätzung und Bewegungsanalyse (Mittelwerte pro Bild (≈ 756 x 278), sofern nicht anders angegeben)

Die abschließende Suche nach Korrespondenzen zwischen den Komponenten zweier aufeinanderfolgender Bilder ist im Verhältnis zu den übrigen Zeiten zu vernachlässigen. Die zuvor stattfindenden Maßnahmen zur Aufwandsreduktion, die im Wesentlichen im Ausschluss zu kleiner Regionen und der Überprüfung von Schwerpunktdistanzen und Größendifferenzen zwischen den Komponenten bestehen, führen dazu, dass nur noch sehr wenige Vergleiche tatsächlich stattfinden.

²gemessen auf einer DEC AlphaStation 500 / 400 MHz

³Die eingeklammerten Werte entsprechen den Mittelwerten unter Auslassung zweier Bildfolgen mit sehr großen bewegten Objekten (Details im Text).

Darüberhinaus ist in den hier untersuchten Bildfolgen die Anzahl bewegter Objekte jeweils klein. Pro Bild resultieren nur wenige Komponenten und ein Vergleich der Teilzusammenhangskomponenten ist nur selten erforderlich. Falls er dennoch durchgeführt werden muss, ist der damit verbundene Aufwand sehr gering, da nur wenige Regionen in den einzelnen Komponenten enthalten sind. In Bildfolgen, in denen mehrere kleine, unter Umständen eng benachbarte Objekte vorliegen, wird er entsprechend höher sein.

Die Laufzeiten für die Bewegungsanalyse liegen in jedem Fall jedoch unter denen für die Parameterschätzung. Während sie bei gleichbleibenden Bildgrößen nur mit der Anzahl detektierter Regionen skalieren, hängt der Aufwand für die Parameterschätzung von der Kameradynamik ab. Bei großen Bewegungen zwischen den Bildern treten Berechnungszeiten von 10 - 15 Sekunden auf, so dass der Aufwand für die Bewegungsanalyse dann zu vernachlässigen ist. Daraus lässt sich im Hinblick auf eine Echtzeitverarbeitung von Bildfolgen die Notwendigkeit ableiten, zunächst vorrangig effizientere Verfahren zur Parameterschätzung zu entwickeln.

Kapitel 8

Zusammenfassung & Ausblick

In der maschinellen Bildverarbeitung gewinnt die aktive Szenenexploration zunehmend an Bedeutung. Angelehnt an die aktive visuelle Wahrnehmung des Menschen werden mit Hilfe bewegter Kameras Bildfolgen des zu untersuchenden Szenarios aufgenommen und verarbeitet. Mit der auf diese Weise gewonnenen Flexibilität in der Akquisition visueller Informationen sind große Datenmengen verbunden, die einer effizienten Szenenanalyse entgegenstehen. Mosaikbilder bieten durch eine kompakte Repräsentation von Bildfolgen einen Ansatzpunkt für eine Aufwandsreduktion, die im Hinblick auf die Entwicklung interaktiver Systeme notwendig ist. Die im Allgemeinen gegebene Beschränkung der Generierungsalgorithmen auf die Verarbeitung von Bildfolgen statischer Szenen schließt dabei allerdings einen Einsatz in vielen Bereichen aus, in denen Bewegungen innerhalb der Szenen nicht ausgeschlossen werden können oder sollen. Die in dieser Arbeit vorgestellten und realisierten Ansätze zur Erweiterung bestehender Verfahren für die Verarbeitung von Bildfolgen bewegter Szenen leisten damit einen wichtigen Beitrag für die Ausdehnung der Einsatzfelder von Mosaikbildern.

