

Hochschule Darmstadt

– Fachbereich Mathematik und
Naturwissenschaften –

Prototypische Umsetzung: Analyse und Vergleich von ausgewählten Methoden zur Berechnung von Bewegungsvektoren in Drohnenvideos

Abschlussarbeit zur Erlangung des akademischen Grades
Master of Science (M. Sc.)

vorgelegt von

Hannah Eleen Schult

Matrikelnummer: 1129010

Referent : Prof. Dr. Elke Hergenröther
Korreferent : Prof. Dr. Timo Schürg
Ausgabedatum : 15.12.2025
Abgabedatum : 25.06.2026

ERKLÄRUNG

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht.

Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen.

Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

Darmstadt, 24. Juni 2026

Hannah Eleen Schult

ABSTRACT

Automated drone surveillance can help to detect and defuse critical situations in large crowds at an early stage. Reliable analysis of movement patterns in drone videos is central to this, for example to prevent mass panics. Optical flow methods for determining motion vectors are an important tool, but their performance in drone scenarios with complex perspectives and dynamics has hardly been systematically investigated so far.

This thesis develops a prototypical and reproducible pipeline for estimating motion vectors in drone videos and compares three established methods: Lucas-Kanade (classical, sparse), Farnebäck (classical, dense), and RAFT (deep learning, dense). The basis consists of seven real drone videos from a festival with varying scene complexities, supplemented by manual ground-truth annotations. All methods are tested under identical conditions. The research question is: *Which of the selected classical and deep learning-based methods delivers the quantitatively and qualitatively best results for drone videos in terms of accuracy, robustness, and runtime when calculating motion vectors?*

The results show that classical methods (Lucas-Kanade and Farnebäck) provide more robust and accurate motion vectors in the examined drone videos than RAFT. Lucas-Kanade achieves the lowest mean endpoint error (approx. 1.29 px), Farnebäck is similarly precise and most reliably detects the diversity of directions in crowd dynamics. RAFT has higher errors and is less robust in complex drone scenarios. In terms of runtime, Lucas-Kanade is the only real-time capable method (approx. 7 ms/frame); Farnebäck and RAFT are considerably slower and unsuitable for time-critical applications.

The analysis provides clear recommendations: For real-time applications and limited resources, Lucas-Kanade is optimal; Farnebäck is suitable for detailed crowd analyses; RAFT can be an addition for offline evaluation and after targeted domain fine-tuning. This work establishes a solid basis for comparison, highlights methodological strengths and weaknesses, and provides practical recommendations for the selection of optical flow methods in drone-based movement scenarios. The outlook emphasizes the importance of larger ground-truth datasets, specific training of deep learning models, and optimizations for real-time and edge applications.

Thus, this master's thesis contributes to the methodological and practical advancement of movement analysis in drone videos and lays the foundation for future, practical applications in automated crowd surveillance.

ZUSAMMENFASSUNG

Automatisierte Drohnenüberwachung kann helfen, kritische Situationen in großen Menschenmengen frühzeitig zu erkennen und zu entschärfen. Die zuverlässige Analyse von Bewegungsmustern in Drohnenvideos ist dabei zentral, etwa zur Prävention von Massenpaniken. Optical-Flow-Methoden zur Bestimmung von Bewegungsvektoren sind ein wichtiges Werkzeug, doch deren Leistungsfähigkeit in Drohnenszenen mit komplexer Perspektive und Dynamik ist bislang kaum systematisch untersucht.

Diese Arbeit entwickelt eine prototypische und reproduzierbare Pipeline zur Schätzung von Bewegungsvektoren in Drohnenvideos und vergleicht drei etablierte Methoden: Lucas-Kanade (klassisch, sparse), Farnebäck (klassisch, dense) und RAFT (Deep Learning, dense). Basis sind sieben reale Drohnenvideos eines Festivals mit unterschiedlichen Szenenkomplexitäten, ergänzt durch manuelle Ground-Truth-Annotationen. Alle Verfahren werden unter identischen Bedingungen getestet. Die Forschungsfrage lautet hierbei: *Welche der ausgewählten klassischen und Deep-Learning-basierten Methoden liefert für Drohnenvideos in Bezug auf Genauigkeit, Robustheit und Rechenzeit die quantitativ und qualitativ hochwertigsten Ergebnisse bei der Berechnung von Bewegungsvektoren?*

Die Ergebnisse zeigen, dass klassische Verfahren (Lucas-Kanade und Farnebäck) in den untersuchten Drohnenvideos robustere und genauere Bewegungsvektoren liefern als RAFT. Lucas-Kanade erzielt den geringsten durchschnittlichen Endpunktfehler (ca. 1,29 px), Farnebäck ist ähnlich präzise und erkennt zudem die Richtungsvielfalt in der Crowd-Dynamik am zuverlässigsten. RAFT hat im Drohnenkontext höhere Fehler und ist weniger robust in komplexen Szene. Hinsichtlich Rechenzeit ist Lucas-Kanade als einziges Verfahren echtzeitfähig (ca. 7 ms/ Frame), Farnebäck und RAFT sind deutlich langsamer und für zeitkritische Anwendungen ungeeignet.

Die Analyse liefert klare Empfehlungen: Für Echtzeitanwendungen und begrenzte Ressourcen ist Lucas-Kanade optimal, Farnebäck eignet sich für detaillierte Crowd-Analysen, RAFT kann bei Offline-Auswertung und nach gezieltem Domänen-Fine-Tuning eine Ergänzung sein. Die Arbeit schafft eine fundierte Vergleichsbasis, zeigt methodische Stärken und Grenzen auf und liefert Handlungsempfehlungen für die Auswahl von Optical-Flow-Verfahren in drohnenbasierten Bewegungsszenarien. Der Ausblick betont die Bedeutung größerer Ground-Truth-Datensätze, spezifisches Training von Deep-Learning-Modellen sowie Optimierungen für Echtzeit- und Edge-Anwendungen.

Die Masterarbeit trägt somit zur methodischen und praktischen Weiterentwicklung der Bewegungsanalyse in Drohnenvideos bei und legt die Basis für zukünftige, praxisnahe Anwendungen in der automatisierten Crowd-Überwachung.

Inhaltsverzeichnis

I	Thesis	
1	Einleitung	2
1.1	Motivation und Problemstellung	2
1.2	Zielsetzung und Forschungsfrage der Arbeit	3
1.3	Aufbau der Arbeit	4
2	Stand der Forschung	5
2.1	Optische Fluss Methoden	5
2.2	Optischer Fluss im Kontext von Menschenmengen	7
2.3	Qualitätsbewertung ohne Ground-Truth-Daten	8
3	Herausforderungen	10
3.1	Herausforderungen dieser Arbeit	10
3.2	Herausforderungen bei Drohnenvideos	11
3.3	Verschiedene Lösungsansätze	14
4	Theoretische Grundlagen	26
4.1	Bewegungsvektoren und optischer Fluss	26
4.2	Verfahren zur Bewegungsvektorenberechnung	28
4.2.1	Lucas-Kanade-Ansatz	28
4.2.2	Farnebäck-Algorithmus	29
4.2.3	RAFT	30
5	Datenbasis und Vorverarbeitung	32
5.1	Beschreibung der verwendeten Daten	32
5.2	Charakterisierung und Ablauf der Szenen	33
5.2.1	Charakterisierung der Videos	33
5.2.2	Ablaufbeschreibung der Videos	34
5.2.3	Zuteilung der Videos zu den Kategorien	36
5.3	Vorverarbeitung der Videodaten	38
5.4	Annotation und Ground-Truth-Daten	39
6	Methodik	42
6.1	Pipeline Konzept (Überblick über das methodische Vorgehen) .	43
6.2	Detailbeschreibung der Pipeline-Schritte	45
6.3	Technische Umsetzung / Realisierung	48
7	Experimente	51
7.1	Übersicht der Experimente	51
7.2	Experiment 1: Quantitative Evaluation anhand von Ground-Truth-Daten	52
7.3	Experiment 2: Quantitativer Methodenvergleich ohne Ground-Truth-Daten	54

7.4	Experiment 3: Qualitätsbewertung mittels Forward-Backward Consistency	55
7.5	Experiment 4: Crowd-Dynamik-Analyse	56
7.6	Experiment 5: Laufzeitvergleich	58
7.7	Zusammenfassende Bewertung	59
8	Ergebnisse	60
8.1	Evaluation anhand von Ground-Truth-Daten	60
8.2	Quantitativer Methodenvergleich ohne Ground-Truth-Daten . .	64
8.3	Qualitätsbewertung mittels Forward-Backward Consistency . .	67
8.4	Crowd-Dynamik-Analyse	70
8.5	Laufzeitvergleich	83
9	Diskussion und Beantwortung der Forschungsfrage	86
9.1	Evaluation gegen Ground-Truth-Daten	86
9.2	Quantitativer Methodenvergleich ohne Ground-Truth-Daten . .	87
9.3	Qualitätsbewertung mittels Forward-Backward-Consistency . .	89
9.4	Crowd-Dynamik-Analyse	90
9.5	Laufzeitvergleich	93
9.6	Kritische Reflexion der Methodik	94
9.6.1	Datenauswahl	94
9.6.2	Manuelle Annotation und potenzielle Fehlerquellen . .	95
9.7	Beantwortung der Forschungsfrage	96
10	Fazit und Ausblick	98
10.1	Fazit	98
10.2	Ausblick	99
 II Appendix		
A	Appendix: Weitere Tabellen	101
	 Literatur	106

Abbildungsverzeichnis

Abbildung 5.1	Jeweils ein stellvertretendes Standbild pro Video. (Eigene Darstellung.)	36
Abbildung 5.2	Screenshots des eigens entwickelten Annotationstools zur Markierung von Punktpaaren zwischen zwei Frames. (Eigene Darstellung.)	40
Abbildung 6.1	Methodische Pipeline der Arbeit. (Eigene Darstellung.)	43
Abbildung 6.2	Prototypische Visualisierung der Bewegungsvektoren aller drei Methoden auf einem Beispielframe. (Eigene Darstellung.)	49
Abbildung 8.1	Gesamtvergleich von Mean, Median und Max EPE der Methoden Lucas-Kanade (LK), Farnebäck (FB) und Recurrent All-Pairs Field Transforms (RAFT) im Vergleich zu den Ground-Truth-Daten. (Eigene Darstellung.)	60
Abbildung 8.2	EPE-Verteilung im Vergleich zu den Ground-Truth-Daten. Links: Boxplot der gesamten EPE-Verteilung für die Methoden LK, FB und RAFT. Rechts: Durchschnittlicher EPE-Wert pro Video und Methode mit Standardabweichung. (Eigene Darstellung.)	61
Abbildung 8.3	Winkelfehler im Vergleich zu den Ground-Truth-Daten. Links: Boxplot der gesamten Winkelfehler für die Methoden LK, FB und RAFT. Rechts: Durchschnittlicher Winkelfehler pro Video und Methode mit Standardabweichung. (Eigene Darstellung.)	62
Abbildung 8.4	Gruppiertes Boxplot des Endpunktfehler (EPE) (in Px) für die Methoden LK, FB und RAFT zu den Ground-Truth-Daten, aufgeteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.)	62
Abbildung 8.5	Gruppiertes Boxplot des Winkelfehlers (in Grad) für die Methoden LK, FB und RAFT zu den Ground-Truth-Daten, aufgeteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.)	63
Abbildung 8.6	Zusammenhang zwischen EPE und Bewegungsamplitude für die Methoden LK, FB und RAFT. (Eigene Darstellung.)	64

Abbildung 8.7	Boxplots der EPE-Abweichungen zwischen allen Methoden (LK, FB, RAFT) im Vergleich, jeweils unterteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.)	66
Abbildung 8.8	Boxplots der Winkelabweichungen zwischen allen Methoden (LK, FB, RAFT) im Vergleich, jeweils unterteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.)	66
Abbildung 8.9	Mittlerer Forward-Backward-Error (in Pixel) für LK, FB und RAFT , jeweils unterteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.) . . .	68
Abbildung 8.10	Consistency-Karten mit stellvertretenden Bildern für LK , FB und RAFT . (Eigene Darstellung.)	70
Abbildung 8.11	Boxplots der Richtungs-Entropie pro Video für die Methode Lucas-Kanade. (Eigene Darstellung.)	71
Abbildung 8.12	Boxplots der Richtungs-Entropie pro Video für die Methode Farneäck. (Eigene Darstellung.)	72
Abbildung 8.13	Boxplots der Richtungs-Entropie pro Video für die Methode RAFT . (Eigene Darstellung.)	72
Abbildung 8.14	Boxplots der Richtungs-Uniformität pro Video für die Methode Lucas-Kanade. (Eigene Darstellung.)	73
Abbildung 8.15	Boxplots der Richtungs-Uniformität pro Video für die Methode Farneäck. (Eigene Darstellung.)	73
Abbildung 8.16	Boxplots der Richtungs-Uniformität pro Video für die Methode RAFT . (Eigene Darstellung.)	74
Abbildung 8.17	Vergleich der Bewegungs-Heatmaps in der Moshpit-Szene für vier Zeitpunkte: Jeweils links Farneäck, rechts RAFT . (Eigene Darstellung.)	75
Abbildung 8.18	Boxplots der Geschwindigkeitsverteilung pro Video für die Methode Lucas-Kanade. (Eigene Darstellung.) .	76
Abbildung 8.19	Boxplots der Geschwindigkeitsverteilung pro Video für die Methode Farneäck. (Eigene Darstellung.) . . .	76
Abbildung 8.20	Boxplots der Geschwindigkeitsverteilung pro Video für die Methode RAFT . (Eigene Darstellung.)	77
Abbildung 8.21	Amplitude-Heatmaps exemplarischen Frames: Links Farneäck, rechts RAFT . (Eigene Darstellung.)	78
Abbildung 8.22	Mittlere Divergenz für alle Videos mit Lucas-Kanade. (Eigene Darstellung.)	79
Abbildung 8.23	Mittlere Divergenz für alle Videos mit Farneäck. (Eigene Darstellung.)	80
Abbildung 8.24	Mittlere Divergenz für alle Videos mit RAFT . (Eigene Darstellung.)	80
Abbildung 8.25	Mittlerer Curl für alle Videos mit Lucas-Kanade. (Eigene Darstellung.)	81
Abbildung 8.26	Mittlerer Curl für alle Videos mit Farneäck. (Eigene Darstellung.)	81

Abbildung 8.27	Mittlerer Curl für alle Videos mit RAFT . (Eigene Darstellung.)	82
Abbildung 8.28	Vergleich des Zusammenhangs zwischen mittlerer Divergenz und mittlerem Curl für alle Videos, jeweils für Lucas-Kanade (links), Farnebäck (Mitte) und RAFT (rechts). (Eigene Darstellung.)	82
Abbildung 8.29	Mittlere Laufzeit pro Frame für alle Methoden. Links: Mittelwert und Standardabweichung der Laufzeit pro Frame über alle Videos. Rechts: Laufzeit pro Video und Methode (Lucas-Kanade, Farnebäck, RAFT). (Eigene Darstellung.)	83
Abbildung 8.30	Mittlere Laufzeit (links) und Verhältnis Komplex/Einfach (rechts) für die Methoden Lucas-Kanade (LK), Farnebäck (FB) und RAFT . (Eigene Darstellung.)	84
Abbildung 8.31	Erreichbare FPS im Vergleich zu den Video-FPS (30 FPS) für die Methoden LK , FB und RAFT und für alle Videos. (Eigene Darstellung.)	85

Tabellenverzeichnis

Tabelle 3.1	Überblick über die betrachteten Lösungsansätze und deren Einsatz in der Arbeit.	15
Tabelle 5.1	Eigenschaften der für die Optical-Flow-Analyse verwendeten Videos	33
Tabelle 5.2	Bewertungskriterien für die Zuordnung der Videos zu den Szenentypen.	37
Tabelle 5.3	Zuordnung der Videos zu den Szenenkategorien. . . .	38
Tabelle 5.4	Übersicht der Annotationen: Anzahl der Frame-Paare und Punktpaare pro Video	41
Tabelle 6.1	Überblick über zentrale Software- und Hardwarekomponenten.	50
Tabelle 7.1	Übersicht aller durchgeführten Experimente.	52
Tabelle 8.1	Zusammenfassung der Evaluation anhand von Ground-Truth-Daten	64
Tabelle 8.2	Zusammengefasste Vergleichsmetriken der Methoden (paarweise, aggregiert über alle Videos)	65
Tabelle 8.3	Vergleich der wichtigsten Metriken zwischen einfachen und komplexen Szenen für verschiedene Methodenpaare.	67
Tabelle 8.4	Forward-Backward-Consistency (FB-Cons.) für LK, FB , RAFT auf allen Videos.	69
Tabelle 8.5	Forward-Backward-Error (FB-Error) für LK, FB , RAFT auf allen Videos.	69
Tabelle 8.6	Laufzeiten (Mittelwerte) pro Frame für jedes Video und Methode.	84
Tabelle 8.7	Hardware-Kontext der Laufzeitanalyse.	85
Tabelle A.1	Kategorisierung der Videos von Person A	101
Tabelle A.2	Kategorisierung der Videos von Person B	102
Tabelle A.3	Mean±Std der EPE-Metriken der Methoden (paarweise) für alle Videos	102
Tabelle A.4	Median der EPE-Metriken der Methoden (paarweise) für alle Videos	103
Tabelle A.5	RMSE der Methoden (paarweise) für alle Videos	103
Tabelle A.6	Fl-all (%) der Methoden (paarweise) für alle Videos . .	103
Tabelle A.7	Winkelvergleiche (°) der Methoden (paarweise) für alle Videos	104
Tabelle A.8	Lucas-Kanade: Entropie, Uniformität, Divergenz und Curl (Mittelwert ± Standardabweichung) pro Video . .	104

Tabelle A.9	Farnebäck: Entropie, Uniformität, Divergenz und Curl (Mittelwert $\pm$ Standardabweichung) pro Video	105
Tabelle A.10	RAFT: Entropie, Uniformität, Divergenz und Curl (Mit- telwert $\pm$ Standardabweichung) pro Video	105

ABKÜRZUNGSVERZEICHNIS

CNN	Convolutional Neural Network
EPE	Endpunktfehler
FB	Farneback
FPS	Frames per second
GRU	Gated Recurrent Unit
HS	Horn-Schunck
LK	Lucas-Kanade
RAFT	Recurrent All-Pairs Field Transforms
YOLO	You Only Look Once

Teil I

THESIS

EINLEITUNG

Dieses Kapitel führt in die Problemstellung, Motivation und Zielsetzung der Arbeit ein. Es erläutert die Relevanz der automatisierten Bewegungsanalyse in Drohnenvideos zur Überwachung von Menschenmengen und stellt die zentrale Forschungsfrage sowie den Aufbau der Arbeit vor.

1.1 MOTIVATION UND PROBLEMSTELLUNG

Große Menschenmengen können schnell zu lebensbedrohlichen Situationen führen, wie die tragischen Ereignisse bei der Loveparade 2010 in Duisburg und der Massenpanik bei einer Feierlichkeit im Rahmen einer Halloween-Party 2022 auf offener Straße in Seoul eindrücklich verdeutlichen. In beiden Fällen führten räumliche Engpässe, Panikreaktionen und unzureichende Überwachungssysteme zu verheerenden Folgen. In Duisburg starben 21 Menschen und es gab Hunderte Verletzte [1]. In Seoul gab es mehr als 150 Todesopfer [2]. Diese Katastrophen unterstreichen die kritische Bedeutung einer effektiven Überwachung und Echtzeitanalyse von Menschenströmen bei Großveranstaltungen, um gefährliche Entwicklungen frühzeitig zu erkennen und präventive Maßnahmen rechtzeitig einzuleiten.

In den letzten Jahren haben sich Drohnen zunehmend in der Sicherheitsüberwachung etabliert [3, 4]. Beispielsweise in der Grenzüberwachung in Echtzeit für eine schnelle Bedrohungserkennung [5], in der Erkennung von Landminen [6] oder auch in der Überwachung sowie Analyse von Menschenmengen [7]. Drohnen ermöglichen eine flexible, kosteneffiziente und großflächige Erfassung des Geschehens aus der Luft und erzeugen Videodaten, welche wertvolle Einblicke in Bewegungsmuster und Dynamiken geben. Eine manuelle Auswertung dieser Daten ist jedoch zeitaufwendig, fehleranfällig und für eine Echtzeitauswertung ungeeignet [4]. Daher ist die Entwicklung automatisierter Verfahren zur Analyse von Bewegungen in Drohnenvideos von großer Bedeutung.

Allerdings sind Drohnenaufnahmen, anders als bodennahe Aufnahmen, oft von starken Perspektiv- und Maßstabsänderungen sowie deutlichen Kamerabewegungen geprägt [8, 9]. Das führt dazu, dass viele etablierte Verfahren der klassischen Bildanalyse bei Drohnenaufnahmen an ihre Grenzen stoßen, da sie spezifische Aufnahmewinkel oder Bildgeometrien voraussetzen, die im realen Einsatz häufig nicht gegeben sind. Dies erschwert insbesondere die zuverlässige Erkennung und Quantifizierung von Bewegungsmustern in dichten Menschenmengen.

Ein entscheidender Bestandteil automatisierter Verfahren zur Auswertung von Drohnenvideos ist die Identifikation und Analyse von Bewegungsvektoren [10–12]. Bewegungsvektoren beschreiben die zeitliche Verschiebung von

Bildinhalten zwischen aufeinanderfolgenden Videoframes [13] und liefern somit quantitative Informationen über Richtung, Dynamik und Geschwindigkeit von Menschen oder Objekten. Dadurch lassen sich Bewegungsmuster systematisch quantifizieren, Anomalien erkennen und potenziell gefährliche Situationen frühzeitig detektieren.

Die Kombination von Drohnenvideos und modernen Deep-Learning-Methoden eröffnet neue Möglichkeiten für die automatische Analyse von Menschenmengen. Durch die Integration solcher Verfahren können Sicherheitsbehörden und Veranstalter in Zukunft datengetriebene Entscheidungen treffen und Risiken besser bewerten.

Es zeigt sich in der Literatur, dass optische Flussverfahren im Vergleich zu anderen Methoden zur Bewegungserkennung besonders effizient und robust sind, insbesondere bei der Analyse von Menschenmengen unter schwierigen Bedingungen [14]. Daher widmet sich diese Arbeit schwerpunktmäßig der Untersuchung von optischen Fluss-Algorithmen und deren Anwendung in der Crowd-Analyse.

Vor diesem Hintergrund ergibt sich die zentrale Problemstellung dieser Arbeit: Es wird systematisch untersucht, wie verschiedene Bewegungsvektorverfahren (sowohl klassische als auch Deep-Learning-basierte) unter unterschiedlichen Szenenbedingungen abschneiden. Ziel ist es, die Grenzen und Potenziale der Methoden zu identifizieren und Empfehlungen für den praktischen Einsatz in drohnenbasierten Bewegungsszenarien abzuleiten.

1.2 ZIELSETZUNG UND FORSCHUNGSFRAGE DER ARBEIT

Ziel dieser Masterarbeit ist die prototypische Implementierung und der systematische Vergleich klassischer und Deep-Learning-basierter Verfahren zur Berechnung von Bewegungsvektoren in Drohnenvideos. Im Mittelpunkt steht die Analyse der Leistungsfähigkeit ausgewählter Methoden unter verschiedenen Szenenbedingungen. Es soll herausgearbeitet werden, welcher Ansatz die zuverlässigsten Ergebnisse im Hinblick auf Genauigkeit, Robustheit und Rechenzeit erzielt.

Der Schwerpunkt der Arbeit liegt primär auf der Berechnung und Analyse der Bewegungsvektoren selbst. Nachgelagerte Aufgaben wie die automatisierte Anomalieerkennung oder die semantische Interpretation von Bewegungsmustern sind nicht Gegenstand dieser Untersuchung. Vielmehr wird die Qualität der Bewegungsvektorschätzung als Grundlage für weiterführende Anwendungen evaluiert.

Die zentrale Forschungsfrage dieser Arbeit lautet:

Welche der ausgewählten klassischen und Deep-Learning-basierten Methoden liefert für Drohnenvideos in Bezug auf Genauigkeit, Robustheit und Rechenzeit die quantitativ und qualitativ hochwertigsten Ergebnisse bei der Berechnung von Bewegungsvektoren?

Durch die Beantwortung dieser Forschungsfrage soll eine fundierte Grundlage für die Auswahl und Weiterentwicklung von Bewegungsvektorverfahren

ren im Kontext drohnenbasierter Videoanalyse geschaffen werden. Die Ergebnisse sollen dazu beitragen, die Praxistauglichkeit sowie Einsatzgrenzen aktueller Methoden kritisch zu bewerten und Empfehlungen für zukünftige Anwendungen zu formulieren.

1.3 AUFBAU DER ARBEIT

Die vorliegende Arbeit ist wie folgt strukturiert: Nach der Einleitung, in der Motivation, Problemstellung und Zielsetzung dargelegt werden, folgt in Kapitel 2 ein umfassender Überblick über den Stand der Forschung. Hier werden sowohl klassische als auch Deep-Learning-basierte optische Fluss-Methoden im Kontext von Drohnenvideos und Crowd-Analyse dargestellt und hinsichtlich ihrer Leistungsfähigkeit diskutiert. Kapitel 3 widmet sich den spezifischen Herausforderungen bei der Bewegungsanalyse in Drohnenvideos und stellt mögliche Lösungsansätze vor. Im Anschluss werden in Kapitel 4 die theoretischen Grundlagen zur Berechnung von Bewegungsvektoren sowie die ausgewählten Methoden detailliert erläutert. Kapitel 5 beschreibt die verwendeten Datensätze, die Vorverarbeitungsschritte sowie die Erstellung des Ground-Truth-Datensatzes. Die methodische Pipeline und deren Umsetzung werden in Kapitel 6 vorgestellt. In Kapitel 7 werden die durchgeführten Experimente und deren Aufbau beschrieben, bevor in Kapitel 8 die Ergebnisse systematisch präsentiert und ausgewertet werden. Kapitel 9 beinhaltet die Diskussion der Ergebnisse, eine kritische Reflexion der Methodik und die Beantwortung der Forschungsfrage. Den Abschluss bildet in Kapitel 10 die zusammenfassende Darstellung der wesentlichen Erkenntnisse und der Ausblick auf zukünftige Forschungsarbeiten.

STAND DER FORSCHUNG

Dieses Kapitel gibt einen Überblick über den aktuellen Stand der Forschung zur Bewegungsanalyse in Drohnenvideos. Es werden sowohl klassische als auch Deep-Learning-basierte optische Fluss Methoden betrachtet, deren Leistungsfähigkeit im Kontext von Menschenmengen analysiert und Methoden zur Qualitätsbewertung, auch ohne Ground-Truth-Daten, diskutiert.

2.1 OPTISCHE FLUSS METHODEN

Optische Fluss Methoden dienen der Schätzung von Bewegungen in Bildfolgen, indem sie Verschiebungen von Bildpunkten zwischen aufeinanderfolgenden Frames berechnen und sind damit ein zentrales Werkzeug für zahlreiche Anwendungen der Computer Vision wie Bewegungsanalyse oder Objekterkennung [13]. Die Schätzung des optischen Flusses ist ein zentrales Thema der Computer Vision und hat in den letzten Jahren durch die Verfügbarkeit großer Benchmark-Datensätze wie dem Middlebury-Optical-Flow-Benchmark [13] und durch Fortschritte bei Methoden und Optimierungstechniken signifikante Verbesserungen erfahren. Zu den klassischen Methoden zählt der Lucas-Kanade-Algorithmus [15], der auf der Annahme basiert, dass die Helligkeit lokaler Bildregionen über aufeinanderfolgende Frames konstant bleibt und die Bewegung innerhalb eines Patches als translationsinvariant angenommen werden kann. Diese Methode ist effizient und besonders für kleine Bewegungen und wenig Rauschen geeignet. In zahlreichen Studien wurde Lucas-Kanade als robust und rechenzeiteffizient beschrieben, insbesondere in Szenarien mit geringem Bildrauschen und moderaten Bewegungen [13, 14].

Der Ansatz von Farnebäck [16] berechnet einen dichten optischen Fluss. Farnebäck approximiert lokale Nachbarschaften durch quadratische Polynome und erlaubt so die Berechnung eines dichten Bewegungsfeldes. Diese Methode ist weniger empfindlich gegenüber Rauschen und kann größere Bewegungen besser abbilden als klassische, lokale Verfahren. Wie mehrere Evaluationen zeigen, liefert Farnebäck robuste Resultate in komplexen Szenen, ist jedoch rechenintensiver als Lucas-Kanade [13, 14].

In den vergangenen Jahren haben Deep-Learning-Verfahren wie RAFT [17] das Feld entscheidend geprägt. RAFT nutzt ein rekurrentes Matching von All-Pairs-Korrelationen zwischen allen Pixelpaaren in zwei Bildern und erreicht damit auf den gängigen Benchmarks wie KITTI und Sintel eine deutlich höhere Genauigkeit als klassische Methoden [18]. Allerdings zeigen aktuelle Benchmark-Studien [18], dass Deep-Learning-Modelle wie RAFT zwar auf i.i.d.-Testdaten sehr gut abschneiden, aber unter Verteilungen außerhalb

des Trainingsdatensatzes (Out-of-Distribution) signifikante Schwächen aufweisen.

Baker et al. [13] etablierten mit Middlebury einen umfassenden Benchmark, der die Leistung von Algorithmen unter verschiedenen Szenarien evaluiert. Ihre Ergebnisse zeigen, dass klassische Methoden wie Lucas-Kanade und Farneäck in vielen praktischen Anwendungen nach wie vor konkurrenzfähig sind, insbesondere wenn Rechenzeit oder Robustheit gegenüber Bildrauschen entscheidend sind. Neuere Untersuchungen, wie in FlowBench [18], vergleichen explizit die Robustheit und Generalisierungsfähigkeit moderner Deep-Learning-Methoden wie RAFT mit klassischen Verfahren. Sie zeigen, dass klassische Methoden unter gezielten Störungen am Input oft stabiler sind, während RAFT in idealen, bekannten Testumgebungen die besten Ergebnisse erzielt.

Für die Analyse von Drohnenvideos, die typischerweise geprägt sind durch hohe Bewegungsdynamik, große Objekte, wechselnde Beleuchtung und Okklusionen [8, 9], ist die Wahl des Flussalgorithmus entscheidend. Untersuchungen [14] belegen, dass klassische Verfahren wie Lucas-Kanade und Farneäck besonders bei hoher Bildrate und moderaten Bewegungen verlässlich sind, während RAFT aufgrund seiner Fähigkeit, große Bewegungen und komplexe Szenen zu modellieren, auch unter schwierigen Bedingungen sehr genaue Ergebnisse liefert. Allerdings erfordert RAFT eine deutlich höhere Rechenleistung und ist weniger robust gegenüber Bildrauschen und Verteilungsverschiebungen.

Für die Auswahl der in dieser Arbeit eingesetzten Algorithmen zur optischen Flussberechnung (Lucas-Kanade, Farneäck und RAFT) wurden verschiedene Kriterien berücksichtigt. Einerseits sollen die Methoden sowohl klassische als auch moderne Ansätze abdecken, andererseits müssen sie hinsichtlich Genauigkeit, Robustheit und Praxistauglichkeit für Drohnenvideos überzeugen. Andere etablierte Verfahren wie Horn-Schunck (HS) [19], FlowNet [20] oder PWC-Net [21] werden bewusst nicht eingesetzt.

Das klassische Horn-Schunck-Verfahren [19] basiert auf einer globalen Optimierung mit quadratischer Glättung. Wie zahlreiche Vergleichsstudien zeigen [13, 14, 22], ist HS gegenüber Bildrauschen und Helligkeitsänderungen sehr empfindlich und neigt zu deutlicher Überschätzung der Glättung, was zu einer Verfälschung insbesondere an Bewegungsgrenzen und bei großen Bewegungen führt. Gerade in realen, komplexen Szenen, wie sie bei Drohnenflügen häufig sind, liefert Horn-Schunck daher deutlich schlechtere Ergebnisse als moderne dichte Verfahren oder robuste lokale Ansätze.

Bei den Deep-Learning-basierten Ansätzen sind FlowNet [20] und PWC-Net [21] wichtige Meilensteine. Jedoch zeigen die aktuellen Benchmarks und Metastudien [18], dass beide Architekturen in Bezug auf Genauigkeit, Robustheit und Generalisierungsfähigkeit modernen Methoden wie RAFT klar unterlegen sind. FlowBench [18] demonstriert explizit, dass FlowNet und PWC-Net sowohl hinsichtlich Zuverlässigkeit als auch Generalisierungsfähigkeit deutlich schlechter abschneiden als RAFT. Insbesondere unter Verteilungsverschiebungen brechen FlowNet und PWC-Net stark ein, während

[RAFT](#) auch bei größeren Bewegungen und komplexeren Szenen eine höhere Genauigkeit bietet.

Zusammenfassend werden Horn-Schunck, FlowNet und PWC-Net in dieser Arbeit nicht berücksichtigt, da sie entweder in Bezug auf Robustheit ([HS](#)) oder auf Genauigkeit und Generalisierungsfähigkeit (FlowNet, PWC-Net) nachweislich schlechtere Resultate liefern als die ausgewählten Methoden Lucas-Kanade, Farnebäck und [RAFT](#). Während Horn-Schunck als klassischer Vertreter globaler, glatter Verfahren in realen Szenarien zu wenig robust ist, sind FlowNet und PWC-Net gegenüber modernen Deep-Learning-Methoden wie [RAFT](#) heute nicht mehr konkurrenzfähig. Die Auswahl orientiert sich daher an aktuellen Literaturergebnissen und Benchmarks [[13](#), [14](#), [18](#), [23](#)].

2.2 OPTISCHER FLUSS IM KONTEXT VON MENSCHENMENGEN

Optische Flussverfahren sind ein zentrales Werkzeug zur Analyse und Überwachung von Menschenmengen, insbesondere zur Schätzung von Bewegungsrichtung und -geschwindigkeit von Personen in dicht besetzten oder komplexen Szenen. In der Literatur werden insbesondere klassische Algorithmen wie Lucas-Kanade und seine Varianten sowie dichte Verfahren wie Farnebäck regelmäßig für Crowd-Analysen evaluiert und eingesetzt [[14](#)]. Deep-Learning-basierte Ansätze sind im Kontext von Crowd-Analysen noch vergleichsweise neu, zeigen aber in jüngsten Studien großes Potenzial [[24](#)].

Der Lucas-Kanade-Algorithmus wird vielfach in Crowd-Szenarien verwendet, um stabile Featurepunkte zu extrahieren und Bewegungsmuster auch bei hoher Dichte und Okklusion zu erfassen. Gang et al. [[25](#)] sowie Solmaz und Moore [[26](#)] zeigen, dass Lucas-Kanade zuverlässig verschiedene Crowd-Verhaltensmuster wie Bottlenecks, Lanes oder Blockaden identifizieren kann, ohne dass explizite Objektverfolgung oder Training notwendig sind. Weiterentwicklungen wie der Kanade-Lucas-Tomasi-Tracker ermöglichen zudem eine effiziente Anomaliedetektion durch Vergleich von aktuellem und erwartbarem Bewegungsverhalten [[27](#), [28](#)]. Die Robustheit gegen Rauschen und die Fähigkeit auch unter schwierigen Bedingungen wie hoher Dichte oder teilweiser Okklusion verlässliche Bewegungsschätzungen zu liefern, werden als Stärken dieses Ansatzes hervorgehoben [[14](#)]. Einschränkungen bestehen insbesondere bei schlechten Lichtverhältnissen und komplexen Szenen mit vielen überlappenden Bewegungen.

Dichte optische Flussverfahren wie der Ansatz von Farnebäck [[16](#)] sind in der Lage, Bewegungen für jedes Pixel zu schätzen und liefern damit detaillierte Bewegungsfelder. In der Studie von Kajo et al. [[14](#)] werden verschiedene dichte Methoden verglichen, wobei Farnebäck als robust gegenüber Rauschen und Beleuchtungsänderungen bewertet wird. Farnebäck eignet sich besonders gut für die Analyse von Crowd-Szenen mit großflächigen Bewegungen und ist weniger anfällig für Fehler bei großen Verschiebungen oder schnellen Bewegungen als viele andere klassische Methoden. Allerdings ist

der Rechenaufwand höher, was insbesondere bei Echtzeitanforderungen berücksichtigt werden muss.

Neuere Deep-Learning-basierte Methoden wie RAFT [17] haben in generischen Optical-Flow-Benchmarks eine deutlich höhere Genauigkeit und Robustheit gegenüber komplexen Bewegungen erzielt. Während RAFT im Bereich Crowd-Analysen noch nicht so umfassend eingesetzt oder evaluiert wurde wie klassische Ansätze, zeigen erste Studien [18], dass diese Methode insbesondere bei großen Bewegungen, Verdeckungen und komplexen Szenenstrukturen Vorteile bietet.

Kajo et al. [14] evaluieren verschiedene optische Fluss-Algorithmen speziell für Crowd-Analysen und zeigen auf, dass die Methoden Lucas-Kanade und Farnebäck in vielen praktischen Szenarien robuste und akkurate Ergebnisse liefern. Während Lucas-Kanade durch seine Effizienz und Robustheit punktet, bietet Farnebäck Vorteile bei der dichten Bewegungsanalyse. In Szenarien mit extremen Herausforderungen (z. B. starke Okklusionen, variable Lichtverhältnisse, große Bewegungen) stoßen beide klassischen Ansätze jedoch an Grenzen, sodass moderne Methoden wie RAFT zunehmend in den Fokus rücken [22].

Insgesamt lässt sich sagen, dass klassische Verfahren wie Lucas-Kanade und Farnebäck weiterhin Standardlösungen für die Bewegungsanalyse in Menschenmengen sind. Diese Methoden werden in der Literatur vielfach als robust, effizient und praxistauglich bewertet. Moderne Deep-Learning-Methoden wie RAFT zeigen großes Potenzial für den Einsatz in komplexen Crowd-Szenarien, sind in der anwendungsnahen Forschung aber noch vergleichsweise wenig verbreitet und werden derzeit hauptsächlich in Benchmark-Studien untersucht.

2.3 QUALITÄTSMETRIKEN OHNE GROUND-TRUTH-DATEN

Da für viele reale Szenarien keine Ground-Truth-Daten für den optischen Fluss existieren [7], ist die Entwicklung von Bewertungsmethoden ohne Referenzdaten sehr relevant. Klassische Evaluierungsbenchmarks wie Middlebury [13] und Sintel liefern zwar umfangreiche Vergleichswerte für Ground-Truth-Situationen, sind aber in realen Anwendungen nur eingeschränkt nutzbar. Daher wurden alternative, indirekte Qualitätsmaße entwickelt, die auf Konsistenzannahmen, Fehlerbildern und selbstüberwachtem Lernen beruhen.

Eine etablierte Methode basiert auf der sogenannten Forward-Backward-Fehleranalyse [29]. Hierbei wird ein Punkt zunächst vorwärts und anschließend rückwärts getrackt. Stimmen die Start- und Endpunkte nicht überein, wird ein Tracking-Fehler bzw. eine unzuverlässige Flussschätzung angenommen. Diese Methode ist einfach zu implementieren, robust gegenüber Drift und eignet sich insbesondere zum automatischen Detektieren von Tracking-Ausfällen und zur Auswahl zuverlässiger Trajektorien. Die Forward-Backward-Fehleranalyse wird mittlerweile auch als Standardmaß

zur Qualitätsbewertung von Feature-Tracking und Flussfeldern eingesetzt [29].

