

Motivation

Die Motivation der Arbeit liegt in der zunehmenden Bedeutung automatisierter Drohnenüberwachung zur frühzeitigen Erkennung kritischer Crowd-Situationen. Die Relevanz dieses Forschungsfeldes wird eindrücklich durch reale Ereignisse verdeutlicht: Bei der Loveparade 2010 in Duisburg starben 21 Menschen und es gab Hunderte Verletzte, weil Engpässe und eine unzureichende Überwachung zu einer Massenpanik führten. Ein weiteres tragisches Beispiel ist die Massenpanik bei einer Halloween-Veranstaltung in Seoul 2022, bei der mehr als 150 Menschen ihr Leben verloren; auch hier zeigte sich, dass große, dicht gedrängte Menschenmengen in Echtzeit zuverlässig beobachtet und gesteuert werden müssen. Solche Vorfälle illustrieren die potenziell verheerenden Folgen von Versäumnissen im Crowd-Management und unterstreichen den dringenden Bedarf an automatisierten, Echtzeit-fähigen Analysewerkzeugen.

Bewegungsvektoren, abgeleitet aus dem optischen Fluss, liefern quantitative Aussagen zu Richtung, Geschwindigkeit und Dynamik von Personenströmen und bilden die Grundlage für weiterführende Analysen. Angesichts der Herausforderungen durch Perspektivwechsel, Parallaxen, Verdeckungen und variierende Beleuchtung in Drohnenaufnahmen ist es wichtig zu verstehen, welche optischen Fluss-Methoden unter realen Drohneneinsatzbedingungen die besten Eigenschaften in Bezug auf Genauigkeit, Robustheit und Rechenzeit zeigen. Die zentrale Forschungsfrage lautet daher: 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?

Methodik

Datenbasis und Vorverarbeitung

Für die Untersuchung wurden sieben Drohnenvideos aus dem JuicyBeats Festival 2024 in Dortmund verwendet. Diese Videos decken verschiedene Szenen ab, von übersichtlichen Aufnahmen Einzelner bis zu dichten Crowd-Szenen einschließlich Moshpits. Die sieben Videos lassen sich somit in zwei Kategorien der Szenenkomplexität einteilen: einfache Szenen, in denen wenige Personen klare Bewegungsmuster zeigen, und komplexe Szenen mit dichter Crowd, vielen Interaktionen und Verdeckungen.

Die Videos wurden auf eine einheitliche räumliche Auflösung (1280x720) und eine einheitliche Bildrate gebracht, um eine faire Vergleichbarkeit der Methoden sicherzustellen. Da Ground-Truth-Daten in den Rohdaten nicht vorhanden waren, wurde ein eigenes Ground-Truth-Dataset erstellt: Ein in Python entwickeltes Annotationstool ermöglicht das manuelle Markieren korrespondierender Punkte in aufeinanderfolgenden Frames. Es wurden insgesamt 3108 Punktpaare auf 624 Frame-Paaren annotiert. Diese Annotationen dienen als Ground Truth zur Berechnung von Endpunktfehlern und Winkelfehlern.

Methoden

Es wurden drei Methoden verglichen: Lucas-Kanade (LK) als sparsamer, klassischer optischer Fluss; Farneback (FB) als dichter-optischer Fluss mit polynomialer Approximation; und RAFT als modernes Deep-Learning-Modell für dichte Flussfelder. LK basiert auf lokalen Gradienteneinschätzungen und der Annahme der Helligkeitskonsistenz innerhalb eines kleinen Fensters; FB nutzt eine polynomiale Modellierung zur Erfassung dichter Flussfelder; RAFT kombiniert Merkmalsextraktion, All-Pairs-Korrelationsvolumen und eine iterative Flussverfeinerung

durch ein rekurrentes Modul. Ground-Truth-Annotationen wurden genutzt, um eine objektive Beurteilung der Genauigkeit zu ermöglichen. Die Evaluierung erfolgte unter identischen Bedingungen, wodurch sich die Stärken und Schwächen der Methoden systematisch gegenüberstellen lassen.

Prototypische Pipeline

Die Implementierung folgt einer prototypischen Pipeline, in der zunächst markante Bildpunkte mittels Shi-Tomasi-Detektion identifiziert werden. Anschließend werden Bewegungsvektoren wie folgt berechnet: LK wird als sparser Fluss mit cv2.calcOpticalFlowPyrLK auf die markierten Punkte angewendet; Farneback erzeugt ein dichtes Flussfeld durch cv2.calcOpticalFlowFarneback und erlaubt eine punktbasierte Sampling-Strategie an den identischen Punkten; RAFT wird als vortrainiertes Deep-Learning-Modell (KITTI-Domäne) auf RGB-Eingaben angewendet, wobei das Sampling ebenfalls an den identischen Punkten erfolgt. Die Ergebnisse werden qualitäts- und zeitbasiert ausgewertet. Die Laufzeitmessung erfolgt pro Frame und berücksichtigt GPU-Beschleunigung (RTX 2070 Super). Die Ergebnisse werden in CSV- und JSON-Formaten gespeichert und zusätzlich visualisiert, beispielsweise über Quiver-Plots, Boxplots oder Heatmaps.