In einem ersten Schritt wurden zunächst drei verschiedene Heuristiken für die Detektion von Bewegungen innerhalb von Bildfolgen untersucht. Es konnte gezeigt werden, dass die Berechnung der Intensitätsdifferenz und des Normal Flow, sowie die Analyse des Grauwertkonvergenzverhaltens bei erfüllten Voraussetzungen eine robuste Erkennung bewegter Bereiche zulassen. Allein in Bildausschnitten, in denen aus Veränderungen innerhalb der Szene keine oder nur geringe Abweichun-

gen in den Intensitäten resultieren, lassen sich Bewegungen nicht registrieren. Durch die Berücksichtigung der gewonnenen Bewegungsdaten bei der Integration der Bilder in ein Mosaikbild wird es möglich, eine qualitativ hochwertige Repräsentation des statischen Szenenhintergrundes zu erzeugen. Die üblicherweise auftretenden Integrationsfehler werden dabei weitgehend vermieden, sofern eine vollständige Erkennung der Objekte möglich ist. Eine Bewegungsdetektion relativ zum aktuellen Mosaikbild im *Frame-To-Mosaic*-Modus bietet eine gute Grundlage, um dieses Ziel zu erreichen. Dabei hat sich allerdings gezeigt, dass die als bewegt klassifizierten Bereiche genauer zu untersuchen sind. Aus Fehldetektionen, die beispielsweise bei den Initialpositionen anfangs statischer Objekte der Szene und bei zwischenzeitlichen Bewegungsunterbrechungen auftreten, resultieren sonst Integrationsfehler. In dieser Hinsicht sind Ergänzungen notwendig, durch die Mosaikbilder garantiert werden können, die den aktuellen Zustand der Szene in *allen* Bereichen korrekt wiedergeben.

Neben den statischen Daten sind im Rahmen der Mosaikbildgenerierung auch die dynamischen Informationen der Bildfolgen zu beachten. Sie gehen bei der Integration der Bilder in ein Mosaik verloren und müssen daher zuvor extrahiert werden. Mit der vorgestellten zeitlichen Korrespondenzanalyse konnte hierfür eine geeignete Vorgehensweise realisiert werden. Zwischen den bewegten Bereichen aufeinanderfolgender Bilder lassen sich auf Basis der zugehörigen Grauwertistogramme Ähnlichkeiten finden und Korrespondenzen etablieren. Die Berechnung der Schnittflächen der normalisierten Histogramme erlaubt dabei den Ausgleich von Größen- und Formveränderungen der detektierten Gebiete, die aus der Segmentierung resultieren und sich im gegebenen Kontext nicht vermeiden lassen. Durch die Gruppierung der Bereiche in Zusammenhangskomponenten konnten ferner der Aufwand bei der Suche erheblich reduziert und Varianzen in den Segmentierungsergebnissen ausgeglichen werden. Diese liegen zum einen in den zur Segmentierung eingesetzten Verfahren begründet. Zum anderen werden sie durch Zerfall oder Verschmelzung von Bereichen auf Grund entsprechender Ereignisse in den Bildern hervorgerufen. Auch diese können somit behandelt werden und lassen eine robuste Verfolgung über die Zeit zu.

Zur Repräsentation der Bewegungsdaten selbst konnte in Form von Trajektorien ein geeigneter Datentyp entwickelt werden. In sie werden die Schwerpunktpositio-

nen der verfolgten Bereiche als Punkte eingefügt und auf diese Weise einer weiteren Verarbeitung zugänglich gemacht. Die hier untersuchten, vorrangig translatorischen Objektbewegungen lassen sich durch sie zutreffend modellieren, ihre Eignung für komplexere Bewegungsmuster sollte jedoch weiter untersucht werden. Zu berücksichtigen ist darüberhinaus, dass aus Schwerpunktverschiebungen in den detektierten Regionen, die nicht durch Bewegungen verursacht werden, unter Umständen größere Abweichungen von der wahren Bewegungsrichtung resultieren können. Hierfür konnte noch keine adäquate Behandlungsmöglichkeit gefunden werden.

Eine detailliertere Analyse der Trajektoriendaten eröffnet im Gegensatz dazu Perspektiven zur Behandlung der verbliebenen Probleme im Hinblick auf die Konsistenz der Mosaikbilder. Die Trajektorien von detektierten Bereichen, die initialen Objektpositionen oder anderen fehlklassifizierten Gebieten zuzuordnen sind, weisen sehr geringe Varianzen innerhalb ihrer Punkte auf. Nach ersten Untersuchungen erscheinen diese zur eindeutigen Identifikation der Bereiche geeignet. Eine nachträgliche lokale Korrektur des Mosaikbildes, durch die die Konsistenz mit der aktuellen Szene wieder hergestellt werden kann, sollte sich damit realisieren und die Qualität der Bilder weiter erhöhen lassen. Auch für bei der Detektion zerfallende Objekte, die lediglich durch unzusammenhängende Teilbereiche in den Bewegungskarten hervortreten, erscheint mit diesem Ansatz eine Rekonstruktion möglich. Es konnte grundsätzlich demonstriert werden, dass Zusammenhangskomponenten, denen dieselbe Bewegungsursache zu Grunde liegt, ähnliche Trajektorien aufweisen und diese somit für eine Gruppierung potenziell herangezogen werden können.