Im Deep-Learning-Bereich werden zunehmend Konsistenz- und Selbstüberwachungsstrategien eingesetzt. Jeong et al. [30] zeigen, dass Konsistenzverluste zwischen vorwärts- und rückwärtsgeschätzten Flussfeldern (Occlusion Consistency, Transformation Consistency) nicht nur zur Fehlerdetektion, sondern auch als Trainingssignal dienen können, auch ohne Ground-Truth-Daten. Solche Methoden ermöglichen es, Modelle auch auf realen, nicht annotierten Daten weiter zu verbessern und die Qualität indirekt zu überwachen.

HERAUSFORDERUNGEN

Dieses Kapitel vertieft die Problemstellung der vorliegenden Arbeit und beleuchtet die spezifischen Herausforderungen, welche bei der Berechnung von Bewegungsvektoren in Drohnenvideos auftreten. Aufbauend darauf werden verschiedene Lösungsansätze evaluiert, die zur Beantwortung der zentralen Fragestellung dieser Arbeit beitragen können. Ziel ist es, die Eignung und Grenzen klassischer sowie Deep-Learning-Methoden zur Bewegungsvektorschätzung kritisch zu untersuchen und eine fundierte Entscheidungsgrundlage für die weitere methodische Ausrichtung dieser Arbeit zu schaffen.

3.1 HERAUSFORDERUNGEN DIESER ARBEIT

Die technische und methodische Analyse von Bewegungen in Drohnenvideos ist mit spezifischen Herausforderungen verbunden, die sich aus den Eigenschaften der Daten und den Anforderungen an die automatisierte Verarbeitung ergeben [8, 9]. Zunächst ist festzustellen, dass Drohnenaufnahmen durch variierende Höhen, Blickwinkel und Bewegungsrichtungen geprägt sind. Diese Faktoren führen zu ständigen Veränderungen der Bildgeometrie, wodurch die Zuordnung von Bewegungsvektoren zu einzelnen Personen oder Gruppen erschwert wird. Besonders die Kombination aus Kamerabewegung und Objektbewegung stellt eine grundlegende Herausforderung dar [8, 9]. Die Trennung von Eigenbewegung der Drohne und tatsächlicher Bewegung im beobachteten Szenenbereich ist für viele klassische Algorithmen problematisch und kann zu fehlerhaften Vektoren führen. Ein detaillierter Einblick in die Herausforderungen bei der Verarbeitung von Drohnenvideos folgt in Abschnitt 3.2.

Eine weitere Schwierigkeit ergibt sich aus der hohen Dichte und Dynamik von Menschenmengen [26]. In solchen Szenen kann es zu Verdeckungen, Überlagerungen und schnellen Richtungswechseln kommen [14, 28]. Diese Bedingungen führen dazu, dass Bildinhalte nicht kontinuierlich oder eindeutig verfolgt werden können, was die Stabilität und Genauigkeit der Bewegungsvektoren beeinträchtigt [14, 28]. Hinzu kommt, dass Drohnenvideos oft in komplexen, unstrukturierten Umgebungen entstehen, in denen Bildartefakte, Schatten, Reflexionen und wechselnde Lichtverhältnisse die Analyse zusätzlich erschweren [14].

Die Auswahl und Anpassung geeigneter Methoden zur Berechnung von Bewegungsvektoren muss daher gezielt auf diese Herausforderungen reagieren. Klassische Verfahren wie die Lucas-Kanade- oder Farneäck-Methode (näher erläutert in Abschnitt 4.2) wurden für strukturierte und ruhige Szenen entwickelt und stoßen deshalb bei den typischen Bedingungen von

Drohnenvideos schnell an ihre Grenzen. Moderne Deep-Learning-Verfahren, welche auf umfangreichen Trainingsdaten basieren, bieten zwar neue Möglichkeiten für die robuste Schätzung von Bewegungen, sind jedoch auf spezifische Domänen und Szenarien angewiesen und können ebenfalls durch ungewohnte Perspektiven und Bildstrukturen beeinträchtigt werden.

Ein weiterer zentraler Aspekt der Problemstellung ist die Evaluierung der Verfahren. In realen Drohnenvideos fehlen oft zuverlässige Ground-Truth-Daten [7], sodass klassische Metriken der Gütebestimmung nicht direkt anwendbar sind. Dies erfordert die Entwicklung und Anwendung alternativer Bewertungsstrategien [29] (näher erläutert in Abschnitt 2.3).

Zusammenfassend zeigt sich, dass die Berechnung von Bewegungsvektoren in Drohnenvideos ein multidimensionales Problem darstellt, welches nicht nur algorithmische und technische, sondern auch datenspezifische und bewertungsspezifische Herausforderungen umfasst. Die Arbeit muss daher sowohl die methodische Robustheit als auch die praktische Anwendbarkeit der untersuchten Verfahren kritisch hinterfragen, um belastbare Empfehlungen für den Einsatz in realen Szenarien ableiten zu können.

3.2 HERAUSFORDERUNGEN BEI DROHNENVIDEOS

Die Schätzung von Bewegungsvektoren und optischem Fluss in Drohnenvideos ist deutlich anspruchsvoller als in statischen oder bodennah aufgenommenen Szenen. Ursache sind vor allem die Eigenbewegung der Kamera, große Tiefenunterschiede in der Szene sowie aufnahmebedingte Artefakte [9, 31–33]. Im Folgenden werden zentrale Herausforderungen im Kontext der Bewegungsvektoren beschrieben.

STARKE UND KOMPLEXE KAMERABEWEGUNG Drohnen erzeugen typischerweise:

- Horizontale Translationen in alle Raumrichtungen (Vorwärts-/ Rückwärtsbewegungen, Seitwärtstranslationen),
- Vertikale Translationen (bei Änderung der Flughöhe),
- Rotationen (Gieren, Nicken, Rollen).

Diese Eigenbewegung führt dazu, dass ein Großteil der im Bild beobachteten Bewegung nicht auf Objektbewegungen, sondern auf Kamerabewegung zurückzuführen ist. Aus Sicht der Bewegungsvektoren entsteht ein globaler (oder teilweise globaler) Bewegungsanteil, der sich über große Teile des Bildes erstreckt. Klassische optische Flussverfahren, die primär lokale Bildgradienten auswerten (z. B. Lucas-Kanade, Farnebäck), können die Bewegung der Kamera und die Bewegung einzelner Objekte nicht ohne zusätzliche Modellannahmen trennen [34]. Für eine semantisch sinnvolle Bewegungsanalyse (z. B. Crowd- oder Fahrzeugbewegungen) ist daher eine Kamerabewegungskompensation oder explizite Modellierung der Kamerapose notwendig.

PARALLAXENEFFEKTE UND GROSSE TIEFENUNTERSCHIEDE Drohnenvideos weisen häufig große Tiefenunterschiede auf (z. B. Boden, Gebäude, hohe Objekte). Durch die perspektivische Projektion führt die Eigenbewegung der Kamera zu *Parallaxeneffekten*: nahe Objekte scheinen sich schneller und stärker zu bewegen als weit entfernte [9, 33]. Für die Bewegungsvektoren bedeutet dies:

- Das Bewegungsfeld ist stark ortsabhängig (kein einfacher globaler Verschiebungsvektor),
- Klassische Modelle wie reine Translation oder Homographien sind oft nur lokal gültig,
- Eine Trennung von Kamerabewegung und Objektbewegung wird erschwert.

In Szenen mit ausgeprägter 3D-Struktur kann die Projektion komplexer Kamerabewegungen daher zu hochgradig nichtlinearen Bewegungsfeldern führen, die von einfachen lokalen Verfahren nur unzureichend erfasst werden.

GROSSE BEWEGUNGEN UND SCHNELLE SZENENDYNAMIK Aufgrund der typischerweise hohen Fluggeschwindigkeiten und großer Bildausschnitte weisen Drohnenvideos regelmäßig folgende charakteristische Szenarien auf:

- Große Pixelverschiebungen zwischen aufeinanderfolgenden Frames,
- Schnelle Bewegungen von Objekten (z. B. Fahrzeuge),
- Plötzliche Richtungsänderungen (z. B. bei Manövern).

Viele klassische Verfahren zur Bewegungsvektorschätzung setzen jedoch kleine Verschiebungen zwischen Frames voraus (lokale Linearitätsannahme). Große Bewegungen müssen dann durch Pyramidenverfahren und Warping approximiert werden, was insbesondere an Kanten und in Regionen mit feinen Details zu Fehlern führen kann [31, 34]. Deep-Learning-Methoden wie [RAFT](#) können zwar prinzipiell größere Bewegungen erfassen, sind aber ebenfalls durch Auflösung, Rechenzeit und Speicherbedarf begrenzt.

BILDQUALITÄT: RAUSCHEN, BEWEGUNGSUNSCHÄRFE UND ROLLING SHUTTER Drohnenaufnahmen sind häufig durch folgende Punkte beeinträchtigt:

- Sensorrauschen (kleine, leichte Kameras mit begrenztem Dynamikumfang),
- Bewegungsunschärfe (schnelle Bewegungen, längere Belichtungszeiten),
- Rolling-Shutter-Effekte (zeilenweise Auslesung des Sensors).

Diese Effekte verletzen zentrale Annahmen vieler optischer Flussverfahren, insbesondere die Helligkeitskonstanz und die lokale Linearität der Intensität [32, 34]. Bewegungsunschärfe kann beispielsweise dazu führen, dass

Objektkanten “verschmieren” und Gradienteninformationen, welche für die Schätzung von Bewegungsvektoren essentiell sind, verloren gehen. Rolling-Shutter-Effekte führen zudem zu geometrischen Verzerrungen, bei denen unterschiedliche Bildzeilen zu leicht unterschiedlichen Zeitpunkten aufgenommen werden.

KLEINE OBJEKTE UND BEGRENZTE AUFLÖSUNG In Luftaufnahmen erscheinen viele für sicherheitsrelevante Anwendungen interessante Objekte (z. B. Personen, kleine Fahrzeuge) relativ klein im Bild. Aus Sicht der Bewegungsvektoren ergeben sich daraus:

- Geringe Anzahl von Pixeln pro Objekt,
- Eingeschränkte Textur und schwache Gradienten,
- Hohe Empfindlichkeit gegenüber Rauschen.

Klassische gradientenbasierte Verfahren benötigen für eine zuverlässige Schätzung in der Regel ausreichend Struktur in der lokalen Nachbarschaft (z. B. Ecken, Kanten). Bei sehr kleinen Objekten ist diese Voraussetzung oft nur eingeschränkt erfüllt, was zu unzuverlässigen Bewegungsvektoren führt [11]. Deep-Learning-basierte Verfahren können dies teilweise kompensieren, sind aber wiederum auf Trainingsdaten mit vergleichbaren Auflösungen angewiesen.

VERDECKUNGEN, SCHATTEN UND SEMANTISCHE AMBIGUITÄTEN In Drohnenvideos treten häufig Verdeckungen und Überlagerungen auf (z. B. Personen in Menschenmengen, Fahrzeuge unter Bäumen oder Brücken). Darüber hinaus können Schatten, Spiegelungen und bewegte Hintergründe (z. B. Vegetation im Wind) Bewegungsvektoren verfälschen:

- Verdeckte Bereiche führen zu nicht beobachtbaren oder abrupt ändernden Bewegungsfeldern,
- Schattenbewegungen erzeugen scheinbare Bewegungsvektoren, die nicht der tatsächlichen Objektbewegung entsprechen,
- Bewegte Hintergründe können fälschlicherweise als relevante Objektbewegung interpretiert werden.

Optischer Fluss beschreibt per Definition die scheinbare Bewegung der Bildintensitäten und unterscheidet zunächst nicht zwischen physikalischer Objektbewegung und anderen Ursachen für Intensitätsänderungen [34]. Für eine sinnvolle Interpretation im sicherheitsrelevanten Kontext sind daher zusätzliche semantische Informationen oder nachgelagerte Analyseschritte erforderlich.

FEHLENDE GROUND-TRUTH UND EVALUATION Ground-Truth-Daten sind Referenzdatensätze, bei denen zu den Messwerten jeweils die zugehörigen, meist manuell oder durch Experten zugewiesenen,

wahren Labels vorliegen und die als Grundlage für das Training oder die Bewertung von überwachten Lernverfahren dienen [35]. Somit ist ein weiterer praktischer Aspekt die fehlenden oder nur schwer verfügbare Ground-Truth-Daten für Bewegungsvektoren in realen Drohnenaufnahmen. Die genaue 3D-Bewegung aller Objekte sowie der Kamera ist in realen Einsatzszenarien selten bekannt.

Dies erschwert:

- Die quantitative Bewertung von Verfahren zur Bewegungsvektorschätzung,
- Das Training überwachter Deep-Learning-Methoden,
- Die Domänenadaption von vortrainierten Modellen.

Häufig wird daher auf synthetische Datensätze, approximative Referenzmethoden oder indirekte Evaluationsmetriken ausgewichen [31, 34]. Dies ist insbesondere für die vorliegende Arbeit relevant, da die Evaluationskonzepte ohne Ground-Truth-Daten speziell an diese Rahmenbedingungen angepasst werden müssen (vgl. Abschnitt 3.3).

Insgesamt machen diese Herausforderungen deutlich, dass die direkte Übertragung von Verfahren zur Bewegungsvektorschätzung aus klassischen Videoszenarien auf Drohnen-Videos nicht trivial ist.

3.3 VERSCHIEDENE LÖSUNGSANSÄTZE

Übersicht möglicher Lösungsansätze

In der folgenden Tabelle 3.1 werden alle in dieser Arbeit untersuchten Ansätze zur Schätzung von Bewegungsvektoren in Drohnenvideos vorgestellt sowie weitere Ansätze, auch wenn sie nicht umgesetzt wurden. Die Übersicht enthält Methoden zur Berechnung von Bewegungsvektoren sowie unterstützende Verfahren, die zur Optimierung der Bewegungsanalyse beitragen können. Für jeden Lösungsansatz enthält die Tabelle eine kurze Beschreibung sowie eine Angabe, ob dieser im Rahmen der experimentellen Umsetzung tatsächlich verwendet wurde. Dadurch soll ein umfassender Überblick über die methodischen Möglichkeiten und deren praktische Relevanz für die vorliegende Arbeit vermittelt werden. In den nachfolgenden Abschnitten werden die in der Tabelle 3.1 aufgeführten Ansätze hinsichtlich ihrer Umsetzung in dieser Arbeit analysiert und die jeweiligen Gründe für deren Auswahl oder Verwerfung detailliert erläutert.

Nr.	Lösungsansatz	Beschreibung	Verwendung in dieser Arbeit
1	Alternative Bewertungsmetriken (z. B. Forward-Backward-Consistency)	Bewertung ohne Ground-Truth-Daten	Ja
2	Kamerabewegungskompensation	Homographie-basierte Entfernung globaler Kamerabewegung	Nein (nicht nötig, da statische Videos)
3	Beschränkung auf ausschließlich statische Drohnenvideos	Vermeidung von Kamerabewegungsproblemen durch Auswahl der Daten	Ja
4	Hintergrundentfernung	Segmentierung zur Fokussierung auf relevante Objekte	Nein
5	Nutzung zusätzlicher semantischer Information (z. B. YOLO)	Personenidentifikation mittels Objektdetektion vor der Flow-Schätzung	Nein
6	Horn-Schunck (dense Optical Flow)	Klassisches dichtes Verfahren für Bewegungsvektoren	Nein
7	Lucas-Kanade (sparse Optical Flow)	Klassisches featurebasiertes Verfahren zur Bewegungsvektorschätzung	Ja
8	Farnebäck (dense Optical Flow)	Klassisches, dichtes Verfahren für Bewegungsvektoren	Ja
9	FlowNet (Deep-Learning Optical Flow)	CNN-basierte Methode für optischen Fluss	Nein
10	RAFT (Deep-Learning Optical Flow)	Modernes, rekurrentes Deep-Learning-Modell für dichte Flussfelder	Ja

Tabelle 3.1: Überblick über die betrachteten Lösungsansätze und deren Einsatz in der Arbeit.

Alternative Bewertungsmetriken

Für die Bewertung der Bewegungsvektorschätzung in Drohnenvideos ist die Verfügbarkeit von Ground-Truth-Daten oft stark eingeschränkt oder gar nicht gegeben [7]. Klassische Gütemaße wie der EPE oder der Winkelfehler, die einen Vergleich mit Referenzdaten ermöglichen, sind daher nur begrenzt anwendbar. Um dennoch eine systematische und aussagekräftige Bewertung der Methoden zu ermöglichen, werden alternative Metriken eingesetzt, welche ohne vollständige Referenz auskommen.

Eine zentrale alternative Bewertungsstrategie ist die *Forward-Backward-Consistency*. Hierbei wird für jeden Bildpunkt der Fluss sowohl vorwärts (Frame $t \rightarrow t + 1$) als auch rückwärts (Frame $t + 1 \rightarrow t$) berechnet und miteinander abgeglichen. Ein konsistenter Fluss bedeutet, dass ein Punkt, welcher vorwärts verschoben wird und anschließend rückwärts verfolgt wird, wieder an seiner ursprünglichen Position landet. Abweichungen zwischen den beiden Richtungen sind ein Indikator für Unsicherheiten, Fehler oder Inkonsistenzen im Bewegungsfeld [29].

Eine weitere Alternative sind qualitative Visualisierungen (z. B. Heatmaps von Richtung und Amplitude). Dieser Ansatz ermöglicht die Identifikation systematischer Fehler, Inkonsistenzen und Ausreißer auch ohne Referenzdaten und unterstützen die methodische Vergleichbarkeit der Verfahren [36].

Bewertung: Der Einsatz alternativer Bewertungsmetriken ist insbesondere für Drohnenvideos unverzichtbar, da sie eine Möglichkeit bieten, die Leistungsfähigkeit und die Robustheit der Methoden unter realen Bedingungen kritisch zu hinterfragen. Sie liefern wichtige Erkenntnisse über die Zuverlässigkeit, Plausibilität und Stabilität der Bewegungsvektoren und sind daher fester Bestandteil der experimentellen Methodik dieser Arbeit.

Kamerabewegungskompensation

Zur Trennung von Objekt- und Kamerabewegung wird häufig eine explizite Kompensation der globalen Kamerabewegung durchgeführt (z. B. mittels Homographie-Schätzung) [37, 38].

Vorteile:

- Verbesserte Analyse lokaler Bewegungen
- Reduzierung von Fehlern durch globale Bewegung
- Notwendig für semantisch sinnvolle Objektverfolgung

Nachteile:

- Starke Abhängigkeit von ausreichend strukturierten Hintergründen
- Parallaxeneffekte und nicht-planare Szenen schränken die Wirksamkeit ein

Bewertung: Eine effektive Kamerabewegungskompensation ist für Drohnenvideos meist unverzichtbar. Sie stellt jedoch eigene methodische Herausforderungen dar und ist nicht immer vollständig lösbar.

In dieser Arbeit wurde darauf verzichtet, da der Schwerpunkt der Arbeit auf dem Methodenvergleich liegen soll. Wie im nächsten Abschnitt 3.3 beschrieben wird, werden statt einer Kamerabewegungskompensation ausschließlich statische Videos genutzt, was die Kompensation überflüssig macht, da somit keine relevante Kamerabewegung vorliegt.

Nutzung ausschließlich statischer Drohnenvideos

Um einen präzisen und unverfälschten Methodenvergleich zu ermöglichen, wurden im Rahmen dieser Arbeit gezielt statische Drohnenvideos ausgewählt. Dadurch kann sich die Analyse vollständig auf die Bewegung innerhalb der Szene, etwa von Menschenmengen, konzentrieren, ohne dass zusätzliche Effekte durch Kamerabewegungen berücksichtigt werden müssen.

Die bewusste Einschränkung auf statische Aufnahmen reduziert die Komplexität vieler typischer Herausforderungen, die bei Drohnenvideos auftreten, wie sie in Abschnitt 3.2 beschrieben werden. Insbesondere entfällt der Bedarf, globale Bildverschiebungen oder Parallaxeneffekte zu kompensieren, da die Kameraposition konstant bleibt. Dies verhindert, dass Fehlerquellen wie Kamerabewegung oder Veränderungen im Aufnahmewinkel die Ergebnisse der Bewegungsvektorschätzung verfälschen.

Ein weiterer Vorteil dieser Herangehensweise ist die verbesserte Vergleichbarkeit und Nachvollziehbarkeit der Ergebnisse. Da alle Methoden unter identischen Aufnahmebedingungen getestet werden, sind Unterschiede im Ergebnis auf die jeweiligen Eigenschaften der Verfahren zurückzuführen. Die Ergebnisse lassen sich dadurch klarer interpretieren und methodische Stärken und Schwächen treten deutlicher hervor.

Zusammengefasst ermöglicht die Konzentration auf statische Drohnenvideos eine gezielte, methodisch objektive Bewertung der untersuchten optischen Fluss Methoden. Potenzielle Störfaktoren durch Kamerabewegungen werden gezielt ausgeschlossen, sodass die Analyse besonders transparent und aussagekräftig bleibt.

Hintergrundentfernung

Die Entfernung des Hintergrunds ist ein gängiger Vorverarbeitungsschritt, um die Bewegungsanalyse gezielt auf relevante Objekte zu fokussieren. Hierbei kommen Methoden wie beispielsweise Segmentierungsalgorithmen oder Deep-Learning-basierte Masken zum Einsatz [39, 40].

Vorteile:

- Verbesserung der Präzision und Robustheit der Bewegungsanalyse
- Reduzierung von Störfaktoren (Hintergrundbewegungen, Schatten)

- Fokussierung auf semantisch bedeutsame Bildregionen

Nachteile:

- Einfluss von Segmentierungsfehlern auf die Genauigkeit
- Erhöhung der Komplexität und Rechenlast durch die Kombination aus Segmentierung und Flussschätzung
- Verlust von Kontext (z. B. Gesamtbewegungen)

Bewertung: Das Entfernen des Hintergrunds kann die Robustheit trotz Komplexität und Parallaxeneffekten erhöhen, hängt aber maßgeblich von der Segmentierungsqualität ab. Für heterogene Videos kann keine konsistente Segmentierung garantiert werden.

Obwohl die Hintergrundentfernung prinzipiell die Genauigkeit erhöhen kann, wurde in dieser Arbeit aus den nachfolgenden Gründen darauf verzichtet. Die zu analysierenden Drohnenvideos weisen zum Teil erhebliche Unterschiede auf. Erste prototypische Voruntersuchungen in der Arbeit haben bestätigt, dass jedes Drohnenvideo eine individuelle Segmentierung benötigt hätte. Die Anwendung eines einheitlichen Segmentierungsverfahrens hätte aufgrund der teilweise erheblichen Unterschiede zwischen den Videos zu inkonsistenten Ergebnissen geführt und somit die Möglichkeit eines validen Methodenvergleichs eingeschränkt. Für einen objektiven und reproduzierbaren Vergleich wurde die gesamte Pipeline daher ohne individuelle Hintergrundentfernung gestaltet.

Nutzung zusätzlicher semantischer Information durch z. B. Objekterkennung

Durch die Kombination von Bewegungsanalyse mit Objektdetektion, Segmentierung oder Tracking lassen sich Bewegungsvektoren gezielter interpretieren [41]. Zum Beispiel kann der Objekterkennungsalgorithmus You Only Look Once (YOLO) als leistungsfähiges Deep-Learning-basiertes Detektionsmodell eingesetzt werden, um Personen zu identifizieren und die Bewegungsanalyse gezielt auf diese Bereiche zu beschränken [42].

Vorteile:

- Verbesserung der Robustheit bei Verdeckungen und Schatten
- Ermöglichung einer differenzierten Analyse von Crowd- und Einzeleinheiten
- Reduzierung von Fehlern durch Hintergrundbewegungen
- Ermöglichung einer semantischen Trennung von Bewegungen (Personen, Fahrzeuge, Hintergrund)

Nachteile:

- Erfordernis zusätzlicher Rechenleistung und Datenannotation

- Auftreten zusätzlicher Fehlerquellen durch Detektionsungenauigkeiten
- Erhöhung der Komplexität und Rechenlast durch die Kombination von Detektion und Flussschätzung

Bewertung: Die Integration semantischer Informationen ist insbesondere für anspruchsvolle Crowd- und Sicherheitsanwendungen sinnvoll und erhöht die Aussagekraft der Bewegungsvektoren deutlich [41]. Die Kombination beider Ansätze ist jedoch mit großem zusätzlichem Aufwand verbunden.

Im Rahmen der Arbeit wurde die Nutzung zusätzlicher semantischer Information mittels YOLO als Personendetektionsmodell zunächst experimentell erprobt. Im Verlauf der Voruntersuchungen zeigt sich, dass dieser Ansatz aus folgenden Gründen nicht geeignet ist. Zum einen ist YOLO nicht auf Drohnenvideos spezialisiert, da die Standardmodelle primär auf bodennahe und klassische Perspektiven trainiert sind. In Drohnenaufnahmen mit ungewohnten Blickwinkeln und kleinen Objekten war die Detektionsqualität daher unzureichend. Zum anderen handelt es sich bei vielen der untersuchten Szenen um dichte Menschenmengen, in denen die Detektion einzelner Personen weder praktikabel noch sinnvoll ist, da die Bewegung als kollektive Menge betrachtet werden sollte. Darüber hinaus hätte die Kombination aus Detektion und anschließender Bewegungsvektorschätzung die Laufzeit der Pipeline erheblich verlängert und somit die Effizienz und potenzielle Echtzeitfähigkeit der Analyse stark eingeschränkt.

Aufgrund dieser praktischen und methodischen Limitationen wurde der Ansatz, die Bewegungsvektoren ausschließlich auf semantisch detektierten Personen zu berechnen, nicht weiterverfolgt. Stattdessen wurde die Analyse auf die gesamte Szene angewendet, um eine robuste und konsistente Bewegungsdetektion zu gewährleisten.

Ausgewählte klassische Modelle zur Bewegungsvektoren Berechnung

1. Horn-Schunck (Dense Optical Flow)

Das Horn-Schunck-Verfahren berechnet Bewegungsvektoren für jedes Pixel basierend auf einer globalen Optimierung. Es setzt voraus, dass die Helligkeit eines Pixels über die Zeit konstant bleibt und das Bewegungsfeld möglichst glatt ist. Die Methode minimiert eine kombinierte Energiefunktion aus Daten- und Glattheitsterm und liefert dadurch dichte Flussfelder [19].

Vorteile:

- Bereitstellung dichter Bewegungsvektoren für das gesamte Bild
- Gewährleistung einer globalen Glättung für gleichmäßige Flussfelder
- Bereitstellung eines einfachen und gut verständlichen mathematischen Modells

Nachteile:

- Hohe Sensitivität gegenüber Bildrauschen und Artefakten
- Glättung feiner Bewegungsdetails
- Erhöhung des Rechenaufwands bei hochauflösenden und komplexen Szenen
- Anfälligkeit für Fehler bei schnellen Bewegungen und abrupten Richtungswechseln

Bewertung: Das Horn-Schunck-Verfahren ist ein grundlegender historischer Ansatz und eignet sich als Referenzmodell (Baseline) für die optische Flussberechnung.

2. Lucas-Kanade (*Sparse Optical Flow*)

Das Lucas-Kanade-Verfahren berechnet Bewegungsvektoren für markante Bildpunkte auf Basis lokaler Helligkeitsänderungen [15]. Die Methode ist effizient und bei klaren Strukturen sowie geringen Bewegungen robust.

Vorteile:

- Ermöglichung einer schnellen Berechnung und eines geringen Ressourcenbedarfs
- Robustheit bei kleinen Bewegungen und klaren Texturen
- Eignung für einfache, strukturierte Szenen

Nachteile:

- Bereitstellung von Bewegungsvektoren nur für ausgewählte Punkte ("Sparse"-Eigenschaft)
- Anfälligkeit für Fehler bei starker Kamerabewegung, Verdeckungen und großen Bewegungen
- Auftreten von Schwierigkeiten bei geringer Auflösung und schwacher Textur

Bewertung: Für statische oder wenig komplexe Szenarien stellt das Lucas-Kanade-Verfahren eine relevante Baseline dar.

3. Farnebäck (*Dense Optical Flow*)

Das Farnebäck-Verfahren berechnet für jedes Pixel einen Bewegungsvektor durch polynomielle Approximationen lokaler Bildbereiche. Es liefert so ein vollständiges Bewegungsfeld [16].

Vorteile:

- Bereitstellung dichter Bewegungsfelder
- Robustheit gegenüber moderaten Bewegungen
- Erzeugung glatter Ergebnisse und gute Generalisierbarkeit in vielen Szenen

Nachteile:

- Anfälligkeit für Parallaxeneffekte und nichtlineare Bewegungen
- Fehleranfälligkeit bei starker Kamerabewegung und Bildartefakten
- Empfindlichkeit gegenüber Rauschen, Unschärfe und geringen Objektgrößen

Bewertung: Farnebäck ist eine solide Wahl für viele Szenarien, stößt jedoch bei den typischen Herausforderungen von Drohnenaufnahmen (Parallaxen, große Bewegungen, Okklusionen) an methodische Limits. Für Menschenmengen-Analysen liefert es aber oft konsistente, glatte Felder.

Vergleich klassischer Verfahren: Horn-Schunck, Lucas-Kanade und Farnebäck

Die klassischen Methoden Horn-Schunck, Lucas-Kanade und Farnebäck sind weit verbreitete Ansätze zur Berechnung des optischen Flusses, unterscheiden sich jedoch grundlegend in Methodik und Eignung für verschiedene Anwendungsszenarien [13, 15, 16, 19].

Das Lucas-Kanade-Verfahren basiert auf einer lokalen, merkmalsbasierten Analyse und berechnet Bewegungsvektoren nur für markante Bildpunkte (sparse optical flow). Es ist besonders effizient und robust bei kleinen Bewegungen und liefert zuverlässige Ergebnisse in strukturierten Szenen [15, 43]. Farnebäck hingegen ist ein klassischer, dichter Ansatz, der für jedes Pixel Bewegungsvektoren berechnet (dense optical flow) und dabei lokale polynomielle Approximationen nutzt. Das Verfahren liefert glatte, konsistente Bewegungsfelder und ist dabei weniger anfällig für Rauschen als Horn-Schunck [13, 16].

Das Horn-Schunck-Verfahren basiert auf einer globalen Optimierung mit einem Glattheitsterm, ist aber sehr empfindlich gegenüber Bildrauschen und eignet sich weniger für Szenen mit schnellen Bewegungen oder komplexen Tiefenstrukturen [19]. In Praxistests und modernen Benchmarks zeigt sich, dass Horn-Schunck typischerweise größere Fehler aufweist und weniger robust gegenüber realen Bildartefakten ist [13, 22].

Auswahl: Auf Basis dieser Überlegungen und gestützt durch die einschlägige Literatur [13, 22] wurden für diese Arbeit Lucas-Kanade und Farnebäck als repräsentative klassische Modelle ausgewählt. Sie bieten eine vorteilhafte Kombination aus Robustheit, Effizienz und Praxistauglichkeit. Lucas-Kanade eignet sich besonders für Szenarien mit klaren Strukturen und geringer Bewegungsdynamik bzw. niedriger Bewegungsintensität, während Farnebäck auch in komplexeren Szenen mit hoher Punktdichte und Dynamik

konsistente Ergebnisse liefert. Die Entscheidung gegen andere klassische Verfahren, wie Horn-Schunck, sowie deren methodische Unterlegenheit wurden im Kapitel 2.1 ausführlich begründet.

Ausgewählte Deep-Learning-Modelle zur Bewegungsvektoren Berechnung

1. FlowNet (Deep-Learning Optical Flow)

FlowNet ist eines der ersten Deep-Learning-Modelle zur direkten Schätzung von optischem Fluss aus Bildpaaren. Es basiert auf einer Convolutional-Neural-Network-Architektur und wurde mit synthetischen Flussdaten trainiert. FlowNet und seine Weiterentwicklungen (FlowNet2) ermöglichen die dichte Berechnung von Bewegungsvektoren für alle Pixel [20, 44].

Vorteile:

- Ermöglichung eines End-to-End-Ansatzes zur Erfassung großer Bewegungen und komplexer Muster
- Bereitstellung dichter Bewegungsfelder für sämtliche Bildbereiche
- Deutliche Geschwindigkeitssteigerung gegenüber vielen anderen Deep-Learning-Ansätzen beim Einsatz auf der GPU
- Verfügbarkeit gut dokumentierter und weit verbreiteter Implementierungen

Nachteile:

- Abhängigkeit der Qualität und Robustheit von den Trainingsdaten
- Auftreten von Generalisierungsproblemen bei Drohnenvideos durch ursprüngliches Training auf synthetischen Daten und Standardperspektiven (ungewohnte Blickwinkel, Parallaxen)
- Geringere Genauigkeit und Robustheit im Vergleich zu neueren Modellen wie RAFT, insbesondere bei feinen Details und schnellen Bewegungen
- Geringere Flexibilität der Modellarchitektur im Vergleich zu aktuellen Ansätzen

Bewertung: FlowNet eignet sich gut für explorative Analysen und Anwendungen mit beschränkten Ressourcen. In anspruchsvollen Drohnenszenarien stößt es jedoch, insbesondere ohne gezieltes Nachtraining auf Drohnen-Daten, in Bezug auf Genauigkeit und Robustheit an Grenzen. Für viele praktische Anwendungsfälle kann FlowNet dennoch eine nützliche Ergänzung oder Baseline sein, insbesondere wenn Rechenzeit ein kritischer Faktor ist.

2. *RAFT* (Deep-Learning Optical Flow)

Das Deep-Learning-basierte Verfahren *RAFT* nutzt neuronale Netze, um Bewegungsvektoren zu schätzen. Es ist in der Lage, komplexe Bewegungsmuster, große Bewegungen und schwierige Bildstrukturen zu erfassen [17].

Vorteile:

- Erreichung einer sehr hohen Genauigkeit auf vielen Standard-Datensätzen
- Robustheit gegenüber großen Bewegungen, Verdeckungen und komplexen Mustern
- Bereitstellung einer dichten Schätzung für alle Bildbereiche
- Teilweise Toleranz gegenüber Bildrauschen und Unschärfe

Nachteile:

- Starke Abhängigkeit von den Trainingsdaten und Auftreten von Generalisierungsproblemen bei stark abweichenden Szenarien (Domain-Shift)
- Hoher Rechen- und Speicherbedarf (GPU-Nutzung nötig)
- Notwendigkeit eines spezifischen Nachtrainings (Drohnendaten) für optimale Ergebnisse
- Mangelnde Transparenz bei der Fehlererklärung ("Black-Box"-Eigenschaft)

Bewertung: *RAFT* bietet großes Potenzial, insbesondere bei komplexer Crowd-Dynamik. Ohne gezieltes Fine-Tuning auf Drohnenvideos zeigen sich jedoch auch Limitierungen, etwa in Form von Fehlern bei ungewohnten Perspektiven oder Szenen. In der Praxis werden vortrainierte Modelle häufig direkt auf neue Datensätze angewendet, um eine erste Baseline zu erhalten, auch wenn sie nicht explizit für diese Daten domänenspezifisch angepasst wurden [45].

*Vergleich *RAFT* und FlowNet*

FlowNet [20] stellt das erste Deep-Learning-Modell zur Schätzung von optischem Fluss dar und basiert auf einer klassischen CNN-Architektur. Es ermöglicht eine schnelle, end-to-end trainierbare Berechnung dichter Bewegungsfelder und eignet sich als solide Baseline, insbesondere wenn die Rechenzeit eine wichtige Rolle spielt. Allerdings zeigt FlowNet Schwächen in der Genauigkeit und Robustheit, insbesondere bei realen, komplexen Szenen wie Drohnenvideos, die durch ungewöhnliche Perspektiven und hohe Dynamik gekennzeichnet sind. Die Qualität der Ergebnisse ist stark von den Trainingsdaten abhängig und Generalisierungsprobleme treten häufig auf.

Im Gegensatz dazu nutzt *RAFT* [17] eine rekursive Architektur mit All-Pairs-Korrelationen und iterativer Verfeinerung. Dadurch werden auch große

Bewegungen, feine Details und komplexe Strukturen präzise erfasst. Benchmark-Studien [11, 17] zeigen, dass RAFT FlowNet und FlowNet2 hinsichtlich Genauigkeit und Robustheit deutlich übertrifft, insbesondere in anspruchsvollen Szenarien wie Drohnenvideos mit Crowd-Dynamik, Parallaxeneffekten und variabler Bildqualität.

Bewertung: Für die in dieser Arbeit adressierten Herausforderungen im Bereich Drohnenvideos ist RAFT nach aktuellem Stand der Forschung die leistungsstärkere und zuverlässigere Architektur. FlowNet eignet sich weiterhin als schnelle Referenzlösung, erreicht jedoch nicht die Präzision und Flexibilität moderner Ansätze. Die methodischen Unterschiede und die Überlegenheit von RAFT wurden im Kapitel 2.1 ausführlich diskutiert.

Finale Modellauswahl

Die Auswahl der optischen Fluss Modelle in dieser Arbeit erfolgte gezielt unter Berücksichtigung der typischen Anforderungen verschiedener Szenarientypen sowie aktueller technischer Entwicklungen. Mit Lucas-Kanade, Farnebäck und RAFT wurden drei grundlegend verschiedene Ansätze einbezogen, die von klassischen, punktebasierten Verfahren bis hin zu modernen, dichten und auf Deep-Learning-basierenden Methoden reichen, um ein möglichst breites Spektrum an Leistungsfähigkeit, Robustheit und Recheneffizienz zu analysieren.

Verzicht auf Nachtraining von RAFT

Im Rahmen dieser Arbeit wird das RAFT-Modell ausschließlich in seiner vortrainierten Form verwendet und nicht auf den eigenen Drohnenvideodaten nachtrainiert oder feinjustiert. Dieser bewusste Verzicht auf ein Nachtraining (Fine-Tuning) basiert auf mehreren methodischen, praktischen und wissenschaftlichen Überlegungen:

- *Fehlende annotierte Trainingsdaten:* Für ein effektives Nachtraining eines Deep-Learning-basierten Optical-Flow-Modells wie RAFT sind umfangreiche und präzise annotierte Ground-Truth-Bewegungsdaten erforderlich. Im Rahmen dieser Arbeit standen jedoch nur vergleichsweise wenige, manuell annotierte Punktepaaare zur Verfügung, welche speziell für die Evaluation, nicht jedoch für robustes supervised Training konzipiert wurden. Ein Nachtraining auf Basis dieses kleinen Datensatzes hätte zu Overfitting und mangelnder Generalisierbarkeit geführt.
- *Methodische Vergleichbarkeit:* Ziel der Arbeit ist ein Methodenvergleich unter weitgehend standardisierten Bedingungen. Die Verwendung des "out-of-the-box" vortrainierten RAFT-Modells ermöglicht, dass die Ergebnisse mit den von der Community verbreiteten Baselines vergleichbar bleiben.
- *Ressourcen und Aufwand:* Das Nachtrainieren von Deep-Learning-Modellen für den optischen Fluss erfordert erhebliche Rechenressourcen

(insbesondere GPUs) und Zeit. Im Rahmen dieser Arbeit lag der Fokus auf der algorithmischen Vergleichbarkeit, nicht auf der Entwicklung eines spezialisierten Deep-Learning-Modells.

