



Fakultät für Informatik

Bildverstehen und wissensbasierte Systeme

Kameraparameter-Nachführung durch natürliche
Landmarken in Sequenzen monokularer Bilder am Beispiel
von Fußballübertragungen mit Anwendungen zu
automatischer Ballbesitz- und Spielereigniserkennung

Martin Hörnig

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität
München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. F. Matthes

Prüfer der Dissertation:

1. Prof. Dr. B. Radig (i. R.)
2. Prof. Dr. D. Burschka

Die Dissertation wurde am 21.12.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 13.05.2016 angenommen.

Zusammenfassung

Eine RGB-Kamera bildet Objekte der Anschauungswelt auf eine zweidimensionale Bildebene ab. Explizite Daten zur Entfernung der Bildobjekte zur Kamera stehen dabei nicht zur Verfügung. Sollen Bildinhalte, etwa für nachfolgende Verarbeitungsschritte, wieder dreidimensionalen Weltkoordinaten zugeordnet werden und soll dies unabhängig von Standpunkt, Blickwinkel und Brennweite der Kamera automatisch erfolgen, so bleibt nur die Möglichkeit, die Weltkoordinaten aus dem Bildinhalt abzuleiten. Da es für Aufnahmen, welche aus der Vergangenheit stammen, nicht möglich ist, nachträglich künstliche Landmarken einzufügen und es nicht immer realistisch oder praktikabel ist, mit künstlichen Landmarken zu arbeiten, können nur natürliche Landmarken in Betracht gezogen werden.

In dieser Dissertation wird ein Ansatz vorgestellt, welcher das schwierige Problem der Bereitstellung von Kameraparametern zu Bildsequenzen für die Berechnung von Weltkoordinaten aufgreift. Änderungen der externen und internen Kameraparameter im Verlauf der Bildsequenz werden im Verfahren berücksichtigt, wodurch auch schwenkende und zoomende Kameras Verwendung finden können. Die Kalibrierung geschieht auf der Basis einer relevanten Umgebung (ortsfeste Landmarken) in ihrer dreidimensionalen Anordnung, für welche ein Modell *a priori* festzulegen ist. Das vorgestellte Verfahren ist in der Lage, Probleme nicht-ortsfester Landmarken zu handhaben und gleichzeitig Fälle zu bewältigen, in welchen bezogen auf ein einzelnes Bild nur kleine Teile der im Modell hinterlegten Landmarken durch die Kamera abgebildet werden. Der Ansatz erlaubt weitestgehend ohne manuelle Eingriffe die Berechnung der notwendigen Parameter für die Abbildung vom Zweidimensionalen aufs Dreidimensionale in über 500 000 Bildern eines anspruchsvollen Bildmaterials. Das dafür ausgewählte Anwendungsgebiet waren Übertragungen von Fußballspielen. Für die Bearbeitung von TV-Material wird eine Erkennung von Schnitten durch Kamerawechsel und eine Zuordnung von Bildern zu den Einstellungen einer Kamera vorgestellt.

Mit der Verfügbarkeit von kameraunabhängigen Weltkoordinaten für Fußballspieler und Ball werden Spielereignisse (wie Pässe, Zweikämpfe, Einwürfe, Frei-, Eck-, Ab- und Anstöße) automatisch erkannt. Dazu werden nicht nur Ergebnisse des zuvor geschilderten Verfahrens genutzt, sondern auch fremde Quellen herangezogen, um eine große Statistik für die Evaluation bereitzustellen. Die Grundlage für die Ereigniserkennung ist ein individueller Ballbesitz. Dieser wird hinsichtlich wesentlicher taktischer Einflussgrade differenziert. Kann ein Spieler in der Situation eines Ballbesitzes eine Wahl zwischen verschiedenen Handlungsalternativen treffen, besitzt er Kontrolle über den Ball. Basierend auf der Ballbesitz-Differenzierung werden taktische Konstrukte aufgestellt. Zusammen mit Verknüpfungen mit den zu ihnen im Bezug stehenden Weltkoordinaten (z. B. des ballführenden Spielers) werden bisher nicht dagewesene Einblicke zur leistungstaktischen Charakterisierung von Spielern und Begegnungen ermöglicht.

Abstract

An RGB-camera maps objects from the perceived world onto a two-dimensional image plane, although explicit data concerning distance between image objects and the camera are not available. The only possible way to match image contents with three-dimensional world coordinates automatically and independently of the position, perspective, and the focal length of the camera, i.e. in preparation for later processing steps, is to derive the world coordinates from the image contents. As it is not possible to insert artificial landmarks into pictures taken in the past, and as working with artificial landmarks does not always seem feasible or realistic, only natural landmarks can be considered.

This doctoral thesis presents an approach to the difficult problem of providing camera parameters to image sequences for the calculation of world coordinates. Changes of external and internal camera parameters during an image sequence are taken into account, enabling the usage of pan-tilt-zoom cameras. The calibration is based on a relevant surrounding (stationary landmarks) in its three-dimensional array, for which a model needs to be determined *a priori*. The algorithm being presented is able to handle problems of non-stationary landmarks and simultaneously solve cases, in which merely small parts of landmarks embedded in the model are depicted in a single camera take. This approach allowed the calculation of necessary parameters for the mapping of two-dimensional onto three-dimensional spaces in more than 500 000 images of a challenging test set, which to the greatest extent proceeded without manual intervention. The field of application chosen was the broadcasting of soccer games. Regarding TV material, which involves changes of camera perspective, this thesis introduces a detector for cuts and an algorithm assigning images to camera shots.

Given the availability of camera-independent world coordinates of soccer players and the ball, events (such as passes, tackles, and throw-ins as well as corner kicks, free kicks, goal kicks, and kickoffs) are detected automatically. Besides the events extracted using the method previously described, external sources are adduced in order to provide large-scale statistics for evaluation. The foundation of the event detection is an individual ball possession, differentiated with respect to significant tactical degrees of influence. A player having ball possession and being able to choose between various actions to take also possesses ball control. Based on the differentiation of ball possession, tactical constructs are being raised. Along with the connections to related world coordinates (i.e. of the player controlling the ball) new, unprecedented insights concerning performance-tactical characteristics of players and encounters are realizable.