Obwohl die bislang erzielten Resultate bereits eine robuste Verarbeitung von Bildfolgen bewegter Szenen zulassen, sollten im Rahmen einer stärkeren Ausrichtung der Mosaikbildgenerierung auf den Einsatz in interaktiven Systemen weitere Verbesserungen erfolgen. Insbesondere geeignete Konfigurationsparameter und Schwellwerte lassen sich in einer schritthaltenden Verarbeitung von Bildfolgen initial nur schwer festlegen. Durchgängig robuste Detektions- und Trackingergebnisse sind damit aktuell nicht notwendigerweise gewährleistet. Eine automatische Schwellwertadaption kann hier Abhilfe schaffen. Des Weiteren sollten auch Möglichkeiten untersucht werden, die Bewegungsinformationen aus den Bildfol-

gen in der Parameterschätzung stärker zu berücksichtigen. Durch den Ausschluss bewegter Pixel bei der Bestimmung der Parameter lassen sich möglicherweise robustere Transformationsparameter ermitteln und verbesserte Mosaikbilder für eine nachfolgende Analyse generieren.

Insgesamt konnte durch die realisierten Verfahren jedoch schon jetzt eine gute Ausgangsbasis für den Einsatz von Mosaikbildern in der Verarbeitung von Bildfolgen bewegter Szenen geschaffen werden. Sie vereinfachen die Analyse und Interpretation von Szenen in der maschinellen Bildverarbeitung und leisten damit einen wichtigen Beitrag für die Weiterentwicklung interaktiver Systeme.

Literaturverzeichnis

- [BEPR94] Moshe Ben-Ezra, Shmuel Peleg und Benny Rousso. Motion Segmentation Using Convergence Properties. *ARPA Image Understanding Workshop*, Seiten 1233–1235, November 1994.
- [Bom97] Thorsten Bomberg. Regionensegmentierung auf Bildfolgen mit Mechanismen des perzeptiven Gruppierens. Diplomarbeit, Universität Bielefeld, Technische Fakultät, AG Angewandte Informatik, 1997.
- [Bra99] Siegmund Brandt. *Datenanalyse*. Spektrum Akademischer Verlag, 4. Auflage, 1999.
- [CM98] Isaac Cohen und Gérard Medioni. Detection and Tracking of Objects in Airborne Video Imagery. *CVPR Workshop on Interpretation of Visual Motion*, 1998.
- [FA94] Cornelia Fermüller und Yiannis Aloimonos. Vision and Action. Technical report, Computer Vision Laboratory, Center for Automation Research, University of Maryland, Juni 1994.
- [GMS95] Michel Goosens, Frank Mittelbach und Alexander Samarin. *Der \LaTeX -Begleiter*. Addison-Wesley, 1995.
- [HS80] Berthold K.P. Horn und Brian G. Schunck. Determining Optical Flow. *Artificial Intelligence Memo No. 572*, April 1980.
- [IRP94] Michal Irani, Benny Rousso und Shmuel Peleg. Computing Occluding and Transparent Motions. *International Journal of Computer Vision (IJCV)*, 12(1):5–16, February 1994. (Kurzfassung in ECCV'92).