- *Relevanz für reale Anwendungen:* In der Praxis werden vortrainierte Modelle wie [RAFT](#) oft direkt auf neue Anwendungsbereiche übertragen, da das Annotieren großer Mengen Ground-Truth-Daten in realen Drohnenvideos kaum praktikabel ist [7]. Der Verzicht auf Nachtraining spiegelt somit auch typische Anwendungsszenarien in der Forschung und Industrie wider [45].

THEORETISCHE GRUNDLAGEN

Dieses Kapitel fasst die theoretischen Grundlagen für die Bewegungsanalyse in Videos zusammen und erläutert zentrale Begriffe wie Bewegungsvektoren und optischer Fluss. Es stellt die Verfahren zur Berechnung von Bewegungsvektoren vor, welche in dieser Arbeit prototypisch realisiert werden.

4.1 BEWEGUNGSVEKTOREN UND OPTISCHER FLUSS

Die Beschreibung von Bewegung in Bildfolgen erfolgt üblicherweise über *Bewegungsvektoren* und den *optischen Fluss*. Beide Begriffe sind eng verwandt, werden in der Literatur jedoch unterschiedlich verwendet und beziehen sich auf verschiedene Abstraktionsebenen der Bewegung [19, 31, 32].

BEWEGUNGSVEKTOREN Unter einem Bewegungsvektor wird im Allgemeinen die Verschiebung eines Bildpunktes oder eines Bildbereichs zwischen zwei aufeinanderfolgenden Frames verstanden. Formal beschreibt ein Bewegungsvektor

$$\mathbf{d}(x, y) = (u(x, y), v(x, y))^T \quad (4.1)$$

die Positionänderung eines Punktes von (x, y) im ersten Frame zu $(x + u, y + v)$ im zweiten Frame.

In vielen praktischen Anwendungen, insbesondere in der Videokodierung werden Bewegungsvektoren *blockbasiert* geschätzt. Das Bild wird in Blöcke fester Größe (z. B. 16×16 Pixel) unterteilt und für jeden Block wird ein Bewegungsvektor bestimmt, der die Verschiebung dieses Blocks zwischen zwei Frames angibt [13]. Diese Vektoren dienen primär der Redundanzreduktion und effizienten Kompression, sind aber gleichzeitig eine grobe Beschreibung der Bildbewegung.

In der Bildverarbeitung und Computer Vision können Bewegungsvektoren auch punktbasiert (z. B. Lucas-Kanade-Tracking von Merkmalspunkten) oder dichtenweise (Vektor an jedem Pixel) betrachtet werden. In diesem Fall ist der Begriff eng mit dem optischen Fluss verknüpft.

OPTISCHER FLUSS Der Begriff *optischer Fluss* bezeichnet das zweidimensionale Bewegungsfeld in der Bildebene, welches die scheinbare Bewegung von Bildintensitäten über die Zeit beschreibt [19, 32]. Optischer Fluss ist damit ein kontinuierliches oder diskretes Vektorfeld

$$\mathbf{v}(x, y, t) = (u(x, y, t), v(x, y, t))^T, \quad (4.2)$$

das jedem Pixel (x, y) zum Zeitpunkt t eine Geschwindigkeitsrichtung und -größe zuordnet.

Wichtig ist die Unterscheidung zwischen:

- **Physikalischem Bewegungsfeld:** der tatsächlichen dreidimensionalen Bewegung von Objekten in der Szene (3D-Translation und -Rotation) [31, 32].
- **Optischem Fluss:** der zweidimensionalen Projektion dieser Bewegung auf die Bildebene, beeinflusst durch Kamerabewegung, Perspektive und Beleuchtung [19, 32, 34].

Optischer Fluss ist also die *sichtbare* bzw. *scheinbare* Bewegung im Bild und muss nicht immer exakt der wahren physikalischen Bewegung entsprechen (z. B. bei Schatten, Spiegelungen oder Beleuchtungswechseln) [19, 32].

DICHTER VS. SPARSAMER OPTISCHER FLUSS In der Literatur wird häufig zwischen *dichtem* und *sparsamen* optischen Fluss unterschieden [31, 34]:

- **Dichter optischer Fluss:** Es wird ein Bewegungsvektor für jedes Pixel (oder nahezu jedes Pixel) geschätzt. Beispiele sind der Farneäck-Algorithmus [16] oder moderne Deep-Learning-Verfahren wie RAFT [17].
- **Sparsamer optischer Fluss:** Bewegungsvektoren werden nur für ausgewählte Bildpunkte bestimmt, etwa für Merkmals- oder Eckpunkte (z. B. Lucas-Kanade-Verfahren [15]).

Dichte Verfahren liefern eine reichhaltigere Bewegungsbeschreibung und eignen sich gut für globale Analysen (z. B. Crowd- oder Verkehrsbewegungen), sind jedoch oft rechenintensiver. Sparsame Verfahren sind typischerweise effizienter und robust in gut strukturierten Regionen, beschreiben aber nur einen Teil der Szene.

EINORDNUNG IM KONTEXT DIESER ARBEIT In dieser Arbeit werden Bewegungsvektoren und optischer Fluss als zentrale Größen zur Beschreibung der Dynamik in Drohnenvideos verwendet. Die betrachteten Methoden (klassische Ansätze wie Lucas-Kanade und Farneäck sowie Deep-Learning-Methoden wie RAFT) liefern jeweils unterschiedliche Formen von Bewegungsfeldern:

- **Klassische Verfahren:** meist lokale, gradientenbasierte Schätzungen, teilweise sparse (Lucas-Kanade) oder dicht (Farneäck),
- **Deep-Learning-Verfahren:** dichte Schätzungen, die aus großen Trainingsdatensätzen gelernt wurden.

Bewegungsvektoren und optischer Fluss bilden damit die Grundlage für die in dieser Arbeit durchgeführte Bewegungsanalyse und den Vergleich der Verfahren hinsichtlich Genauigkeit, Robustheit und Eignung für sicherheitsrelevante Szenarien.

4.2 VERFAHREN ZUR BEWEGUNGSVEKTORENBERECHNUNG

4.2.1 Lucas-Kanade-Ansatz

Der Lucas-Kanade-Ansatz gehört zu den klassischen gradientenbasierten Verfahren zur Schätzung des optischen Flusses und wird in vielen praktischen Anwendungen als Referenzverfahren eingesetzt [15, 31]. Im Rahmen dieser Arbeit dient er als Vertreter klassischer Methoden.

INTUITION UND GRUNDANNAHMEN Ziel des Verfahrens ist es, für zwei aufeinanderfolgende Bilder die scheinbare Bewegung der Bildpunkte in Form eines Bewegungsvektors $\mathbf{v} = (u, v)^\top$ zu schätzen. Lucas-Kanade beruht dabei auf zwei zentralen Annahmen:

1. **Helligkeitskonstanz:** Die Intensität eines Bildpunktes bleibt über die Zeit hinweg (lokal) konstant. Für zwei aufeinanderfolgende Frames t und $t + \Delta t$ gilt näherungsweise

$$I(x, y, t) = I(x + u, y + v, t + \Delta t), \quad (4.3)$$

wobei $I(x, y, t)$ die Bildintensität an der Position (x, y) und (u, v) der gesuchte Bewegungsvektor ist [13].

2. **Lokale Konstanz der Bewegung:** Innerhalb eines kleinen Fensters um einen Pixel wird angenommen, dass alle Pixel denselben Bewegungsvektor besitzen. Damit kann die Bewegung aus den Intensitätsänderungen mehrerer benachbarter Pixel gemeinsam geschätzt werden [13].

Diese Annahmen sind für viele realweltliche Szenen in guter Näherung erfüllt, insbesondere wenn keine starken Beleuchtungswechsel oder Bewegungsunschärfen auftreten.

PYRAMIDENBASIERTER LUCAS-KANADE-ANSATZ Beim Lucas-Kanade-Ansatz wird vorausgesetzt, dass die Verschiebungen zwischen zwei Frames klein sind (typischerweise deutlich kleiner als ein Pixel). In realen Szenen, insbesondere in Drohnenvideos, treten jedoch häufig größere Bewegungen durch schnelle Kamerafahrten oder sich schnell bewegende Objekte auf.

Um dennoch robuste Schätzungen zu ermöglichen, wird Lucas-Kanade üblicherweise in einem *pyramidenbasierten* Verfahren eingesetzt [46]:

1. Aufbau einer Gaussian-Pyramide für beide Frames (mehrere Skalen von grob nach fein).
2. Schätzung des optischen Flusses auf der größten Ebene, auf der die Verschiebungen durch die Skalierung kleiner erscheinen.
3. Hochskalierung der dort geschätzten Bewegungsvektoren auf die nächst feinere Ebene.

4. Verfeinerung der Schätzung durch erneute Anwendung von Lucas-Kanade auf der aktuellen Ebene.
5. Wiederholung bis zur Originalauflösung [46].

Dieses hierarchische Vorgehen erlaubt es, auch größere Bewegungen effizient und relativ robust zu erfassen und wird in gängigen Bibliotheken (z. B. OpenCV) standardmäßig verwendet.

4.2.2 Farneback-Algorithmus

Der Farneback-Algorithmus ist ein klassisches Verfahren zur dichten Schätzung des optischen Flusses, das auf einer polynomiellen Modellierung der lokalen Bildintensität basiert [16]. Im Gegensatz zum Lucas-Kanade-Ansatz, der typischerweise für ausgewählte Merkmalspunkte eingesetzt wird (sparse optical flow), liefert der Farneback-Algorithmus standardmäßig einen Bewegungsvektor für jedes Pixel (dense optical flow) und eignet sich damit besonders für die flächendeckende Analyse von Bewegungsmustern.

GRUNDPRINZIP Die zentrale Idee ist, die lokale Nachbarschaft eines Pixels durch ein quadratisches Polynom zu modellieren:

$$f(\mathbf{x}) \approx \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{x} + c, \quad (4.4)$$

Zwischen zwei aufeinanderfolgenden Frames wird die Verschiebung der lokalen Intensitätsstruktur geschätzt, indem die Änderungen der Modellparameter ausgewertet werden. Dadurch kann der Bewegungsvektor direkt aus der Veränderung der polynomiellen Koeffizienten bestimmt werden. Diese Methode ist robust gegenüber Rauschen und Intensitätsschwankungen und ermöglicht eine dichte Bewegungsfeldschätzung [16].

MEHRSKALIGKEIT UND WARPING Um auch größere Bewegungen zu erfassen, wird das Verfahren (wie auch bei Lucas-Kanade) pyramidenbasiert implementiert: Die Schätzung erfolgt von groben zu feinen Auflösungsebenen, wobei auf jeder Ebene das Bewegungsfeld schrittweise verfeinert wird. Ein sogenanntes Warping hilft dabei, verbleibende Bewegungen zu reduzieren und die Genauigkeit zu erhöhen. Diese Strategie ist in gängigen Bibliotheken wie OpenCV implementiert.

PARAMETER UND PRAKTISCHE ASPEKTE Die wichtigsten Parameter sind die Fenstergröße der lokalen Approximation, die Anzahl der Pyramidenstufen sowie die Glättung. Größere Fenster und stärkere Glättung erhöhen die Robustheit, können jedoch feine Bewegungsdetails verwischen. Die sorgfältige Parametrierung ist insbesondere für Drohnenvideos entscheidend.

4.2.3 RAFT

RAFT ist ein Deep-Learning-basierter Ansatz zur Schätzung des dichten optischen Flusses, der 2020 von Teed und Deng vorgestellt wurde [17]. Im Vergleich zu früheren Deep-Learning-basierten Verfahren (z. B. FlowNet [20], PWC-Net [21]) erreicht **RAFT** eine deutlich höhere Genauigkeit auf mehreren Standard-Datensätzen (u. a. Sintel, KITTI), bei gleichzeitig konzeptionell vergleichsweise einfacher Architektur.

Im Rahmen dieser Arbeit dient **RAFT** als moderner Vertreter Deep-Learning-basierter optischer Flussverfahren und wird hinsichtlich Genauigkeit, Robustheit und Rechenzeit mit klassischen Verfahren wie Lucas-Kanade und Farneback verglichen.

ARCHITEKTURÜBERBLICK Die **RAFT**-Architektur lässt sich grob in drei Komponenten unterteilen [17]:

1. **Merkmalextraktion:** Aus den beiden Eingabebildern werden mittels eines Convolutional Neural Network (**CNN**)s Merkmalsrepräsentationen extrahiert.
2. **All-Pairs-Korrelationsvolumen:** Für alle Paare von Pixelpositionen in den beiden Bildern wird eine Korrelation berechnet und in einem hochdimensionalen Korrelationsvolumen gespeichert.
3. **Iterative Flussverfeinerung:** Ein rekurrentes Modul (Gated Recurrent Unit (**GRU**)) aktualisiert in mehreren Iterationsschritten ein dichten Flussfeld, das auf das Korrelationsvolumen zugreift.

Anstatt den optischen Fluss in einem einzigen Vorwärtsthroughlauf zu schätzen, wie dies bei vielen früheren Netzen der Fall ist, modelliert **RAFT** den Schätzprozess explizit als iterative Verfeinerung. Dies ähnelt konzeptionell klassischen Optimierungsverfahren, wird jedoch vollständig gelernt.

ITERATIVE FLUSSVERFEINERUNG Der optische Fluss wird in **RAFT** nicht direkt aus den Merkmalen vorhergesagt, sondern in mehreren Schritten verfeinert. Ausgangspunkt ist ein initiales Flussfeld (z. B. Nullfluss), das in jeder Iteration durch ein rekurrentes Netzwerk aktualisiert wird:

- Ein **Kontext-Netzwerk** extrahiert zusätzliche kontextuelle Informationen aus dem ersten Frame (z. B. lokale Struktur, Semantik).
- In jeder Iteration betrachtet das **GRU**-basierte Update-Modul:
 - das aktuelle Flussfeld,
 - das Korrelationsvolumen (abgetastet an Positionen, die dem aktuellen Fluss entsprechen),
 - den Kontext aus dem ersten Frame.
- Auf Basis dieser Informationen berechnet das Modul eine *inkrementelle* Flusskorrektur, die zum aktuellen Flussfeld addiert wird.

Dieser Prozess wird typischerweise mehrmals (oft 10-32 Mal) wiederholt, sodass das Flussfeld Schritt für Schritt verfeinert wird. Dadurch erreicht **RAFT** eine hohe Genauigkeit, insbesondere an Objektkanten und in Bereichen mit komplexen Bewegungen. Zudem wird das gleiche Update-Modul in jeder Iteration wiederverwendet (Gewichtssharing), was die Anzahl der Modellparameter begrenzt.

TRAININGSSTRATEGIE UND GENERALISIERUNG **RAFT** wird in der Originalarbeit auf einer Kombination mehrerer synthetischer und realer Datensätze trainiert (u. a. FlyingChairs, FlyingThings3D, Sintel, KITTI). Dabei kommen überwiegend überwachte Loss-Funktionen zum Einsatz, bei denen der geschätzte Fluss mit der bekannten Ground Truth verglichen wird (z. B. End-Point-Error). Durch Datenaugmentation (z. B. Random Crops, Helligkeits- und Geometrie-Transformationen) wird die Generalisierungsfähigkeit verbessert.

Ein zentrales Ergebnis von dem Paper [17] ist, dass **RAFT** auch auf Datensätzen, die sich deutlich von den Trainingsdaten unterscheiden, vergleichsweise gut generalisiert. Für Domänen wie Drohnenvideos ist dies besonders relevant, da annotierte Ground-Truth-Flussdaten dort typischerweise nur eingeschränkt verfügbar sind [7].

In diesem Kapitel werden die verwendeten Drohnenvideos, deren charakteristische Szenen sowie die Schritte der Datenvorverarbeitung systematisch beschrieben. Ziel ist es, die experimentelle Grundlage für die Analyse und Bewertung der Bewegungsvektorverfahren transparent darzustellen. Dies umfasst die Auswahl und Kategorisierung der Videos, die Vereinheitlichung der Eingabedaten sowie die Erstellung eines eigenen Ground-Truth- Datensatzes durch manuelle Annotation.

5.1 BESCHREIBUNG DER VERWENDETEN DATEN

Für die systematische Analyse und den Vergleich der Bewegungsvektorverfahren wurden in dieser Arbeit verschiedene Drohnenvideos herangezogen, welche unterschiedliche Szenarien, Komplexitätsgrade und Bewegungsmuster abbilden. Die Auswahl dieser Videos erfolgte mit dem Ziel, sowohl einfache als auch komplexe realitätsnahe Bedingungen abzudecken und eine belastbare Evaluation der Verfahren zu ermöglichen.

Bei den vom Fraunhofer Institut für Optronik, Systemtechnik und Bildverarbeitung (IOSB) zur Verfügung gestellten Daten handelt es sich um Drohnenvideos, die unter realen Bedingungen bei einer Veranstaltung aufgenommen wurden. Die Veranstaltung ist das JuicyBeats Festivals im Jahre 2024 in der Stadt Dortmund. Die Videos umfassen sowohl einfache Szenen mit wenigen Personen und deutlichen Bewegungen als auch komplexere Szenen mit dichten Menschenmengen und heterogenen Bewegungsrichtungen.

Für die Analyse der Bewegungsvektoren wurden insgesamt sieben Videos verwendet. Die Videos dienen als Grundlage für die Berechnung der Bewegungsvektoren mittels Optical-Flow-Verfahren.

Die wichtigsten Kennzahlen der verwendeten Videos sind in Tabelle [5.1](#) zusammengefasst. Für jedes Video wurden Anzahl der Frames, Dauer, Auflösung, Framerate und die Dateigröße dokumentiert.

Video	Frames	Dauer (s)	Auflösung	Framerate	Größe (MB)
Bühnen fern	219	7,30	1280x720	30	17,88
JuicyBeats Crowd fern	173	5,77	1280x720	30	6,61
JuicyBeats Crowd nah	124	4,13	1280x720	30	4,64
JuicyBeats nah	217	7,23	1280x720	30	8,03
Moshpit	296	9,87	1280x720	30	24,17
Moshpit trennen	237	7,90	1280x720	30	19,29
Moshpit zusam- men	236	7,87	1280x720	30	19,23

Tabelle 5.1: Eigenschaften der für die Optical-Flow-Analyse verwendeten Videos

5.2 CHARAKTERISIERUNG UND ABLAUF DER SZENEN

5.2.1 Charakterisierung der Videos

Die verwendeten Drohnenvideos lassen sich hinsichtlich Bewegungsdynamik, Personendichte und struktureller Eigenschaften in zwei Hauptkategorien einteilen: *einfache Szenen* und *komplexe Szenen*.

Einfache Szenen zeigen wenige Personen oder Objekte, die sich klar voneinander getrennt und mit deutlich erkennbaren Bewegungsmustern bewegen. Typische Beispiele sind Aufnahmen einzelner Personen oder kleiner Gruppen, welche sich über eine offene Fläche bewegen.

Komplexe Szenen hingegen zeigen dichte Menschenmengen vor der Bühne. In diesen Szenen sind die Bewegungsrichtungen oft heterogen, es treten zahlreiche Verdeckungen auf und die Einzelbewegungen sind semantisch schwer trennbar. Ein besonders herausforderndes Beispiel ist das Auftreten eines Moshpits.

DEFINITION MOSHPIT Ein Moshpit ist ein Bereich bei Konzerten, meist im Genre Heavy Metal oder Punk, in dem sich die Teilnehmer energisch und zufällig bewegen und dabei oft miteinander kollidieren. Die Bewegungen sind ungerichtet und erinnern an das Verhalten von Teilchen in einem Gas. Moshpits entstehen aus der kollektiven Stimmung und Energie der Musik sowie des Publikums [47].

Moshpits eignen sich besonders für die Analyse von Bewegungsvektoren in Drohnenvideos, da sie ein natürlich entstandenes, hochdynamisches und dichtes Mehrkörpersystem repräsentieren, in welchem individuelle Bewegungen stark durch lokale Interaktionen beeinflusst werden. Dadurch entstehen komplexe, nichtlineare Bewegungsmuster mit kontinuierlichen Richtungs- und Geschwindigkeitsänderungen. Gerade die hohe Variabilität bei gleichzeitig klarer Gruppenbildung macht Moshpits zu einem idealen Untersu-

chungsobjekt für die Validierung von Methoden zur Bewegungsvektorberechnung. Die Analyse dieser Szene ermöglicht es, die Grenzen und Potenziale der Methoden im Hinblick auf die Detektion und Quantifizierung komplexer Bewegungsmuster herauszuarbeiten.

5.2.2 Ablaufbeschreibung der Videos

Im Nachfolgenden wird jedes Video kurz beschrieben. Zusätzlich ist in Abbildung 5.1 für jedes Video ein stellvertretendes Standbild abgebildet.

Bühnen_fern.mp4

Das Video *Bühnen_fern.mp4* zeigt eine Drohnenaufnahme des Juicy Beats Festivals aus der Vogelperspektive. Im oberen linken Bereich des Bildes ist eine Bühne zu erkennen, während sich auf der rechten Seite, außerhalb des sichtbaren Bereichs, eine weitere Bühne befindet. Die beiden Bühnen werden durch ein offenes Gelände, mehrere Büsche sowie zwei Straßen voneinander getrennt. Zwischen diesen beiden Arealen ist eine Bewegung von Menschen zu beobachten, die sich entlang der Wege zwischen den Bühnen bewegen. Zudem sind auf dem Gelände zahlreiche Zelte, Fahrzeuge und kleinere Gebäude verteilt. Rechts im Bild sind Foodtrucks zu sehen. Vor diesen Foodtrucks haben sich Menschengruppen gebildet.

JuicyBeats_Crowd_fern.mp4

Das Video *JuicyBeats_Crowd_fern.mp4* zeigt eine Drohnenaufnahme des Festivalgeländes. Im Zentrum des Bildes befindet sich eine große Bühne, vor der sich eine dichte Menschenmenge versammelt hat. Um die Bühne herum sind verschiedene Bereiche erkennbar: offene Grünflächen, Bäume, Wege und kleinere Gebäude. Mehrere weiße Zelte und Fahrzeuge sind auf dem Gelände verteilt.

JuicyBeats_Crowd_nah.mp4

Im Video *JuicyBeats_Crowd_nah.mp4* ist eine große Menschenmenge, welche sich vor der Bühne versammelt hat, zu sehen. Die Bühne ist mit verschiedenen bunten Elementen und großen Bildschirmen ausgestattet. Auf der Bühne stehen mehrere Künstler. Die Zuschauerinnen und Zuschauer sind eng gedrängt und blicken in Richtung Bühne. Im Vordergrund sind technische Anlagen und Absperrungen zu erkennen, welche den Bühnenbereich vom Publikum trennen.

JuicyBeats_nah.mp4

Im Bild sind mehrere Foodtrucks und kleinere Verkaufsstände zu sehen, vor denen sich jeweils Menschengruppen gebildet haben. Besucher stehen oder sitzen auf dem offenen Gelände, einige sind in Gruppen versammelt. Die

Wege und Grünflächen zwischen den Ständen werden von Personen und Personengruppen genutzt, welche sich frei bewegen. Im Hintergrund sind weitere Veranstaltungsstrukturen und kleine Gebäude erkennbar.

Moshpit_trennen.mp4

Das Video zeigt eine Menschenmenge vor der Festivalbühne. Zwei Personen stehen im vorderen Bereich und trennen aktiv die Menge, um einen sogenannten "Moshpit" zu ermöglichen. Dabei teilen sich die Zuschauer nach links und rechts, sodass eine freie Fläche in der Mitte entsteht.

Moshpit_zusammen.mp4

Das Video zeigt, eine Menschenmenge vor der Bühne. In der Mitte der Menschenmenge ist ein freier Bereich. Ab der Mitte des Videos entsteht ein Moshpit: Die Menschenmenge rennt von links und von rechts aufeinander zu und trifft sich in der Mitte.

Moshpit.mp4

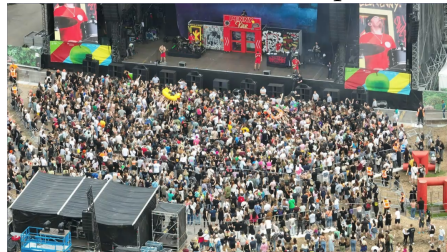
Das Video zeigt eine Menschenmenge vor der Bühne. In der Mitte der Menge ist ein Moshpit. Darin befinden sich Menschen, die sich dynamisch bewegen und miteinander kollidieren.



(a) Bühnen_fern.mp4



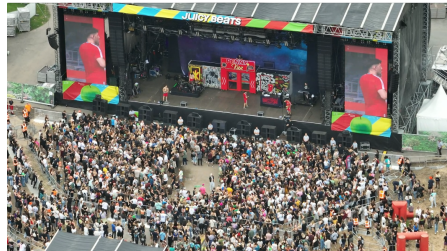
(b) JuicyBeats_Crowd_fern.mp4



(c) JuicyBeats_Crowd_nah.mp4



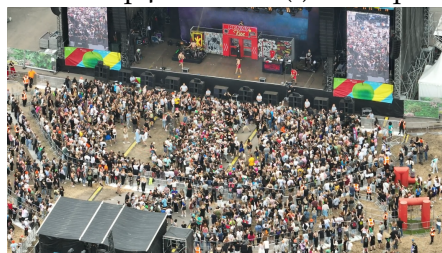
(d) JuicyBeats_nah.mp4



(e) Moshpit_trennen.mp4



(f) Moshpit_zusammen.mp4



(g) Moshpit.mp4

Abbildung 5.1: Jeweils ein stellvertretendes Standbild pro Video. (Eigene Darstellung.)

5.2.3 Zuteilung der Videos zu den Kategorien

Die sieben Videos wurden nach Sichtung und inhaltlicher Bewertung von der Versuchsleitung den beiden Kategorien zugeordnet. Die Bewertung erfolgte nach dem Vier-Augen-Prinzip. Dafür wurden die Bewertungskriterien aus Tabelle 5.2 genutzt. Die Bewertung jedes Videos anhand der Kriterien ist im Anhang in den Tabellen A.1 und A.2 zu finden. Jedes Video wurde der Kategorie zugeordnet, deren Kriterien überwiegend erfüllt wurden.

Kriterium	Beschreibung	Typische Ausprägung "einfach"	Typische Ausprägung "komplex"
Personendichte	Wie viele Personen pro Bildbereich/ bewegter Fläche?	Gering	Hoch
Bewegungsdynamik	Wie unterschiedlich oder schnell sind die Bewegungen?	Gering, gleichförmig	Hoch, heterogen, abrupt
Interaktionen	Gibt es viele Interaktionen oder Kollisionen?	Wenige, klar getrennte Personen	Viele Interaktionen, Überlagerungen
Sichtbarkeit/ Trennbarkeit einzelner Personen	Sind einzelne Personen oder Gruppen erkennbar und verfolgbar?	Ja, klar trennbar	Nein, oft überdeckt/ verschmolzen
Szenenstruktur	Aufbau der Szene (offene Fläche, Bühne, Hindernisse etc.)	Offen, übersichtlich	Dicht, komplex strukturiert
Okklusionen	Werden Personen häufig verdeckt?	Selten	Häufig
Bewegungsrichtung	Einheitliche oder viele verschiedene Bewegungsrichtungen?	Einheitlich oder wenige Richtungen	Viele, wechselnde Richtungen
Ereignisdichte	Gibt es besondere Ereignisse (z. B. Moshpit)?	Nein	Ja

Tabelle 5.2: Bewertungskriterien für die Zuordnung der Videos zu den Szenentypen.

Die Bewertung der sieben Videos anhand der Kriterien führte zu der Zuordnung in Tabelle 5.3. Zwei Videos wurden als einfache Szenen klassifiziert und fünf Videos wurden als komplex eingestuft.

Dateiname	Szenenkategorie
Bühnen_fern.mp4	Einfach
JuicyBeats_Crowd_fern.mp4	Komplex
JuicyBeats_Crowd_nah.mp4	Komplex
JuicyBeats_nah.mp4	Einfach
Moshpit.mp4	Komplex
Moshpit_trennen.mp4	Komplex
Moshpit_zusammen.mp4	Komplex

Tabelle 5.3: Zuordnung der Videos zu den Szenenkategorien.

Insgesamt bestehen die *einfachen Szenen* aus Übersichtsaufnahmen mit geringer Personenzahl und wenig Interaktion (z. B. JuicyBeats_nah.mp4), während die *komplexen Szenen* (wie Moshpit.mp4 oder JuicyBeats_Crowd_fern.mp4) durch hohe Dichte, zahlreiche Interaktionen und dynamische Bewegungsmuster charakterisiert sind.

Diese Einteilung bildet die Grundlage für die experimentelle Analyse und den Vergleich der Bewegungsvektorverfahren unter unterschiedlichen Komplexitätsgraden.

5.3 VORVERARBEITUNG DER VIDEODATEN

Für einen objektiven und reproduzierbaren Vergleich verschiedener Verfahren zur Schätzung von Bewegungsvektoren ist eine Vereinheitlichung der Eingabedaten hinsichtlich räumlicher und zeitlicher Auflösung erforderlich. Die Rohdaten lagen mit unterschiedlichen Bildauflösungen und Bildraten vor, abhängig von Kamerasensor, Flugprofil und Aufnahmeeinstellungen. Diese Unterschiede können jedoch Einfluss auf die Eigenschaften der geschätzten Bewegungsfelder und damit auf die Leistungsbewertung der untersuchten Methoden haben.

Die räumliche Auflösung bestimmt unter anderem die Anzahl verfügbarer Pixel pro Objekt sowie die Detail- und Texturinformation im Bild. Verfahren zur Bewegungsvektorschätzung reagieren sensibel auf diese Faktoren, da Gradientenstärke, Kantenschärfe und lokale Struktur maßgeblich von der Auflösung abhängen. Ein Vergleich von Ergebnissen bei variierender Auflösung würde daher nicht ausschließlich methodische Unterschiede widerspiegeln, sondern auch Effekte der Bildskalierung überlagern.

Analog beeinflusst die Bildrate die zeitliche Distanz zwischen aufeinanderfolgenden Frames und damit die Größe der zu schätzenden Bewegungen. Niedrige Bildraten führen zu größeren Pixelverschiebungen, während hohe Bildraten tendenziell kleinere, besser approximierbare Bewegungen erzeugen. Da viele Verfahren implizit Annahmen über die maximale Bewegungsgröße oder zeitliche Linearität treffen, würde eine unterschiedliche Bildrate die Vergleichbarkeit der Resultate zusätzlich einschränken.

Durch die Vereinheitlichung von Auflösung und Bildrate wird sichergestellt, dass alle untersuchten Verfahren unter identischen räumlich-zeitlichen Bedingungen evaluiert werden. Auf diese Weise lassen sich beobachtete Leistungsunterschiede eindeutig auf die Eigenschaften der jeweiligen Methoden zurückführen und nicht auf variierende Aufnahmeparameter. Die gewählten Parameter stellen somit eine kontrollierte experimentelle Basis für die vergleichende Analyse dar.

5.4 ANNOTATION UND GROUND-TRUTH-DATEN

Für die in dieser Arbeit verwendeten Videos existiert kein Ground-Truth-Datensatz, welcher für die Evaluierung von Bewegungsvektoren geeignet wäre. Daher wurde eine eigene manuelle Annotation durchgeführt, um einen geeigneten Vergleichsmaßstab zu schaffen.

Zu diesem Zweck wurde ein Annotationstool in Python entwickelt, welches die manuelle Markierung von korrespondierenden Punkten in aufeinanderfolgenden Frames ermöglicht. Das Tool zeigt jeweils zwei aufeinanderfolgende Videoframes an. Zum Annotieren werden im ersten Frame Zielpunkte ausgewählt, die dann entsprechend im zweiten Frame auch markiert werden. Die annotierten Punktpaare dienen als Referenz für die Berechnung des optischen Flusses und werden als Ground-Truth-Daten gespeichert.

Das Tool unterstützt verschiedene Komfortfunktionen, darunter Zoom, Pan und einen Blinkmodus zur besseren Orientierung zwischen den Frames. Der Blinkmodus springt mehrmals zwischen Frame 1 und Frame 2 hin und her, damit der Annotator sehen kann, wo auf dem Bild Bewegung stattfindet und so die passenden Punkte setzen kann. Die Annotationsergebnisse werden im JSON-Format abgespeichert und enthalten für jedes Frame-Paar die Quell- und Zielkoordinaten sowie die berechneten Vektoren.

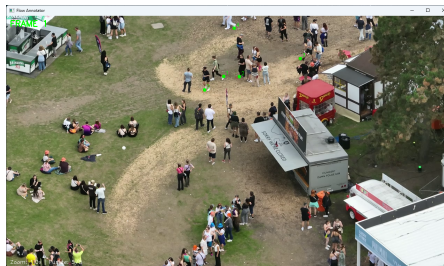
Die folgende Abbildung zeigt einen Screenshot des Annotationstools während der Arbeit an einem Frame-Paar.



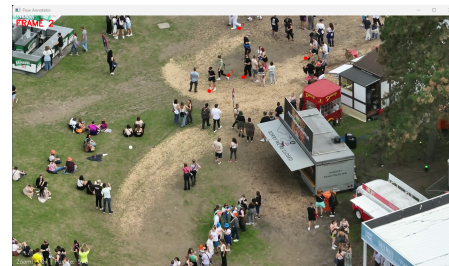
(a) Auswahl eines Punktes auf dem vergrößerten Frame 1 zur Vorbereitung der korrespondierenden Markierung im darauf folgenden Frame 2.



(b) Visualisierung der markierten Punkte: grüne Markierungen repräsentieren Frame 1, rote Markierungen den korrespondierenden Punkt auf Frame 2.



(c) Übersicht aller auf Frame 1 annotierter Punkte.



(d) Übersicht der zu Frame 1 korrespondierenden Punkte auf Frame 2.

Abbildung 5.2: Screenshots des eigens entwickelten Annotationstools zur Markierung von Punktpaaren zwischen zwei Frames. (Eigene Darstellung.)

Die Tabelle 5.4 gibt eine Übersicht über die Anzahl der annotierten Frame-Paare und die Gesamtanzahl der annotierten Punktpaare für jedes Video sowie die Gesamtsummen aller Videos:

Video	Anzahl annotierter Frame-Paare	Anzahl Punktpaare (gesamt)
JuicyBeats nah (Einfach)	100	501
Bühnen fern (Einfach)	96	483
JuicyBeats Crowd nah (Komplex)	59	293
JuicyBeats Crowd fern (Komplex)	83	416
Moshpit trennen (Komplex)	96	465
Moshpit zusammen (Komplex)	96	480
Moshpit (Komplex)	94	470
Gesamt	624	3108

Tabelle 5.4: Übersicht der Annotationen: Anzahl der Frame-Paare und Punktpaare pro Video

Die Annotationen decken damit eine große Bandbreite an Bewegungsmustern und Szenenkomplexitäten ab und ermöglichen eine fundierte quantitative Evaluation der Optical-Flow-Verfahren im Verlauf der Arbeit.

METHODIK

Dieses Kapitel beschreibt das methodische Vorgehen in der Arbeit, das als mehrstufige Pipeline konzipiert ist. Es erläutert die einzelnen Schritte von der Literaturrecherche über Datenvorbereitung über die Implementierung und Anwendung verschiedener Optical-Flow-Verfahren bis hin zur systematischen Durchführung und Auswertung der Experimente. Ziel ist es, eine transparente und reproduzierbare Berechnung und Analyse der Bewegungsvektoren in Drohnenvideos zu ermöglichen.

6.1 PIPELINE KONZEPT (ÜBERBLICK ÜBER DAS METHODISCHE VORGEHEN)

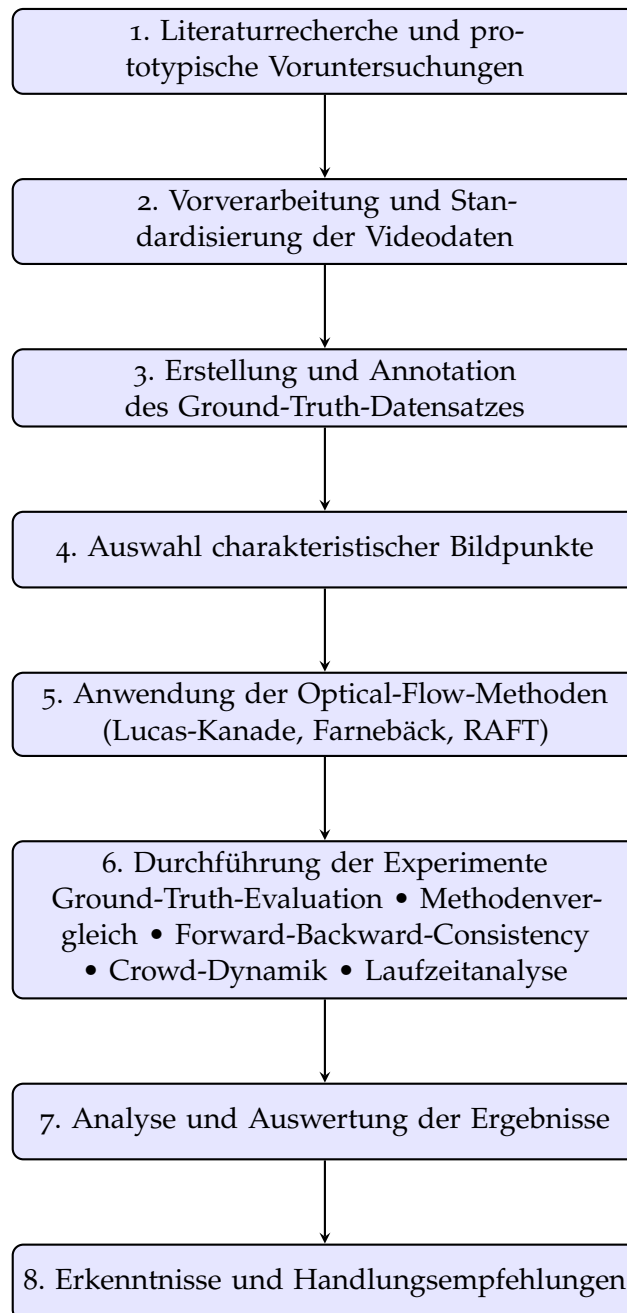


Abbildung 6.1: Methodische Pipeline der Arbeit. (Eigene Darstellung.)

Das methodische Vorgehen dieser Arbeit ist wie in Abbildung 6.1 dargestellt als achtstufige Pipeline aufgebaut.

1. Literaturrecherche und prototypische Voruntersuchungen

Zu Beginn stand eine ausführliche Literaturrecherche und erste Metho-

den wurden prototypisch getestet, um die finale Umsetzung festzulegen.

2. Vorverarbeitung und Standardisierung der Videodaten

Die Videodaten werden vorverarbeitet und standardisiert, um eine einheitliche und vergleichbare Ausgangsbasis für alle weiteren Analysen zu schaffen.

3. Erstellung und Annotation des Ground-Truth-Datensatzes

Da die Rohdaten keine Referenzannotationen enthalten, wird ein eigenes Tool entwickelt, um die notwendigen Ground-Truth-Daten zu generieren und damit eine objektive Evaluierung der Methoden zu ermöglichen.

4. Auswahl charakteristischer Bildpunkte

Es werden charakteristische Bildpunkte ausgewählt, die als Grundlage für die spätere Berechnung der Bewegungsvektoren dienen und sicherstellen, dass alle Optical-Flow-Methoden auf denselben Daten operieren.