Danksagung

Herr Prof. Dr. Bernd Radig ermöglichte mir, durch die Überlassung dieses hochinteressanten Themas in seiner Forschungsgruppe Bildverstehen und wissensbasierte Systeme wertvolle Erfahrungen zu sammeln. Für seine Anleitung und Unterstützung bedanke ich mich besonders. Weiterhin gilt ein herzlicher Dank auch meinen Kollegen — insbesondere Dr. Christoph Mayer, Michael Herrmann und Andreas Bigontina — für das angenehme Arbeitsklima, die Stunden fachlicher Diskussionen und viele nützliche Hinweise.

Inhaltsverzeichnis

Zusammenfassung	iii
Danksagung	v
Inhaltsverzeichnis	vii
Notation und Nomenklatur	xi
1 Einführung	1
1.1 Gegenstand der Dissertation	1
1.2 Berücksichtigte Datentypen und Kamerasysteme	2
1.3 Zielsetzung	4
1.4 Stand der Forschung	6
2 Aufbau und Überblick	7
2.1 Einleitung	7
2.2 Ein System zur Fußballanalyse	7
2.3 Vorausgegangene Veröffentlichungen und wissenschaftlicher Beitrag	10
2.4 Implementation als Client-Server-Architektur	12
3 Pixelweise Spielfeldanalyse	17
3.1 Einleitung	17
3.2 Clipping eines konvexen Vierecks in einem Rechteckfenster	19
3.3 Berechnung der Spielfeldhülle	20
3.4 Berechnung der Grasmasken	26
3.5 Ergänzungen zur Grasmasken	30
3.5.1 Kontinuierliche Wertebereiche	30
3.5.2 Adaptive Schwellwerte	31
3.6 Stand der Forschung	32
3.7 Evaluation	33
3.8 Anwendungen	36
3.9 Diskussion und Ausblick	37
4 Erkennung und Klassifikation von Schnitten	39
4.1 Einleitung	39
4.2 Das System zur Klassifikation von Einstellungen	41

4.3	Erkennung von Schnitten	43
4.3.1	Merkmalsgenerierung	43
4.3.2	Kalkulation von Distanzen zwischen Bildern	44
4.3.3	Berechnung von Distanzen für mehrere Zeitachsen	45
4.3.4	Erkennung harter Schnitte	46
4.3.5	Erkennung weicher Schnitte	47
4.3.6	Behandlung von Wischblenden und Blitzlichtern	50
4.4	Berechnung von Schnitten aus Bildklassen	51
4.5	Schnitt-Klassifikation	52
4.6	Stand der Forschung	53
4.7	Evaluation	54
4.7.1	Evaluation der Schnitt-Erkennung	54
4.7.2	Evaluation der Schnitt-Klassifikation	56
4.8	Diskussion und Ausblick	57
5	Erkennung von Geradensegmenten	59
5.1	Einleitung	59
5.2	Bestimmung der Geradheit von Punktfolgen	61
5.3	Bestimmung eines adaptiven Schwellwerts	63
5.4	Erkennung von gerader Punktfolgen	68
5.5	Algorithmische Betrachtung	71
5.6	Stand der Forschung	72
5.7	Evaluation	74
5.8	Diskussion und Ausblick	78
6	Kameraparameter-Nachführung	79
6.1	Einleitung	79
6.2	Grundlagen zur projektiven Geometrie	80
6.3	Quasi-affine Homographien	83
6.4	Das Spielfeld-Modell	86
6.5	Automatische Zuordnung zwischen Bild und Spielfeldmodell	89
6.6	Quasi-affine Kameraparameter-Nachführung	93
6.7	Kombination der Startposition	97
6.8	Evaluation	99
6.8.1	Evaluation auf Basis einer simulierten Kamerabewegung	100
6.8.2	Evaluation auf Basis realer Kamerabegegnungen	104
6.8.3	Evaluation am Beispiel eines Fußgängerüberwegs	109
6.9	Stand der Forschung	111
6.10	Diskussion und Ausblick	114
7	Spieler- und Ballbehandlung	117
7.1	Einleitung	117
7.2	Monokulare Spieler- und Ball-Nachführung	118
7.2.1	Mehrspieler-Nachführung	118
7.2.2	Ballerkennung	120
7.2.3	Ausblick	122
7.3	Spieler-Nachführung in CIP-Systemen	122

7.3.1	Analyse eines exemplarischen Datensatzes	123
7.3.1.1	Fehlerbetrachtung des Running Flags	123
7.3.1.2	Fehlerbetrachtung der Spieler- und Ball-Nachführung	126
7.4	Evaluation des monokularen Systems	127
7.5	Stochastische Glättung	131
7.6	Diskussion und Ausblick	134
8	Ballbesitz und Spielereignisse	135
8.1	Einleitung	135
8.2	Stand der Forschung	136
8.3	Einführung spielerbezogener Ballbesitzmodelle	137
8.3.1	Ball Possession	139
8.3.2	Ball Action	141
8.3.3	Ball Control	143
8.3.4	Berechnung der Ballbesitz-Zeiträume	144
8.3.5	Mannschaftsballbesitz	144
8.3.6	Evaluation und Training	146
8.3.6.1	Bestimmung der Modellparameter	147
8.3.6.2	Evaluation von Ball Possession und Ball Action	147
8.3.6.3	Evaluation der Ball Control	149
8.3.7	Vorstellung exemplarischer Anwendungen	150
8.4	Erkennen von Spielereignissen	154
8.5	Diskussion und Ausblick	157
9	Abschließende Diskussion	159
A	Kapitelresulte zur Kameraparameter-Nachführung für Bildbeispiele	163
B	Einzelbilderergebnisse zur Kameraparameter-Nachführung	167
	Quellenverzeichnis	173

Notation und Nomenklatur

$\mathbf{x} = (x_1, \dots, x_n)^T$	(Spalten-)Vektor
\mathbf{A}, \mathbf{B}	Matrix, Menge
$\mathbf{x}_{1:n} = (x_1, \dots, x_n)$	Tupel, Folge
$\ \mathbf{x}\ _p, \ \mathbf{A}\ _p$	p-Norm
$\text{atan2}(x, y)$	$:= \begin{cases} \arctan \frac{y}{x} & \text{für } x > 0 \\ \arctan \frac{y}{x} + \pi & \text{für } x < 0, y \geq 0 \\ \arctan \frac{y}{x} - \pi & \text{für } x < 0, y < 0 \\ +\pi/2 & \text{für } x = 0, y > 0 \\ -\pi/2 & \text{für } x = 0, y < 0 \\ 0 & \text{für } x = 0, y = 0 \end{cases}$
#	(A) Nachgestellt: Nummerierung (B) Vorangestellt: Kardinalitätsoperator (card) z. B.: (A) „Bild # i “ für i -tes Bild, (B) # \mathbf{B}
[s], [m]	physikalische Größen (SI)
[m:s]	Minuten : Sekunden-Formatierung
[0; 1]	Intervall
$\text{Cov}_{\mathbf{x} \in \mathbf{B}}$	Kovarianzmatrix der $\mathbf{x} \in \mathbf{B}$
fps	Bilder pro Sekunde
$\text{RANSAC}(\phi(x \theta \in \Theta), \Theta, \mathbf{x}_{1:n}, \mathbf{y}_{1:n})$	Berechnung einer robusten Schätzung der Funktionsparameter für ϕ aus Θ in Annäherung der Punkte (x_j, y_j) , $j = 1, \dots, n$ durch RANSAC [Fischler u. Bolles, 1981]
Precision	Relation korrekter Messungen zur Gesamtzahl an Messungen
Recall	Relation korrekter Messungen zur Gesamtzahl relevanter Entitäten
F1-Score	Harmonisches Mittel aus Precision und Recall