- [Jäh97] B. Jähne. *Digitale Bildverarbeitung*. Springer Verlag, 4. Auflage, 1997.
- [Lan00] Sebastian Lang. Perzeptives Gruppieren von Regionen. Diplomarbeit, Universität Bielefeld, Technische Fakultät, AG Angewandte Informatik, 2000.
- [MF92] Margaret W. Matlin und Hugh J. Foley. *Sensation and Perception*. 3. Auflage, 1992.
- [MS99] Rémi Mégret und Caterina Saraceno. Building the background mosaic of an image sequence. Technical Report PRIP-TR-060, TU Wien, 1999.
- [Nie81] Heinrich Niemann. *Pattern Analysis and Understanding*. Springer, Berlin, 2. Auflage, 1981.
- [PBA99] Robert Pless, Tomáš Brodský und Yiannis Aloimonos. Independent Motion: The Importance of History. In *ICVPR*, 1999.
- [PTVF92] William H. Press, Saul A. Teukolsky, William T. Vetterling und Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2. Auflage, 1992.
- [Reh98] Volker Rehrmann. Object Oriented Motion Estimation in Color Image Sequences. *Proc. of 5th ECCV*, 2.-6.Juni 1998. Freiburg.
- [SA96] Harpreet S. Sawhney und Serge Ayer. Compact Representations of Videos Through Dominant and Multiple Motion Estimation. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 18(8):814–830, August 1996.
- [Sch98] Werner X. Schneider. Visual-spatial working memory, attention, and scene representation: A neuro-cognitive theory. *Psychological Research*, 62:220–236, 1998.
- [Shu95] J. Kenneth Shultis. *TEX-Tips*. Prentice Hall Verlag GmbH, 1995.
- [Sin91] Ajit Singh. *Optic Flow Computation*. IEEE Computer Society Press, 1991.

- [ST99] Dirk Sypli und Heiko Tappe. Konstruktion von Mosaikbildern für die Bildanalyse. Diplomarbeit, Universität Bielefeld, Technische Fakultät, AG Angewandte Informatik, 1999.
- [Teu96] *Teubner - Taschenbuch der Mathematik*. B.G.Teubner Verlagsgesellschaft, Stuttgart/Leipzig, 1996.
- [WW88] Allen M. Waxman und Kwangyoen Wohn. Image Flow Theory: A Framework for 3D-Inference from Time-Varying Imagery. In *Advances in Computer Vision 1*. Lawrence Erlbaum Associates, 1988.

LITERATURVERZEICHNIS

Anhang A

Theorie des χ^2 -Tests

Nachfolgend werden die theoretischen Grundlagen des χ^2 -Tests erläutert. Weiterführende Details sind [Bra99] zu entnehmen.

Ausgangspunkt für den Test ist eine diskrete Zufallsvariable X . Aus einer Stichprobe vom Umfang n wird eine Verteilung für X gewonnen. In den einzelnen Intervallen i , $i = 1 \dots r$, werden dabei jeweils n_i Treffer gezählt, deren Gesamtsumme n entspricht. Zusätzlich wird eine weitere Verteilung ausgewählt, von der angenommen wird, dass die Stichprobe ihr folgt. Aus dieser Verteilung lassen sich erwartete Trefferzahlen t_i für die einzelnen Intervalle ermitteln:

$$t_i = n \cdot p_i \tag{A.1}$$

p_i entspricht der Wahrscheinlichkeit von Intervall i .

Zur Verifikation der Hypothese, dass die Stichproben-Verteilung der erwarteten Verteilung entspricht (Nullhypothese), wird zunächst die Größe u_i^2 berechnet:

$$u_i^2 = \frac{(n_i - t_i)^2}{n_i} \tag{A.2}$$

Falls die aufgestellte Hypothese zutrifft, so ist die Summe

$$\chi^2 = \sum_{i=1}^r u_i^2 \tag{A.3}$$

über die normalisierten quadratischen Abweichungen u_i^2 in den einzelnen Intervallen χ^2 -verteilt, wobei der Freiheitsgrad der Verteilung durch r festgelegt wird.