5. Anwendung der Optical-Flow-Methoden

Die Verfahren Lucas-Kanade, Farnebäck und [RAFT](#) werden auf identischen Daten und Referenzpunkten angewendet, um eine objektive und vergleichbare Gegenüberstellung der Ergebnisse zu gewährleisten.

6. Durchführung der Experimente

Es werden verschiedene Experimente durchgeführt. Dazu zählen die Evaluation gegen die Ground-Truth-Daten, ein quantitativer Methodenvergleich ohne Ground-Truth-Daten, die qualitative Bewertung mittels Forward-Backward-Consistency, eine Analyse der Crowd-Dynamik sowie ein Laufzeitvergleich der Ansätze.

7. Analyse und Auswertung der Ergebnisse

Alle Ergebnisse werden gesammelt, mithilfe unterschiedlicher Bewertungsmetriken analysiert und die Methoden hinsichtlich ihrer Leistungsfähigkeit verglichen.

8. Ableitung von Erkenntnissen und Empfehlungen

Die gewonnenen Erkenntnisse werden zusammengeführt und konkrete Empfehlungen für den Einsatz der verschiedenen Optical-Flow-Verfahren abgeleitet.

Die gesamte Pipeline ist so konzipiert, dass sie reproduzierbar ist und sich auf verschiedene Datensätze oder Anwendungsszenarien übertragen lässt. Alle Zwischenschritte und Ergebnisse werden umfassend dokumentiert und visualisiert, um Transparenz und Nachvollziehbarkeit sicherzustellen.

6.2 DETAILBESCHREIBUNG DER PIPELINE-SCHRITTE

SCHRITT 1: LITERATURRECHERCHE UND PROTOTYPISCHE VORUNTERSUCHUNGEN Zu Beginn stand eine ausführliche Literaturrecherche, um die Methodik zur Beantwortung der Forschungsfrage festzulegen. Dabei gefundene Ansätze wurden prototypisch getestet und verglichen, um die finale Umsetzung festzulegen. Dieser Prozess ist in Abschnitt 3.3 ausführlich beschrieben.

SCHRITT 2: DATENVORBEREITUNG UND STANDARDISIERUNG Die vorliegende Arbeit basiert auf Drohnenaufnahmen mit statischer Kameraposition, wodurch Kamerabewegungen als Fehlerquelle entfallen und beobachtete Bewegungen echten Objektbewegungen entsprechen. Zunächst werden Videos anhand von Qualitätskriterien (Beleuchtung, Sichtbarkeit, Stabilität) geprüft und unbrauchbare Aufnahmen ausgeschlossen. Um faire Vergleiche zu ermöglichen, werden alle Sequenzen räumlich auf eine einheitliche Auflösung und zeitlich auf dieselbe Bildrate gebracht. Danach werden die Videos in Graustufen umgewandelt. Die Datenvorbereitung und Standardisierung wurde ausführlich in Abschnitt 5.3 beschrieben.

SCHRITT 3: ERSTELLUNG UND ANNOTATION DES GROUND-TRUTH-DATENSATZES Für die objektive Bewertung und den quantitativen Vergleich der Optical-Flow-Methoden ist ein annotierter Ground-Truth-Datensatz unerlässlich. Da die verwendeten Drohnenvideos keine eingebetteten Bewegungsvektoren oder Referenzdaten enthalten, wird ein spezifisches Annotationstool entwickelt, um an ausgewählten Bildpunkten manuelle Bewegungsvektoren zu erfassen. Diese Vorgehensweise entspricht dem Stand der Forschung, da in vielen Optical-Flow-Benchmarks Ground-Truth-Vektoren entweder durch manuelle Annotation [22] oder halbautomatische Verfahren [13] erzeugt werden, wenn keine synthetischen oder simulierten Daten zur Verfügung stehen. Die sorgfältige Erstellung von Ground-Truth-Datensätzen wird als zentrale Voraussetzung für die aussagekräftige Evaluation von Bewegungsdetektionsverfahren angesehen [20, 48]. Der detaillierte Ablauf der Annotation ist in Abschnitt 5.4 beschrieben.

SCHRITT 4: AUSWAHL CHARAKTERISTISCHER BILDPUNKTE Die Auswahl charakteristischer Bildpunkte ist ein grundlegender Schritt, um eine effiziente und robuste Bewegungsanalyse zu gewährleisten. In dieser Arbeit werden markante Bildpunkte mit dem Shi-Tomasi-Algorithmus extrahiert. Die Auswahl solcher Featurepunkte ist in der Optical-Flow-Literatur etabliert, da sie eine stabile Basis für die Schätzung von Bewegungsvektoren bieten und sowohl in klassischen als auch modernen Verfahren verwendet werden [13, 15].

Um die Vergleichbarkeit der verschiedenen Optical-Flow-Methoden sicherzustellen, werden die gleichen Featurepunkte für alle Verfahren genutzt. Dadurch wird verhindert, dass Unterschiede in der Punktselektion die Evalu-

ierung verfälschen. Die Forschung hebt hervor, dass die Konsistenz der betrachteten Punktmenge entscheidend für objektive Methodenvergleiche ist [13, 31]. Alternativen wie Harris-Corner oder SIFT werden zwar in anderen Kontexten eingesetzt, sind jedoch für dichte oder schnelle Bewegungsanalysen weniger geeignet, da sie entweder weniger selektiv oder deutlich rechenintensiver sind [49, 50].

Die resultierenden Featurepunkte werden für alle nachfolgenden Optical-Flow-Schätzungen gespeichert und dienen als Grundlage für das Sampling der Bewegungsvektoren bei dichten Verfahren wie Farneback und RAFT. So wird eine objektive und reproduzierbare Vergleichbarkeit über alle Methoden und Szenen hinweg erreicht.

SCHRITT 5: ANWENDUNG DER OPTICAL-FLOW-METHODEN Die Anwendung mehrerer Optical-Flow-Methoden auf denselben Datensatz dient dem Ziel, die Stärken, Schwächen und Einsatzbereiche unterschiedlicher Ansätze systematisch zu vergleichen. Unterschiedliche Methoden nutzen verschiedene algorithmische Prinzipien, was zu abweichender Genauigkeit, Robustheit und Laufzeit führen kann [13]. Ein direkter Vergleich ist daher nur möglich, wenn alle Verfahren unter identischen Bedingungen getestet werden, also mit denselben Eingangsdaten und Featurepunkten. Dies wird in der Optical-Flow-Forschung als essentielle Voraussetzung für objektive und aussagekräftige Evaluierungen hervorgehoben [22, 31].

Klassische Verfahren wie Lucas-Kanade (feature-basiert) und Farneback (dicht, polynomiell) gelten als zuverlässig in Standardszenarien und sind in vielen wissenschaftlichen Arbeiten als Vergleichsmethoden etabliert [13, 31]. Moderne Deep-Learning-Ansätze wie RAFT erzielen insbesondere bei komplexen Bewegungen und schwierigen Bildstrukturen teils deutlich bessere Ergebnisse [17]. Durch die gleichzeitige Anwendung aller drei Verfahren lassen sich Leistungsunterschiede in Bezug auf Genauigkeit, Robustheit gegenüber Störungen und Rechenaufwand systematisch untersuchen.

Die wissenschaftliche Literatur empfiehlt explizit, verschiedene Optical-Flow-Algorithmen unter kontrollierten, identischen Bedingungen zu vergleichen, um ihre Eignung für spezifische Anwendungsfälle bewerten zu können [13, 48]. Nur so lassen sich belastbare Aussagen über die Praxistauglichkeit und die Grenzen der jeweiligen Methoden treffen.

SCHRITT 6: DURCHFÜHRUNG DER EXPERIMENTE Die gezielte Durchführung verschiedener quantitativer und qualitativer Experimente ist entscheidend, um die Forschungsfrage umfassend und aus unterschiedlichen Perspektiven zu beantworten. Jedes Experiment adressiert dabei einen spezifischen Aspekt der Leistungsfähigkeit der untersuchten Optical-Flow-Methoden. Ein mehrstufiges experimentelles Vorgehen ist in der Optical-Flow-Forschung üblich und notwendig, um sowohl Genauigkeit als auch Robustheit, Laufzeit und Praxistauglichkeit systematisch zu bewerten [13, 48].

- **Experiment 1 (Evaluation gegen Ground-Truth-Daten):** Die Gegenüberstellung mit manuell annotierten Referenzvektoren ermöglicht ei-

ne objektive Bestimmung der Genauigkeit und ist Standard für die Algorithmenbewertung [22].

- **Experiment 2 (Quantitativer Methodenvergleich ohne Ground-Truth-Daten):** Da in vielen realen Szenarien keine Ground-Truth-Daten vorliegen, wird der direkte Vergleich der Methoden untereinander empfohlen, um relative Stärken und Schwächen offenzulegen [31].
- **Experiment 3 (Qualitätsbewertung mittels Forward-Backward-Consistency):** Die Forward-Backward-Consistency gilt als etabliertes internes Qualitätsmaß, um Fehler und Unsicherheiten zu erkennen, wenn keine Referenzdaten verfügbar sind [13, 29].
- **Experiment 4 (Crowd-Dynamik-Analyse):** Die Analyse von Bewegungsordnung, Geschwindigkeit und Gruppenverhalten ist für Crowd-Szenarien zentral und wird in der Forschung als wichtige Erweiterung klassischer Optical-Flow-Evaluationen angesehen [51, 52].
- **Experiment 5 (Laufzeitvergleich):** Die Bewertung der Rechenzeit ist ein wesentliches Praxiskriterium für die Anwendbarkeit von Optical-Flow-Verfahren, insbesondere bei zeitkritischen Anwendungen wie Videoanalyse [13, 22].

Die Kombination dieser Experimente ermöglicht eine ganzheitliche Bewertung und folgt anerkannten Standards für die Evaluierung von Bewegungsdetektionsverfahren.

Die konkrete Durchführung und Auswertung der Experimente wird in Kapitel 7 detailliert beschrieben.

SCHRITT 7: SYSTEMATISCHE AUSWERTUNG DER EXPERIMENTE Nach der Durchführung der Experimente erfolgt eine systematische Auswertung aller erhobenen Daten. Ziel ist es, die Leistungsfähigkeit der verschiedenen Optical-Flow-Methoden anhand objektiver Metriken und nachvollziehbarer Kriterien umfassend zu bewerten. Eine strukturierte, quantitative und qualitative Analyse wird in der Optical-Flow-Forschung als zentrale Voraussetzung für belastbare und vergleichbare Ergebnisse angesehen [13, 48]. Die systematische Auswertung ist in Kapitel 8 zu finden.

Zur Auswertung werden die in den Experimenten erhobenen Kennzahlen, wie Endpunktfehler (EPE), Winkelfehler, RMSE, Forward-Backward-Consistency, Richtungsentropie und Laufzeit, herangezogen. Diese Metriken gelten als Standard in der Bewertung von Bewegungsvektoren und sind in Optical-Flow-Benchmarks etabliert [22, 31, 48]. Die Analyse erfolgt sowohl aggregiert (z. B. Mittelwerte, Mediane, Verteilungen über alle Szenen) als auch differenziert nach Szenentypen und Methodenpaaren, um Unterschiede und Gemeinsamkeiten systematisch herauszuarbeiten.

Ergänzend werden qualitative Methoden wie Visualisierungen (z. B. Flowfelder, Amplitude- und Richtungs-Heatmaps, Fehlerkarten) eingesetzt, um

die Ergebnisse anschaulich darzustellen und Besonderheiten einzelner Szenen oder Methoden sichtbar zu machen [13]. Die Kombination aus quantitativer und qualitativer Auswertung ermöglicht eine ganzheitliche Bewertung der untersuchten Algorithmen und dient als Grundlage für die Beantwortung der Forschungsfrage.

Die gesamte Auswertung erfolgt reproduzierbar, indem sämtliche Analyseschritte, Parameter und Visualisierungen dokumentiert und archiviert werden [48]. Dadurch wird eine transparente und überprüfbare Bewertung der Optical-Flow-Methoden gewährleistet.

SCHRITT 8: ZUSAMMENFÜHRUNG DER ERKENNTNISSE UND ABLEITUNG VON EMPFEHLUNGEN Im letzten Schritt werden die Ergebnisse aller Analysen in einen Gesamtzusammenhang gestellt und hinsichtlich der Forschungsfrage bewertet. Neben dem reinen Vergleich der Methoden steht hier besonders die Interpretation der Resultate im Vordergrund. Es werden zentrale Muster, Stärken und Schwächen der einzelnen Optical-Flow-Verfahren identifiziert und die wichtigsten Erkenntnisse priorisiert [13].

Darauf aufbauend werden gezielt Empfehlungen für die Auswahl und den Einsatz der untersuchten Methoden in ähnlichen Anwendungsszenarien formuliert. Diese Empfehlungen richten sich nach den in der Auswertung identifizierten Bedingungen (etwa Szenenkomplexität, Echtzeit-Anforderungen oder Genauigkeitsbedarf) und helfen, fundierte Entscheidungen für den praktischen Einsatz zu treffen [22]. Besonderes Augenmerk gilt dabei auch der Übertragbarkeit der Erkenntnisse auf andere Datensätze und Anwendungsbereiche sowie den praktischen Grenzen der jeweiligen Ansätze.

Abschließend werden offene Fragen, Limitationen und Ansatzpunkte für weiterführende Forschung benannt, um den Erkenntnisfortschritt transparent einzuordnen und Impulse für zukünftige Arbeiten zu geben [48].

6.3 TECHNISCHE UMSETZUNG / REALISIERUNG

Die gesamte Pipeline wurde modular in Python (Version 3.10) umgesetzt. Die technische Realisierung erfolgte in Form von Jupyter Notebooks und Python-Skripten, wobei zentrale Bibliotheken wie OpenCV, NumPy, PyTorch und Matplotlib verwendet wurden. Die Deep-Learning-Komponenten (RAFT) wurden GPU-beschleunigt auf einer NVIDIA RTX 2070 SUPER (CUDA 12.1) ausgeführt.

Vorverarbeitung und Ground-Truth-Annotation

Die Videoframes werden einheitlich vorverarbeitet, indem sie als BGR-Bilder eingelesen und für klassische Methoden in Graustufen konvertiert werden. Die Rohvideos werden hinsichtlich Auflösung und Framerate vereinheitlicht. Eine Kompensation der globalen Kamerabewegung mittels Homographie-Schätzung wurde in den ausgewählten Szenen aufgrund der stabilen Flugphasen der Drohnen nicht benötigt. Die manuelle Ground-Truth-Annotation

erfolgt mittels eines eigenen grafischen Tools, das eine interaktive Erfassung von korrespondierenden Bildpunkten in Framepaaren ermöglicht. Das Tool speichert die annotierten Punktpaare sowie die daraus berechneten Bewegungsvektoren als JSON-Dateien ab.

Optical-Flow-Implementierung und Methodenvergleich

Für jeden Frame werden markante Bildpunkte mittels Shi-Tomasi-Algorithmus (`cv2.goodFeaturesToTrack`) detektiert. Anschließend werden für diese Punkte Bewegungsvektoren berechnet:

- **Lucas-Kanade (LK):** Sparse Optical Flow mit `cv2.calcOpticalFlowPyrLK`, nur für die detektierten Featurepunkte.
- **Farnebäck:** Dichtes Flowfeld mit `cv2.calcOpticalFlowFarneback`, Sampling an den gleichen Featurepunkten.
- **RAFT:** GPU-beschleunigtes Deep-Learning-Modell (PyTorch, vortrainiert auf KITTI), Input als RGB-Tensoren, ebenfalls Sampling an den Featurepunkten.

Für die Auswertung werden für alle Methoden und identische Punktpositionen zentrale Metriken wie RMSE (Root Mean Squared Error), Winkeldifferenz oder [EPE](#) berechnet. Die Ergebnisse werden als CSV, JSON und Visualisierungen (z. B. Quiver-Plots, Boxplots, Heatmaps) automatisiert abgelegt.

In Abbildung [6.2](#) ist ersichtlich, dass alle drei Methoden an den gleichen Punkten gesampelt werden. Jeder Pfeil steht dabei für eine Methode.



Abbildung 6.2: Prototypische Visualisierung der Bewegungsvektoren aller drei Methoden auf einem Beispielframe. (Eigene Darstellung.)

Software- und Hardware-Umgebung

Programmiersprache:	Python 3.10
Kernbibliotheken:	OpenCV, NumPy, PyTorch, Matplotlib, Pandas, tqdm
Deep-Learning-Modell:	RAFT (pretrained, KITTI, PyTorch)
Hardware (GPU):	NVIDIA RTX 2070 SUPER (CUDA 12.1)
Hardware (CPU):	AMD Ryzen 7 3700X 8-Core Prozessor
RAM:	32 GB

Tabelle 6.1: Überblick über zentrale Software- und Hardwarekomponenten.

Reproduzierbarkeit und Dokumentation

Alle Parameter (z. B. Featureanzahl, Modellpfade, Hardwareauswahl) sind zentral konfigurierbar. Die Pipeline ist modular aufgebaut, sodass alle Zwischenschritte, Auswertungen und Visualisierungen reproduzierbar sind. Die Ground-Truth-Annotation, die Methodenvergleiche und die Evaluationen sind als eigenständige, aber kompatible Notebooks dokumentiert.

EXPERIMENTE

Dieses Kapitel beschreibt die Durchführung und Auswertung aller Experimente, die zur systematischen Bewertung klassischer und Deep-Learning-basierter Optical-Flow-Methoden in Drohnenvideos dienen. Es werden Aufbau, Metriken und Ergebnisse der einzelnen Analysen vorgestellt, um die Forschungsfrage fundiert zu beantworten.

7.1 ÜBERSICHT DER EXPERIMENTE

Um die Forschungsfrage dieser Arbeit zu beantworten, wurden eine Reihe gezielter Experimente durchgeführt, die es erlauben, die Leistungsfähigkeit der ausgewählten klassischen und Deep-Learning-basierten Methoden zur Bewegungsdetektion in Drohnenvideos umfassend zu bewerten. Ziel war es, die Methoden hinsichtlich Genauigkeit, Robustheit und Rechenzeit systematisch zu vergleichen.

Das experimentelle Untersuchungsdesign dieser Arbeit besteht aus mehreren, aufeinander abgestimmten Teil-Analysen, welche gezielt unterschiedliche Leistungsaspekte der untersuchten Optical-Flow-Verfahren adressieren. Diese Experimente wurden auf einer gemeinsamen Datengrundlage (vereinheitlichte Drohnenvideos, zentrale Featurepunkt-Auswahl) durchgeführt und decken sowohl quantitative als auch qualitative Bewertungsdimensionen ab.

Nr.	Experiment / Teil-Analyse	Zweck / Zielsetzung
1	Evaluation anhand von Ground-Truth-Daten	Quantitative Bestimmung von Genauigkeit (EPE, Winkelfehler) der Methoden anhand manueller Referenzvektoren.
2	Quantitativer Methodenvergleich ohne Ground-Truth-Daten	Vergleich der Ergebnisse der Methoden direkt miteinander (EPE, RMSE, Fl-all, Winkeldifferenz), unabhängig von externer Referenz.
3	Qualitätsbewertung mittels Forward-Backward-Consistency	Analyse der Konsistenz der Bewegungsvektoren (vorwärts/rückwärts), als Proxy-Maß für Zuverlässigkeit auch ohne Ground-Truth-Daten.
4	Crowd-Dynamik-Analyse	Evaluation der Fähigkeit der Methoden, dynamische Gruppenbewegungen und deren Eigenschaften (Richtungsentropie, Geschwindigkeit, Divergenz, Rotation (Curl)) im Crowd-Kontext zu erfassen.
5	Laufzeitvergleich	Vergleich der Rechenzeiten der Methoden, Ermittlung der Echtzeitfähigkeit und Analyse des Einflusses der Szenenkomplexität.

Tabelle 7.1: Übersicht aller durchgeführten Experimente.

7.2 EXPERIMENT 1: QUANTITATIVE EVALUATION ANHAND VON GROUND-TRUTH-DATEN

Zur objektiven Bewertung der Genauigkeit der untersuchten Optical-Flow-Methoden wurde ein quantitatives Experiment anhand von manuell annotierten Ground-Truth-Bewegungsvektoren durchgeführt. Ziel dieses Experiments ist es, die Leistung von Lucas-Kanade, Farnebäck und [RAFT](#) anhand etablierter Fehlermaße systematisch zu vergleichen.

Die objektive Bewertung von Optical-Flow-Algorithmen anhand manuell annotierter Ground-Truth-Daten und etablierter Fehlermaße ist Standard im Bereich der Bewegungsanalyse. Die Verwendung von Endpunktfehler (EPE) und Winkelfehler (Angular Error, AE) als zentrale Gütekriterien ist sowohl historisch als auch aktuell breit akzeptiert und in allen relevanten Bench-

marks und Übersichtsarbeiten zu Optical Flow verankert [13, 14, 18, 22, 31, 53].

Baker et al. [13] und die darauf aufbauenden Arbeiten (z. B. FlowBench [18]) haben gezeigt, dass nur durch die Evaluation auf echten Ground-Truth-Daten mit pixel-genauen Referenzvektoren eine objektive und differenzierte Bewertung von Genauigkeit, Robustheit und Generalisierungsfähigkeit möglich ist. Die Kombination von EPE und Winkelfehler erlaubt dabei eine umfassende Erfassung von Fehlern in Betrag und Richtung der geschätzten Bewegungsvektoren und ist als Best-Practice für Optical-Flow-Benchmarks empfohlen [13, 22, 53]. Die Anwendung dieser Metriken ist nicht nur für klassische Algorithmen wie Lucas-Kanade oder Farnebäck, sondern auch für moderne Deep-Learning-Verfahren etabliert und ermöglicht die direkte Vergleichbarkeit und Reproduzierbarkeit zwischen verschiedenen Ansätzen [18, 23]. Hinsichtlich ihrer Aussagekraft und Relevanz wurden beide Fehlermaße in umfassenden Metastudien, methodischen Übersichtsarbeiten und aktuellen Benchmarks (auch im Kontext von Crowd-Analytics) als unverzichtbar zur objektiven Quantifizierung der Methodenleistung bewertet [11, 13, 14].

Versuchsaufbau

Für ausgewählte Framepaare aus allen Drohnenvideos wurden charakteristische Bildpunkte manuell annotiert, um Ground-Truth-Vektoren zu erhalten. Alle drei Optical-Flow-Methoden wurden auf identische Punkte und Framepaare angewandt. Die Experimente wurden standardisiert durchgeführt, um Vergleichbarkeit und Reproduzierbarkeit zu gewährleisten.

Metriken

Die Methoden wurden anhand folgender Metriken beurteilt:

- **Endpunktfehler (EPE):** Euklidischer Abstand zwischen geschätztem und Ground-Truth-Vektor.
- **Winkelfehler:** Abweichung der Bewegungsrichtung in Grad.

Beide Fehlerwerte wurden pro Bildpunkt berechnet und anschließend aggregiert (Mittelwert, Median, Maximum) sowie in ihrer Verteilung analysiert.

Durchführung

Für jedes Framepaar und jeden Bildpunkt wurden EPE und Winkelfehler bestimmt. Die aggregierten Fehlerwerte sowie deren Verteilungen wurden in Boxplots und Balkendiagrammen visualisiert. Zusätzlich wurden die Ergebnisse nach Szenenkomplexität (einfach/komplex) und Bewegungsstärke

(Amplitude) gruppiert, um den Einfluss dieser Faktoren auf die Methodenleistung zu untersuchen.

7.3 EXPERIMENT 2: QUANTITATIVER METHODENVERGLEICH OHNE GROUND-TRUTH-DATEN

Zur Bewertung der Konsistenz und Übereinstimmung der Optical-Flow-Verfahren wurde ein experimenteller Methodenvergleich ohne Bezug auf manuelle Ground-Truth-Daten durchgeführt. Ziel dieses Experiments ist es, systematische Unterschiede und Gemeinsamkeiten zwischen Lucas-Kanade, Farnebäck und [RAFT](#) anhand quantitativer Metriken zu identifizieren.

Das experimentelle Vorgehen, Optical-Flow-Methoden anhand ihrer gegenseitigen Übereinstimmung und Konsistenz zu vergleichen, ist wissenschaftlich fundiert und in der Literatur anerkannt. Da in vielen realen Szenarien keine verlässlichen Referenzdaten vorliegen, gilt der Vergleich von Methoden untereinander (beispielsweise durch Differenzmetriken, Unsicherheits- und Konfidenzmaße) als sinnvolle Strategie, um die Robustheit, Stärken und Schwächen verschiedener Ansätze objektiv zu bewerten. Zahlreiche Arbeiten zeigen, dass solche Analysen wichtige Hinweise auf die Zuverlässigkeit und Praxistauglichkeit der Verfahren liefern und damit eine legitime Alternative zur klassischen Evaluation anhand von Ground-Truth-Daten darstellen. Barron et al. [31] zeigen, dass direkte, quantitative Methodenvergleiche notwendig sind, um systematische Unterschiede und Stärken verschiedener Ansätze sichtbar zu machen. Kybic & Nieuwenhuis [54] empfehlen explizit unter anderem Konsistenzanalysen zur Bewertung der Zuverlässigkeit von Optical Flow einzusetzen, besonders wenn keine Ground-Truth-Daten vorliegen. Wannenwetsch et al. [55] verdeutlichen zudem, dass Unsicherheits- und Konsistenzanalysen ein wichtiges Werkzeug sind, um die Praxistauglichkeit und Robustheit von Optical-Flow-Verfahren auch ohne Referenzdaten objektiv zu bewerten.

Versuchsaufbau

Für sämtliche ausgewählten Drohnenvideos wurden die Bewegungsvektoren an identischen Bildpunkten mit allen drei Methoden berechnet. Die Analyse erfolgte paarweise, wobei die jeweiligen Differenzen zwischen den Methoden als Fehlerwerte ausgewertet wurden.

Metriken und Analyse

Folgende Metriken wurden zur Bewertung der Methodenunterschiede eingesetzt:

- **Endpunktfehler (EPE):** Euklidischer Abstand zwischen den geschätzten Vektoren zweier Methoden.
- **RMSE:** Quadratwurzel des mittleren quadratischen Fehlers.

- **Fl-all:** Fl-all steht für „Flow-all“ und bezeichnet den Prozentsatz der Bildpunkte, bei denen der Fehler des geschätzten Bewegungsvektors einen bestimmten Schwellenwert überschreitet. Als Fehlermaß wird der Endpunktfehler (EPE) verwendet. Ein niedriger Wert bedeutet kaum Ausreißer und eine sehr gute Schätzgenauigkeit des Flows. Ein hoher Wert bedeutet viele grobe Ausreißer und eine tendenziell unzuverlässige Methode.
- **Winkelabweichung:** Differenz der Bewegungsrichtung in Grad.

Alle Metriken wurden sowohl für alle Videos gemeinsam als auch getrennt nach Szenenkomplexität (einfach/komplex) berechnet. Die statistischen Kennwerte (Verteilungen, Mittelwerte, Streuungen) wurden in Boxplots und Tabellen visualisiert.

Durchführung

Die Auswertung erfolgte für die Methodenpaare **LK** vs. **RAFT**, **FB** vs. **RAFT** und **LK** vs. **FB**. Um den Einfluss der Szenenkomplexität auf die Unterschiede zwischen den Methoden zu analysieren, wurden die Ergebnisse zusätzlich nach einfachen und komplexen Szenen gruppiert.

7.4 EXPERIMENT 3: QUALITÄTSBEWERTUNG MITTELS FORWARD-BACKWARD CONSISTENCY

Zur Einschätzung der Qualität der berechneten Bewegungsvektoren wurde die *Forward-Backward-Consistency* als Proxy-Maß eingesetzt. Ihre Relevanz ergibt sich daraus, dass in vielen realen Szenarien keine Ground-Truth-Daten vorliegen. Die Forward-Backward-Consistency ermöglicht dennoch eine interne Qualitätsprüfung, denn sie prüft, ob ein Punkt nach Vorwärts- und Rückwärts-Verfolgung auf nahezu denselben Bildpunkt zurückgeführt wird. Ein Schwellenwert von 1,5 px wurde gewählt, um konsistente von inkonsistenten Punkten zu unterscheiden. Dieser Wert orientiert sich an etablierten Benchmarks in der Optical-Flow-Literatur und gewährleistet eine robuste Fehlergrenze. Außerdem wurde er an den durchschnittlichen **EPE** aus Experiment 1 angepasst. Hohe Forward-Backward-Consistency-Werte deuten auf robuste und verlässliche Schätzungen hin und helfen, fehlerhafte oder instabile Bewegungsvektoren zu identifizieren.

Die Anwendung der Forward-Backward-Consistency als Qualitätsmaß für Bewegungsvektoren ist in der Optical-Flow-Literatur etabliert und wird insbesondere dann empfohlen, wenn keine Ground-Truth-Daten verfügbar sind. Baker et al. [13] beschreiben die Forward-Backward-Consistency explizit als gängige interne Evaluationsmetrik und empfehlen diesen Ansatz zur robusten Fehlererkennung bei realen Benchmarks. Kalal et al. zeigen empirisch, dass der Forward-Backward-Error ein robustes Kriterium zur Erkennung von Tracking-Fehlern ist und empfehlen einen Schwellenwert im Bereich von 1 Pixel [29]. Alfano et al. stellen Forward-Backward-Consistency als

Standard zur internen Qualitätskontrolle vor [11]. Jeong et al. zeigen, dass Konsistenzmetriken wie Forward-Backward-Consistency zentrale Bestandteile moderner, selbstüberwachter Optical-Flow-Verfahren sind und explizit zur Qualitätsbewertung und Fehlerdetektion beitragen [30]. Xu et al. nutzen das Forward-Backward-Prinzip als Teil ihrer Konfidenzbewertung und belegen so die Relevanz von Konsistenzprüfungen zur Bewertung der Vertrauenswürdigkeit optischer Flussfelder [56].

Versuchsaufbau

Für alle Framepaare wurden die Bewegungsvektoren mit Lucas-Kanade, Farneback und RAFT berechnet. Für jeden Bildpunkt wurde zunächst der Bewegungsvektor vom ersten zum zweiten Frame bestimmt (Vorwärtsrichtung), anschließend der Rückwärtsvektor vom zweiten zurück zum ersten Frame. Die Differenz zwischen Ausgangs- und Rückkehrpunkt wurde als Forward-Backward-Error erfasst.

Metriken

- **Forward-Backward-Consistency (%)**: Anteil der Punkte mit Forward-Backward-Error unter 1,5 px.
- **Forward-Backward-Error (px)**: Mittlerer Fehlerwert pro Video und Szene.

Analyse und Visualisierung

Die Auswertung erfolgte als Heatmaps und räumliche Konsistenzkarten. Zusätzlich wurde die Forward-Backward-Consistency getrennt nach Szenenkomplexität (einfach/komplex) betrachtet, um den Einfluss unterschiedlicher Bedingungen auf die Konsistenz der Methoden zu untersuchen.

7.5 EXPERIMENT 4: CROWD-DYNAMIK-ANALYSE

Ziel des Experiments ist es, Unterschiede in der Bewegungsordnung und -intensität zwischen verschiedenen Szenen und Optical-Flow-Methoden zu erfassen und zu vergleichen. Zur systematischen Charakterisierung der Bewegungsdynamik wurden verschiedene statistische Kennzahlen und Visualisierungsmethoden eingesetzt.

Das Vorgehen im Experiment orientiert sich am Stand der Forschung zur systematischen Analyse von Crowd-Dynamik. Gu et al. nutzen Richtungs-Entropie und Geschwindigkeitsverteilung, um Bewegungsordnung und -intensität in Menschenmengen quantitativ zu charakterisieren und verschiedene Szenen sowie Optical-Flow-Methoden zu vergleichen. Die Kombination quantitativer und qualitativer Auswertungen auf Basis von Optical-Flow-Daten entspricht dem empfohlenen methodischen Vorgehen zur Untersuchung kom-

plexer Crowd-Szenarien [51]. Auch Zhang et al., Choudhary et al. und Altalbi et al. setzen Entropie gezielt ein, um zwischen normalem Verhalten und Panik bzw. abnormalem Verhalten zu unterscheiden und Anomalien frühzeitig zu detektieren [57–59]. Abdelwahab et al. betonen, dass Geschwindigkeit ein zentraler Aspekt bei der Analyse von Crowd-Verhalten ist [60]. Chen et al. legitimieren die gezielte Analyse von Richtung, Geschwindigkeit und Intensität mittels Optical Flow sowie deren Visualisierung als etablierten Ansatz zur Unterscheidung geordneter und chaotischer Bewegungsmuster [61]. Die Messung des Auseinanderlaufens benachbarter Partikel im Flussfeld, wie von Ali & Shah beschrieben, ist ein anerkanntes Verfahren zur Erkennung unterschiedlicher Dynamiken und Instabilitäten in Crowd-Flows [52], weshalb im Experiment eine Divergenz-Analyse integriert wurde. Auch Xu et al. legitimieren die Verwendung von Heatmaps zur Visualisierung und Analyse der Crowd-Verteilung, adressieren die Bedeutung von Metriken zur Erfassung von Heterogenität und Aggregation (z. B. über Entropie und Uniformität) sowie die Analyse räumlicher und bewegungsbedingter Unterschiede [36]. Diese Aspekte werden im Experiment durch die Auswertung von Entropie, Uniformität, Divergenz und Curl systematisch quantifiziert.

Versuchsaufbau

Die Analyse erfolgte auf Basis der mit Lucas-Kanade, Farnebäck und RAFT berechneten Bewegungsvektoren aus ausgewählten Drohnenvideos. Die folgenden Parameter wurden pro Video und Methodenpaar ausgewertet:

Metriken

- **Richtungsentropie:** Quantitative Messung der Bewegungsordnung. Hohe Werte kennzeichnen chaotische, niedrige Werte gerichtete Bewegungen.
- **Richtungsuniformität:** Maß für die Dominanz einer Bewegungsrichtung im Video. Hohe Uniformitäts-Werte bedeuten eine dominante Bewegungsrichtung.
- **Geschwindigkeitsverteilung:** Analyse der mittleren Bewegungsgeschwindigkeit.
- **Divergenz und Rotation(Curl):** Messung von räumlicher Expansion und Rotation (Drehbewegungen) im Bewegungsfeld.

Ergänzend wurden qualitative Heatmaps zur Visualisierung der Bewegungsmuster und Amplitude-Heatmaps zur räumlichen Darstellung der Geschwindigkeitsverteilung erstellt.

Analyse und Auswertung

Die ausgewählten Metriken wurden für verschiedene Szenariotypen (z. B. Crowd vs. Moshpit) und Methoden (Lucas-Kanade, Farneback, RAFT) berechnet und in Boxplots, Heatmaps und Scatterplots dargestellt. Dadurch konnten Unterschiede in Bewegungsordnung, Geschwindigkeit und Gruppendynamik systematisch erfasst und verglichen werden. Die experimentelle Auswertung ermöglicht eine differenzierte Bewertung der Leistungsfähigkeit der Optical-Flow-Verfahren im Kontext komplexer Crowd-Szenarien.

7.6 EXPERIMENT 5: LAUFZEITVERGLEICH

Das Ziel dieses Experiments ist die Analyse und der Vergleich der Verarbeitungszeiten der drei untersuchten Optical-Flow-Methoden (Lucas-Kanade, Farneback, RAFT) unter praxisnahen Bedingungen. Besonderes Augenmerk lag auf der Bewertung der Echtzeitfähigkeit sowie der Skalierbarkeit mit zunehmender Szenenkomplexität.

Der Laufzeitvergleich der Optical-Flow-Methoden (Lucas-Kanade, Farneback, RAFT) ist durch den aktuellen Forschungsstand gut begründet. Ju et al. unterstreichen die Bedeutung der Echtzeitfähigkeit und GPU-Beschleunigung für den praktischen Einsatz in dynamischen und komplexen Szenarien [62]. Qiu et al. bestätigen, dass die Wahl und Kombination von Optical-Flow-Methoden direkt die Berechnungszeit und Praxistauglichkeit beeinflusst [63], weshalb die Analyse der Laufzeiten zentral für die Bewertung der Methoden im Anwendungskontext ist. Darüber hinaus zeigen Kajo et al. in ihrer umfassenden Übersicht und Evaluation von Optical-Flow-Algorithmen, dass die Laufzeit ein zentrales Vergleichskriterium darstellt und insbesondere für Echtzeitanwendungen wie Crowd-Analytics von entscheidender Bedeutung ist [14]. Die Autoren führen einen quantitativen Vergleich der Rechenzeiten verschiedener Methoden durch und zeigen, dass klassische Ansätze wie Horn & Schunck oder Lucas-Kanade trotz ihrer Einfachheit weiterhin relevant für zeitkritische Anwendungen bleiben. Auch Sun et al. analysieren die Laufzeiten moderner Verfahren detailliert und stellen tabellarisch dar, wie sich verschiedene Modellierungen und Optimierungstechniken auf die Rechenzeit auswirken [22]. Baker et al. haben in ihrer Methodensammlung und Online-Benchmark die Ausführungszeiten zahlreicher Optical-Flow-Algorithmen auf standardisierten Datensätzen veröffentlicht und zeigen, dass Laufzeit und Genauigkeit entscheidende Parameter für die Praxistauglichkeit und die Auswahl geeigneter Methoden im Forschungskontext sind [13].

Versuchsaufbau

Die Laufzeiten wurden auf einem System mit GPU-Beschleunigung (NVIDIA GeForce RTX 2070 SUPER) gemessen. Für jede Methode wurden die mittleren Verarbeitungszeiten pro Frame sowie die erreichbaren Bilder pro

Sekunde (FPS) erfasst. Die Auswertung erfolgte getrennt nach einfachen und komplexen Szenen, um den Einfluss der Szenenkomplexität abzubilden.

Metriken

- **Mittlere Laufzeit pro Frame (ms):** Durchschnittliche Rechenzeit für die Verarbeitung eines Frames pro Methode.
- **Erreichbare FPS:** Vergleich zwischen berechneter Bildrate und den Video-FPS (30 FPS) zur Bewertung der Echtzeitfähigkeit.

Analyse und Auswertung

Die Ergebnisse wurden als Balkendiagramme, Boxplots und Tabellen dargestellt, um Unterschiede zwischen Methoden und Szenarien transparent zu machen. Zusätzlich wurden die Laufzeiten in einfachen und komplexen Szenen verglichen und der Hardware-Kontext dokumentiert.

Echtzeitfähigkeit

Zur Einschätzung der Praxistauglichkeit wurden die erreichten Frames per second (FPS) mit der tatsächlichen Video-Framerate verglichen. Als echtzeitfähig gilt das Verfahren, wenn die Verarbeitungszeit pro Frame den Kehrwert der angestrebten Bildrate nicht überschreitet. Die generierten FPS müssen folglich größer oder gleich der Video-Framerate sein.

7.7 ZUSAMMENFASSENDE BEWERTUNG

Alle beschriebenen Experimente wurden auf einer einheitlichen Datengrundlage und unter standardisierten, reproduzierbaren Bedingungen durchgeführt. Jede Teil-Analyse liefert dabei eine eigenständige Perspektive auf die Leistungsfähigkeit der untersuchten Optical-Flow-Methoden und adressiert gezielt unterschiedliche Qualitätsdimensionen: Genauigkeit, Robustheit, Konsistenz, Praxistauglichkeit und Anwendbarkeit in realen Crowd-Szenarien.