Kapitel 1

Einführung

1.1 Gegenstand der Dissertation

Im Entstehungsprozess von RGB-Bildern wird die dreidimensionale Anschauungswelt auf eine zweidimensionale Bildebene abgebildet. Mathematisch wird diese Abbildung durch ein *Kameramodell* erfasst, im Rahmen dieser Arbeit durch das Modell einer Lochkamera zur Beschreibung geläufiger projektiver Kameras. Ist ein *Kamerazentrum* sowie eine *Bildebene* gegeben, kann einem Punkt der Anschauungswelt ein Punkt der Bildebene zugeordnet werden. Dieser ist der Schnittpunkt der Geraden durch das Kamerazentrum und den Punkt der Anschauungswelt mit der Bildebene. Die Umkehrfunktion bildet einen Punkt der Bildebene auf einen Strahl ab, welcher ausgehend vom Kamerazentrum die Bildebene in jenem Punkt passiert. Ist durch ein *a priori* festgelegtes Modell bekannt, an welcher Stelle der Strahl auf eine Oberfläche trifft, können die dreidimensionalen Koordinaten aus den zweidimensionalen Bildpunkten berechnet werden. Sind die Kameraparameter (Kamerazentrum, Lage der Bildebene, ...) bekannt, können somit Aussagen zu Lagebeziehungen zwischen Objekten in Weltkoordinaten aus Bildern abgeleitet werden. Die Bestimmung der Parameter einer Kamera heißt Kalibrierung.

In der Literatur kann eine Reihe von Verfahren zur Bestimmung von Kameraparametern vorgefunden werden. Bekanntheit genießen vor allem der in [Hartley u. Zisserman, 2003] beschriebene Ansatz die Kameraparameter durch *Direct Linear Transformation (DLT)* zu bestimmen sowie der ebenfalls gebräuchliche Ansatz von [Zhang, 2000]. Für beide Verfahren kann festgehalten werden, dass die Kenntnis von Punkten innerhalb nur einer Kalibrierebene nicht ausreichend ist. Diese Dissertation widmet sich dem Problem der robusten Berechnung von Kameraparametern aus Bildsequenzen, in welchen eine Kalibrierung der externen sowie internen Kameraparameter bildweise zu erfolgen hat und in welchen eine Kalibriertafel in den Bildern nicht gegeben ist. Weiterhin wird die

Kalibrierung auf Basis einer relevanten Umgebung (ortsfester Landmarken) in ihrer dreidimensionalen Anordnung erfolgen und damit keine künstlichen Kalibriermarkierungen voraussetzen. Zudem werden auch Fälle betrachtet, in welchen eine für die Kalibrierung nicht ausreichende Zahl von Landmarken für eine gewisse Zahl aufeinanderfolgender Bilder innerhalb einer Bildsequenz vorgefunden wird und das Verfahren zur Kalibrierung somit auf Modelllandmarken ausweichen muss.

Eben das Problem der Nachführung der Kameraparameter und einer damit verbundenen nicht sichergestellten Sichtbarkeit des Modells ist häufig bei der Berechnung dreidimensionaler Koordinaten der Spieler in Sportübertragungen anzutreffen. Mit dem Spielfeld als Modellfläche können Distanzen und Winkel zwischen Spielern und anderen Spielfeldobjekten in Weltkoordinaten bestimmt werden. Speziell dem Fußball als Anwendungsgebiet wird sich diese Dissertation in besonderem Maße widmen. In Übertragungen dieser Sportart ist eine Vielzahl von Feldlinien oft nicht sichtbar, was eine genaue Verortung der im Bild dargestellten Spielfeldobjekte erschwert.

Vom Ende des bayerischen Fußballverbots im Jahr 1927 bis zur 25. Meisterschaft des FC Bayern München im Jahr 2015 erlebte der Fußball auch in Deutschland ein rasantes Wachstum. Technische Entwicklungen zogen rasch in verschiedene Aspekte wie die Spieler-, Trainer- und Schiedsrichterausbildung, TV-Übertragungen und statistische Datenauswertung ein. In jüngster Zeit begünstigt das Verbot technischer Hilfsmittel auf dem Rasen die Entwicklung bildverarbeitender Verfahren zur spielbegleitenden Erhebung von Spieler- und Ballpositionen, welche die individuelle Erhebung von raumzeitlichen Statistiken erlaubt. Während diese Systeme finanziell in einem vertretbaren Rahmen für die Lizenzclubs der 1. und 2. Bundesliga sind, stehen über 25 000 im DFB organisierten Vereinen [DFB, 2014b] oft nicht die Mittel für derartige Maßnahmen zur Verfügung. Ebenso liegen unterklassigen Vereinen (meist) keine Spielstatistiken zu Pässen, Toren, Ballbesitz, Eck- oder Strafstoßen vor, die in den professionellen Ligen manuell geführt werden. In Antwort auf diese Einschränkung werden im Rahmen dieser Dissertation auch Lösungen zur automatisierten Erfassungen von Spielereignissen vorgestellt.