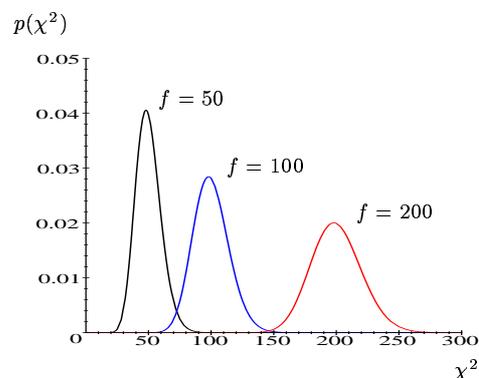


Abbildung A.1: χ^2 -Verteilungsfunktionen für verschiedene Freiheitsgrade f

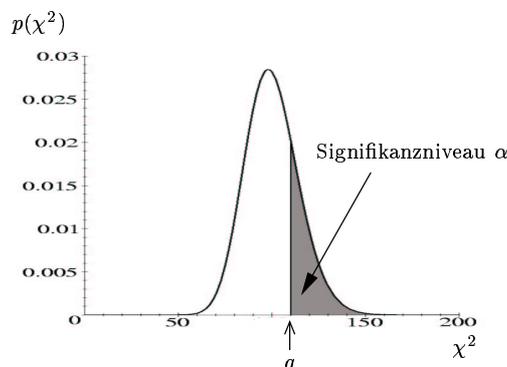


Abbildung A.2: Das Signifikanzniveau α beim χ^2 -Test

In Abb. A.1 sind exemplarisch Wahrscheinlichkeitsdichten der χ^2 -Verteilung für verschiedene Freiheitsgrade f dargestellt.

Anhand eines zu wählenden Signifikanzniveaus α , das der Wahrscheinlichkeit entspricht, mit der eine gegebene Abweichung q unter der Nullhypothese überschritten wird (Abb. A.2), lässt sich diese schließlich bestätigen oder zurückweisen. Mit jedem α ist ein maximal zulässiger Wert q (Quantil) für χ^2 verbunden, den die nach Gleichung A.3 errechnete Abweichung bei zutreffender Nullhypothese nicht übersteigen sollte. Falls der ermittelte Wert dennoch größer ist, so ist die Annahme, dass die Stichprobe der angenommenen Verteilung folgt, mit einer Irrtumswahrscheinlichkeit von α zurückzuweisen. Andernfalls besteht kein Grund dazu, die Nullhypothese abzulehnen. Die Quantile der χ^2 -Verteilung können entsprechenden Tabellen entnommen (z.B. [Teu96]) oder mit Hilfe von Computer-Algebra-Systemen (z.B. Maple) berechnet werden.

Durch Abänderung der Berechnungsformel A.2 für die u_i^2 kann der Test auch zum Vergleich von zwei, jeweils aus Stichproben hervorgegangenen Verteilungen eingesetzt werden.

Anhang B

Bildfolgen

Bildfolge "Baufix1"



(a) Bild 35



(b) Bild 40



(c) Bild 45



(d) Bild 48



(e) Bild 51



(f) Bild 54



(g) Bild 57



(h) Bild 60



(i) Bild 65

Abbildung B.1: Bildfolge Baufix1

Diese Bildfolge zeigt eine Szene aus dem Konstruktionszenario des SFB 360. Parallel zu einem horizontalen Schwenk der Kamera nach rechts bewegt sich die zentral in der Szene platzierte Holzleiste diagonal nach rechts oben.

Bildfolge "Baufix2"



(a) Bild 15



(b) Bild 22



(c) Bild 26



(d) Bild 30



(e) Bild 34



(f) Bild 38



(g) Bild 42



(h) Bild 46



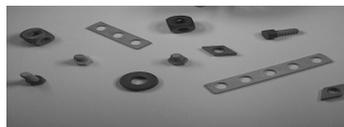
(i) Bild 49

Abbildung B.2: Bildfolge Baufix2

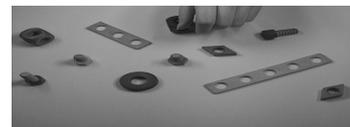
Bildfolge "Baufix3"



(a) Bild 0



(b) Bild 30



(c) Bild 35



(d) Bild 40



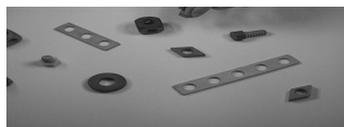
(e) Bild 45



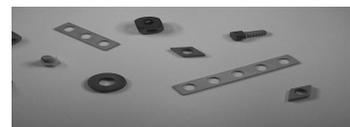
(f) Bild 50



(g) Bild 55



(h) Bild 60



(i) Bild 65

Abbildung B.3: Bildfolge Baufix3

Bildfolge "Baufix4"