Die Kombination aus quantitativen (Ground-Truth-basiert, Methodenvergleich, Crowd-Analyse, Laufzeit) und qualitativen (Forward-Backward Consistency, Visualisierungen) Experimenten ermöglicht eine umfassende und mehrdimensionale Bewertung klassischer und Deep-Learning-basierter Optical-Flow-Verfahren. Damit werden sowohl Stärken als auch Schwächen der Methoden unter realitätsnahen, für Drohnenvideos typischen Bedingungen transparent herausgearbeitet.

ERGEBNISSE

Dieses Kapitel präsentiert die Ergebnisse der in Kapitel 7 beschriebenen Experimente. Dazu gehören die Evaluation anhand von den Ground-Truth-Daten, der quantitative Methodenvergleich ohne Referenzdaten, die qualitative Bewertung mittels Forward-Backward-Consistency, die Analyse der Crowd-Dynamik sowie der Laufzeitvergleich. Die Ergebnisse geben einen umfassenden Überblick über die Leistungsfähigkeit und Grenzen der Methoden in unterschiedlichen Szenarien.

8.1 EVALUATION ANHAND VON GROUND-TRUTH-DATEN

Die drei untersuchten Methoden wurden anhand manueller Ground-Truth-Annotationen quantitativ evaluiert. Das Vorgehen ist in Abschnitt 7.2 beschrieben. Hierbei wurden sowohl der Endpunktfehler (EPE) als auch der Winkelfehler pro Methode berechnet.

Endpunktfehler (EPE)

Abbildung 8.1 zeigt einen Gesamtvergleich der mittleren, medianen und maximalen EPE-Werte für Lucas-Kanade, Farneback und RAFT. Die mittleren EPE-Werte liegen bei 1,29 px, für Farneback bei 1,32 px und für RAFT bei 2,09 px. Die Medianwerte sind jeweils niedriger (1,06 px, 1,1 px, 1,47 px), während die Maximalwerte 2,49 px (Lucas-Kanade), 2,44 px (Farneback) und 3,44 px (RAFT) betragen.

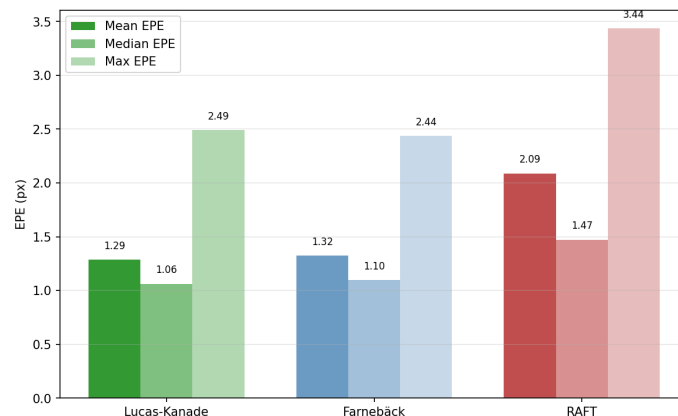


Abbildung 8.1: Gesamtvergleich von Mean, Median und Max EPE der Methoden [LK](#), [FB](#) und [RAFT](#) im Vergleich zu den Ground-Truth-Daten. (Eigene Darstellung.)

Die Verteilung der EPE-Werte über alle Videos und Methoden ist in Abbildung 8.2 dargestellt. Alle Methoden zeigen Ausreißer, die Mediane der Fehlerwerte liegen im Bereich unter 1,5 Pixeln Abweichung. Lucas-Kanade und Farneback weisen sowohl im Gesamtergebnis als auch pro Video sehr ähnliche Ergebnisse auf. Im Vergleich dazu zeigt RAFT in 6 von 7 Videos höhere Fehlerwerte (außer im Video *JuicyBeats Crowd nah*) sowie eine breitere Streuung.

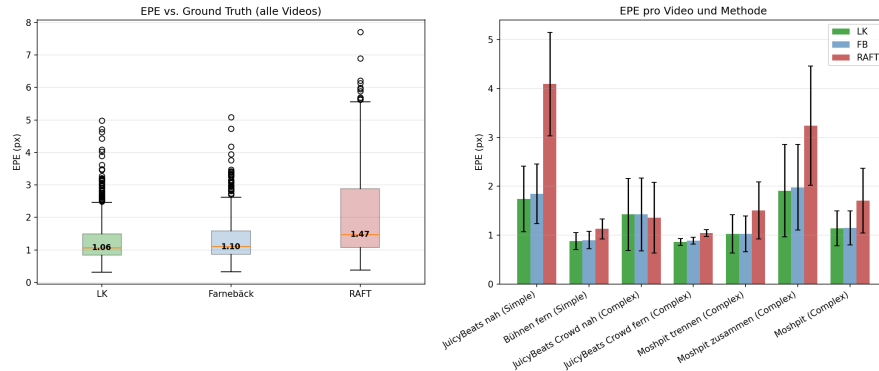


Abbildung 8.2: EPE-Verteilung im Vergleich zu den Ground-Truth-Daten. Links: Boxplot der gesamten EPE-Verteilung für die Methoden LK, FB und RAFT. Rechts: Durchschnittlicher EPE-Wert pro Video und Methode mit Standardabweichung. (Eigene Darstellung.)

Winkelfehler

Der Winkelfehler wurde ebenfalls anhand der Ground-Truth-Daten untersucht. Die Boxplots in Abbildung 8.3 zeigen, dass RAFT größere Winkelfehler aufweist (Median bei $58,39^\circ$ und durchschnittlicher Winkelfehler von 63°) als die Methoden von Lucas-Kanade und Farneback (Median jeweils etwa 20° und durchschnittlicher Winkelfehler von jeweils etwa 30°). Insgesamt zeigt sich, dass Lucas-Kanade und Farneback eine präzisere Richtungsbestimmung liefern, während RAFT häufiger größere Abweichungen in der Bewegungsrichtung produziert.

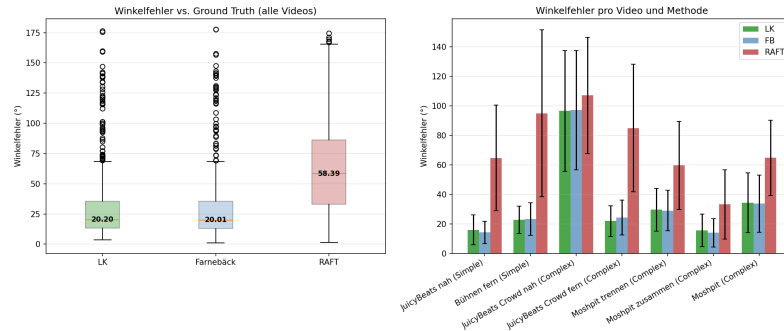


Abbildung 8.3: Winkelfehler im Vergleich zu den Ground-Truth-Daten. Links: Boxplot der gesamten Winkelfehler für die Methoden LK, FB und RAFT. Rechts: Durchschnittlicher Winkelfehler pro Video und Methode mit Standardabweichung. (Eigene Darstellung.)

Szenenkomplexität

Im gruppierten Boxplot in Abbildung 8.4 ist der EPE (in Pixel) für die Methoden Lucas-Kanade, Farneback und RAFT jeweils für einfache und komplexe Szenen dargestellt. Lucas-Kanade und Farneback erzielen in einfachen und komplexen Szenen sehr ähnliche Abweichungen (Mediane alle im Bereich 1,06 bis 1,14), wobei die Streuung bei einfachen Szenen etwas breiter ist. RAFT erzielt in einfachen Szenen (Median 1,66) etwas höhere Abweichungen und hat eine deutlich breitere Streuung im Vergleich zu den komplexen Szenen (Median 1,45).

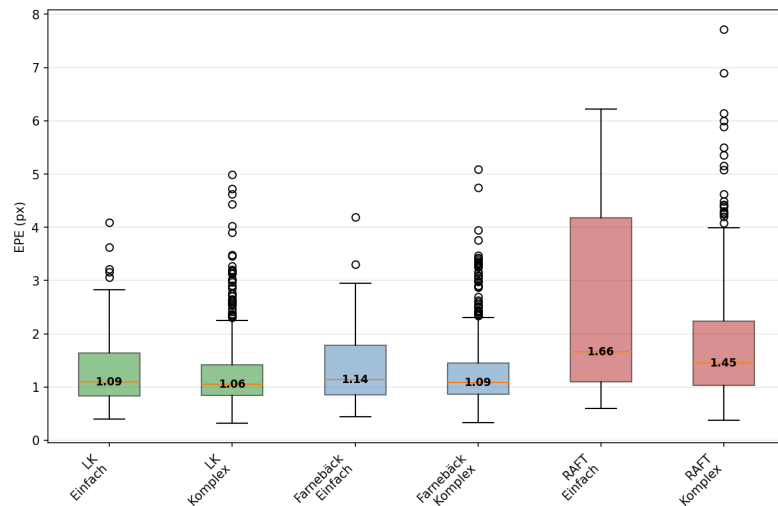


Abbildung 8.4: Gruppiertes Boxplot des EPE (in Px) für die Methoden LK, FB und RAFT zu den Ground-Truth-Daten, aufgeteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.)

Im gruppierten Boxplot in Abbildung 8.5 ist der Winkelfehler (in Grad) für die Methoden Lucas-Kanade, Farneback und RAFT jeweils für einfache und komplexe Szenen dargestellt. LK und Farneback erzielen in einfachen Sze-

nen (Median jeweils $\sim 17^\circ$) geringere Winkelfehler als in komplexen (Median jeweils $\sim 24^\circ$). Bei **LK** steigt der durchschnittliche Winkelfehler von einfach zu komplexen Videos um $15,73^\circ$ und bei Farneback um $16,23^\circ$. **RAFT** erzielt in beiden Szenarien ähnliche Werte (Median $60^\circ/58^\circ$). Dort sinkt der durchschnittliche Winkelfehler von einfachen zu komplexen Szenen um $3,7^\circ$. Es ist ersichtlich, dass **RAFT** im Vergleich zu Lucas-Kanade und Farneback größere Winkelfehler und eine größere Streuung aufweist. Lucas-Kanade und Farneback zeigen hingegen in beiden Szenentypen eine vergleichsweise geringe Streuung.

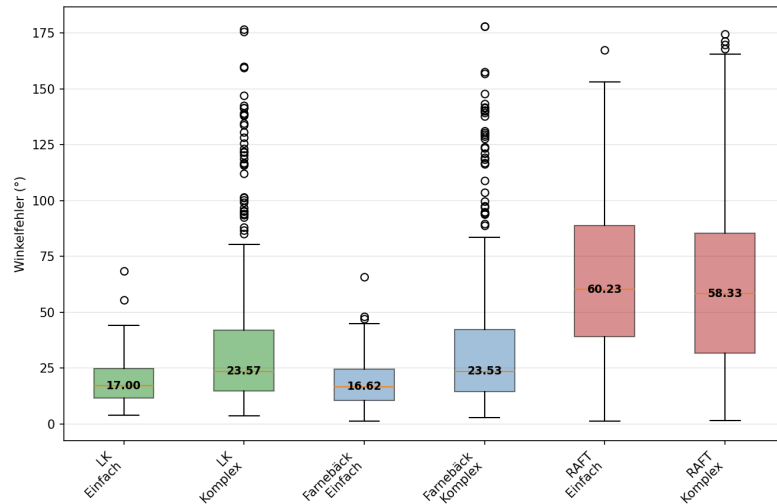


Abbildung 8.5: Gruppiertes Boxplot des Winkelfehlers (in Grad) für die Methoden LK, FB und RAFT zu den Ground-Truth-Daten, aufgeteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.)

Einfluss der Bewegungsamplitude

Die Abbildung 8.6 zeigt den Zusammenhang zwischen dem Endpunktfehler (EPE) und der mittleren Bewegungsstärke (Bewegungsamplitude der Ground-Truth-Daten) für die drei Methoden. Bei allen Verfahren steigt der EPE mit zunehmender Bewegungsamplitude an. Besonders bei **RAFT** ist der Zusammenhang ausgeprägt mit einer Steigung von $m=0,7$, wobei der EPE mit steigender Bewegungsamplitude zunimmt. Lucas-Kanade und Farneback zeigen ebenfalls eine Zunahme, aber auf insgesamt niedrigerem Niveau (LK mit $m=0,26$ und FB mit $m=0,28$).

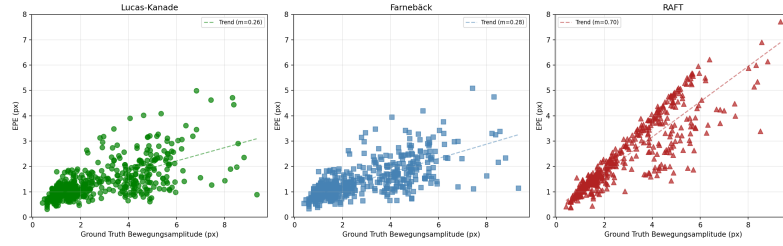


Abbildung 8.6: Zusammenhang zwischen EPE und Bewegungsamplitude für die Methoden LK, FB und RAFT. (Eigene Darstellung.)

Zusammenfassung der Ergebnisse

Die Kennzahlen der Evaluation anhand von den Ground-Truth-Daten sind in der Tabelle 8.1 dargestellt. Lucas-Kanade und Farneback erreichen nahezu identische Mittelwerte beim Endpunktfehler (EPE) und Winkelfehler, mit durchschnittlichen EPE-Werten unter 1,5 Pixel und Winkelfehlern um 30 Grad. Beide klassischen Methoden zeigen zudem eine geringe Streuung (Standardabweichung von 0,68) und niedrige Maximalwerte.

Im Vergleich dazu weist RAFT einen höheren mittleren EPE von 2,09 Pixel sowie einen höheren Winkelfehler von durchschnittlich 63 Grad auf. Die Streuung ist bei RAFT mit einer Standardabweichung von 1,36 ebenfalls größer.

Methode	EPE Mean (px)	EPE Median (px)	EPE Max (px)	Winkelfehler (°)
Lucas-Kanade	$1,29 \pm 0,68$	1,08	2,49	$30,0 \pm 28,3$
Farneback	$1,32 \pm 0,68$	1,14	2,44	$29,7 \pm 28,3$
RAFT	$2,09 \pm 1,36$	1,96	3,44	$63,0 \pm 37,5$

Tabelle 8.1: Zusammenfassung der Evaluation anhand von Ground-Truth-Daten

8.2 QUANTITATIVER METHODENVERGLEICH OHNE GROUND-TRUTH-DATEN

In diesem Abschnitt werden die drei untersuchten Methoden (Lucas-Kanade, Farneback, RAFT) hinsichtlich ihrer gegenseitigen Übereinstimmung verglichen. Die Analyse erfolgt anhand verschiedener Metriken ohne Bezug auf manuelle Ground-Truth-Daten.

Paarweiser Vergleich der Methoden über alle Videos

Die Tabelle 8.2 zeigt die aggregierten Vergleichsmetriken der Methodenpaare LK vs. RAFT, FB vs. RAFT und LK vs. FB. Beim EPE beträgt der Mittelwert für LK vs. RAFT 0,84 px, für FB vs. RAFT 0,38 px und für LK vs. FB 0,61 px. Der Unterschied zwischen LK und RAFT ist somit um 0,46 px größer als zwischen FB und RAFT. Für die Fl-all-Metrik liegt der Mittelwert bei den Vergleichen LK

vs. **RAFT** und **LK** vs. **FB** bei etwa 0,93%, während er für **FB** vs. **RAFT** geringer ausfällt (0,08%). Die mittlere Winkeldifferenz beträgt 102,7° beim Vergleich **LK** vs. **RAFT**, 78,8° für **FB** vs. **RAFT** und 44,4° für **LK** vs. **FB**. Eine ausführliche tabellarische Darstellung der Vergleichsmetriken, aufgeteilt nach einzelnen Videos, ist im Anhang in den Tabellen A.3, A.4, A.5, A.6, A.7 zu finden. Zusammenfassend sind die Unterschiede zwischen **LK** und **RAFT** mit einem durchschnittlich um 0,46 px größerem **EPE** und einem durchschnittlichen um 23,96° höherem Winkelfehler größer als zwischen **FB** und **RAFT**.

Metrik	Mean	Std	Median
EPE LK ↔ RAFT (px)	0.84	7.21	0.67
EPE FB ↔ RAFT (px)	0.38	0.78	0.30
EPE LK ↔ FB (px)	0.61	7.54	0.34
RMSE LK ↔ RAFT (px)	0.89	7.83	0.71
RMSE FB ↔ RAFT (px)	0.43	0.92	0.35
RMSE LK ↔ FB (px)	0.64	8.16	0.36
Fl-all LK ↔ RAFT (%)	0.93	9.28	0.00
Fl-all FB ↔ RAFT (%)	0.08	2.69	0.00
Fl-all LK ↔ FB (%)	0.93	9.47	0.00
Winkel LK ↔ RAFT (°)	102.72	33.89	106.62
Winkel FB ↔ RAFT (°)	78.76	25.49	77.68
Winkel LK ↔ FB (°)	44.44	22.91	41.97

Tabelle 8.2: Zusammengefasste Vergleichsmetriken der Methoden (paarweise, aggregiert über alle Videos)

Paarweiser Vergleich der Methoden gruppiert nach Videokomplexität

Die aggregierten Boxplots in Abbildung 8.7 und Abbildung 8.8 zeigen die Verteilung des Endpunktfehlers (EPE) und der Winkelabweichung für die drei Methodenvergleiche **LK** vs. **RAFT**, **FB** vs. **RAFT** und **LK** vs. **FB**, jeweils getrennt nach einfachen und komplexen Szenen. Die Fehlerwerte für den **EPE** sind bei komplexen Szenen größer als bei einfachen Szenen. Dies gilt für alle Methodenvergleiche. Eine Zunahme des durchschnittlichen Fehlers von einfachen zu komplexen Szenen zeigt sich bei Lucas-Kanade vs. **RAFT** (von 0,48 auf 1), bei Lucas-Kanade vs. Farnebäck (von 0,2 auf 0,79) sowie bei Farnebäck vs. **RAFT** (**EPE** von 0,35 auf 0,39) (siehe Tabelle 8.3).

Bei der Winkelabweichung ist der Medianwert bei **LK** vs. **RAFT** (komplex: 99,16) und **FB** vs. **RAFT** (komplex: 75,64) in komplexen Szenen niedriger als in einfachen Szenen (**LK** vs. **RAFT**: 118,03 / **FB** vs. **RAFT** 81,51). Der durchschnittliche Winkelfehler bei **LK** vs. **RAFT** sinkt von 113,9° (einfach) zu 97,6° (komplex). Bei **FB** vs. **RAFT** sinkt der durchschnittliche Winkelfehler von 82,7° (einfach) zu 77° (komplex). Ausschließlich bei dem Vergleich **LK** vs. **RAFT** (grün) ist der Median bei den komplexen Szenen (46,13°) höher als bei den einfachen Szenen (36,54°). Dort steigt die durchschnittliche Winkelabweichung von 36,8°

(einfach) zu $47,9^\circ$ (komplex). Die Streuung ist bei allen drei Methoden bei den komplexen Szenen breiter (siehe Std. in Tabelle 8.3).

Die mittleren Fehlerabweichungen und Winkelabweichungen sowie der Fl-all Wert sind in Tabelle 8.3 nochmal zusammengefasst.

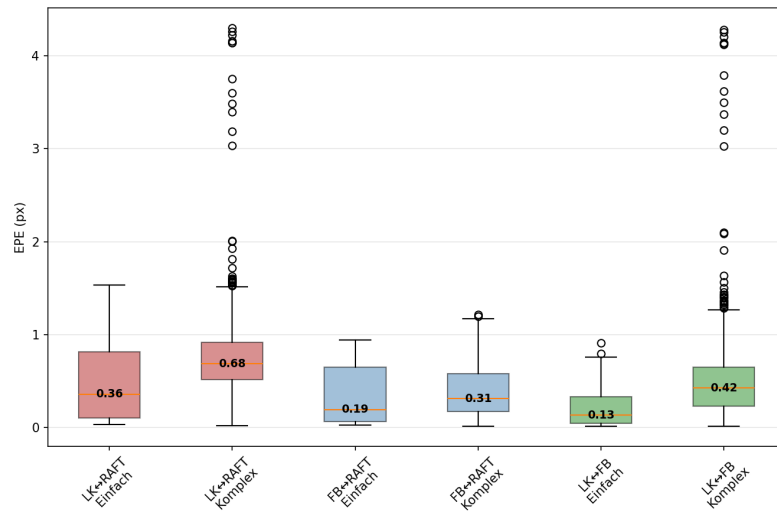


Abbildung 8.7: Boxplots der EPE-Abweichungen zwischen allen Methoden (LK, FB, **RAFT**) im Vergleich, jeweils unterteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.)

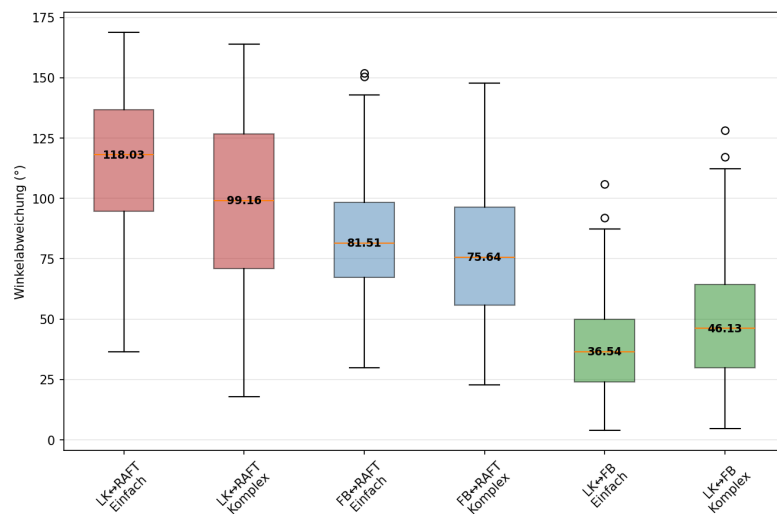


Abbildung 8.8: Boxplots der Winkelabweichungen zwischen allen Methoden (LK, FB, **RAFT**) im Vergleich, jeweils unterteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.)

Metrik	Einfach (Mean±Std)	Komplex (Mean±Std)
EPE LK↔RAFT	0,48 ± 0,39	1,00 ± 8,68
EPE LK↔FB	0,20 ± 0,18	0,79 ± 9,09
EPE FB↔RAFT	0,35 ± 0,30	0,39 ± 0,92
RMSE LK↔RAFT	0,52 ± 0,41	1,05 ± 9,44
RMSE LK↔FB	0,22 ± 0,19	0,83 ± 9,84
RMSE FB↔RAFT	0,39 ± 0,33	0,44 ± 1,09
Winkel LK↔RAFT (°)	113,9 ± 28,3	97,6 ± 35,0
Winkel LK↔FB (°)	36,8 ± 19,1	47,9 ± 23,7
Winkel FB↔RAFT (°)	82,7 ± 22,3	77,0 ± 26,6
Fl-all LK↔RAFT (%)	0,0 ± 0,0	1,4 ± 11,2
Fl-all LK↔FB (%)	0,0 ± 0,0	1,4 ± 11,4
Fl-all FB↔RAFT (%)	0,0 ± 0,0	0,1 ± 3,2

Tabelle 8.3: Vergleich der wichtigsten Metriken zwischen einfachen und komplexen Szenen für verschiedene Methodenpaare.

Zusammenfassung

Die Übereinstimmung zwischen Farneback und Lucas-Kanade ist in den einfachen Szenen am größten. Die Unterschiede in Fehlerwerten und Winkelabweichungen für diesen Methodenvergleich liegen überwiegend näher beieinander als bei den Vergleichen mit RAFT. Bei den komplexen Szenen liegen FB und RAFT näher beieinander. Zusätzlich weisen RAFT und Farneback eine höhere Übereinstimmung auf als RAFT und Lucas-Kanade. In komplexen Szenen werden die Unterschiede der EPEs zwischen den Methoden größer.

8.3 QUALITÄTSBEWERTUNG MITTELS FORWARD-BACKWARD CONSISTENCY

Die Forward-Backward-Consistency wird als Proxy-Maß zur Qualitätsbewertung der berechneten Bewegungsvektoren eingesetzt. Ein Punkt gilt in diesem Experiment als konsistent, wenn die Differenz zwischen Vorwärts- und Rückwärts-Tracking unter 1,5 px liegt. Dieser Wert orientiert sich wie in Abschnitt 7.4 beschrieben an etablierten Benchmarks in der Optical-Flow-Literatur.

Mittlerer Forward-Backward-Fehler im Szenarienvergleich

Ein Szenarienvergleich des mittleren Forward-Backward-Errors ist in Abbildung 8.9 dargestellt. Je niedriger der Wert ist, desto konsistenter ist die Methode. Die Heatmap zeigt, dass sowohl Lucas-Kanade als auch Farneback in einfachen Szenen niedrige mittlere Forward-Backward-Error-Werte (jeweils 0,01 Px) aufweisen. RAFT hingegen hat in einfachen Videos bereits einen mittleren Fehler von 0,21 Pixel. In komplexen Szenen steigt der mittlere Forward-

Backward-Error bei **RAFT** auf 0,79 Pixel an, während Lucas-Kanade auf 0,03 Pixel ansteigt und Farnebäck bei 0,01 Pixel Abweichung bleibt.



Abbildung 8.9: Mittlerer Forward-Backward-Error (in Pixel) für LK, **FB** und **RAFT**, jeweils unterteilt nach Szenenkomplexität (Einfach und Komplex). (Eigene Darstellung.)

Quantitative Übersicht

Die Ergebnisse in Tabelle 8.4 zeigen, dass die Forward-Backward-Consistency für alle Methoden und Szenarien auf sehr hohem Niveau liegt. Lucas-Kanade (LK) und Farnebäck (FB) erreichen über alle Videos hinweg nahezu fehlerfreie Konsistenzwerte von $100,0 \pm 0,1 \%$ (z. B. "JuicyBeats nah") bzw. $100,0 \pm 0,0 \%$ (z. B. "Bühnen fern"). Farnebäck erzielt im Video "Juicy Beats Crowd nah" niedrigere Werte mit $99,0 \pm 8,9 \%$. Die RAFT-Methode weist in komplexen Szenen leichte Schwankungen und insgesamt geringere Konsistenz auf. Der größte Unterschied ist im Video "Moshpit zusammen (Komplex)" mit dem niedrigsten Konsistenzwert von $96,9 \pm 4,6 \%$ und dem Video "Juicy Beats Crowd nah (Komplex)" mit dem Wert von $98,7 \pm 11,2 \%$.

Der Forward-Backward-Error (Tabelle 8.5) verdeutlicht diese Werte. Die Fehlerwerte für **LK** und **FB** sind in allen Videos sehr niedrig und zeigen kaum Varianz, beispielsweise **LK** mit $0,00 \pm 0,00 \text{ px}$ und **FB** mit $0,00 \pm 0,00 \text{ px}$ bei "Bühnen fern". In den meisten Videos (bei **LK** in 5/7 und bei **FB** in 4/7 Videos) liegen die Fehler für beide Methoden bei kleiner/gleich $0,1 \text{ px}$. Bei **RAFT** treten höhere Fehler und Standardabweichungen auf. Der Fehler bei dem Video "JuicyBeats Crowd nah (Komplex)" ist mit $4,89 \pm 39,02 \text{ px}$ für **RAFT** am größten.

Video	FB-Cons. LK (%)	FB-Cons. FB (%)	FB-Cons. RAFT (%)
JuicyBeats nah (Ein-fach)	100,0 $\pm$ 0,1	100,0 $\pm$ 0,1	99,3 $\pm$ 1,2
Bühnen fern (Ein-fach)	100,0 $\pm$ 0,0	100,0 $\pm$ 0,0	100,0 $\pm$ 0,0
JuicyBeats Crowd nah (Komplex)	100,0 $\pm$ 0,1	99,0 $\pm$ 8,9	98,7 $\pm$ 11,2
JuicyBeats Crowd fern (Komplex)	100,0 $\pm$ 0,0	100,0 $\pm$ 0,0	100,0 $\pm$ 0,0
Moshpit trennen (Komplex)	100,0 $\pm$ 0,1	100,0 $\pm$ 0,0	100,0 $\pm$ 0,2
Moshpit zusammen (Komplex)	99,8 $\pm$ 0,3	100,0 $\pm$ 0,0	96,9 $\pm$ 4,6
Moshpit (Komplex)	99,9 $\pm$ 0,2	100,0 $\pm$ 0,0	99,9 $\pm$ 0,3

Tabelle 8.4: Forward-Backward-Consistency (FB-Cons.) für LK, FB, RAFT auf allen Videos.

Video	FB-Error LK (px)	FB-Error FB (px)	FB-Error RAFT (px)
JuicyBeats nah (Ein-fach)	0,01 $\pm$ 0,01	0,02 $\pm$ 0,01	0,41 $\pm$ 0,18
Bühnen fern (Ein-fach)	0,00 $\pm$ 0,00	0,00 $\pm$ 0,00	0,02 $\pm$ 0,01
JuicyBeats Crowd nah (Komplex)	0,01 $\pm$ 0,00	0,06 $\pm$ 0,43	4,89 $\pm$ 39,02
JuicyBeats Crowd fern (Komplex)	0,00 $\pm$ 0,00	0,00 $\pm$ 0,00	0,01 $\pm$ 0,00
Moshpit trennen (Komplex)	0,01 $\pm$ 0,01	0,01 $\pm$ 0,00	0,44 $\pm$ 0,20
Moshpit zusammen (Komplex)	0,06 $\pm$ 0,14	0,02 $\pm$ 0,01	0,64 $\pm$ 0,26
Moshpit (Komplex)	0,03 $\pm$ 0,07	0,01 $\pm$ 0,00	0,45 $\pm$ 0,21

Tabelle 8.5: Forward-Backward-Error (FB-Error) für LK, FB, RAFT auf allen Videos.

Räumliche Visualisierung

Die räumliche Verteilung konsistenter und inkonsistenter Punkte ist in den Abbildungen 8.10 exemplarisch visualisiert. Konsistente Punkte sind grün, inkonsistente rot eingefärbt. Die Werte aus der Tabelle 8.4 und 8.5 werden hier qualitativ gezeigt. Während die stellvertretenden Bilder von der Lucas-Kanade und Farneback Methode nur grüne und damit konsistente Punkte aufweisen, ist im Gegensatz bei der RAFT Methode ein anderes Bild sichtbar. Dort sind überall auch rote Punkte sichtbar.

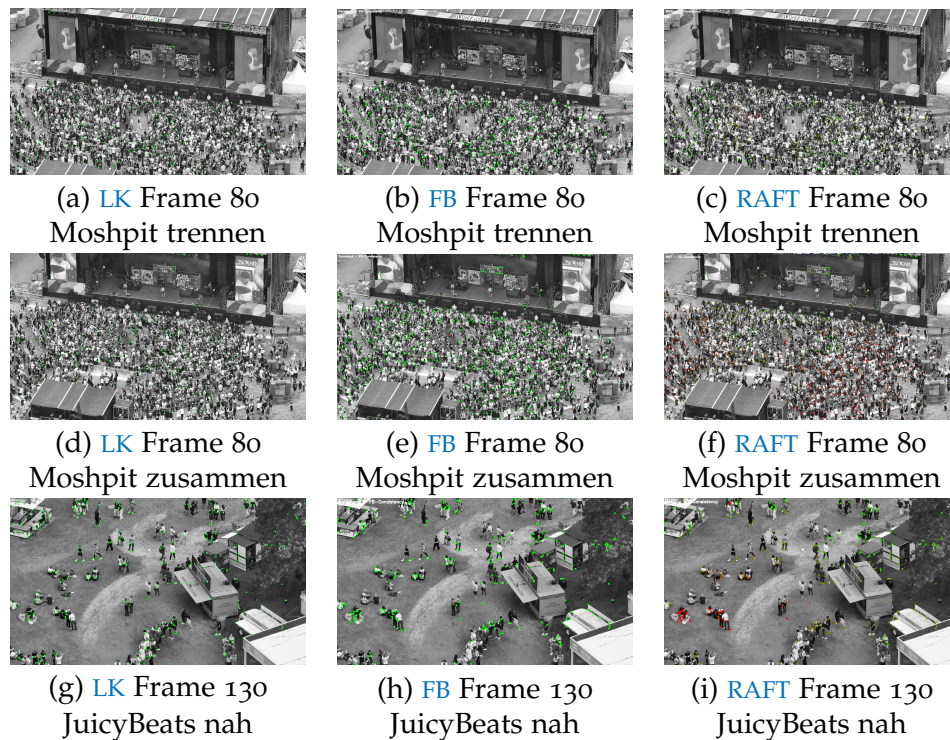


Abbildung 8.10: Consistency-Karten mit stellvertretenden Bildern für LK, FB und RAFT. (Eigene Darstellung.)

Zusammenfassung

Zusammenfassend liefern beide klassischen Methoden (LK und FB) sehr hohe Forward-Backward-Consistency-Werte. Farnebäck zeigt eine nahezu fehlerfreie Konsistenz. RAFT erreicht im Vergleich in komplexen Szenen geringere, aber immer noch hohe Werte.

8.4 CROWD-DYNAMIK-ANALYSE

Zur Charakterisierung der Bewegungsdynamik in den Drohnenvideos wurden verschiedene statistische Kennzahlen und Visualisierungen herangezogen und im folgenden Kapitel beschrieben.

Richtungs-Entropie und Richtungs-Uniformität

Die Richtungs-Entropie dient als Maß für den Ordnungsgrad der Bewegung. Hohe Entropiewerte deuten auf chaotische, niedrigere Werte auf gerichtete Bewegungen hin.

Die quantitative Auswertung der Richtungs-Entropie pro Video zeigt Unterschiede zwischen den betrachteten Verfahren (Abbildungen 8.11, 8.12, 8.13).

Beim Lucas-Kanade-Verfahren (Abbildung 8.11) bewegen sich die Entropiewerte (Mediane im Bereich 0,883 bis 1,395) insgesamt auf moderatem Niveau (ca. 0,75–1,75 Bits). In den Crowd-Szenen (*JuicyBeats Crowd nah/fern*) ist die Entropie tendenziell etwas höher als in den Moshpit-Szenen, wobei sich die Wertebereiche überlappen. Die durchschnittliche Entropie der Crowd-Videos liegt bei 1,08 bzw. 1,26, während die durchschnittliche Entropie der Moshpit-Videos bei 0,94 bis 1,15 liegen (siehe A.8).

Das Farneäck-Verfahren (Abbildung 8.12) liefert insgesamt höhere Entropiewerte (Mediane im Bereich 1,61 bis 2,81) und zeigt eine größere Streuung (ca. 1,5–3,1 Bits). In den Crowd-Szenen (*JuicyBeats Crowd nah/fern*) sind die durchschnittlichen Entropiewerte mit 1,9 bzw. 2,72 höher als in den Moshpit-Szenen mit durchschnittlichen Entropiewerten von 1,71 bis 1,85 (siehe Tabelle A.9).

Beim RAFT-Verfahren (Abbildung 8.13) fallen die Entropiewerte insgesamt kompakter (ca. 0,75–1,75 Bits) und niedriger (Mediane im Bereich 0,75 bis 1,39) aus. Hier zeigen sich keine klaren Unterschiede in den Werten zwischen Crowd- und Moshpit-Szenen. Teilweise sind die durchschnittlichen Entropiewerte in den Crowd-Szenen (0,92 bzw. 1,36) niedriger als in den Moshpit-Szenen (1,15 bis 1,35) (siehe Tabelle A.10).

Alle Mittelwerte und Standardabweichungen sind im Anhang in den Tabellen A.8, A.9 und A.10 zu finden.

Zusammenfassend sind bei Farneäck die größten Unterschiede zwischen den Szenentypen sichtbar, während RAFT insgesamt geringere und weniger differenzierte Entropiewerte liefert.

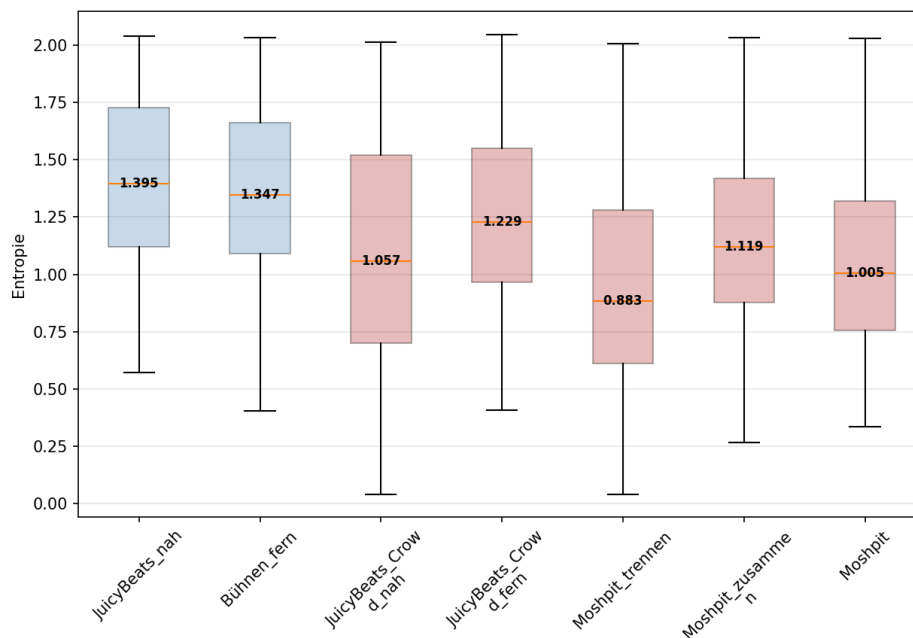


Abbildung 8.11: Boxplots der Richtungs-Entropie pro Video für die Methode Lucas-Kanade. (Eigene Darstellung.)

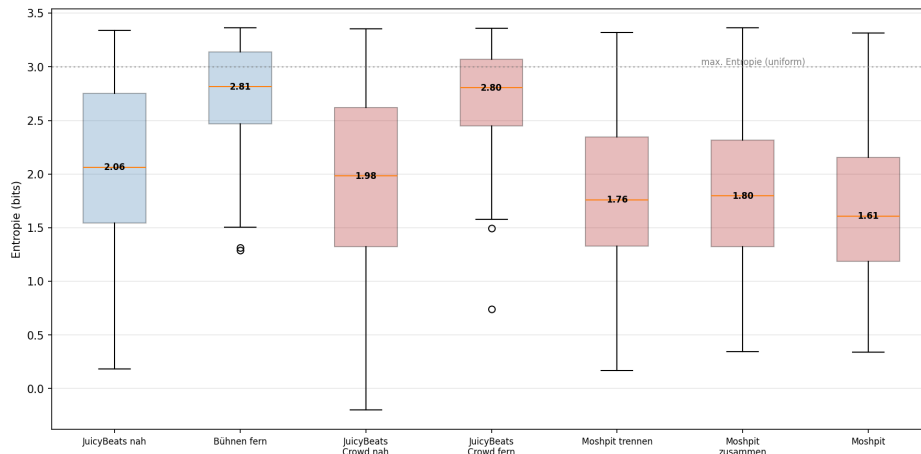


Abbildung 8.12: Boxplots der Richtungs-Entropie pro Video für die Methode Farnebäck. (Eigene Darstellung.)