1.2 Berücksichtigte Datentypen und Kamerasysteme

Im Rahmen dieser Arbeit werden verschiedene Datentypen und Kamerasysteme bei der Betrachtung von Eingangsdaten Berücksichtigung finden. Dadurch kann eine breite Palette von Systemen für die Vermessung zum Einsatz kommen. Diese schließt Kameras von Mobiltelefonen ebenso ein wie statische Kameras oder präzise — und während einer Begegnung untersagte — Tracking-Westen [Catapult, 2015]. Konkret wird nach folgenden Typen unterschieden:



Bildquelle: [Kabel eins, 2013]



TV-Bilder

Die Auswertung von Fußballübertragungen durch Fernsehbilder stellt aufgrund ihrer leichten Zugänglichkeit einen Gegenstand häufiger Untersuchungen dar [Alemán-Flores et al., 2014, Ekin et al., 2003, Kim et al., 2010]. Neben Bildern der Begegnung ist in einem Video mit Unterbrechungen, Wiederholungen und Schnitten zu rechnen. Es kommt eine Reihe von Kameras zum Einsatz, deren Parameter (z. B. Blickwinkel, Brennweite) laufend verändert werden.

Scouting Feed

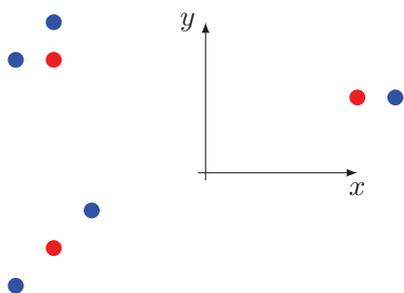
Im Rahmen bedeutender Begegnungen werden Aufnahmen zur gezielten Bewertung des Verhaltens der Spieler angefertigt. Um dabei alle Feldspieler im Kamerabild zu halten und zugleich die Spieler möglichst groß im Bild dargestellt zu wissen, werden derartige Aufnahmen durch einen Kameramann im Stadion durchgeführt. Entsprechend ist mit einer Veränderung des Bildausschnitts zu rechnen, Schnitte oder Wiederholungen kommen jedoch nicht vor. Es besteht die Vorgabe, während gewisser Ereignisse, wie Eckbälle oder erfolgreicher Torschüsse, die beteiligten Spieler großformatig im Bild darzustellen.

Panorama Feed

Eine weitere Möglichkeit der Aufzeichnung von Begegnungen sind statische Kameras. In bedeutenden Begegnungen, zum Beispiel Spielen der Bundesliga, wird das Geschehen durch zwei Kameras aufgezeichnet, welche eine Übersicht über das gesamte Spielfeld bieten. Deren Bilder werden zu einem gemeinsamen Video verbunden (Stitching).

Positionsdaten

Liegen Daten zu den Positionen der Spieler, des Balls und eventuell des Schiedsrichters und der Linienrichter in Weltkoordinaten vor, wird von Positionsdaten gesprochen. Da im Fußball während einer Begegnung der Gebrauch elektronischer Kommunikationssysteme untersagt ist [FIFA, 2014], wird für die Bereitstellung dieser Daten auf optische Systeme zurückgegriffen. Hierbei werden Systeme aus einer Vielzahl von Kameras fest auf ein Spielfeld kalibriert. Die Systeme arbeiten semi-automatisch. Fehler in der Spieler-Nachführung werden während einer Begegnung von Hand behoben. Dabei übersehene Fehler werden in einem zweiten Arbeitsschritt nach dem Spiel bereinigt. Bekannte Anbieter für die Erfassung von Positionsdaten, sogenannte *Content Information Provider (CIP)*, sind Tracab [ChyronHego, 2005], [Prozone, 2015] und [Deltatre AG, 2015].



Da jeder der betrachteten Datentypen behandelt werden wird, besteht auch der Vorteil, dass eine statische Kamera keine zusätzliche Befestigung zwecks Windschutz benötigt, wobei fortlaufend eine neue Kalibrierung am Spielfeld erfolgen kann. Neben der Erfassung von Positionsdaten stellen gewisse CIPs (wie Opta [Perform, 2015]) auch höherwertige, manuell erfasste, Spielereignisdaten bereit. Die manuelle Aufnahme des Spielverlaufs kann hierbei alle entscheidenden Ereignisse in einer Begegnung einschließen.

1.3 Zielsetzung

Zu den Zielen dieser Dissertation gehört es, das Problem der Bereitstellung von Kameraparametern zu Bildsequenzen für die Berechnung von Weltkoordinaten aufzugreifen. Im Rahmen der genannten Einschränkungen bzgl. der Sichtbarkeit der im Modell hinterlegten Landmarken werden folgende Teilziele festgehalten:

- Robuste Klassifikation der Punkte eines Bilds bzgl. ihrer Zugehörigkeit zur Kalibrierebene
- Extraktion natürlicher Landmarken unter dem Aspekt des Ausschlusses nicht-ortsfester und der Erhaltung ortsfester Landmarken
- Nachführung der Kameraparameter in Bildsequenzen mit einer einzigen Kalibrierebene, unter Berücksichtigung wechselnder intrinsischer und extrinsischer Kameraparameter unter Berücksichtigung der Nicht-Sichtbarkeit der im Modell hinterlegten Landmarken

Durch Bezugnahme auf das Anwendungsgebiet Fußball erweitert sich diese Liste um die folgenden Punkte:

- Erkennung der Teilmenge relevanter Bilder in Bildsequenzen, aus welchen sich Objekt- (Spieler-, Ball-)positionen potenziell bestimmen lassen. Dies ist vor allem für TV-Bilder von Bedeutung, da in diesen eine hohe Anzahl von Bildern enthalten ist, welche keine im Sinn der Auswertung relevanten Informationen enthalten (z. B. Nahaufnahmen).
- Zuordnung von Bild- zu Spielfeldkoordinaten durch eine automatische Kameraparameter-Nachführung auf Grundlage der Spielfeldmerkmale auch in langen Bildsequenzen (Größenordnung 100 000 Bilder).
- Erkennung von taktischen Strukturen und Spielereignissen aus Positions- (Raum-Zeit-)Daten.

Ausschlaggebend für die Bereitstellung von Positionsdaten ist ebenso das Tracking von Objekten im Bild. Aufgrund der Symmetrie des Spielfelds ist anzumerken, dass ein Spielfeldobjekt ohne weitere Annahmen nicht eindeutig einer Weltkoordinate zugeordnet werden kann. Eine feste Seitenzuordnung wird genauso wie die Annahme einer sich über dem Boden befindenden Kamera verwendet, um jene Mehrdeutigkeiten aufzulösen.

Durch die Erkennung von Spielereignissen, wie Pässen, Einwüfen, Torschüssen, Eck- und Strafstoßen mit den involvierten Spielern, ihren Positionen und den Zeitpunkten des Ereignisses, kann eine Begegnung umfassend charakterisiert werden. Dies wird im Weiteren als *Spieljournal* bezeichnet. Neben Statistiken zu einzelnen Spielern und Mannschaften lassen sich automatische Sportreporter ebenso realisieren wie Simulationen von Spielsituationen. In der Bereitstellung eines Spieljournals kann ein weiteres Ziel dieser Dissertation gesehen werden.