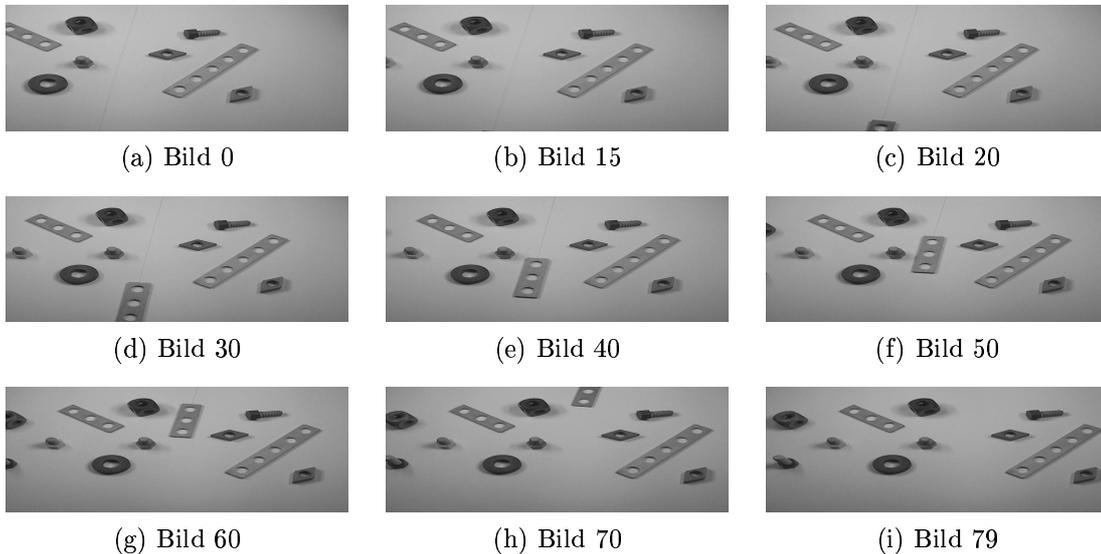


Abbildung B.4: Bildfolge Baufix4

Die drei auf dieser Doppelseite gezeigten Bildfolgen enthalten ebenfalls jeweils Objekte aus dem angesprochenen Konstruktionszenario. In der Bildfolge "Baufix2" bewegt sich der dunkle Ring in der Mitte der Bilder parallel zur 5-Lochleiste nach rechts oben, während die Kamera ebenfalls nach rechts schwenkt.

Innerhalb der Bildfolge "Baufix3" ist die Kamerabewegung dagegen sehr gering. Durch die ins Bild greifende Hand ist allerdings eine deutliche Objektbewegung in der Szene gegeben. Diese führt zu Verdeckungen von einzelnen Baufixteilen während die Rautenmutter in der Mitte der Szene entfernt wird.

In der Bildfolge "Baufix4" schließlich ist die Bewegung einer Leiste gezeigt, die initial nicht in den Bildern der Folge enthalten ist, und die den Aufnahmebereich der Kamera am Ende auch wieder verlässt. Sie bewegt sich von unten her in vertikaler Richtung durch die Szene.

Bildfolge "Baufix5"