Experimente

Es wurden fünf Experimente durchgeführt: Erstens eine Evaluation gegen Ground-Truth-Daten zur quantitativen Bestimmung von Endpunktfehler (EPE) und Winkelfehlern; zweitens ein quantitativer Methodenvergleich ohne Ground-Truth-Daten zur Beurteilung der gegenseitigen Konsistenz und Unterschiede der Methoden; drittens eine Qualitätsbewertung mittels Forward-Backward-Consistency als interne Plausibilitätsprüfung in Abwesenheit externer Ground-Truth; viertens eine Crowd-Dynamik-Analyse zur Charakterisierung von Richtungsentropie, Richtungs-Uniformität, Geschwindigkeit, Divergenz und Curl; fünftens ein Laufzeitvergleich zur Einschätzung der Echtzeitfähigkeit und Hardware-Abhängigkeit der Methoden. Zur Stützung der Ergebnisse werden zusätzlich qualitative Visualisierungen eingesetzt, darunter Heatmaps der Crowd-Dynamik und Amplitude-Heatmaps der Geschwindigkeitsverteilung.

Ergebnisse

Die quantitative Evaluation gegen Ground-Truth-Daten zeigt, dass Lucas-Kanade und Farneback in der Gesamtbewertung robuster und genauer sind als RAFT. Der mittlere Endpunktfehler liegt für LK bei etwa 1,29 Pixel, für FB bei 1,32 Pixel, während RAFT einen höheren Wert von ca. 2,09 Pixel erreicht. Die Winkelfehler liegen bei LK und FB im Durchschnitt um die 30°, während RAFT hier schlechtere Werte mit 63° zeigt. Die Verteilungen der Fehler deuten darauf hin, dass Lucas-Kanade und Farneback sowohl konsistent als auch zuverlässig arbeiten, während RAFT in komplexeren Szenen stärker variiert. In Bezug auf die Szenenkomplexität zeigen Lucas-Kanade und Farneback in einfachen Szenen vergleichsweise geringe Fehler, während RAFT auch dort größere Abweichungen aufweist; in komplexen Szenen bleiben Lucas-Kanade und Farneback tendenziell robuster, RAFT zeigt dagegen größere Streuung.

Methode	EPE Mean (px)	EPE Median (px)	EPE Max (px)	Winkel (Grad, Mean)
Lucas-Kanade (LK)	1,29 ± 0,68	1,08	2,49	30,0 ± 28,3
Farneback	1,32 ± 0,68	1,14	2,44	29,7 ± 28,3
RAFT	2,09 ± 1,36	1,96	3,44	63,0 ± 37,5

Tabelle 1: Ergebnisse gegen Ground-Truth

In Bezug auf die weiteren Analysen ist die Forward-Backward-Consistency nahezu durchgehend hoch bei Lucas-Kanade und Farneback, während RAFT in komplexen Szenen gelegentlich Abweichungen zeigt. Die räumlichen Konsistenzkarten verdeutlichen, dass Lucas-Kanade und Farneback nahezu vollständige Konsistenz liefern, RAFT hingegen vereinzelt inkonsistente Bereiche zeigt, besonders in dynamischen Szenen. Die Crowd-Dynamik-Analyse zeigt, dass Farneback stärker die Richtungsvielfalt in dichten Crowd-Szenen erfasst, während RAFT dominante Bewegungsrichtungen robuster abbildet, aber feinere Richtungswechsel glättet. Die Divergenz- und Curl-Werte sind bei Farneback oft höher als bei Lucas-Kanade bzw. RAFT. Hinsichtlich der Laufzeit ist Lucas-Kanade das einzige Verfahren, das Echtzeit-Fähigkeiten zeigt (ungefähr 7 ms pro Frame), Farneback und RAFT benötigen deutlich mehr Zeit (im Mittel ca. 623 ms bzw. 1540 ms pro Frame) und sind somit auf der getesteten Hardware nicht für Echtzeit-Anwendungen geeignet. Insgesamt ergibt sich damit eine klare Empfehlung: Für Echtzeit-Anwendungen oder bei eingeschränkten Ressourcen ist Lucas-Kanade die bevorzugte Wahl; Farneback eignet sich gut für detaillierte Crowd-Analysen; RAFT kann als Ergänzung für Offline-Evaluationen oder domänenspezifisches Fine-Tuning eingesetzt werden.

Fazit und Ausblick

Zusammenfassend liefert die Arbeit eine fundierte, replizierbare Vergleichsbasis für Optical-Flow-Methoden in Drohnenvideos. Zwei klassische Methoden (Lucas-Kanade und Farneback) und ein modernes Deep-Learning-Verfahren (RAFT) wurden hinsichtlich Genauigkeit, Robustheit, Crowd-Dynamik-Erfassung und Laufzeit bewertet. Die Ergebnisse zeigen, dass klassische Methoden in den betrachteten Drohnenszenarien robust und genau sind, insbesondere wenn Rechenzeit eine wesentliche Rolle spielt. RAFT bietet Potenzial erfordert aber Domänenanpassung (Fine-Tuning) und signifikante Rechenressourcen. Auf Basis dieser Ergebnisse lassen sich praktikable Empfehlungen ableiten: Für Echtzeit-Analysen und begrenzte Ressourcen empfiehlt sich Lucas-Kanade; Farneback eignet sich für detailliertere Crowd-Analysen; RAFT kann alternativ oder ergänzend für Offline-Bewertungen oder domänen-spezifisches Fine-Tuning genutzt werden. Der Ausblick betont die Notwendigkeit größerer Ground-Truth-Datensätze, domänenspezifischer Trainingsansätze für Deep-Learning-Modelle sowie Optimierungen für Edge- und Echtzeit-Anwendungen.