1.4 Stand der Forschung

In der Literatur existieren wenige Ansätze im sportlichen Kontext, welche die Aufbereitung von Bilddaten zu Positionsdaten anstreben und die so gewonnenen Informationen anschließend im Rahmen einer Ereigniserkennung verwenden. Die fußballspezifische Arbeit von [Ekin et al., 2003] ist in dieser Reihe an erster Stelle zu nennen. Wie auch im vorgestellten Ansatz werden die betrachteten TV-Bildsequenzen hier durch eine Schnitterkennung auf aussagekräftige Bilder eingegrenzt. In diesen werden erzielte Tore, die Strafräume und der Schiedsrichter erkannt. Da die Auswertung nicht direkt über die Positionen der Spieler erfolgt, sondern auf Basis des im Bild sichtbaren Spielfeldausschnitts, bleibt diesem Ansatz eine Vielzahl an Ereignissen verborgen. Im ASpoGAMo-System [Gedikli, 2009], welches in erster Linie auf die Betrachtung von Fußball ausgelegt ist, geschieht die Zuordnung von Bild- zu Weltkoordinaten durch eine mehrstufige Erkennung der Kameraparameter. Ebenfalls im Fußballkontext angesiedelt ist die Dissertation von [Schlipsing, 2014]. Das Ziel dieser Arbeit liegt in der bildbasierten Bereitstellung der Positionen der Spieler in Weltkoordinaten. Dafür wird der Panorama Feed herangezogen, wodurch keine Kameraparameter-Nachführung zu berücksichtigen ist. Eine taktische Untersuchung der berechneten Positionsdaten ist in diesem Ansatz für weitere Betrachtungen offen gelassen.

Für TV-Übertragungen von Baseball-Spielen stellen [Hung u. Hsieh, 2008] ein System zur Ereigniserkennung auf Basis einer Detektion von Schnitten und Einblendungen vor. Vergleichbar mit [Ekin et al., 2003] wird dabei der Schritt über Positionsdaten übergangen. Passender zur Zielsetzung ist die Arbeit von [Lu et al., 2013] im Basketball-Kontext. Ihr Verfahren erkennt und trackt Spieler automatisch und ist ebenso in der Lage, unter der Voraussetzung einer gegebenen Startposition die Kameraparameter automatisch nachzuführen. Da das Fußballspielfeld aufgrund seiner größeren Abmessungen, der geringeren Anzahl markanter Elemente ([Lu et al., 2013] nutzten Werbeflächen im Spielfeld) und kleineren Spielern auf seiner Oberfläche (im Durchschnitt und im Bild) eine andere Herausforderung darstellt, können die erzielten Ergebnisse nicht ohne Weiteres verwendet werden. Ebenso spielen bei Hallensportarten Umwelteinflüsse keine Rolle. In den weiteren Kapiteln wird auf die jeweils relevante Literatur im Einzelnen eingegangen.

Kapitel 2

Aufbau und Überblick

2.1 Einleitung

Die im vorigen Kapitel formulierten Ziele werden nun in Teilaufgaben zerlegt und den einzelnen Kapiteln der Dissertation zugeordnet. Aus der Reihenfolge, welche Daten bei der Analyse im vorgestellten System durchlaufen, ergibt sich auch die natürliche Reihenfolge der vorliegenden Schrift. Diese wird in Abschnitt 2.2 näher erläutert. Die vorausgegangenen Veröffentlichungen werden in Abschnitt 2.3 um eine Erklärung zu weiteren Erkenntnissen ergänzt. Das Kapitel schließt mit der Verknüpfung der theoretischen Betrachtung zu einer Implementation der aufgezeigten Verfahren in Abschnitt 2.4.

2.2 Ein System zur Fußballanalyse

Wie in Kapitel 1.2 festgehalten, soll das Analysesystem nicht nur mit unterschiedlichem Videomaterial arbeiten können; auch Raum-Zeit-Informationen, wie sie unter Umständen von CIPs zur Verfügung gestellt werden, sollen verarbeitet werden können. Um dies in einer einheitlichen Plattform ermöglichen zu können, werden die Daten, je nach ihrer Herkunft, zu höherwertigen Informationen transformiert.

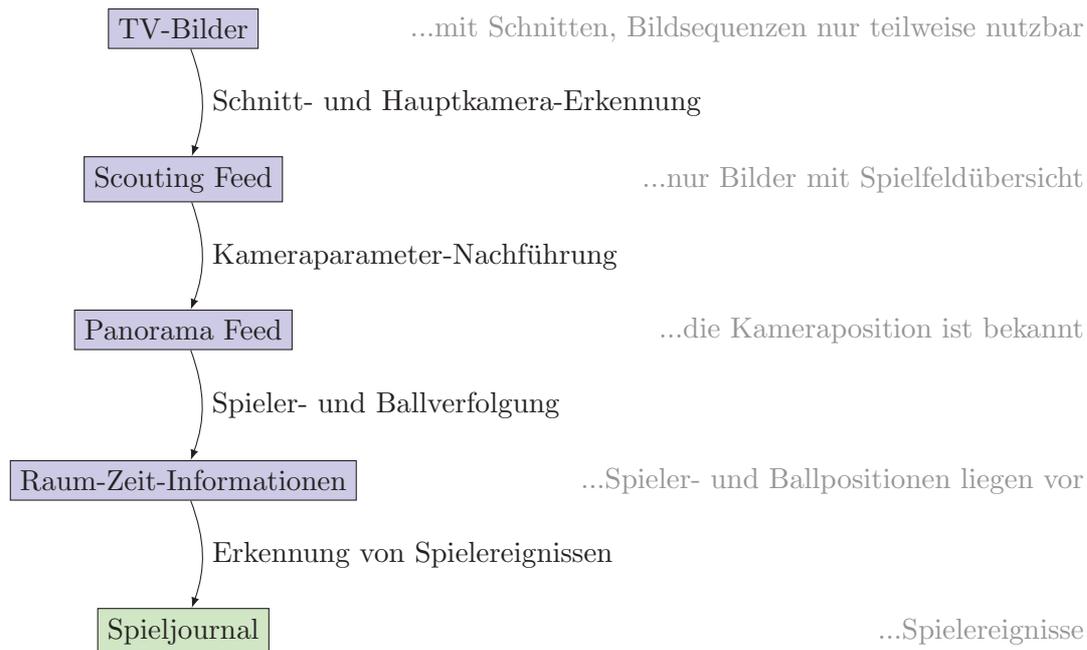


ABBILDUNG 2.1: Transformation von verschiedenen Eingangsmedien. Durch die gewählte Verarbeitungshierarchie können die Verfahren auf verschiedene Eingangsmedien angewandt werden.