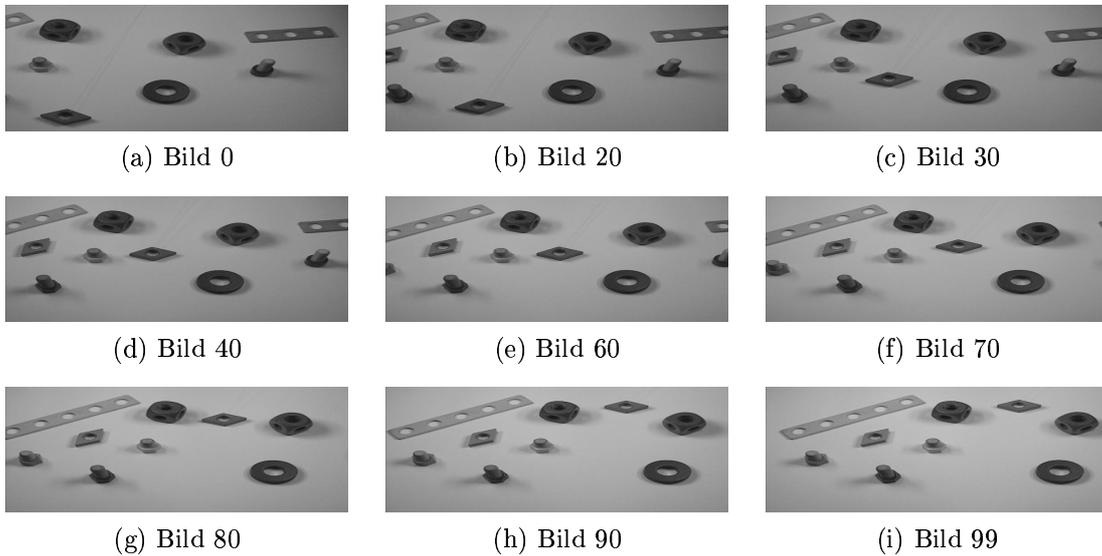


Abbildung B.5: Bildfolge Baufix5

Die letzte Folge des Baufixszenarios zeigt eine Szene, in der sich die anfangs unten links in den Bildern sichtbare Rautenmutter zentral im Aufnahmebereich der Kamera nach oben bewegt. Hier liegt im Gegensatz zu den vorangegangenen Bildfolgen keine stetige Bewegung vor. Die Objektbewegung wird etwa zwischen den Bildern 40 und 60 der Folge unterbrochen und dann fortgesetzt.

Auf der rechten Seite sind die Bildfolgen "Desk1" und "Desk2" abgebildet. Sie zeigen jeweils Aufnahmen von Büroutensilien. In der Folge "Desk1" wird die im oberen Teil der Szene enthaltene Streichholzschachtel im Verlauf der Aufnahme auseinandergezogen. Die Kamera bewegt sich zunächst horizontal nach rechts und später nach oben. In der zweiten Sequenz erfolgt lediglich ein Kameraschwenk nach links. Die Szene ist zu Beginn statisch. Erst später rollt von links oben ein Bleistift in den Aufnahmebereich, der an einen in der Szene enthaltenen Stift anstößt und dort liegenbleibt.

Bildfolge "Desk1"



(a) Bild 30



(b) Bild 40



(c) Bild 45



(d) Bild 50



(e) Bild 55



(f) Bild 60



(g) Bild 70



(h) Bild 80



(i) Bild 90

Abbildung B.6: Bildfolge Desk1

Bildfolge "Desk2"



(a) Bild 20



(b) Bild 25



(c) Bild 30



(d) Bild 35



(e) Bild 40



(f) Bild 45



(g) Bild 50



(h) Bild 60



(i) Bild 70

Abbildung B.7: Bildfolge Desk2

Bildfolge "Office1"



Abbildung B.8: Bildfolge Office1

In der letzten Bildfolge "Office1" ist die Bewegung einer Person in einem Büro-szenario gezeigt, die die Kamera verfolgt. Das bewegte Objekt ist hier groß im Verhältnis zum jeweils aufgenommenen Szenenausschnitt. Zusätzlich wird durch die Bildfolge ein deutlich größerer Bereich der zu Grunde liegenden Szene erfasst als in den Beispielen zuvor.

Anhang C

Hinweise zum Mosaikprogramm

Die Implementierung der Bewegungsanalyse wurde durch eine Ergänzung des von Dirk Sypli und Heiko Tappe erstellten Mosaikprogramms realisiert. Neben den in Kapitel 6 erläuterten, algorithmischen Erweiterungen und den neuen Datenstrukturen wurde für das Programm eine graphische Oberfläche erstellt (Abbildung C.1). Sie ermöglicht die komfortable Einstellung der Konfigurationsparameter für die Mosaikbildgenerierung und die Bewegungsanalyse. In Abhängigkeit der jeweils gewählten Optionen werden entsprechende Fenster eingeblendet, die die zur Verfügung stehenden Konfigurationsparameter enthalten. Des Weiteren lassen sich optional Zwischenergebnisse der Bewegungsdetektion in verschiedenen Fenstern ansehen und parallel zur Verarbeitung der Bildfolge evaluieren. Dazu werden elementare Zoomfunktionen bereitgestellt.