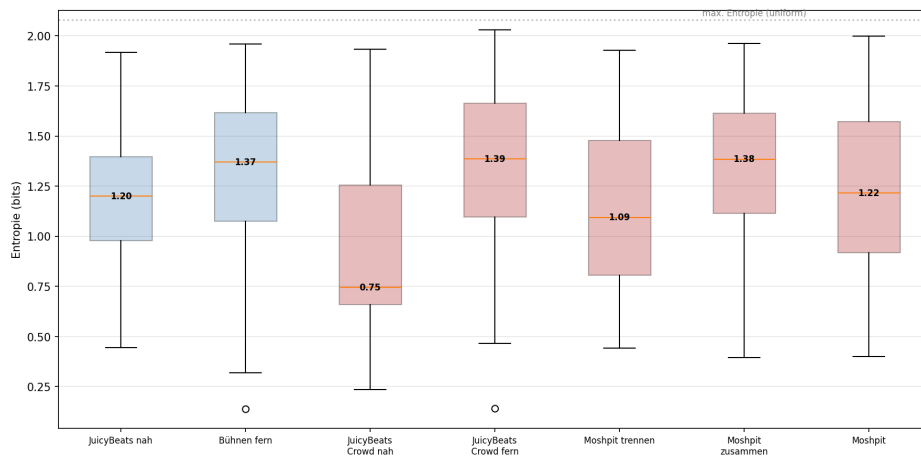


Abbildung 8.13: Boxplots der Richtungs-Entropie pro Video für die Methode RAFT. (Eigene Darstellung.)

Im Folgenden werden die Ergebnisse der Richtungs-Uniformität für die Methoden LK, FB und RAFT präsentiert. Hohe Uniformitäts-Werte bedeuten eine dominante Bewegungsrichtung.

Die Boxplots der Richtungs-Uniformität für LK und FB in Abbildung 8.14 und 8.15 zeigen, dass in den Crowd-Szenen (*JuicyBeats Crowd nah/fern*) die Richtungs-Uniformität niedriger (Mediane bei 4,2 und 4,533) ist als in den Moshpit-Szenen (Mediane im Bereich 4,827 bis 5,027). Bei LK liegen die durchschnittlichen Uniformitätswerte in den Crowd Videos bei 4,64 bzw. 4,2 und bei den Moshpit Videos bei 4,78 bis 5,1. Bei FB liegen die durchschnittlichen Uniformitätswerte der Crowd-Szenen bei 2,68 bzw 4,14 und bei den Moshpit Szenen bei 4,35 bis 4,56.

Die Boxplots der Richtungs-Uniformität in Abbildung 8.16 für RAFT zeigen diesen Unterschied nicht. Hier überlappen die durchschnittlichen Richtungs-Uniformitätswerte der Crowd und Moshpit Szenen in einem Bereich von

3,66 bis 4,83. Alle Mittelwerte und Standardabweichungen sind im Anhang in den Tabellen A.8, A.9 und A.10 zu finden.

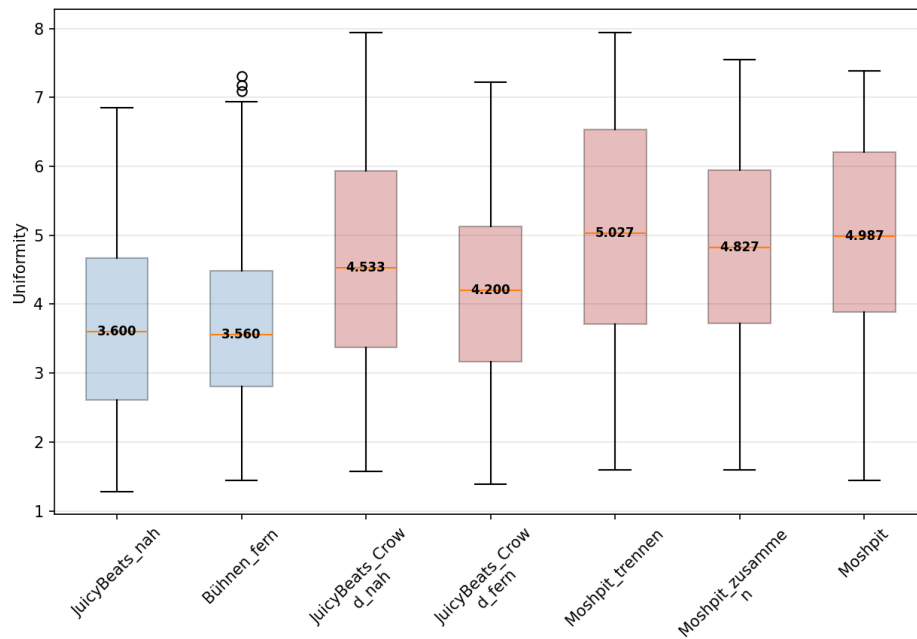


Abbildung 8.14: Boxplots der Richtungs-Uniformität pro Video für die Methode Lucas-Kanade. (Eigene Darstellung.)

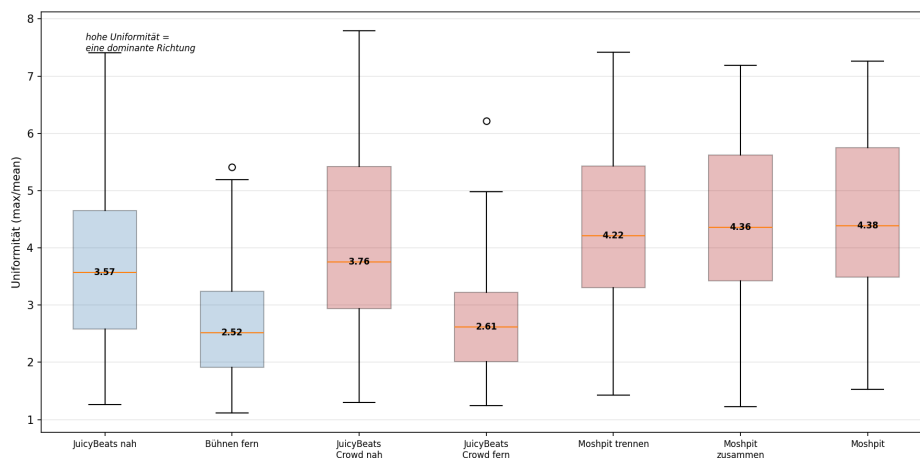


Abbildung 8.15: Boxplots der Richtungs-Uniformität pro Video für die Methode Farnebäck. (Eigene Darstellung.)

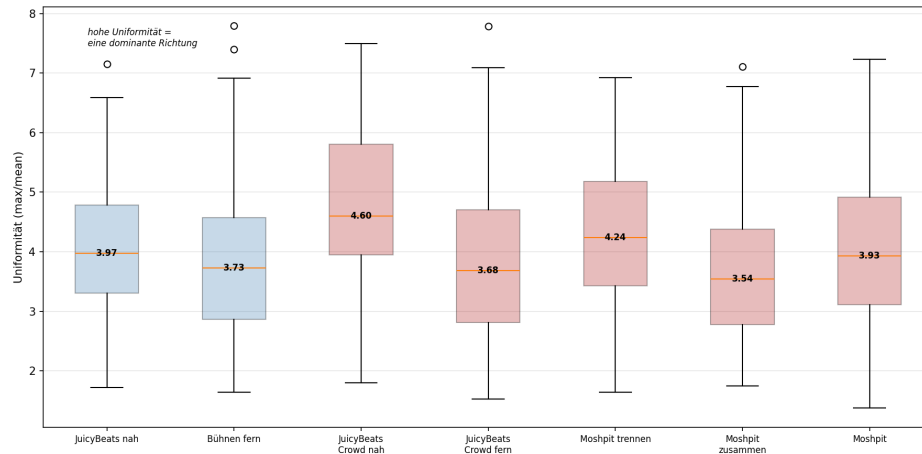


Abbildung 8.16: Boxplots der Richtungs-Uniformität pro Video für die Methode [RAFT](#). (Eigene Darstellung.)

Qualitative Analyse: Visuelle Heatmaps der Crowd-Dynamik

Zur Veranschaulichung der berechneten Bewegungsmuster und zur Ergänzung der quantitativen Auswertungen wurden für ausgewählte Frames Heatmaps der Crowd-Dynamik erstellt (Abbildung 8.17). Die Frames stammen aus dem Video *Moshpit_zusammen.mp4*. In diesem Video rennt die linke Menschenmenge nach rechts los und die rechte Menschenmenge rennt nach links los. Die Bilder in der linken Spalte zeigen die Ergebnisse der Farnebäck-Methode, die Bilder in der rechten Spalte stammen aus der [RAFT](#)-Auswertung. Für Lucas-Kanade gibt es diese Visualisierung nicht, da die Visualisierung auf den Ergebnissen der dense-Bewegungsvektoren basieren und [LK](#) nur die sparse Punkte berechnet. In allen Heatmaps ist die Szene entsprechend der berechneten Bewegungsrichtung eingefärbt.

Die Farnebäck-Heatmaps (linke Seite) erfassen die Bewegungsrichtungen, aber die Bildbereiche sind weniger intensiv und kontrastreich eingefärbt. Die Bereiche mit klaren Bewegungsrichtungen sind sichtbar, wirken aber diffuser und weniger abgegrenzt.

Im Gegensatz dazu zeigen die [RAFT](#)-Heatmaps (rechte Seite) eine deutlich stärkere und klarere Einfärbung. Die Bewegungsrichtungen und -grenzen innerhalb der Menge sind deutlicher herausgearbeitet. So sind beispielsweise die Zonen, in denen sich die Menge aufeinander zubewegt, sehr prägnant sichtbar. Allerdings sind die Frames auch etwas verrauscht. Im Bild (f) und (h) ist erkennbar, dass RAFT das ganze Bild leicht einfärbt.

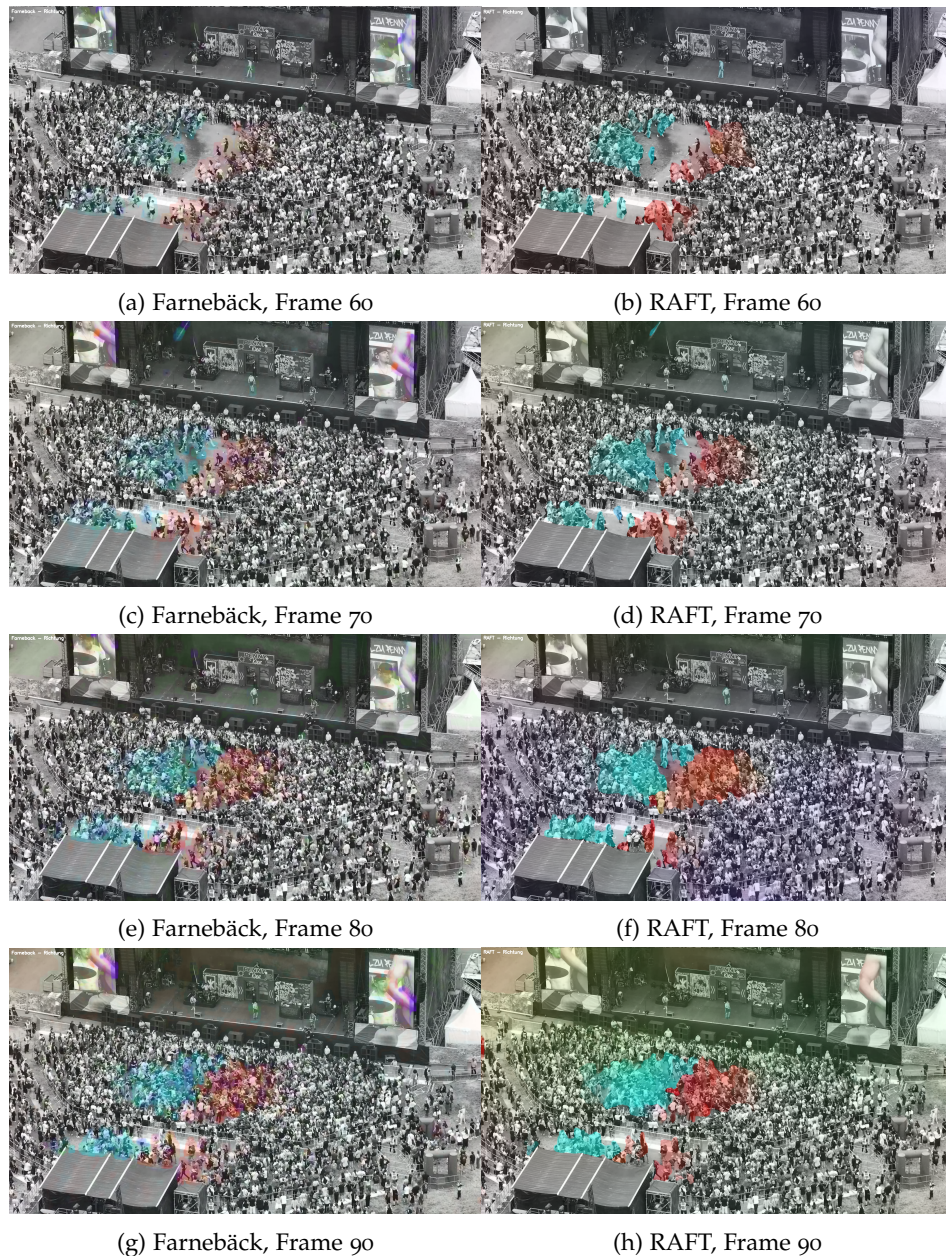


Abbildung 8.17: Vergleich der Bewegungs-Heatmaps in der Moshpit-Szene für vier Zeitpunkte: Jeweils links Farneback, rechts RAFT. (Eigene Darstellung.)

Geschwindigkeitsverteilung

Die Abbildungen 8.18, 8.19 und 8.20 zeigen die Verteilung der mittleren Geschwindigkeiten für die analysierten Videos in Pixel pro Sekunde. Alle Verfahren erzeugen ähnliche Ergebnisse. Die Videos *Bühnen fern* und *Juicy-Beats Crowd fern* zeigen jeweils niedrige Geschwindigkeitswerte mit Medianen unter 3px/s bei jeder Methode. Die anderen Videos zeigen höhere Mediangeschwindigkeiten mit Medianen zwischen 6 und 24 px/s.

Das Video mit den höchsten Geschwindigkeiten ist bei allen Methoden das Video Moshpit_zusammen (Mediane LK: 24,03/ FB: 23,96 / RAFT: 16,25 px/s).

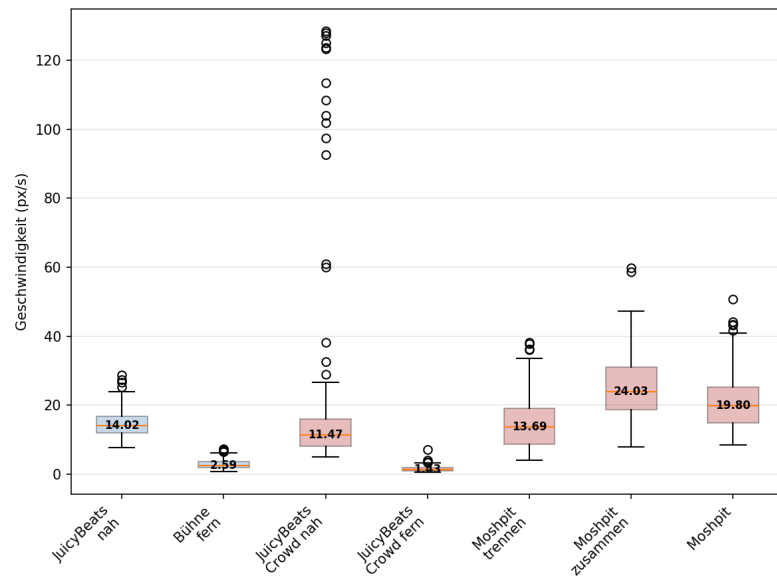


Abbildung 8.18: Boxplots der Geschwindigkeitsverteilung pro Video für die Methode Lucas-Kanade. (Eigene Darstellung.)

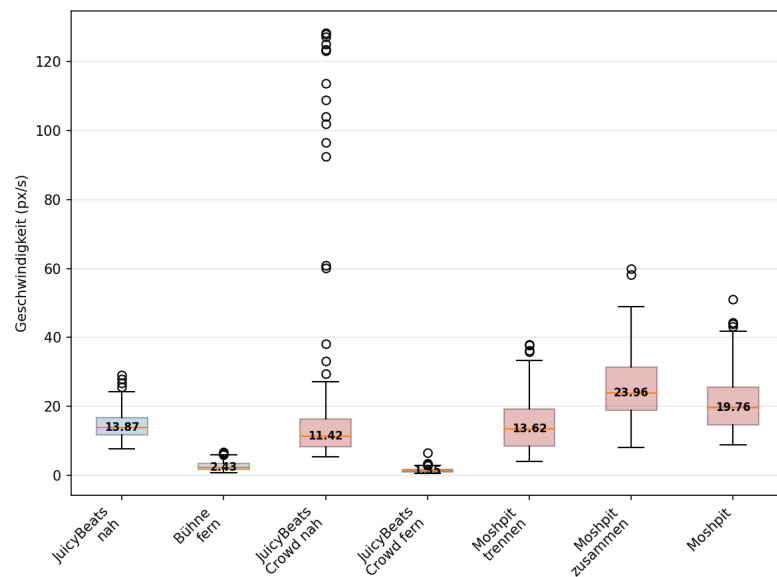


Abbildung 8.19: Boxplots der Geschwindigkeitsverteilung pro Video für die Methode Farneback. (Eigene Darstellung.)

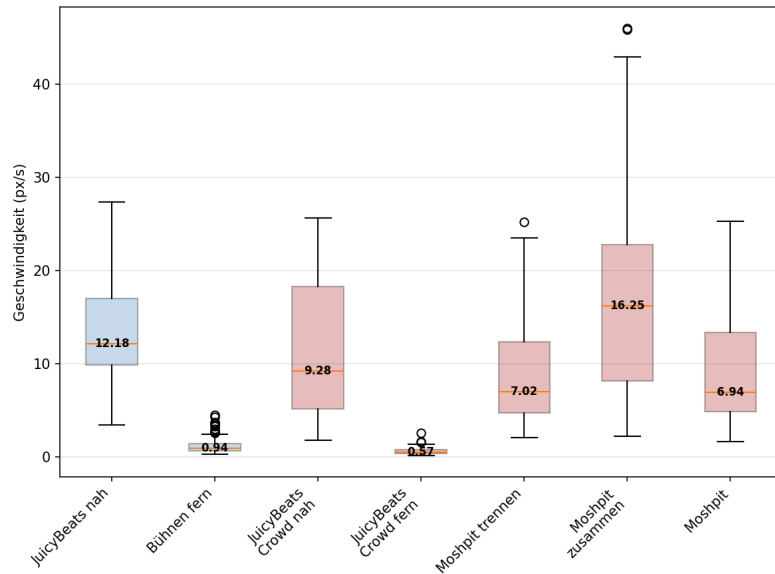


Abbildung 8.20: Boxplots der Geschwindigkeitsverteilung pro Video für die Methode RAFT. (Eigene Darstellung.)

Räumliche Visualisierung der Geschwindigkeitsverteilung mittels Amplitude-Heatmaps

Zur Ergänzung der quantitativen Analyse der Geschwindigkeitsverteilung in Abbildung 8.19 und 8.20 wurden Amplitude-Heatmaps für mehrere exemplarische Frames erstellt (Abbildung 8.21). Diese zeigen für die Verfahren FB und RAFT, wo innerhalb der Menschenmenge die Bewegungen besonders stark ausgeprägt sind. Für Lucas-Kanade gibt es diese Visualisierung nicht, da die Visualisierung auf den Ergebnissen der dense-Bewegungsvektoren basieren und LK nur die sparse Punkte berechnet. Die Farben kodieren die Geschwindigkeit (Betrag der Bewegung) an jedem Pixel, unabhängig von der Bewegungsrichtung: Blau/Grün stehen für geringe, Gelb/Rot für starke Bewegung.

Es ist ersichtlich, dass sich die stärkste Dynamik in Bild (a) und (b) im Bereich des Moshpits konzentriert, während der restliche Crowd-Bereich weitgehend ruhig bleibt. Die Hotspots (rote Bereiche) markieren Zonen mit besonders hoher Bewegung, etwa ausgelöst durch kollektive Crowd-Wellen oder intensive Interaktionen. Graue Bereiche zeigen stehende oder kaum bewegte Zuschauer. Auch in Bild (c) und (d) ist zu sehen, dass die zwei Personen in der Mitte, die den Moshpit aufteilen, deutlich eingefärbt sind. In Bild (e) bis (h) sind zudem die Trajektorien der sich fortbewegenden Personen eingefärbt. Auch hier färbt RAFT in Bild (h) Bildbereiche (Ecke oben links) ein, die keine Bewegung aufweisen.

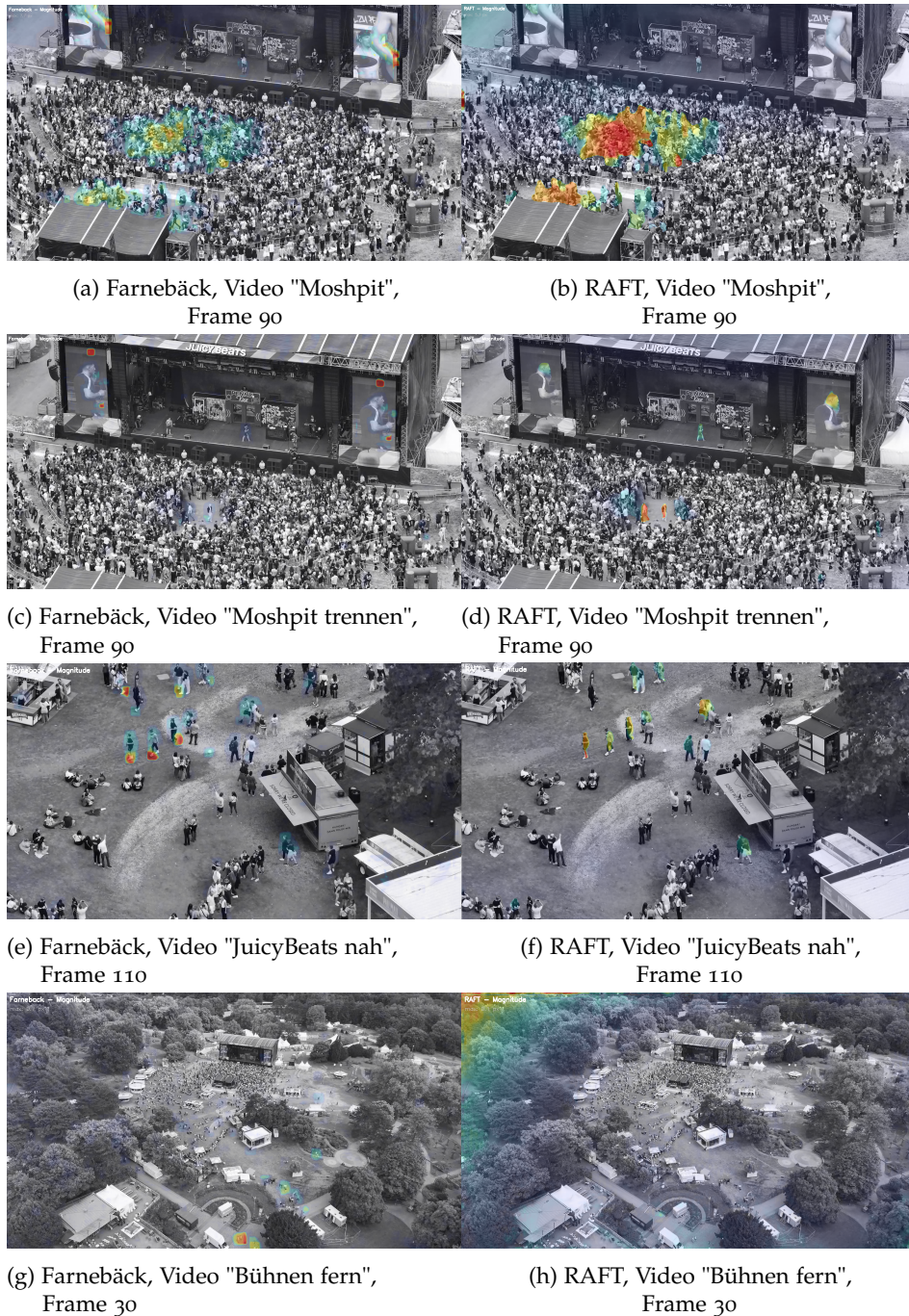


Abbildung 8.21: Amplitude-Heatmaps exemplarischen Frames: Links Farneback, rechts [RAFT](#). (Eigene Darstellung.)

Divergenz und Curl

Die Divergenz misst das Auseinander- oder Zusammenlaufen der Bewegungen und Curl die Rotationskomponenten im Flow-Feld.

Abbildung 8.22 zeigt die Verteilung der Divergenzwerte für die Lucas-Kanade-Methode. Hier wird ersichtlich, dass die Divergenz bei den meisten Videos nahe Null liegt. Insbesondere bei Szenen mit wenig Bewegung, wie

„Bühnen_fern“ und „Juicybeats_Crowd_fern“, sind sowohl Median als auch Streuung gering. Lediglich in dynamischeren Szenen wie „Juicybeats_nah“ und den Moshpit-Videos ist sowohl die Streuung als auch der Medianwert erhöht.

Ein Vergleich der Methoden Farneäck in Abbildung 8.23 und RAFT in Abbildung 8.24 zeigt Unterschiede in der Divergenz. Farneäck weist in allen Videos höhere Medianwerte und eine breitere Streuung auf, insbesondere in den Moshpit-Szenen. Die RAFT-Methode liefert insgesamt niedrigere und gleichmäßigere Divergenzwerte, unabhängig von der Szenendynamik. Die durchschnittlichen Mittelwerte bei FB bei den Moshpit-Videos liegen im Bereich 0,016 bis 0,022, während sie bei RAFT im Bereich von 0,005 bis 0,008 liegen.

Alle durchschnittlichen Mittelwerte und Standardabweichungen sind im Anhang in den Tabellen A.8, A.9 und A.10 zu finden.

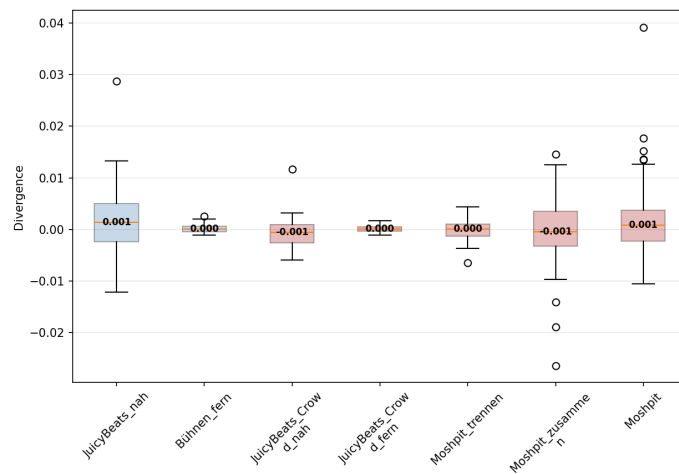


Abbildung 8.22: Mittlere Divergenz für alle Videos mit Lucas-Kanade. (Eigene Darstellung.)

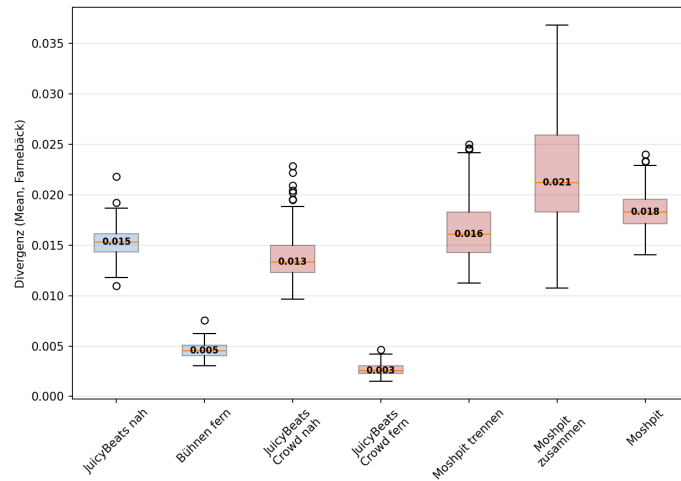


Abbildung 8.23: Mittlere Divergenz für alle Videos mit Farneback. (Eigene Darstellung.)

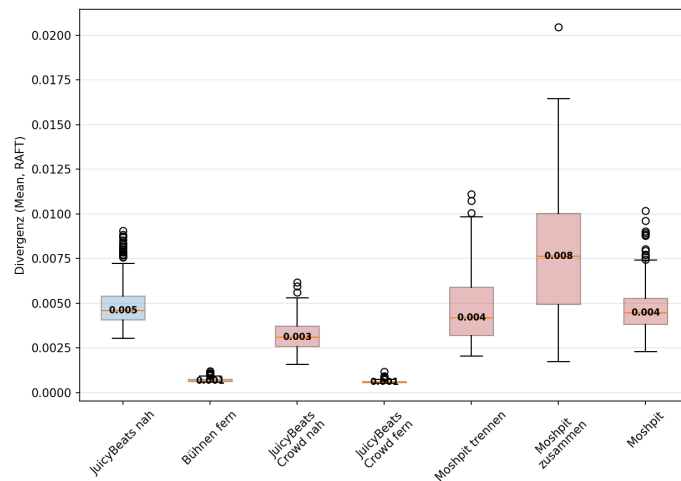


Abbildung 8.24: Mittlere Divergenz für alle Videos mit RAFT. (Eigene Darstellung.)

Zur Bewertung rotierender Bewegungen im optischen Fluss wurde der mittlere Curl für alle Videos mit den Methoden Lucas-Kanade, Farneback und RAFT berechnet.

Die Ergebnisse der Lucas-Kanade-Methode (Abbildung 8.25) zeigen für alle Videos Curl-Werte nahe Null. Lediglich bei „Juicybeats_nah“ und in den Moshpit-Szenen ist eine etwas Streuung zu erkennen, der Median verbleibt stets in unmittelbarer Nähe zur Nulllinie.

Ein Vergleich der Methoden Farneback (Abbildung 8.26) und RAFT (Abbildung 8.27) verdeutlicht Unterschiede in der Sensitivität gegenüber Rotation. Die Farneback-Methode weist für alle Videos, insbesondere für die dynamischen Moshpit-Szenen, höhere und stärker streuende Curl-Werte auf (durchschnittliche Curl-Werte von 0,016-0,022). Im Gegensatz dazu liefert RAFT durchgehend niedrige und wenig streuende Curl-Werte, auch in den bewegungsintensiven Szenen (durchschnittliche Curl Werte alle kleiner als

0,01). Alle durchschnittlichen Mittelwerte und Standardabweichungen sind im Anhang in den Tabellen A.8, A.9 und A.10 zu finden.

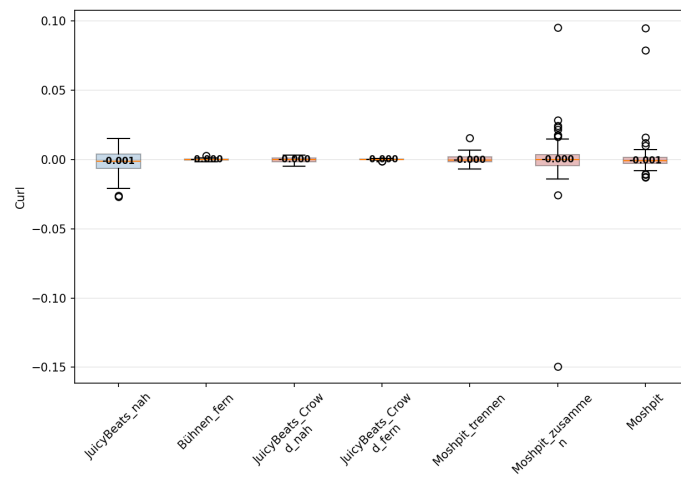


Abbildung 8.25: Mittlerer Curl für alle Videos mit Lucas-Kanade. (Eigene Darstellung.)

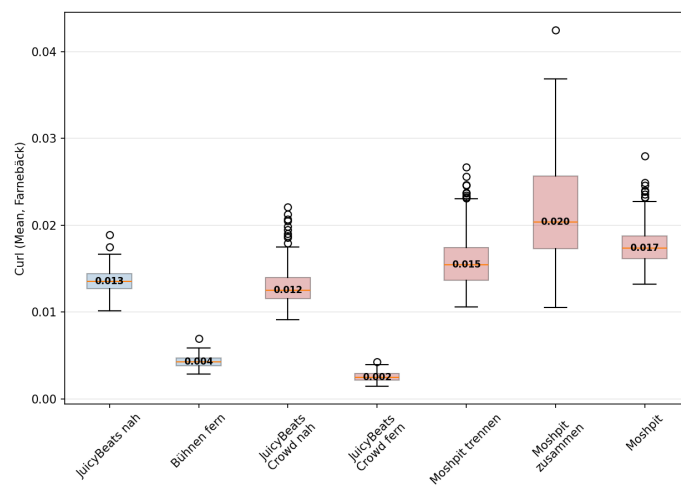


Abbildung 8.26: Mittlerer Curl für alle Videos mit Farneback. (Eigene Darstellung.)

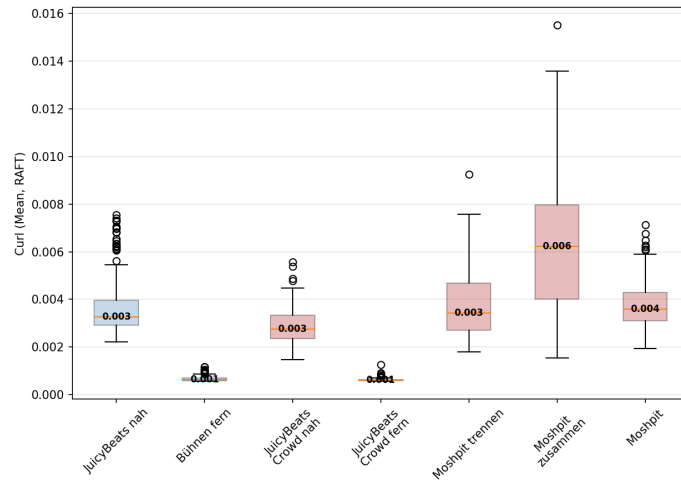


Abbildung 8.27: Mittlerer Curl für alle Videos mit RAFT. (Eigene Darstellung.)

Die Abbildung 8.28 zeigt den Zusammenhang zwischen mittlerer Divergenz und mittlerem Curl für alle analysierten Videos, jeweils getrennt für die Methoden Lucas-Kanade (links), Farneäck (mitte) und RAFT (rechts). Bei Farneäck und RAFT ist ein linearer Zusammenhang zwischen Divergenz und Curl erkennbar. Mit zunehmender Divergenz steigt auch der Curl-Wert. Bei Lucas-Kanade ist kein linearer Zusammenhang ersichtlich.

Die Moshpit-Videos (lila, rosa, braun) weisen sowohl bei Farneäck als auch bei RAFT die höchsten Werte auf und befinden sich im oberen rechten Bereich der Plots, was auf komplexe und sehr dynamische Bewegungsmuster hinweist. Die Crowd-Videos (rot, grün) liegen eher im unteren linken Bereich und zeigen geringere Werte für beide Parameter.

Ersichtlich ist, dass Farneäck im Vergleich zu RAFT insgesamt höhere Werte produziert.

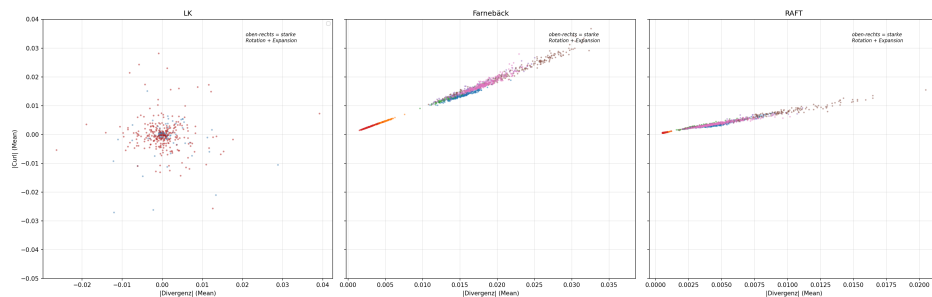


Abbildung 8.28: Vergleich des Zusammenhangs zwischen mittlerer Divergenz und mittlerem Curl für alle Videos, jeweils für Lucas-Kanade (links), Farneäck (Mitte) und RAFT (rechts). (Eigene Darstellung.)

Zusammenfassung

Die quantitative Analyse der Bewegungsparameter zeigt Unterschiede zwischen den untersuchten Szenarien und den optischen Flussverfahren. Farneäck liefert insgesamt höhere und streuungsreichere Werte für Richtungs-

Entropie, Divergenz und Curl, während **RAFT** konsistent kompaktere und niedrigere Werte aufweist. Das Lucas-Kanade-Verfahren (LK) fügt sich mitig ein. Insbesondere bei den Moshpit-Videos sind sowohl Divergenz als auch Curl bei **FB** erhöht. Die Crowd-Videos zeigen geringere Werte und eine stärkere Bewegungsordnung, vor allem bei **RAFT**. Der Zusammenhang zwischen Divergenz und Curl ist bei Farneäck und **RAFT** linear ausgeprägt. **RAFT** liefert hier robustere und klarere Muster. Die Auswertung der Richtungs-Entropie und Uniformität unterstreicht zudem, dass Farneäck eine größere Richtungsvielfalt in dichten Crowd-Szenen erfasst, während **RAFT** dominante Bewegungsrichtungen besser abbildet. Visuelle Heatmaps bestätigen dies. Mit **RAFT** werden Bewegungsmuster und -grenzen in der Menge deutlicher und kontrastreicher dargestellt als mit Farneäck, allerdings erzeugt **RAFT** dabei auch leicht verrauschte Bilder.

8.5 LAUFZEITVERGLEICH

Die Laufzeiten der untersuchten Verfahren wurden umfassend analysiert. Dabei wurden die mittlere Verarbeitungszeit pro Frame und Methode sowie die Echtzeitfähigkeit der Methoden ausgewertet. Die Experimente wurden auf einem System mit NVIDIA GeForce RTX 2070 SUPER (GPU-Beschleunigung, siehe Tabelle 8.7) durchgeführt.

Mittlere Laufzeit pro Frame

Die Abbildung 8.29 zeigt die mittlere Laufzeit pro Frame für Lucas-Kanade (LK), Farneäck (FB) und **RAFT**. **LK** ist mit durchschnittlich 7.1 ms pro Frame am schnellsten, gefolgt von Farneäck mit 623.6 ms. **RAFT** benötigt mit 1540.4 ms pro Frame deutlich mehr Zeit. Auch innerhalb einzelner Videos zeigen sich Unterschiede zwischen den Methoden. So ist der Unterschied zwischen **RAFT** und den anderen zwei Methoden beispielsweise in den Videos "Bühnen fern" und den drei Moshpit Videos deutlich größer.