Dieser Prozess der Transformation ist in Abbildung 2.1 dargestellt. Die grundlegende Idee hierbei ist, dass es für einen Arbeitsschritt keine Rolle spielen muss, woher die Daten letztendlich stammen. In TV-Bildern gilt es, die Teilmenge an Bildern zu finden und auszuschließen, aus welchen keine gewinnbringenden Informationen entnommen werden können (genauer: Szenen, welche nicht von einer Übersichts-Kamera aufgezeichnet wurden). Natürlich führen die daraus entstehenden Schnitte zu Lücken in den Daten, doch auch kommerzielle Tracking-Systeme enthalten Datenfehler und -lücken, welche im Gegensatz zu TV-Aufnahmen aufgrund des geringeren Informationsgehalts (es liegen nur noch die abgetasteten Positionen vor) schwieriger in der Identifikation sein können. Sind die Bilder der Übersichts-Kamera bestimmt, werden die Kameraparameter für jedes Bild berechnet. Nach diesem Schritt kann das Video ähnlich der Aufnahme einer statischen Kamera behandelt werden. Die anschließende Spieler- und Ballverfolgung transformiert die Daten letztlich in ein einheitliches Positionsdatenformat, welches sowohl für die monokulare Spieler-Nachführung als auch die CIP-Positionsdaten verwendet wird.

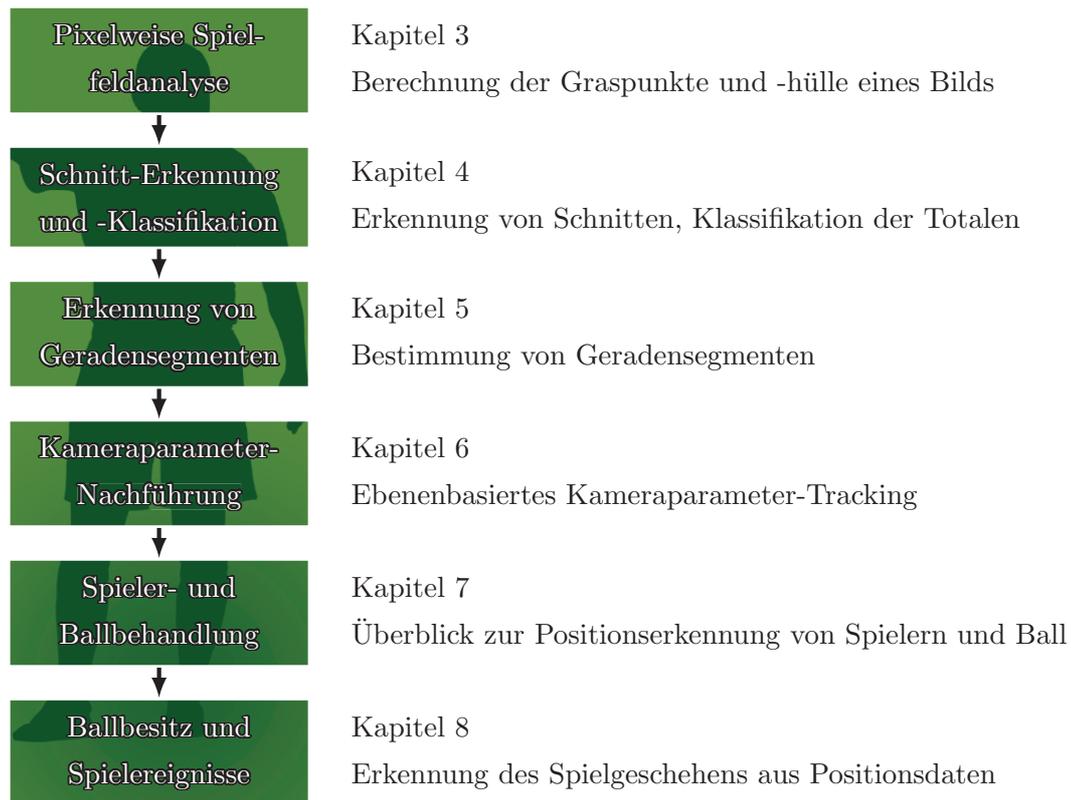


ABBILDUNG 2.2: Die nachfolgenden Kapitel im Überblick

Die vier in Abbildung 2.1 dargestellten Transformationsschritte zwischen den verschiedenen Eingangs-Medien wurden für Abbildung 2.2 um zwei Schritte erweitert. Diese stellen grundlegende Vorverarbeitungen dar, welche für folgende Verfahren eine Rolle spielen werden. Die damit geschaffene Kette stellt gleichzeitig zu den Bearbeitungsschritten für TV-Bilder die Reihenfolge der Kapitel dieser Arbeit dar und wird diese Schrift als roter Faden begleiten. Zum Anfang eines jeden Themenkapitels wird ein Bezug zu Abbildung 2.2 die vorgestellten Techniken thematisch einordnen. Eine Darstellung, welche Kapitel für die verschiedenen Quellmedien von Relevanz sind, ist in Tabelle 2.1 gegeben.

	TV-Bilder	Scouting Feed	Panorama Feed	Positionsdaten
Kapitel 3. Pixelweise Spielfeldanalyse	x	x	x	
Kapitel 4. Schnitt-Erkennung und -Klassifikation	x			
Kapitel 5. Erkennung von Geradensegmenten	x	x		
Kapitel 6. Kameraparameter-Nachführung	x	x	(x)	
Kapitel 7. Spieler- und Ballbehandlung	x	x	x	
Kapitel 8. Ballbesitz und Spielereignisse	x	x	x	x

TABELLE 2.1: Übersicht zur Relevanz der Kapitel in Abhängigkeit von verschiedenen Quellmedien. Für den Panorama Feed besitzt Kapitel 6 nur eingeschränkt Bedeutung. Hierbei kommen lediglich Transformationen von Koordinaten und gegebenenfalls eine Kamera-Kalibrierung in Frage.

Nach einer Evaluation bzgl. der Erkennung von Spielereignissen im Spieljournal (Kapitel 8) folgt eine generelle Diskussion und Einordnung in Kapitel 9. Hier werden die Erkenntnisse der Dissertation gegenüber der Zielsetzung bewertet.