Da die nebenläufig implementierte Oberfläche selbst Rechenzeit verbraucht, ist das Programm im Hinblick auf eine effiziente Mosaikbildberechnung auch weiterhin in einer Kommandozeilenversion verfügbar. Diese wurde um die neuen Parameter zur Steuerung der Bewegungsanalyse erweitert. Sie sind in Tabelle C.1 aufgeführt, werden aber auch bei einem Aufruf des Programms ohne Argumente angezeigt. Die Konfigurationsdialoge der graphischen Oberfläche sind selbsterklärend.

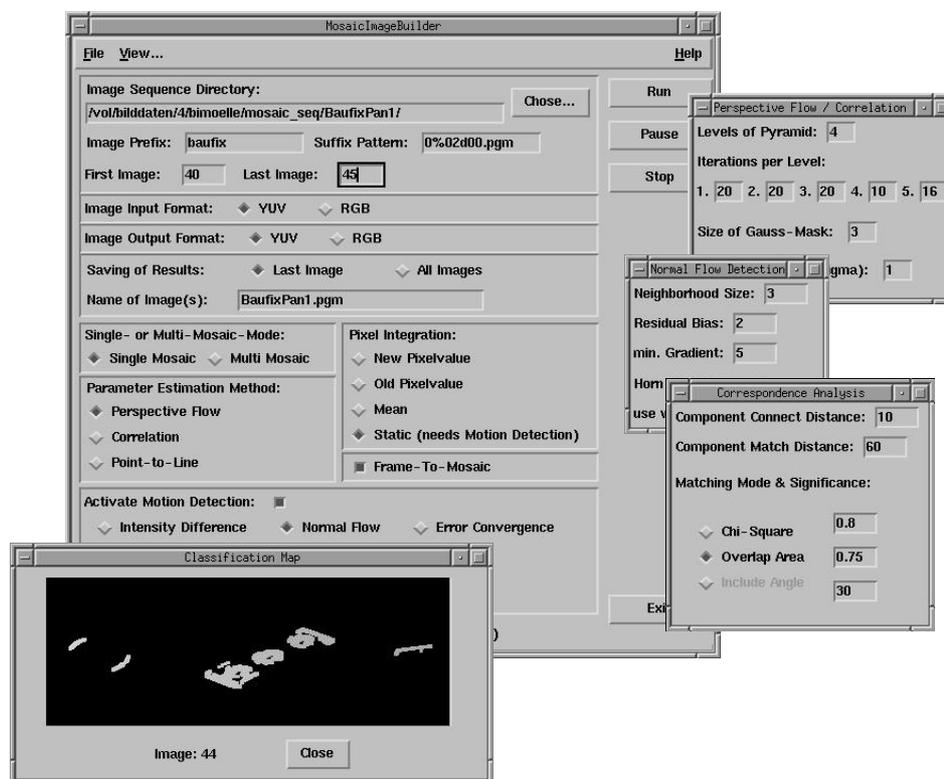


Abbildung C.1: Screenshot der graphischen Oberfläche

Kommandozeilen- schalter	Erläuterung
allgemein:	
-integrate <i>mode</i>	Integrationsmodus: "new", "old", "mean", "static"
Bewegungsdetektion:	
-motion <i>mode</i>	Detektionsverfahren: "intensityDiff", "normalFlow" oder "errorConvergence"
-motionMosaic	erstellt ein Mosaikbild auf Basis der bewegten Pixel
-neighborhood <i>size</i>	Größe der Nachbarschaft
-bias <i>value</i>	Binarisierungsschwellwert
-minGrad <i>value</i>	min. Gradient in der Normal Flow Berechnung
-weightedAverage	gewichteter mittlerer Normal Flow (Kap. 3.2.3)
-hornDerivatives	Ableitungsberechnung nach Horn (Kap. 3.4)
Korrespondenzanalyse:	
-compConnect <i>distance</i>	max. Abstand benachbarter Regionen
-compMatch <i>distance</i>	max. Schwerpunkt Abstand korrespondierender Komponenten
-matchMode <i>mode</i>	Vergleichsheuristik für die Regionen und Komponenten : "chi" oder "overlap"
-significance <i>value</i>	Mindestähnlichkeit für eine erfolgreiche Zuordnung

Tabelle C.1: Kommandozeilenschalter für die Bewegungsanalyse