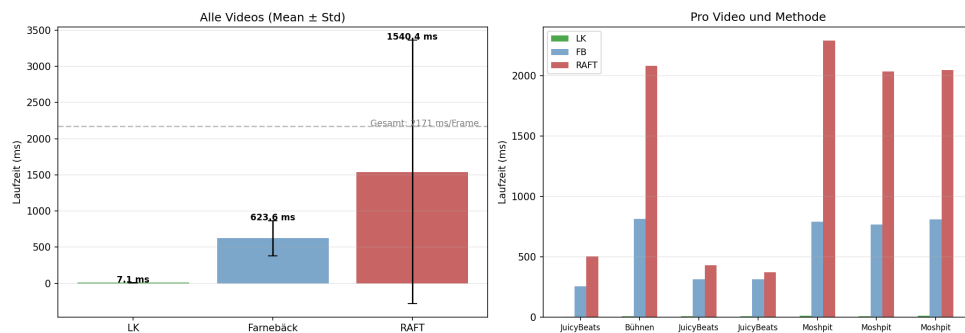


Abbildung 8.29: Mittlere Laufzeit pro Frame für alle Methoden. Links: Mittelwert und Standardabweichung der Laufzeit pro Frame über alle Videos. Rechts: Laufzeit pro Video und Methode (Lucas-Kanade, Farneäck, **RAFT**). (Eigene Darstellung.)

Vergleich Einfach vs. Komplex

Die Abbildung 8.30 zeigt den Vergleich der Laufzeiten in einfachen und komplexen Szenen. In komplexen Szenen steigt die mittlere Laufzeit bei allen Methoden um etwa 24-31 % an. Dies ist vor allem auf mehr Bewegungsvektoren zurückzuführen.

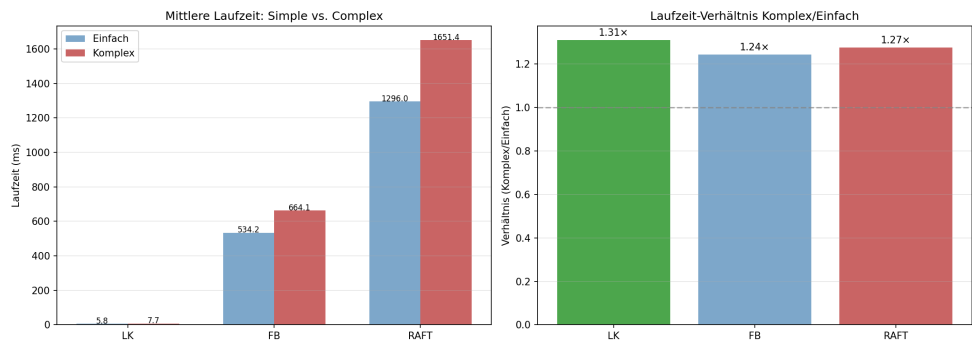


Abbildung 8.30: Mittlere Laufzeit (links) und Verhältnis Komplex/Einfach (rechts) für die Methoden Lucas-Kanade (LK), Farneback (FB) und RAFT. (Eigene Darstellung.)

Tabellarische Übersicht

Video	LK (ms)	FB (ms)	RAFT (ms)
JuicyBeats nah	3,5	253,1	504,3
Bühnen fern	8,1	812,8	2080,4
JuicyBeats Crowd nah	6,3	312,7	430,8
JuicyBeats Crowd fern	5,4	314,4	370,4
Moshpit trennen	8,3	789,5	2290,5
Moshpit zusammen	8,1	767,3	2033,3
Moshpit	8,6	807,2	2048,0

Tabelle 8.6: Laufzeiten (Mittelwerte) pro Frame für jedes Video und Methode.

Echtzeitfähigkeit

Die erreichbaren FPS wurden den jeweiligen Video-FPS (30 FPS) gegenübergestellt (siehe Abbildung 8.31). Die grünen Balken zeigen die erreichten Bildraten der jeweiligen Methode, die gestrichelte Linie markiert die Echtzeitgrenze. Nur Lucas-Kanade ermöglicht eine Echtzeitverarbeitung auf GPU, während Farneback und insbesondere RAFT die Echtzeitfähigkeit deutlich unterschreiten.

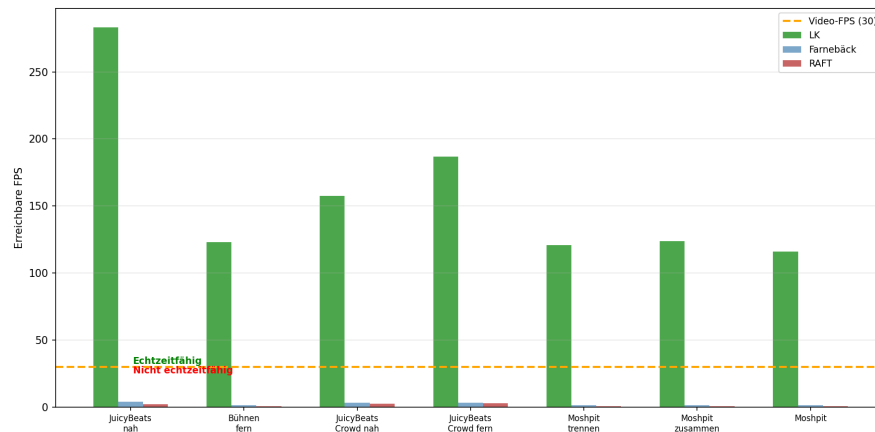


Abbildung 8.31: Erreichbare FPS im Vergleich zu den Video-FPS (30 FPS) für die Methoden LK, FB und RAFT und für alle Videos. (Eigene Darstellung.)

Hardware-Kontext

Die Auswertungen wurden auf folgender Hardware durchgeführt (mit GPU-Beschleunigung):

Komponente	Wert
GPU	NVIDIA GeForce RTX 2070 SUPER
CUDA Version	12.1
Compute Capability	7.5
PyTorch Version	2.5.1+cu121

Tabelle 8.7: Hardware-Kontext der Laufzeitanalyse.

Zusammenfassung

Insgesamt zeigt sich, dass lediglich Lucas-Kanade (LK) auf GPU bei diesen Experimenten eine Echtzeitverarbeitung ermöglicht. Farneback und RAFT sind bei diesen Experimenten auf GPU für Echtzeit-Anwendungen nicht geeignet. Komplexere Szenen führen bei allen Methoden zu erhöhten Laufzeiten.

DISKUSSION UND BEANTWORTUNG DER FORSCHUNGSFRAGE

In diesem Kapitel werden die einzelnen Ergebnisteile (Quantitativer Methodenvergleich ohne Ground-Truth-Daten, Qualitätsbewertung mittels Forward-Backward-Consistency, Crowd-Dynamik-Analyse sowie der Laufzeitvergleich) nacheinander näher beleuchtet. Für jede Teilanalyse erfolgt zunächst eine Interpretation der Ergebnisse und anschließend eine kritische Reflexion möglicher Limitationen. Danach folgt die kritische Reflexion der Methodik. Abschließend wird die Forschungsfrage der Arbeit auf Basis aller Erkenntnisse beantwortet.

9.1 EVALUATION GEGEN GROUND-TRUTH-DATEN

Interpretation der Ergebnisse

Die Auswertung zeigt, dass die klassischen Methoden Lucas-Kanade und Farnebäck hinsichtlich Endpunktfehler (EPE) und Winkelfehler insgesamt bessere und robustere Ergebnisse erzielen als das Deep-Learning-basierte Verfahren [RAFT](#). Wie in [Abbildung 8.1](#) ersichtlich, liegen die mittleren EPE-Werte von Lucas-Kanade (1,29 px) und Farnebäck (1,32 px) deutlich unter dem Wert von [RAFT](#) (2,09 px). Auch Median- und Maximalwerte sind für die klassischen Ansätze günstiger.

Die Verteilung des [EPE](#) über alle Videos ([Abbildung 8.2](#)) zeigt, dass [RAFT](#) eine größere Streuung aufweist, während Lucas-Kanade und Farnebäck sehr ähnliche und stabile Fehlerverteilungen besitzen. Beim Winkelfehler schneiden die klassischen Methoden ebenfalls besser ab. Beide liefern durchschnittliche Winkelfehler um 30° , während [RAFT](#) mit 63° einen deutlich höheren Wert zeigt (vgl. [Abbildung 8.3](#)).

Die nahezu identische Performance von Lucas-Kanade und Farnebäck in Bezug auf mittlere Fehler und Winkelfehler zeigt, dass klassische, auf lokalen Bildmerkmalen basierende, Verfahren nach wie vor konkurrenzfähig sind.

Die tendenziell höheren Fehler und Richtungsabweichungen von [RAFT](#) gegenüber den klassischen Verfahren sind vermutlich darauf zurückzuführen, dass das Modell auf generischen Datensätzen trainiert wurde, welche nicht spezifisch auf Crowd-Szenarien oder Drohnenperspektiven zugeschnitten sind. In diesen Situationen kann [RAFT](#) daher auch Fehlabschätzungen produzieren, da es stärker von Trainingsverteilungen beeinflusst wird.

In Bezug auf die Szenenkomplexität ([Abbildung 8.5](#)) zeigt sich, dass Lucas-Kanade und Farnebäck in einfachen Szenen geringere Winkelfehler erzielen,

während [RAFT](#) sowohl in einfachen als auch komplexen Szenen ähnliche Winkelfehler und insgesamt größere Abweichungen produziert.

Der Zusammenhang zwischen Bewegungsintensität und Fehler ist in Abbildung 8.6 dargestellt. Während bei allen Methoden mit zunehmender Bewegungsstärke der Fehler steigt, ist dieses Verhalten bei [RAFT](#) am stärksten ausgeprägt. Für die Anwendungspraxis bedeutet dies, dass in Szenarien mit sehr dynamischen Bewegungen, wie sie beispielsweise bei Menschenmengen oder plötzlichen Richtungswechseln auftreten, die Verfahren fehleranfälliger sind und unzuverlässigere Schätzungen liefern.

Kritische Reflexion und Limitationen

Die Evaluation basiert auf einer begrenzten Auswahl an Drohnenvideos und spezifischen Szenarien. Die manuelle Ground-Truth-Annotation kann zu subjektiven Fehlern führen. Zudem wurde [RAFT](#) nicht speziell für Drohnenspektiven oder die hier vorkommenden Bewegungsmuster trainiert, was die Generalisierungsfähigkeit einschränkt.

Die klassischen Verfahren profitieren möglicherweise davon, dass sie weniger von Trainingsdaten abhängig sind und robuste Schätzungen in Szenen mit moderaten Bewegungen liefern. Für sehr schnelle oder ungewöhnliche Bewegungen könnten auch sie an ihre Grenzen stoßen. Die Aussagekraft der Ergebnisse ist auf die genutzten Kameraperspektiven, Auflösungen und Bewegungsmuster beschränkt. Für andere Drohnenaufnahmen oder Anwendungskontexte könnten sich die relativen Leistungen der Methoden verschieben.

Fazit

Die Analyse belegt, dass Lucas-Kanade und Farneback in den ausgewerteten Drohnenszenarien zuverlässigere und genauere Schätzungen liefern als das Deep-Learning-basierte Verfahren [RAFT](#). Die klassischen Methoden profitieren von ihrer Unabhängigkeit gegenüber Trainingsdaten und zeigen besonders bei moderaten Bewegungen stabile Ergebnisse. Allerdings ist die Übertragbarkeit der Resultate durch die begrenzte Datenauswahl und Szenarienvielfalt eingeschränkt.

9.2 QUANTITATIVER METHODENVERGLEICH OHNE GROUND-TRUTH-DATEN

Interpretation der Ergebnisse

Der quantitative Methodenvergleich ohne Ground-Truth-Daten zeigt, dass die Übereinstimmung zwischen den klassischen Methoden Lucas-Kanade und Farneback in den einfachen Szenen am größten ist. Die Fehlerwerte und Winkelabweichungen für diesen Methodenvergleich liegen näher beieinander als bei Vergleichen mit [RAFT](#). Bei den komplexen Szenen sind beim [EPE](#) und RSME [FB](#) und [RAFT](#) näher beieinander. Nur beim Winkelfehler sind auch

bei den komplexen Szenen **LK** und **FB** näher beieinander. Die Tabelle 8.2 verdeutlicht, dass insbesondere der Unterschied zwischen Lucas-Kanade und **RAFT** deutlich ausgeprägter ist als zwischen Farnebäck und **RAFT**.

Die Auswertung nach Szenenkomplexität zeigt, dass die Fehlerwerte (EPE) in komplexen Szenen für alle Methodenvergleiche steigen (siehe Abbildung 8.7 und Tabelle 8.3). Besonders die Unterschiede zwischen Lucas-Kanade und **RAFT** nehmen mit zunehmender Komplexität zu. Die Winkelabweichung zwischen **LK** und **FB** steigt in komplexen Szenen ebenfalls, während sie für die Vergleiche mit **RAFT** leicht sinken (Abbildung 8.8).

Die Ergebnisse spiegeln die methodischen Grundprinzipien der eingesetzten Verfahren wider. Die geringere Fehlerdifferenz und Winkeldifferenz zwischen Lucas-Kanade und Farnebäck im Vergleich zu den jeweiligen Vergleichen mit **RAFT** lässt sich auf die algorithmischen Gemeinsamkeiten der klassischen Verfahren zurückführen. Beide Methoden verwenden lokale, bildbasierte Ansätze. Lucas-Kanade trackt markante Features, während Farnebäck eine dichte, auf lokalen Bildstrukturen basierende Flusschätzung vornimmt. Dadurch reagieren beide Verfahren ähnlich auf die Bildinhalte und die vorhandenen Textur- und Bewegungsmuster.

Im Gegensatz dazu nutzt **RAFT** als Deep-Learning-Verfahren globale Kontextinformationen und ist auf anderen Datensätzen trainiert worden. Dadurch kann **RAFT** Bewegungsrichtungen in komplexen, texturarmen oder chaotischen Szenen anders interpretieren. Dies führt zu größeren Abweichungen in der Richtungsbestimmung gegenüber den klassischen Verfahren, vor allem in Situationen, in denen die Bildstruktur atypisch oder die Bewegung schwer erfassbar ist. Die höhere Fehler- und Winkeldifferenz spiegelt somit methodische Unterschiede und unterschiedliche Sensitivität gegenüber Bildartefakten wider.

Zusätzlich resultiert die höhere Übereinstimmung zwischen Farnebäck und **RAFT** im Vergleich zu Lucas-Kanade auch aus der grundlegenden Natur der Algorithmen. Beide liefern dichte Bewegungsfelder und sind darauf ausgelegt, auch in komplexen Bildregionen Bewegungen zu schätzen.

Kritische Reflexion und Limitationen

Die hier durchgeführte Analyse basiert auf einem indirekten Methodenvergleich ohne Referenz zu Ground-Truth-Daten. Dadurch können keine Aussagen über die absolute Fehlerhaftigkeit der Methoden getroffen werden, sondern lediglich deren gegenseitige Übereinstimmung wird betrachtet. Es ist möglich, dass eine hohe Übereinstimmung lediglich auf ähnlicher Fehlerstruktur beruht, nicht aber zwingend auf einer besseren Schätzqualität. Durch die Wahl weiterer vielfältiger Metriken und die detaillierte Analyse in den folgenden Abschnitten wird jedoch die Aussagekraft der Ergebnisse gestärkt und es lässt sich die Forschungsfrage am Ende in Abschnitt 9.7 fundiert beantworten.

Zudem können die Ergebnisse durch die spezifische Auswahl der Videos und die Szenenkomplexität der Methoden beeinflusst sein.

Eine Übertragbarkeit der Ergebnisse auf andere Datensätze oder Anwendungskontexte ist nur eingeschränkt möglich. Der Methodenvergleich liefert jedoch wertvolle Hinweise auf die Robustheit und Konsistenz der Verfahren in unterschiedlichen Szenarien.

Fazit

Die Ergebnisse legen nahe, dass die klassischen Methoden Lucas-Kanade und Farnebäck insgesamt eine bessere gegenseitige Konsistenz aufweisen als Vergleiche mit [RAFT](#). In komplexen Szenen verstärken sich die Unterschiede zwischen den Verfahren, insbesondere im Vergleich zwischen Lucas-Kanade und [RAFT](#). Für eine umfassende Bewertung der Methoden ist eine zusätzliche Analyse mit Ground-Truth-Daten (wie sie in Abschnitt [8.1](#) erfolgte) unerlässlich.

9.3 QUALITÄTSBEWERTUNG MITTELS FORWARD-BACKWARD-CONSISTENCY

Interpretation der Ergebnisse

Die Qualitätsbewertung mittels Forward-Backward-Consistency zeigt, dass alle Methoden einen überwiegenden Anteil konsistenter Bewegungsvektoren liefern.

Die Heatmap des mittleren Forward-Backward-Errors (Abbildung [8.9](#)) zeigt jedoch, dass Lucas-Kanade und Farnebäck in beiden Szenarientypen durchgehend sehr niedrige Fehlerwerte erzielt, während [RAFT](#) insbesondere in komplexen Szenen einen Anstieg des mittleren Forward-Backward-Errors aufweist. Dies wird durch die Übersicht der Kennzahlen in Tabelle [8.4](#) bestätigt. Lucas-Kanade und Farnebäck erreichen nahezu fehlerfreie Forward-Backward-Consistency-Werte, wohingegen [RAFT](#) insbesondere in komplexen Szenen leicht abfällt, aber trotzdem weiterhin hohe Werte liefert. Dies deutet darauf hin, dass [LK](#) und Farnebäck in Bezug auf die Vorwärts-Rückwärts-Konsistenz robuster gegenüber steigender Szenenkomplexität sind, während [RAFT](#) in komplexen Szenen an Konsistenz verliert. Wichtig hierbei ist, dass ein Punkt bereits ab einer Abweichung von 1,5 Pixeln als inkonsistent zählt. Der Wert ist sehr restriktiv gesetzt.

Die räumlichen Visualisierungen (Abbildung [8.10](#)) zeigen, dass inkonsistente Punkte bei [LK](#) und Farnebäck kaum auftreten und fast ausschließlich konsistente Schätzungen vorliegen. Bei [RAFT](#) sind in komplexen Szenen einzelne Bereiche mit inkonsistenten Schätzungen (rote Punkte) erkennbar. Insgesamt bestätigt die Forward-Backward-Consistency-Analyse die Robustheit von [LK](#) und Farnebäck und weist auf punktuelle Schwächen von [RAFT](#) in komplexen Szenen hin.

Kritische Reflexion und Limitationen

Die Forward-Backward-Consistency ist ein nützliches Proxy-Maß zur Bewertung der internen Plausibilität von Bewegungsvektoren. Allerdings ersetzt sie keine Validierung gegen externe Ground-Truth-Daten. Ein hoher Consistency-Wert bedeutet lediglich, dass die Vorwärts- und Rückwärtsberechnungen übereinstimmen, nicht aber zwingend, dass die Bewegungsvektoren korrekt sind. Die Wahl des Schwellenwerts (hier 1,5 px) beeinflusst zudem die Klassifikation von Punkten als konsistent oder inkonsistent. Eine weitere Limitationen bestehen darin, dass die Forward-Backward-Consistency keine systematischen Fehler erkennt. Ein konsistenter, aber falscher Flow kann dennoch als "gut" bewertet werden. Daher sollte die Forward-Backward-Consistency stets im Zusammenspiel mit weiteren Metriken und Plausibilitätschecks verwendet werden.

In sehr dynamischen oder unstrukturierten Szenen verliert die Forward-Backward-Consistency zusätzlich an Aussagekraft, da auch durch systematische Fehler verfälschte Ergebnisse als konsistent erscheinen können.

Fazit

Die Analyse der Forward-Backward-Consistency belegt, dass [LK](#) und Farneäck unter den untersuchten Bedingungen eine sehr hohe interne Robustheit aufweisen. [RAFT](#) liefert ebenfalls konsistente Ergebnisse, zeigt jedoch in komplexen Szenen punktuelle Schwächen. Insgesamt erweitert die Forward-Backward-Consistency die Möglichkeiten zur Qualitätsbewertung erheblich, ersetzt jedoch nicht die sorgfältige Validierung durch Ground-Truth-Daten, wo immer diese verfügbar sind.

9.4 CROWD-DYNAMIK-ANALYSE

Interpretation der Ergebnisse

Die Analyse der Crowd-Dynamik anhand von Richtungs-Entropie, Richtungs-Uniformität, Geschwindigkeit, Divergenz und Curl zeigt deutliche Unterschiede sowohl zwischen den verwendeten Methoden als auch zwischen den untersuchten Szenarientypen.

Die Richtungs-Entropie ist beim Farneäck-Verfahren in den Crowd-Szenen (*JuicyBeats Crowd nah/fern*) höher als in den Moshpit-Szenen (Abbildung 8.12). Das deutet auf eine größere Bewegungsvielfalt und weniger geordnete Dynamik in dichten Menschenmengen hin, wohingegen die geringeren Entropiewerte der Moshpit Szenen darauf hinweisen, dass dort die Bewegungen geordneter und stärker in eine Richtung verlaufen. Bei [RAFT](#) (Abbildung 8.13) sind die Entropiewerte insgesamt niedriger und weniger gestreut. Dies weist darauf hin, dass [RAFT](#) dominantere Bewegungsrichtungen erkennt und weniger Richtungsvielfalt abbildet. Beim Lucas-Kanade-Verfahren bewegen sich

die Richtungs-Entropiewerte insgesamt auf moderatem Niveau und zeigen überlappende Wertebereiche zwischen Crowd- und Moshpit-Szenen.

Ein ähnliches Bild zeigt sich bei der Richtungs-Uniformität. **LK** und **Farnebäck** messen in Crowd-Szenen geringere Uniformität als in Moshpit-Szenen (Abbildung 8.15 und Abbildung 8.14), während **RAFT** (Abbildung 8.16) kaum Unterschiede zwischen den Szenen erkennt und insgesamt homogenere Werte liefert. Dies deutet darauf hin, dass **LK** und **FB** die Bewegungsrichtungen in der Menschenmenge weniger einheitlich sieht und dort keine dominante Richtung berechnet. Im Gegensatz dazu weisen bei **LK** und **FB** die Moshpit-Szenen höhere Uniformitätswerte auf, was dort auf eine gemeinsame und dominante Bewegungsrichtung der Gruppe schließen lässt.

Die qualitativen Heatmaps der Bewegungsrichtungen verdeutlichen diese Unterschiede. Mit **RAFT** werden dominante Bewegungsmuster und -grenzen kontrastreicher und klarer dargestellt als mit **Farnebäck**. Dafür sind die Heatmaps zum Teil verrauscht. Dies ist stimmig mit den Werten der Richtungsentropie und Uniformität der zwei Methoden. Niedrige Entropie- und hohe Uniformitätswerte, wie sie insbesondere bei **RAFT** und in Moshpit-Szenen auftreten, deuten auf geordnete und dominante Bewegungsrichtungen hin. Dies zeigt sich in kontrastreichen Heatmaps. Im Gegensatz dazu führen höhere Entropiewerte und geringere Uniformität, wie sie bei **Farnebäck** in komplexen Szenen beobachtet werden, zu diffuseren Heatmaps. Die visuelle Analyse der Heatmaps ergänzt somit die numerischen Kennzahlen und macht Unterschiede in Bewegungsordnung und Richtungsvielfalt direkt sichtbar.

Die Geschwindigkeitsverteilungen zeigen, dass in den Videos „Bühnen fern“ und „JuicyBeats Crowd fern“ sehr niedrige Geschwindigkeiten auftreten (Mediane unter 3 px/s). Dies ist dadurch zu erklären, dass die Drohne in diesen Videos deutlich höher fliegt als in den anderen Videos und die Bewegung deswegen deutlich kleiner sind. In den anderen Videos, insbesondere in Moshpit-Szenen, werden höhere Geschwindigkeiten erreicht (Mediane 6 bis 24 px/s).

Die räumlichen Heatmaps (vgl. Abbildungen 8.21) machen sichtbar, dass die stärkste Dynamik im Zentrum der Moshpit-Bereiche vorliegt, während der restliche Crowd-Bereich relativ ruhig bleibt. In einfachen, geordneten Szenen wie „Bühnen fern“ zeigen die Heatmaps überwiegend blaue und grüne Bereiche mit niedrigen Geschwindigkeiten. In komplexen Szenen (insbesondere in verschiedenen Moshpit-Situationen) treten klar abgegrenzte rote und gelbe Zonen auf, die Hotspots intensiver Dynamik markieren.

Die qualitativen Heatmap-Ergebnisse zeigen, dass diese Visualisierung die rein numerische Analyse sinnvoll ergänzt. Auffällige Muster in den Kennzahlen können im Bildkontext nachvollzogen und plausibilisiert werden. Außerdem ermöglichen die Heatmaps ein schnelles Screening großer Datenmengen, da kritische Dynamikbereiche sofort ins Auge fallen. Für die Praxis (z. B. etwa im Crowd-Management oder in der Sicherheitsüberwachung) bieten sie damit eine intuitive und effektive Form der Lagebewertung. Insgesamt unterstreichen die Ergebnisse, dass Amplitude-Heatmaps dazu beitra-

gen, die Validität und Aussagekraft der quantitativen Bewegungsmetriken zu erhöhen.

Die Divergenz- und Curl-Analysen zeigen, dass beide Werte tendenziell in dynamischen Szenen (Moshpit) erhöht sind. Allerdings waren die Werte über alle Methoden hinweg auf einem sehr niedrigen Niveau. Lucas-Kanade zeigt insgesamt Werte nahe Null, mit leicht erhöhter Streuung und Mittelwerten in den dynamischeren Szenen. Farnebäck produziert dabei höhere Mittelwerte und eine größere Streuung, insbesondere in den Moshpit-Videos, während RAFT insgesamt niedrigere und gleichmäßigere Werte liefert. Ein linearer Zusammenhang zwischen Divergenz und Curl zeigt sich bei Farnebäck und RAFT, ist bei RAFT jedoch kompakter ausgeprägt. Dies erschafft das Potenzial, die Kombination von Divergenz und Curl als Indikator für chaotische Dynamiken zu nutzen. Durch die gemeinsame Betrachtung lassen sich Moshpit-Situationen tendenziell automatisiert und objektiv von gewöhnlichem Crowd-Verhalten abgrenzen. Für die Praxis bedeutet dies, dass diese Parameter als Frühwarnindikatoren genutzt werden können, um kritische Entwicklungen schnell zu detektieren und gezielt Maßnahmen einzuleiten.

Die Ergebnisse zeigen, dass die gewählten Methoden unterschiedliche Schwerpunkte in der Bewegungsmustererkennung setzen. Farnebäck erfasst eine größere Richtungsvielfalt und ist sensibler für chaotische, ungerichtete Bewegungen und produziert eine größere Streuung der Bewegungsparameter. RAFT liefert homogenere und robustere Schätzungen, glättet jedoch tendenziell kleinere Bewegungsdetails und reduziert damit die Sensitivität für subtile Richtungswechsel. Die Interpretation der Bewegungsparameter hängt zudem stark von der Szenenauswahl und der Qualität der Eingangsdaten ab.

Zudem zeigt sich, dass in komplexen Szenen wie Moshpits die Vielzahl an individuellen Bewegungen, Richtungswechseln und Rotationen zu einer deutlichen Zunahme von Divergenz und Curl führen sowie einer Abnahme der Richtungsentropie (weil geordnete Bewegung beim Trennen oder Aufeinanderzurennen).

Kritische Reflexion und Limitationen

Die Ergebnisse zeigen, dass alle Methoden unterschiedliche Charakteristika in der Bewegungsmustererkennung aufweisen. Farnebäck detektiert eine größere Richtungsvielfalt und produziert eine größere Streuung der Bewegungsparameter. RAFT liefert homogenere und robustere Schätzungen, glättet aber eventuell kleinere Bewegungsdetails und reduziert damit die Sensitivität für subtile Richtungswechsel. Zusätzlich sind bei RAFT die Bilder etwas verrauscht.

Fazit

Die Crowd-Dynamik-Analyse zeigt, dass Farnebäck und **RAFT** komplementäre Stärken besitzen. Farnebäck liefert detaillierte, aber streuungsreiche Informationen zu Bewegungsvielfalt und -dynamik, während **RAFT** dominante Bewegungsrichtungen und makroskopische Muster klarer und robuster abbildet. Amplitude-Heatmaps ermöglichen zudem eine gezielte und räumlich differenzierte Analyse der Crowd-Dynamik, was insbesondere bei der Überwachung großer Menschenmengen von zentraler Bedeutung ist. Insgesamt liefern Optical-Flow-basierte Methoden einen wichtigen Beitrag zur wissenschaftlichen Analyse und praktischen Überwachung von Crowd-Situationen, wobei die Wahl der Methode je nach Anwendungsfall und gewünschtem Detailgrad sorgfältig getroffen werden sollte.

9.5 LAUFZEITVERGLEICH

Interpretation der Ergebnisse

Die Analyse der Laufzeiten zeigt deutliche Unterschiede zwischen den untersuchten Methoden. Wie aus Abbildung 8.29 hervorgeht, ist Lucas-Kanade (LK) mit durchschnittlich 7,1 ms pro Frame mit Abstand das schnellste Verfahren. Farnebäck benötigt im Mittel rund 624 ms pro Frame, während **RAFT** mit 1540 ms deutlich längere Zeiten aufweist. Dieses Muster setzt sich über alle untersuchten Videos hinweg fort.

Ein Vergleich zwischen einfachen und komplexen Szenen (Abbildung 8.30) zeigt, dass die mittlere Laufzeit bei allen Methoden in komplexen Szenen ansteigt. Bei **LK** steigt die Zeit um etwa 31%, bei **FB** um etwa 24% und bei **RAFT** um etwa 27%. Dieser Anstieg ist auf komplexere Bildbereiche und einer höheren Dichte an Bewegungsvektoren zurückzuführen.

Die ausgeprägten Unterschiede bei der Verarbeitungszeit spiegeln direkt die algorithmische Komplexität und die Zielsetzung der Optical-Flow-Verfahren wider. Lucas-Kanade ist als klassischer, sparsamer Feature-Tracking-Algorithmus auf Geschwindigkeit und Ressourceneffizienz optimiert. Daher eignet er sich besonders für Echtzeitanwendungen, bei denen die Hardware begrenzt ist und nur eine schnelle, grundsätzliche Bewegungsanalyse benötigt wird, beispielsweise zur schnellen Vorab-Überwachung oder als Trigger für detailliertere Analysen.

Farnebäck und insbesondere **RAFT** stehen für den modernen Trend zu dichten, globalen Bewegungsfeldschätzungen, welche deutlich mehr Rechenressourcen beanspruchen. Die Laufzeitsteigerung in komplexen Szenen ist plausibel, da mehr Bildpunkte, mehr Vielfalt im Bewegungsmuster und eine höhere Dynamik zusätzliche Berechnungen erfordern. Dies betrifft alle Methoden, ist jedoch bei dichten Ansätzen besonders ausgeprägt.

Hinsichtlich der Echtzeitfähigkeit (Abbildung 8.31) erreicht ausschließlich Lucas-Kanade die für eine Echtzeitverarbeitung nötigen **FPS**-Werte (mind. 30 FPS), während Farnebäck und **RAFT** weit darunter liegen und damit für

Echtzeitanwendungen (auf der in dieser Arbeit genutzten Hardware) ungeeignet sind. Allerdings wäre eine Möglichkeit die Berechnungen bei Farneäck und RAFT nicht auf jedem Frame sondern beispielsweise nur auf jedem 10. Frame durchzuführen, um eine schnellere Berechnung zu gewährleisten.

Kritische Reflexion und Limitationen

Die Ergebnisse unterstreichen, dass die Wahl der Optical-Flow-Methode erhebliche Auswirkungen auf die praktische Einsetzbarkeit in zeitkritischen Anwendungen hat. Während Lucas-Kanade für Live-Anwendungen geeignet ist, schränken die langen Laufzeiten von Farneäck und insbesondere RAFT deren Nutzbarkeit im Echtzeitkontext stark ein.

Dabei ist zu berücksichtigen, dass alle Experimente auf einem leistungsfähigen GPU-System durchgeführt wurden. Auf weniger potenter Hardware wären die Laufzeitunterschiede noch ausgeprägter. Zudem steigt die Rechenzeit in komplexen Szenen an, was in realen Einsatzszenarien mit variabler Szenenkomplexität zu weiteren Laufzeitproblemen führen kann.

Zudem ist zu beachten, dass die Laufzeit nur ein Kriterium unter mehreren ist. Genauigkeit, Robustheit und Interpretierbarkeit der Ergebnisse sollten bei der Methodenwahl stets mitberücksichtigt werden. Für Offline-Analysen oder Anwendungen mit geringerem Zeitdruck kann der Einsatz rechenintensiver Verfahren wie RAFT dennoch sinnvoll sein.

Ein weiterer Punkt ist, dass die Laufzeiten je nach Implementierung, verwendeter Bibliothek und spezifischen Systemeinstellungen variieren können. Hinzu kommt, dass die Laufzeitanalyse keine Aussagen über den Energieverbrauch oder die Skalierbarkeit bei sehr großen Datensätzen trifft. Die Wahl des Verfahrens sollte daher immer unter Berücksichtigung der verfügbaren Ressourcen, der praktischen Anforderungen und der geplanten Hardware erfolgen.

Fazit

Insgesamt zeigt sich, dass die Echtzeitfähigkeit bei Optical-Flow-Analysen eine kritische Einschränkung darstellt und die Auswahl der Methode eng an die Anforderungen und Rahmenbedingungen der Zielanwendung gekoppelt sein sollte.

9.6 KRITISCHE REFLEXION DER METHODIK

9.6.1 Datenauswahl

Die Auswahl und Ausgestaltung des Datensatzes bilden eine solide Grundlage für die Methodenbewertung. Durch die Kombination von einfachen und komplexen Drohnenvideos, wird eine breite Szenenvielfalt abgedeckt. Die Vereinheitlichung von Auflösung und Framerate gewährleisten methodische Vergleichbarkeit. Es wurden allerdings ausschließlich statische Droh-

nenvideos genutzt. Dies verringert die Wahrscheinlichkeit von Verzerrungen durch Aufnahmeparameter, führt aber dazu, dass reale Einschränkungen durch Drohnenbewegungen nicht evaluiert wurden. Ein weiterer limitierender Faktor ist, dass ausschließlich diese sieben Videos verwendet wurden, welche möglicherweise nicht alle für die Aufgabenstellung relevanten Szenarien abbilden. Die Auswahl der Videos erfolgte nach eigenen Bewertungskriterien, was potenziell zu einer unbewussten Verzerrung führen kann. Dadurch könnten bestimmte Bewegungsmuster oder Szenentypen unterrepräsentiert sein, was die Generalisierbarkeit der Ergebnisse weiter einschränkt.

9.6.2 Manuelle Annotation und potenzielle Fehlerquellen

Die manuelle Erstellung von Ground-Truth-Bewegungsvektoren stellt eine essenzielle Grundlage für die objektive Bewertung und Kalibrierung der Optical-Flow-Verfahren dar. In dieser Arbeit wurden insgesamt 3108 Punktpaare auf 624 Frames manuell annotiert, um eine möglichst hochwertige Referenz für unterschiedliche Szenarientypen zu schaffen. Trotz großer Sorgfalt kann dieser Prozess jedoch einigen systematischen und zufälligen Fehlerquellen unterliegen, die die Aussagekraft der Evaluation beeinflussen können.

Interpretation und mögliche Fehlerquellen:

- *Subjektivität und Ermüdung:* Die manuelle Bestimmung von Punktkorrespondenzen ist eine kognitive und visuelle Herausforderung. Besonders bei kleinen, schnellen oder sich überlagernden Bewegungen kann es zu subjektiven Fehleinschätzungen kommen. Längere Annotationsitzungen (wie sie beim Annotieren in dieser Arbeit vorkamen) erhöhen zudem das Risiko von Konzentrationsfehlern und inkonsistenter Bewertung.
- *Unklare Korrespondenzen:* In komplexen Szenen mit vielen verdeckten oder sich ähnelnden Objekten (z. B. bei Crowd- oder Moshpit-Situationen) ist es oft schwierig, einen Punkt im ersten Frame eindeutig im Folgebild zuzuordnen. Dies kann zu systematischen Fehlern in der Vektor-Richtung und -Länge führen oder dazu, dass einzelne Bewegungen gar nicht korrekt erfasst werden.
- *Rundungs- und Eingabefehler:* Die Auswahl oder Platzierung der Punkte erfolgt meist per Maus, sodass es zu Pixel-Rundungen, Abweichungen durch Mauszittern oder versehentlichen Fehlklicks kommen kann.
- *Verdeckungen:* Vor allem in den Drohnenvideos mit hohen Crowd-Dichten kommt es vor, dass ein annotierter Punkt im Folgeframe durch andere Personen oder Objekte verdeckt ist. In solchen Fällen muss die Bewegung geschätzt oder interpoliert werden, was die Genauigkeit verringert.

Diese Fehlerquellen sind beim manuellen Annotieren bekannt, lassen sich aber nie vollständig ausschließen. Sie führen dazu, dass auch die manuellen Ground-Truth-Daten dieser Arbeit nur eine Annäherung an die „wahre“ Bewegung darstellen und insbesondere in komplexen, dynamischen Szenen mit erhöhter Unsicherheit behaftet sind.

Für die Interpretation der Evaluationsmetriken in dieser Arbeit bedeutet dies, dass ein Teil der beobachteten Fehler zwischen Ground Truth und automatischer Schätzung auch auf Unsicherheiten bzw. Fehlern in der Annotation zurückzuführen sein kann. Ein sehr niedriger Endpunktfehler sollte daher nicht als absolute Untergrenze der Methodenleistung interpretiert werden. Umgekehrt können scheinbar größere Fehler teilweise aus Unschärfen in der Referenz resultieren.

9.7 BEANTWORTUNG DER FORSCHUNGSFRAGE

Die Forschungsfrage lautete: *„Welche der ausgewählten klassischen und Deep-Learning-basierten Methoden liefert für Drohnenvideos in Bezug auf Genauigkeit, Robustheit und Rechenzeit die quantitativ und qualitativ hochwertigsten Ergebnisse bei der Berechnung von Bewegungsvektoren?“*

Die Ergebnisse der verschiedenen Analysen ermöglichen eine differenzierte Beantwortung entlang der drei zentralen Bewertungskriterien:

Genauigkeit

Die Evaluation gegen die manuellen Ground-Truth-Daten zeigt, dass die klassischen Methoden Lucas-Kanade und Farnebäck in Bezug auf den Endpunktfehler (EPE) und den Winkelfehler insgesamt die genauesten Resultate liefern. Beide erreichen mittlere EPE-Werte unter 1,5 Pixel und Winkelfehler um 30 Grad, während RAFT einen höheren mittleren EPE (2,09 Pixel) und Winkelfehler (63°) aufweist. Die klassischen Methoden zeigen zudem eine geringere Streuung in den Fehlerverteilungen. Besonders in einfachen und gut strukturierten Szenen liefern Lucas-Kanade und Farnebäck sehr ähnliche und verlässliche Schätzungen. RAFT dagegen ist empfindlicher gegenüber Bewegungsstärke und Szenenkomplexität.

Fazit: In Bezug auf die Genauigkeit sind die klassischen Methoden Lucas-Kanade und Farnebäck für die untersuchten Drohnenvideos überlegen.

Robustheit

Die Robustheit der Methoden wurde auf mehreren Ebenen untersucht. Im quantitativen Methodenvergleich und mittels Forward-Backward Consistency zeigte sich, dass Farnebäck insgesamt die höchste innere Konsistenz und Stabilität aufweist, gefolgt von Lucas-Kanade. RAFT ist vor allem in komplexen Szenen weniger konsistent und weist mehr inkonsistente Schätzungen auf. Die klassischen Methoden sind zudem weniger anfällig für stark variierende Bewegungen und Bildinhalte.