2.3 Vorausgegangene Veröffentlichungen und wissenschaftlicher Beitrag

Einige Kapitel dieser Monographie stützen sich auf peer-reviewte Konferenz- oder Journal-Veröffentlichungen, während andere (hier sind in erster Linie die Erkennung von Geradensegmenten und die Kameraparameter-Nachführung zu nennen) noch nicht einzeln aufbereitet wurden. Zu ersterer Kategorie gehören:

HOERNIG, Martin ; BIGONTINA, Andreas ; RADIG, Bernd: A Comparative Evaluation of Current HTML5 Web Video Implementations. In: *In Open Journal of Web Technologies (OJWT)* 1 (2014), Nr. 2, S. 1-9

→ Den Fähigkeiten moderner HTML5 Videos werden Unwägbarkeiten in den gängigen Implementationen gegenübergestellt. Mehr hierzu kann im Unterpunkt SoccerGapScripts des Abschnitts 2.4 vorgefunden werden.

HOERNIG, Martin ; HERRMANN, Michael ; RADIG, Bernd: Real Time Soccer Field Analysis from Monocular TV Video Data. In: *11th International Conference on Pattern Recognition and Image Analysis (PRIA-11-2013)*, Samara 2 (2013), S. 567-570

→ Grundlagenpapier zum nachfolgenden Journal Paper.

HOERNIG, Martin ; HERRMANN, Michael ; RADIG, Bernd: Real-Time Segmentation Methods for Monocular Soccer Videos. In: *Pattern Recognition and Image Analysis* 25 (2015), Nr. 2, S. 327-337

→ Ein Verfahren zur Berechnung einer konvexen Hülle mit der Bedeutung der Spielfeldregion wird ebenso wie ein Verfahren zur Berechnung von Graspixeln in Einzelbildern vorgestellt. Grundlage für Kapitel 3.

BIGONTINA, Andreas ; HERRMANN, Michael ; HOERNIG, Martin ; RADIG, Bernd: Human Body Part Classification in Monocular Soccer Images. In: *9-th Open German-Russian Workshop on Pattern Recognition and Image Understanding* (2014)

→ Darstellung eines Verfahrens zur Klassifikation von Körperteilen aus monokularen Einzelbildern im Fußball, basierend auf schwachen Klassifikatoren in Random Forests und der Graserkennung aus Kapitel 3. Details hierzu in Kapitel 3.8.

HOERNIG, Martin ; HERRMANN, Michael ; RADIG, Bernd: Multi Temporal Distance Images for Shot Detection in Soccer Games. In: *EUSIPCO 2014 (22nd European Signal Processing Conference 2014)*. IEEE, 2013, S. 2420-2424

→ Eine Schnitterkennung für die automatische Verarbeitung von TV-Fußball-Bildsequenzen. Vorgestellt in Kapitel 4, dort außerdem um eine Klassifikation der Übersichtseinstellungen erweitert.

HERRMANN, Michael ; HOERNIG, Martin ; RADIG, Bernd: Online Multi-player Tracking in Monocular Soccer Videos. In: *AASRI Procedia* 8 (2014), S. 30-37

→ Eine für Fußballspiele optimierte Personenerkennung und -nachführung. Besprochen in Kapitel 7.

HOERNIG, Martin ; LINK, Daniel ; HERRMANN, Michael ; RADIG, Bernd ; LAMES, Martin: Detection of Individual Ball Possession in Soccer. In: *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)* 392. Springer, 2016, S. 103-107

→ Einführung neuer Maße zur Erfassung von individuellem Spielerballbesitz im Fußball. Implementation und Anwendungstests, aufgegriffen in Kapitel 8.

Im Rahmen der Projektarbeiten sind neben den bereits mit den Veröffentlichungen erwähnten Neuerungen weitere Untersuchungsergebnisse hervorzuheben. Dazu zählen:

- ein Verfahren zur Erkennung von Geradensegmenten durch lineare Regression (Kapitel 5). Die Schwellwertbestimmung kann durch Hough-Parameter intuitiv gestaltet werden. Das Verfahren gehört zu den schnellsten Linienerkennungsverfahren und erzielt Ergebnisse oberhalb des State-of-the-Art-Niveaus.
- ein Verfahren zur quasi-affinen Kameraparameter-Nachführung (Kapitel 6). Neben einem angepassten ICP-Algorithmus zur Kamera-Kalibrierung liegt ein Schwerpunkt auf der Erkennung von Homographien, welche Transformationen darstellen, die nicht durch eine andere Kameraperspektive erreicht werden können (nicht-quasi-affine Homographien). Die erreichte Robustheit bietet die Möglichkeit der Kameraparameter-Nachführung auch für lange Bildsequenzen.
- die Verwendung des individuellen Ballbesitzes, um Spielereignisse abzuleiten (Kapitel 8.4). Neben dem Spieljournal und quantitativen Merkmalen können auch neue spielbeschreibende Auswertungen, wie Ballbesitz-Heatmaps, erzeugt werden.

2.4 Implementation als Client-Server-Architektur

Alle im vorangegangenen Abschnitt aufgezeigten (und in den nächsten Kapiteln beschriebenen) Analyseverfahren wurden in einer Bibliothek - genannt *Soccer Game Analysis Platform (SoccerGAP)* - implementiert. Allein diese Bibliothek enthält über 70 000 Quelltextzeilen, wobei hierin ausschließlich bildverarbeitungsspezifische Aufgaben abgebildet werden. Durch ein Web-Interface wird eine Zugriffsmöglichkeiten für einer Vielzahl von Geräten ermöglicht, ohne client-seitig hohen Rechenaufwand zu verursachen. Um SoccerGAP als C++-Bibliothek für diesen Zweck sinnvoll einsetzen zu können, wurden die maßgeblichen Funktionen in Python, der Sprache des *Content Management Systems (CMS)*, durch eine Schnittstelle (Wrapper) verfügbar gemacht.