Die Crowd-Dynamik-Analyse bestätigte, dass Farnebäck eine größere Richtungsvielfalt und eine höhere Sensitivität für dynamische Veränderungen bietet. RAFT liefert homogenere, robustere Bewegungsfelder, kann jedoch feine Bewegungsdetails und Richtungswechsel glätten und erzeugt zum Teil verrauschte Ergebnisse.

Fazit: Bezüglich Robustheit ist Farnebäck am stabilsten, RAFT folgt, während LK in komplexen, dynamischen Szenen Schwächen zeigt, da das Verfahren keine dichte Heatmaps zur Visualisierung erzeugen kann.

Rechenzeit

Die Laufzeitanalyse belegt deutliche Unterschiede in der Effizienz. Lucas-Kanade ist mit durchschnittlich 7,1 ms pro Frame das mit Abstand schnellste Verfahren und als einziges für Echtzeitanwendungen geeignet. Farnebäck benötigt pro Frame etwa das 90-fache an Rechenzeit, RAFT sogar mehr als das 200-fache. Die Laufzeiten steigen in komplexen Szenen weiter an. Echtzeitverarbeitung ist mit Farnebäck und insbesondere RAFT auf der eingesetzten Hardware nicht möglich.

Fazit: In Bezug auf Rechenzeit und Echtzeitfähigkeit ist Lucas-Kanade klar überlegen. Farnebäck und vor allem RAFT sind für zeitkritische Anwendungen ungeeignet.

Gesamtbewertung

Die besten Bewegungsvektoren im Sinne eines ausgewogenen Verhältnisses von Genauigkeit, Robustheit und Rechenzeit liefern in der vorliegenden Untersuchung die klassischen Methoden Lucas-Kanade und Farnebäck. Lucas-Kanade ist die Methode der Wahl, wenn Echtzeitfähigkeit und Effizienz im Vordergrund stehen. Farnebäck bietet Vorteile bei der detaillierten Analyse von Bewegungsvielfalt und Dynamik, ist aber rechenintensiver. RAFT kann, obwohl es ein moderner Deep-Learning-Ansatz ist und in anderen Kontexten oft führt, in den Drohnenvideos mit komplexen, crowd-basierten Szenen weder in Genauigkeit, noch in Robustheit oder Geschwindigkeit mit den klassischen Verfahren mithalten. Grund dafür könnte sein, dass die Trainingsdaten des Modells für den vorliegenden Anwendungsfall nicht ausreichend sind.

Praktische Implikation: Für die automatisierte Analyse von Drohnenvideos in Crowd-Szenarien sind klassische Optical-Flow-Verfahren weiterhin hochrelevant und zeigen sogar Vorteile gegenüber aktuellen Deep-Learning-Lösungen. Die Wahl des Verfahrens sollte sich neben der gewünschten Genauigkeit und Robustheit insbesondere an den Anforderungen der Laufzeit und der verfügbaren Hardware orientieren.

FAZIT UND AUSBLICK

Dieses Kapitel fasst die Erkenntnisse der Arbeit zusammen und gibt einen Ausblick auf zukünftige Forschungs- und Entwicklungsrichtungen im Bereich der Berechnung von Bewegungsvektoren in Drohnenvideos.

10.1 FAZIT

Diese Arbeit liefert eine fundierte, methodisch nachvollziehbare Vergleichsbasis für Optical-Flow-Methoden in Drohnenvideos unter realistischen Bedingungen. Zwei klassische und ein modernes Deep-Learning-Verfahren wurden systematisch im Kontext von Genauigkeit, Crowd-Dynamik, Robustheit und Echtzeittauglichkeit quantitativ und qualitativ verglichen. Die entwickelte Pipeline, die Ground-Truth-Annotation sowie die verwendeten Evaluationsmetriken sind offen dokumentiert und können für weiterführende Forschung und Praxisprojekte genutzt werden.

Ergebnisse:

- **Genauigkeit:** Lucas-Kanade und Farnebäck erzielen mit durchschnittlichen EPE-Werten von ca. 1,29 px bzw. 1,32 px und Winkelfehlern um 30° die besten Resultate. RAFT liefert mit einem EPE von 2,09 px und einem Winkelfehler von 63° ungenauere Ergebnisse.
- **Robustheit:** Forward-Backward-Consistency liegt für Lucas-Kanade und Farnebäck in nahezu allen Szenen bei $\sim 100\%$, RAFT zeigt in komplexen Szenen leichte Einbußen. Farnebäck erkennt Richtungsvielfalt und chaotische Dynamik am sensibelsten, RAFT liefert homogenere, aber geglättete Bewegungsfelder.
- **Crowd-Dynamik:** Farnebäck weist die höchste Richtungsentropie und Varianz bei Divergenz/Curl auf, was ideal zur Detektion chaotischer Bewegungen ist. RAFT zeigt homogenere Felder und erkennt dominante Bewegungsrichtungen, glättet aber kleinere Details und erzeugt teilweise verrauschte Ergebnisse.
- **Laufzeit:** Lucas-Kanade ist mit durchschnittlich 7 ms/Frame auf GPU als einziges Verfahren echtzeitfähig. Farnebäck (ca. 624 ms/Frame) und RAFT (ca. 1540 ms/Frame) sind (zumindest mit der in dieser Arbeit verwendeten Hardware) nur für Offline-Analysen nutzbar.

Die Ergebnisse ermöglichen konkrete Empfehlungen für die Auswahl und den Einsatz von Optical-Flow-Algorithmen in drohnenbasierten Überwachungs- und Analyseanwendungen: Für Echtzeit- und robuste Bewegungsanalysen

sind klassische Verfahren weiterhin erste Wahl, während Deep-Learning-Ansätze wie RAFT gezielt weiterentwickelt und auf drohnenspezifische Szenarien angepasst werden sollten.

Insgesamt zeigt die Arbeit, dass klassische Optical-Flow-Methoden für bewegungsbasierte Analysen in Drohnenvideos nach wie vor sehr leistungsfähig und praxistauglich sind. Deep-Learning-Verfahren bieten Potenzial, erfordern jedoch weiteres Finetuning für den spezifischen Einsatz. Die entwickelte Pipeline schafft hierfür eine solide Grundlage und kann als Referenz für zukünftige Studien dienen.

10.2 AUSBLICK

Für die Zukunft ergeben sich mehrere wichtige Forschungs- und Entwicklungsperspektiven:

- **Domänenspezifisches Training:** Deep-Learning-Modelle wie RAFT sollten gezielt auf annotierten Drohnen- und Crowd-Datensätzen nachtrainiert werden, um ihre Generalisierungsfähigkeit und Robustheit zu verbessern.
- **Kamerabewegung und Parallaxen:** Künftige Arbeiten sollten die Kompensation von Eigenbewegung und Parallaxeneffekten in Drohnenvideos adressieren und Optical-Flow-Methoden auch unter bewegter Kameraposition evaluieren.
- **Semantische Integration:** Die Kombination optischer Flussfelder mit semantischer Segmentierung (z. B. Personenerkennung mittels YOLO) und Hintergrundentfernung kann die Robustheit und Interpretierbarkeit der Ergebnisse weiter steigern, insbesondere bei dichten Crowd-Szenen.
- **Edge-Optimierung:** Die Optimierung klassischer und moderner Verfahren für ressourcenbeschränkte Edge-Geräte ist ein wichtiger Schritt, um die Echtzeitanalyse direkt auf Drohnen zu ermöglichen.
- **Automatisierte Anomalieerkennung:** Aufbauend auf den Bewegungsvektoren können Methoden zur automatischen Erkennung von Anomalien, z. B. Panik, Flucht oder gefährlicher Dynamik, entwickelt werden.

Teil II

APPENDIX

APPENDIX: WEITERE TABELLEN

- Video 1: Bühnen fern
- Video 2: JuicyBeats Crowd fern
- Video 3: JuicyBeats Crowd nah
- Video 4: JuicyBeats nah
- Video 5: Moshpit
- Video 6: Moshpit trennen
- Video 7: Moshpit zusammen

Kriterium	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6	Video 7
Personen- dichte	E	K	K	E	K	K	K
Bewegungs- dynamik	E	K	K	E	K	K	K
Interaktionen	E	K	K	E	K	K	K
Sichtbarkeit	E	K	K	E	K	K	K
Szenen- struktur	E	K	K	E	K	K	K
Okklusionen	E	K	K	E	K	K	K
Bewegungs- richtung	E	K	K	E	K	K	K
Ereignis- dichte	E	K	K	E	K	K	K

Tabelle A.1: Kategorisierung der Videos von Person A

Kriterium	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6	Video 7
Personen- dichte	E	K	K	E	K	K	K
Bewegungs- dynamik	E	K	K	E	K	K	K
Interaktionen	E	K	K	E	K	K	K
Sichtbarkeit	E	K	K	E	K	K	K
Szenen- struktur	E	K	K	E	K	K	K
Okklusionen	E	K	K	E	K	K	K
Bewegungs- richtung	E	K	K	E	K	K	K
Ereignis- dichte	E	K	K	E	K	K	K

Tabelle A.2: Kategorisierung der Videos von Person B

Video		EPE LK vs. RAFT (Mean±Std)	EPE FB vs. RAFT (Mean±Std)	EPE LK vs. FB (Mean±Std)
JuicyBeats nah (Ein- fach)		0,85 ± 0,16	0,63 ± 0,16	0,35 ± 0,15
Bühnen fern (Ein- fach)		0,11 ± 0,07	0,07 ± 0,04	0,05 ± 0,03
JuicyBeats Crowd nah (Komplex)		4,66 ± 29,91	0,79 ± 3,05	4,46 ± 31,35
JuicyBeats Crowd fern (Komplex)		0,06 ± 0,04	0,04 ± 0,02	0,03 ± 0,02
Moshpit trennen (Komplex)		0,74 ± 0,24	0,34 ± 0,20	0,52 ± 0,23
Moshpit zusammen (Komplex)		0,91 ± 0,31	0,54 ± 0,26	0,59 ± 0,31
Moshpit (Komplex)		0,79 ± 0,25	0,41 ± 0,19	0,56 ± 0,30

Tabelle A.3: Mean±Std der EPE-Metriken der Methoden (paarweise) für alle Videos

Video	EPE LK vs. RAFT (Median)	EPE FB vs. RAFT (Median)	EPE LK vs. FB (Median)
JuicyBeats nah (Einfach)	0,82	0,58	0,35
Bühnen fern (Einfach)	0,11	0,06	0,05
JuicyBeats Crowd nah (Komplex)	4,65	0,66	4,42
JuicyBeats Crowd fern (Komplex)	0,06	0,04	0,03
Moshpit trennen (Komplex)	0,74	0,33	0,52
Moshpit zusammen (Komplex)	0,88	0,48	0,58
Moshpit (Komplex)	0,78	0,39	0,55

Tabelle A.4: Median der EPE-Metriken der Methoden (paarweise) für alle Videos

Video	RMSE LK vs. RAFT	RMSE FB vs. RAFT	RMSE LK vs. FB
JuicyBeats nah (Einfach)	0,91	0,70	0,38
Bühnen fern (Einfach)	0,13	0,08	0,06
JuicyBeats Crowd nah (Komplex)	4,96	0,88	4,76
JuicyBeats Crowd fern (Komplex)	0,06	0,04	0,03
Moshpit trennen (Komplex)	0,76	0,36	0,53
Moshpit zusammen (Komplex)	0,97	0,64	0,63
Moshpit (Komplex)	0,81	0,46	0,58

Tabelle A.5: RMSE der Methoden (paarweise) für alle Videos

Video	Fl-all LK vs. RAFT (%)	Fl-all FB vs. RAFT (%)	Fl-all LK vs. FB (%)
JuicyBeats nah (Einfach)	0,0	0,0	0,0
Bühnen fern (Einfach)	0,0	0,0	0,0
JuicyBeats Crowd nah (Komplex)	15,6	1,2	16,0
JuicyBeats Crowd fern (Komplex)	0,0	0,0	0,0
Moshpit trennen (Komplex)	0,0	0,0	0,0
Moshpit zusammen (Komplex)	0,2	0,1	0,1
Moshpit (Komplex)	0,0	0,0	0,0

Tabelle A.6: Fl-all (%) der Methoden (paarweise) für alle Videos

Video	Winkel LK vs. RAFT (°)	Winkel FB vs. RAFT (°)	Winkel LK vs. FB (°)
JuicyBeats nah (Einfach)	99,4°	72,2°	32,7°
Bühnen fern (Einfach)	128,4°	93,1°	41,0°
JuicyBeats Crowd nah (Komplex)	112,4°	93,9°	50,2°
JuicyBeats Crowd fern (Komplex)	123,5°	88,5°	43,3°
Moshpit trennen (Komplex)	95,3°	77,0°	46,4°
Moshpit zusammen (Komplex)	85,9°	64,3°	50,9°
Moshpit (Komplex)	89,3°	75,8°	48,5°

Tabelle A.7: Winkelvergleiche (°) der Methoden (paarweise) für alle Videos

Video	Entropie (M±S)	Uniformität (M±S)	Divergenz (M±S)	Curl (M±S)
JuicyBeats nah	1,422 ± 0,352	3,685 ± 1,273	0,0016 ± 0,0075	-0,0024 ± 0,0086
Bühnen fern	1,347 ± 0,379	3,751 ± 1,302	0,0002 ± 0,0008	-0,0000 ± 0,0007
JuicyBeats Crowd nah	1,075 ± 0,505	4,641 ± 1,642	-0,0005 ± 0,0033	-0,0002 ± 0,0022
JuicyBeats Crowd fern	1,255 ± 0,388	4,199 ± 1,346	0,0001 ± 0,0007	-0,0001 ± 0,0006
Moshpit trennen	0,937 ± 0,441	5,104 ± 1,592	-0,0002 ± 0,0020	0,0002 ± 0,0033
Moshpit zusammen	1,154 ± 0,374	4,781 ± 1,410	-0,0003 ± 0,0065	-0,0000 ± 0,0214
Moshpit	1,047 ± 0,374	4,988 ± 1,383	0,0014 ± 0,0066	0,0012 ± 0,0137

Tabelle A.8: Lucas-Kanade: Entropie, Uniformität, Divergenz und Curl (Mittelwert ± Standardabweichung) pro Video

Video	Entropie (M±S)	Uniformität (M±S)	Divergenz (M±S)	Curl (M±S)
JuicyBeats				
nah	2,1320 ± 0,7120	3,6559 ± 1,2919	0,0153 ± 0,0015	0,0135 ± 0,0013
Bühnen				
fern	2,7459 ± 0,4592	2,6520 ± 0,9039	0,0045 ± 0,0007	0,0043 ± 0,0007
JuicyBeats				
Crowd	1,8994 ± 0,8649	4,1351 ± 1,5620	0,0142 ± 0,0027	0,0134 ± 0,0028
nah				
JuicyBeats				
Crowd	2,7211 ± 0,4584	2,6844 ± 1,346	0,0027 ± 0,0006	0,0025 ± 0,0005
fern				
Moshpit				
trennen	1,8268 ± 0,7054	4,3501 ± 1,4376	0,0160 ± 0,0033	0,0160 ± 0,0033
Moshpit				
zusammen	1,8537 ± 0,6584	4,4469 ± 1,4007	0,0184 ± 0,0019	0,0177 ± 0,0022
Moshpit	1,7062 ± 0,6449	4,5605 ± 1,3676	0,0219 ± 0,0050	0,0215 ± 0,0054

Tabelle A.9: Farneback: Entropie, Uniformität, Divergenz und Curl (Mittelwert ± Standardabweichung) pro Video

Video	Entropie (M±S)	Uniformität (M±S)	Divergenz (M±S)	Curl (M±S)
JuicyBeats				
nah	1,1969 ± 0,2722	4,0956 ± 1,0449	0,0037 ± 0,0012	0,0037 ± 0,0012
Bühnen				
fern	1,3181 ± 0,3723	3,8408 ± 1,2594	0,0007 ± 0,0001	0,0007 ± 0,0001
JuicyBeats				
Crowd	0,9234 ± 0,4135	4,8315 ± 1,3566	0,0033 ± 0,0010	0,0029 ± 0,0008
nah				
JuicyBeats				
Crowd	1,3604 ± 0,3598	3,8299 ± 1,2732	0,0006 ± 0,0001	0,0006 ± 0,0001
fern				
Moshpit				
trennen	1,1479 ± 0,3887	4,2194 ± 1,2086	0,0047 ± 0,0019	0,0038 ± 0,0015
Moshpit				
zusammen	1,3467 ± 0,3353	3,6566 ± 1,1136	0,0047 ± 0,0013	0,0038 ± 0,0009
Moshpit	1,2422 ± 0,3900	4,0522 ± 1,2693	0,0078 ± 0,0035	0,0063 ± 0,0027

Tabelle A.10: RAFT: Entropie, Uniformität, Divergenz und Curl (Mittelwert ± Standardabweichung) pro Video

LITERATUR

- [1] *Loveparade*. URL: <https://www.stadtmuseum-duisburg.de/ausstellungen/erinnern/loveparade/> (besucht am 15.05.2026).
- [2] Kathrin Erdmann. "Schock nach verheerender Massenpanik im Zentrum von Seoul". In: (März 2026). URL: <https://www.tagesschau.de/ausland/asien/suedkorea-gedraenge-halloween-105.html> (besucht am 15.05.2026).
- [3] Khalifa AL-Dosari, Ziad Hunaiti und Wamadeva Balachandran. "Systematic Review on Civilian Drones in Safety and Security Applications". In: *Drones* 7.3 (2023). ISSN: 2504-446X. DOI: [10.3390/drones7030210](https://doi.org/10.3390/drones7030210). URL: <https://www.mdpi.com/2504-446X/7/3/210>.
- [4] Noof Ali. "A Review of Smart Drone Technologies for Security Surveillance and Search and Rescue". In: *Journal of Al-Qadisiyah for Computer Science and Mathematics* 17 (Dez. 2025). DOI: [10.29304/jqcs.2025.17.42549](https://doi.org/10.29304/jqcs.2025.17.42549).
- [5] Sara Khalid Alqahtani und Ghulam E Mustafa Abro. "Autonomous Drone-Based Border Surveillance Using Real-Time Object Detection with Yolo". In: *2025 IEEE 15th Symposium on Computer Applications in Industrial Electronics (ISCAIE)*. 2025, S. 564–569. DOI: [10.1109/ISCAIE64985.2025.11080940](https://doi.org/10.1109/ISCAIE64985.2025.11080940).
- [6] J. Colorado, C. Devia, M. Perez, I. Mondragon, D. Mendez und C. Parra. "Low-altitude autonomous drone navigation for landmine detection purposes". In: *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*. 2017, S. 540–546. DOI: [10.1109/ICUAS.2017.7991303](https://doi.org/10.1109/ICUAS.2017.7991303).
- [7] Muhammad Afif Husman, Waleed Albattah, Zulkifli Zainal Abidin, Yasir Mohd. Mustafah, Kushsairy Kadir, Shabana Habib, Muhammad Islam und Sheroz Khan. "Unmanned Aerial Vehicles for Crowd Monitoring and Analysis". In: *Electronics* 10.23 (2021). ISSN: 2079-9292. DOI: [10.3390/electronics10232974](https://doi.org/10.3390/electronics10232974). URL: <https://www.mdpi.com/2079-9292/10/23/2974>.
- [8] Asif Ali Laghari, Awais Khan Jumani, Rashid Ali Laghari, Hang Li, Shahid Karim und Abudllah Ayub Khan. "Unmanned aerial vehicles advances in object detection and communication security review". In: *Cognitive Robotics* 4 (Jan. 2024), S. 128–141. DOI: [10.1016/j.cogr.2024.07.002](https://doi.org/10.1016/j.cogr.2024.07.002). URL: <https://doi.org/10.1016/j.cogr.2024.07.002>.
- [9] I. Colomina und P. Molina. "Unmanned aerial systems for photogrammetry and remote sensing: A review". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 92 (Apr. 2014), S. 79–97. DOI: [10.1016/j.isprsjprs.2014.02.013](https://doi.org/10.1016/j.isprsjprs.2014.02.013). URL: <https://doi.org/10.1016/j.isprsjprs.2014.02.013>.

- [10] Wenyao Zhu und Ken Chen. "Real-time object detection for unmanned aerial vehicles based on vision transformer and edge computing". In: *Scientific Reports* 16.1 (Jan. 2026), S. 6814. DOI: [10.1038/s41598-026-37938-5](https://doi.org/10.1038/s41598-026-37938-5). URL: <https://doi.org/10.1038/s41598-026-37938-5>.
- [11] Andrea Alfarano, Luca Maiano, Lorenzo Papa und Irene Amerini. "Estimating optical flow: A comprehensive review of the state of the art". In: *Computer Vision and Image Understanding* 249 (Sep. 2024), S. 104160. DOI: [10.1016/j.cviu.2024.104160](https://doi.org/10.1016/j.cviu.2024.104160). URL: <https://doi.org/10.1016/j.cviu.2024.104160>.
- [12] Anton Milan, Laura Leal-Taixe, Ian Reid, Stefan Roth und Konrad Schindler. *MOT16: A Benchmark for Multi-Object Tracking*. 2016. arXiv: [1603.00831](https://arxiv.org/abs/1603.00831) [cs.CV]. URL: <https://arxiv.org/abs/1603.00831>.
- [13] Simon Baker, Daniel Scharstein, J. P. Lewis, Stefan Roth, Michael J. Black und Richard Szeliski. "A Database and Evaluation Methodology for Optical Flow". In: *International Journal of Computer Vision* 92.1 (Nov. 2010), S. 1–31. DOI: [10.1007/s11263-010-0390-2](https://doi.org/10.1007/s11263-010-0390-2). URL: <https://doi.org/10.1007/s11263-010-0390-2>.
- [14] Ibrahim Kajo, Aamir Saeed Malik und Nidal Kamel. "An evaluation of optical flow algorithms for crowd analytics in surveillance system". In: *Bd. 2016. Aug. 2016*, S. 1–6. DOI: [10.1109/iciias.2016.7824064](https://doi.org/10.1109/iciias.2016.7824064). URL: <https://doi.org/10.1109/iciias.2016.7824064>.
- [15] Bruce D. Lucas und Takeo Kanade. "An iterative image registration technique with an application to stereo vision". In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2. IJ-CAI'81*. Vancouver, BC, Canada: Morgan Kaufmann Publishers Inc., 1981, 674–679.
- [16] Gunnar Farnebäck. *Two-Frame Motion Estimation Based on Polynomial Expansion*. Jan. 2003, S. 363–370. DOI: [10.1007/3-540-45103-x_50](https://doi.org/10.1007/3-540-45103-x_50). URL: https://doi.org/10.1007/3-540-45103-x_50.
- [17] Zachary Teed und Jia Deng. *RAFT: Recurrent All-Pairs Field Transforms for Optical Flow*. Jan. 2020, S. 402–419. DOI: [10.1007/978-3-030-58536-5_24](https://doi.org/10.1007/978-3-030-58536-5_24). URL: https://doi.org/10.1007/978-3-030-58536-5_24.
- [18] Shashank Agnihotri, Julian Yuya Caspary, Luca Schwarz, Xinyan Gao, Jenny Schmalfluss, Andres Bruhn und Margret Keuper. "FlowBench: Benchmarking Optical Flow Estimation Methods for Reliability and Generalization". In: *Transactions on Machine Learning Research* (2025). ISSN: 2835-8856. URL: <https://openreview.net/forum?id=Kh4bj6YDNm>.
- [19] Berthold K.P. Horn und Brian G. Schunck. "Determining optical flow". In: *Artificial Intelligence* 17.1-3 (Aug. 1981), S. 185–203. DOI: [10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2). URL: [https://doi.org/10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2).

- [20] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers und Thomas Brox. "FlowNet: Learning Optical Flow with Convolutional Networks". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (Dez. 2015), S. 2758–2766. DOI: [10.1109/iccv.2015.316](https://doi.org/10.1109/iccv.2015.316). URL: <https://doi.org/10.1109/iccv.2015.316>.
- [21] Deqing Sun, Xiaodong Yang, Ming-Yu Liu und Jan Kautz. "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, S. 8934–8943. DOI: [10.1109/CVPR.2018.00931](https://doi.org/10.1109/CVPR.2018.00931).
- [22] Deqing Sun, Stefan Roth und Michael J. Black. "A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them". In: *International Journal of Computer Vision* 106.2 (Sep. 2013), S. 115–137. DOI: [10.1007/s11263-013-0644-x](https://doi.org/10.1007/s11263-013-0644-x). URL: <https://doi.org/10.1007/s11263-013-0644-x>.
- [23] Deqing Sun, Stefan Roth und Michael J. Black. "Secrets of optical flow estimation and their principles". In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Juni 2010), S. 2432–2439. DOI: [10.1109/cvpr.2010.5539939](https://doi.org/10.1109/cvpr.2010.5539939). URL: <https://doi.org/10.1109/cvpr.2010.5539939>.
- [24] Lifang Zhou und Zhen Hu. "Enhanced crowd counting with weighted attention network and multi-scale feature integration". In: *Image and Vision Computing* 163 (2025), S. 105750. ISSN: 0262-8856. DOI: <https://doi.org/10.1016/j.imavis.2025.105750>. URL: <https://www.sciencedirect.com/science/article/pii/S0262885625003385>.
- [25] Zhao Gang, Wang Xiaoli und Wang Lirong. "Motion Analysis and Research of Local Navigation System for Visual-Impaired Person Based on Improved LK Optical Flow". In: *2012 Fifth International Conference on Intelligent Networks and Intelligent Systems* (Nov. 2012), S. 348–351. DOI: [10.1109/icinis.2012.80](https://doi.org/10.1109/icinis.2012.80). URL: <https://doi.org/10.1109/icinis.2012.80>.
- [26] Berkan Solmaz, Brian E. Moore und Mubarak Shah. "Identifying Behaviors in Crowd Scenes Using Stability Analysis for Dynamical Systems". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.10 (Mai 2012), S. 2064–2070. DOI: [10.1109/tpami.2012.123](https://doi.org/10.1109/tpami.2012.123). URL: <https://doi.org/10.1109/tpami.2012.123>.
- [27] Shu Wang und Zhenjiang Miao. "Anomaly detection in crowd scene". In: *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS* (Okt. 2010), S. 1220–1223. DOI: [10.1109/icosp.2010.5655356](https://doi.org/10.1109/icosp.2010.5655356). URL: <https://doi.org/10.1109/icosp.2010.5655356>.
- [28] Hannah M. Dee und Alice Caplier. "Crowd behaviour analysis using histograms of motion direction". In: *2010 IEEE International Conference on Image Processing* (Sep. 2010), S. 1545–1548. DOI: [10.1109/icip.2010.5653573](https://doi.org/10.1109/icip.2010.5653573). URL: <https://doi.org/10.1109/icip.2010.5653573>.

- [29] Zdenek Kalal, Krystian Mikolajczyk und Jiri Matas. "Forward-Backward Error: Automatic Detection of Tracking Failures". In: *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)* (Aug. 2010), S. 2756–2759. DOI: [10.1109/icpr.2010.675](https://doi.org/10.1109/icpr.2010.675). URL: <https://doi.org/10.1109/icpr.2010.675>.
- [30] Jisoo Jeong, Jamie Menjay Lin, Fatih Porikli und Nojun Kwak. "Imposing Consistency for Optical Flow Estimation". In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Juni 2022), S. 3171–3181. DOI: [10.1109/cvpr52688.2022.00318](https://doi.org/10.1109/cvpr52688.2022.00318). URL: <https://doi.org/10.1109/cvpr52688.2022.00318>.
- [31] J. L. Barron, D. J. Fleet und S. S. Beauchemin. "Performance of optical flow techniques". In: *International Journal of Computer Vision* 12.1 (Feb. 1994), S. 43–77. DOI: [10.1007/bf01420984](https://doi.org/10.1007/bf01420984). URL: <https://doi.org/10.1007/bf01420984>.
- [32] S. S. Beauchemin und J. L. Barron. "The computation of optical flow". In: *ACM Computing Surveys* 27.3 (Sep. 1995), S. 433–466. DOI: [10.1145/212094.212141](https://doi.org/10.1145/212094.212141). URL: <https://doi.org/10.1145/212094.212141>.
- [33] Francesco Nex und Fabio Remondino. "UAV for 3D mapping applications: a review". In: *Applied Geomatics* 6.1 (Nov. 2013), S. 1–15. DOI: [10.1007/s12518-013-0120-x](https://doi.org/10.1007/s12518-013-0120-x). URL: <https://doi.org/10.1007/s12518-013-0120-x>.
- [34] D. Fleet und Y. Weiss. *Optical Flow Estimation*. Dez. 2005, S. 237–257. DOI: [10.1007/0-387-28831-7_15](https://doi.org/10.1007/0-387-28831-7_15). URL: https://doi.org/10.1007/0-387-28831-7_15.
- [35] Ronny Hänsch und Olaf Hellwich. "The Truth About Ground Truth: Label Noise in Human-Generated Reference Data". In: *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. 2019, S. 5594–5597. DOI: [10.1109/IGARSS.2019.8898003](https://doi.org/10.1109/IGARSS.2019.8898003).
- [36] Mingliang Xu, Zhaoyang Ge, Xiaoheng Jiang, Gaoge Cui, Pei Lv, Bing Zhou und Changsheng Xu. "Depth Information Guided Crowd Counting for complex crowd scenes". In: *Pattern Recognition Letters* 125 (Feb. 2019), S. 563–569. DOI: [10.1016/j.patrec.2019.02.026](https://doi.org/10.1016/j.patrec.2019.02.026). URL: <https://doi.org/10.1016/j.patrec.2019.02.026>.
- [37] Feng Liu und Michael Gleicher. "Learning color and locality cues for moving object detection and segmentation". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, S. 320–327. DOI: [10.1109/CVPR.2009.5206857](https://doi.org/10.1109/CVPR.2009.5206857).
- [38] Zhongyu Chen, Rong Zhao, Xindong Guo, Jianbin Xie und Xie Han. "Moving Object Detection in Freely Moving Camera via Global Motion Compensation and Local Spatial Information Fusion". In: *Sensors* 24.9 (2024). ISSN: 1424-8220. DOI: [10.3390/s24092859](https://doi.org/10.3390/s24092859). URL: <https://www.mdpi.com/1424-8220/24/9/2859>.

- [39] Mohammadreza Babaei, Duc Tung Dinh und Gerhard Rigoll. "A deep convolutional neural network for video sequence background subtraction". In: *Pattern Recognition* 76 (2018), S. 635–649. ISSN: 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2017.09.040>. URL: <https://www.sciencedirect.com/science/article/pii/S0031320317303928>.
- [40] Jhony Giraldo, Ton Le und Thierry Bouwmans. "Deep Learning based Background Subtraction: A Systematic Survey". In: März 2020.
- [41] B. Ganga, Lata B.T. und Venugopal K.R. "Object detection and crowd analysis using deep learning techniques: Comprehensive review and future directions". In: *Neurocomputing* 597 (2024), S. 127932. ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2024.127932>. URL: <https://www.sciencedirect.com/science/article/pii/S0925231224007033>.
- [42] Joseph Redmon, Santosh Divvala, Ross Girshick und Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
- [43] Simon Baker und Iain Matthews. "Lucas-Kanade 20 Years On: A Unifying Framework". In: *International Journal of Computer Vision* 56.3 (Jan. 2004), S. 221–255. DOI: [10.1023/b:visi.0000011205.11775.fd](https://doi.org/10.1023/b:visi.0000011205.11775.fd). URL: <https://doi.org/10.1023/b:visi.0000011205.11775.fd>.
- [44] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy und Thomas Brox. "FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Juli 2017), S. 1647–1655. DOI: [10.1109/cvpr.2017.179](https://doi.org/10.1109/cvpr.2017.179). URL: <https://doi.org/10.1109/cvpr.2017.179>.
- [45] Lahav Lipson, Zachary Teed und Jia Deng. "RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching". In: *2021 International Conference on 3D Vision (3DV)*. 2021, S. 218–227. DOI: [10.1109/3DV53792.2021.00032](https://doi.org/10.1109/3DV53792.2021.00032).
- [46] Jean-Yves Bouguet. *Pyramidal Implementation of the Lucas Kanade Feature Tracker: Description of the Algorithm*. Techn. Ber. Intel Corporation, Microprocessor Research Labs, Mai 2001. URL: <https://graphics.stanford.edu/courses/cs448a-00-fall/bouget00.pdf>.
- [47] Jesse L. Silverberg, Matthew Bierbaum, James P. Sethna und Itai Cohen. "Collective Motion of Humans in Mosh and Circle Pits at Heavy Metal Concerts". In: *Physical Review Letters* 110.22 (Mai 2013). ISSN: 1079-7114. DOI: [10.1103/PhysRevLett.110.228701](https://doi.org/10.1103/PhysRevLett.110.228701). URL: <http://dx.doi.org/10.1103/PhysRevLett.110.228701>.
- [48] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley und Michael J. Black. "A Naturalistic Open Source Movie for Optical Flow Evaluation". In: *Computer Vision – ECCV 2012*. Hrsg. von Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato und Cordelia Schmid. Berlin,

- Heidelberg: Springer Berlin Heidelberg, 2012, S. 611–625. ISBN: 978-3-642-33783-3.
- [49] C. Harris und M. Stephens. “A Combined Corner and Edge Detector”. In: *Proceedings of the 4th Alvey Vision Conference* (Jan. 1988), S. 147–151. DOI: [10.5244/c.2.23](https://doi.org/10.5244/c.2.23). URL: <https://doi.org/10.5244/c.2.23>.
 - [50] David G. Lowe. “Distinctive Image Features from Scale-Invariant Key-points”. In: *International Journal of Computer Vision* 60.2 (Juni 2004), S. 91–110. DOI: [10.1023/b:visi.0000029664.99615.94](https://doi.org/10.1023/b:visi.0000029664.99615.94). URL: <https://doi.org/10.1023/b:visi.0000029664.99615.94>.
 - [51] Xuxin Gu, Jinrong Cui und Qi Zhu. “Abnormal crowd behavior detection by using the particle entropy”. In: *Optik* 125.14 (Apr. 2014), S. 3428–3433. DOI: [10.1016/j.ijleo.2014.01.041](https://doi.org/10.1016/j.ijleo.2014.01.041). URL: <https://doi.org/10.1016/j.ijleo.2014.01.041>.
 - [52] Saad Ali und Mubarak Shah. “A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis”. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition* (Juni 2007), S. 1–6. DOI: [10.1109/cvpr.2007.382977](https://doi.org/10.1109/cvpr.2007.382977). URL: <https://doi.org/10.1109/cvpr.2007.382977>.
 - [53] Thomas Brox, Andrés Bruhn, Nils Papenberg und Joachim Weickert. *High Accuracy Optical Flow Estimation Based on a Theory for Warping*. Jan. 2004, S. 25–36. DOI: [10.1007/978-3-540-24673-2_{_}3](https://doi.org/10.1007/978-3-540-24673-2_{_}3). URL: https://doi.org/10.1007/978-3-540-24673-2_3.
 - [54] Jan Kybic und Claudia Nieuwenhuis. “Bootstrap optical flow confidence and uncertainty measure”. In: *Computer Vision and Image Understanding* 115.10 (Juli 2011), S. 1449–1462. DOI: [10.1016/j.cviu.2011.06.008](https://doi.org/10.1016/j.cviu.2011.06.008). URL: <https://doi.org/10.1016/j.cviu.2011.06.008>.
 - [55] Anne S. Wannenwetsch, Margret Keuper und Stefan Roth. “ProbFlow: Joint Optical Flow and Uncertainty Estimation”. In: *2017 IEEE International Conference on Computer Vision (ICCV)* (Okt. 2017), S. 1182–1191. DOI: [10.1109/iccv.2017.133](https://doi.org/10.1109/iccv.2017.133). URL: <https://doi.org/10.1109/iccv.2017.133>.
 - [56] Li Xu, Jianing Chen und Jiaya Jia. *A Segmentation Based Variational Model for Accurate Optical Flow Estimation*. Jan. 2008, S. 671–684. DOI: [10.1007/978-3-540-88682-2_{_}51](https://doi.org/10.1007/978-3-540-88682-2_{_}51). URL: https://doi.org/10.1007/978-3-540-88682-2_51.
 - [57] Xuguang Zhang, Xiaohu Shu und Zhen He. “Crowd panic state detection using entropy of the distribution of enthalpy”. In: *Physica A Statistical Mechanics and its Applications* 525 (Apr. 2019), S. 935–945. DOI: [10.1016/j.physa.2019.04.033](https://doi.org/10.1016/j.physa.2019.04.033). URL: <https://doi.org/10.1016/j.physa.2019.04.033>.

- [58] Sourabh Choudhary, Chunqiang Hu, Syed Murtoza Mushrul Pasha, Md Tanvir Islam, Rashedin Islam und Arnob Barua Himo. *An Early Detection of Risky Crowd Dynamics Scheme Based on Motion Entropy and Scene Semantics*. Jan. 2026, S. 215–234. DOI: [10.1007/978-981-95-6203-9_{_}12](https://doi.org/10.1007/978-981-95-6203-9_{_}12). URL: https://doi.org/10.1007/978-981-95-6203-9_12.
- [59] Ali Abid Hussan Altalbi, Shaimaa Hameed Shaker und Akbas Ezaldeen Ali. “Localization of Strangeness for Real Time Video in Crowd Activity Using Optical Flow and Entropy”. In: *International Journal of Online and Biomedical Engineering (iJOE)* 19.07 (Juni 2023), S. 52–68. DOI: [10.3991/ijoe.v19i07.38869](https://doi.org/10.3991/ijoe.v19i07.38869). URL: <https://doi.org/10.3991/ijoe.v19i07.38869>.
- [60] Khaled M. Abdelwahab, Fathi E. Abd El-Samie, Ashraf A. M. Khalaf, Ahmed M. ElSawy und Wafaa A. Shalaby. “Optical flow techniques for crowd behavior analysis”. In: *Journal of Optics* (Feb. 2026). DOI: [10.1007/s12596-025-02938-7](https://doi.org/10.1007/s12596-025-02938-7). URL: <https://doi.org/10.1007/s12596-025-02938-7>.
- [61] Huafeng Chen, Angelina Pashkevich, Shiping Ye, Rykhard Bohush und Sergey Ablameyko. “Crowd Movement Type Estimation in Video by Integral Optical Flow and Convolution Neural Network”. In: *Pattern Recognition and Image Analysis* 34.2 (Juni 2024), S. 266–274. DOI: [10.1134/s1054661824700068](http://dx.doi.org/10.1134/s1054661824700068). URL: <http://dx.doi.org/10.1134/s1054661824700068>.
- [62] Wei Ju, Yiyang Dong, Junlin Zhu, Zhen Tian und Qinghao Li. *Research on Real-Time Vehicle Detection and Tracking Algorithm Based on Dense Optical Flow*. Juni 2025. DOI: [10.3233/atde250249](https://doi.org/10.3233/atde250249). URL: <https://doi.org/10.3233/atde250249>.
- [63] Haiyang Qiu, Xu Zhang, Hui Wang, Dan Xiang, Mingming Xiao, Zhiyu Zhu und Lei Wang. “A Robust and Integrated Visual Odometry Framework Exploiting the Optical Flow and Feature Point Method”. In: *Sensors* 23.20 (Okt. 2023), S. 8655. DOI: [10.3390/s23208655](https://doi.org/10.3390/s23208655). URL: <https://doi.org/10.3390/s23208655>.