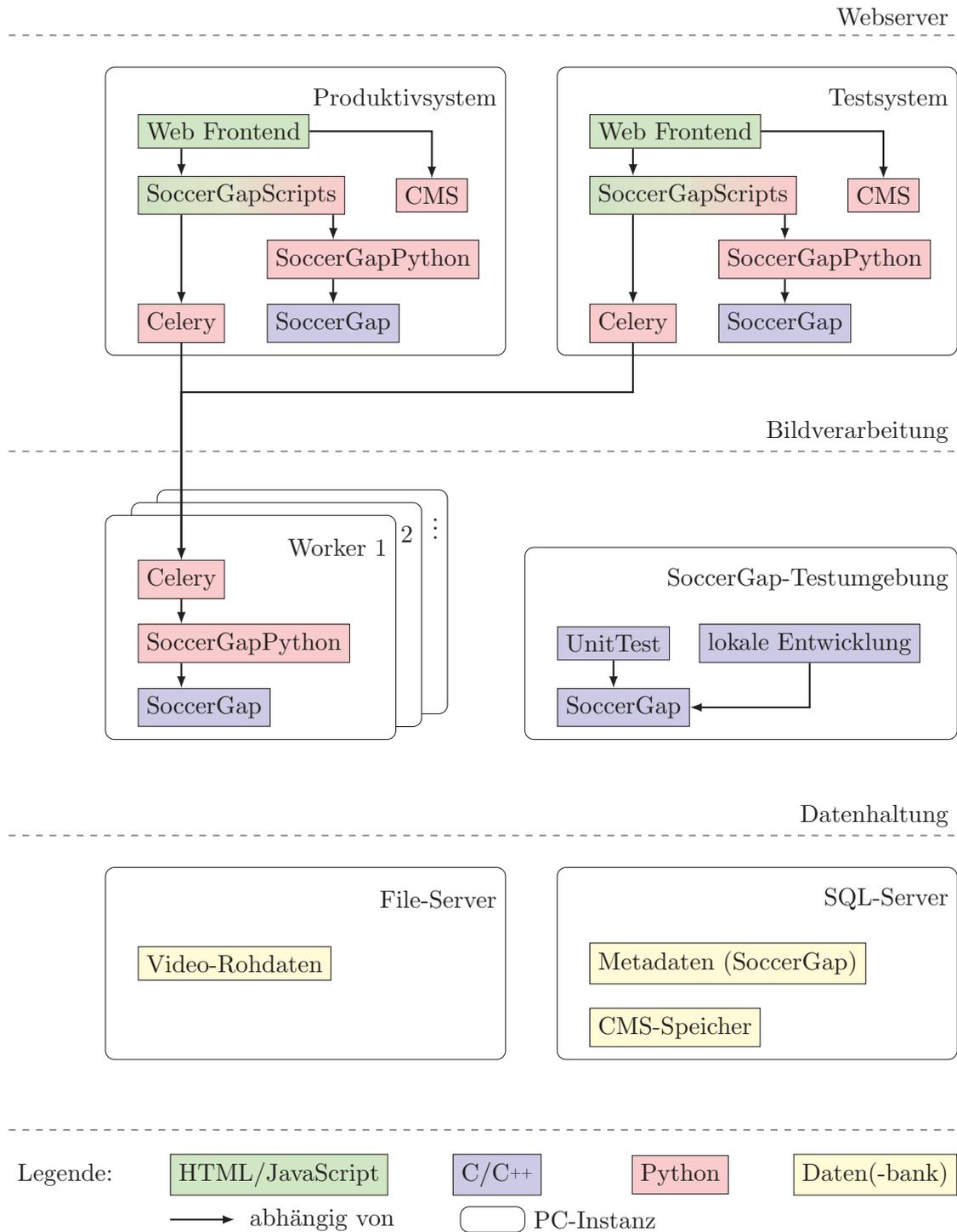


ABBILDUNG 2.3: Schematische Darstellung zum Aufbau der webbasierten Bildverarbeitungsumgebung. Zusammenhängende Abschnitte sind funktional getrennt einzelnen PC-Instanzen zugeordnet. Aufgrund der hohen Anzahl von Abhängigkeiten bzgl. der Datenhaltung wurde auf entsprechende Verknüpfungen verzichtet.

Außerhalb SoccerGaps ist das System über verschiedene PC-Instanzen verteilt. Die Trennung erfolgt dabei nach funktionalen Bausteinen und entsprechend ihren Hardware- und Software-Anforderungen. Wie in Abbildung 2.3 dargestellt, sind Webservereinstanzen

getrennt von Bildverarbeitungsrechnern oder Datenbankservern. Obwohl alle Aufgaben auch auf einem einzigen PC erfüllt werden können, wird es so möglich, ein deutlich größeres Leistungsspektrum abzubilden.

In den Webserver werden die Erwartungen von hoher Zuverlässigkeit und Sicherheit gesetzt, Caching ist wünschenswert. Ein solcher PC kann mühelos virtualisiert werden. Oft empfiehlt sich dies, etwa zum Bereitstellen (Duplizieren) eines Testsystems, wie in Abbildung 2.3 aufgezeigt. Ähnlich verhält es sich mit den Anforderungen an die Datenhaltung, wo Backups (evtl. von sehr großen Datenbeständen) eine Rolle spielen.

Die Verarbeitung von Daten unter dem Aspekt der Bildverarbeitung profitiert oftmals deutlich von GPU-Unterstützung hinsichtlich der Reduzierung von Programmlaufzeiten. Eine entsprechende GPU-Virtualisierung ist derzeit jedoch nur eingeschränkt möglich. Ebenfalls spielt vielmehr Rechenleistung als Zuverlässigkeit eine Rolle. Fällt ein Worker aus, kann ein anderer einspringen (mehr dazu im Unterpunkt Celery).

Auf die Bestandteile der Architektur aus Abbildung 2.3 wird im Folgenden eingegangen:

Web Frontend, CMS, CMS-Speicher

Die für den Betrieb einer Webpräsenz grundlegenden Elemente (Webseiten, Benutzerkontensteuerung, usw.) werden durch ein Content Management System (CMS) realisiert.

Celery

Celery [Solem, 2015] stellt die Schnittstelle zwischen der Aufforderung des Anwenders, ein Bildverarbeitungsverfahren auszuführen, und einem Worker, welcher dieser Tätigkeit nachkommt, dar. Bei Celery handelt es sich um eine asynchrone Warteschlange. Von Seiten des Webservers werden Aufgaben erteilt, welche den Workern (je nach aktueller Belegung) zugewiesen werden. Dabei können auch mehrere Worker auf einem PC gleichzeitig arbeiten. Nach Fertigstellung einer Aufgabe steht ihr Resultat im Web bereit.

Nicht im Bild dargestellt ist die Tatsache, dass Aufgaben im Worker asynchron bearbeitet werden, um (an vordefinierten Exit-Punkten) abgebrochen werden zu können. Erhält ein Worker das Signal zum Abbruch, etwa durch den Anwender, wird dies im CMS-Speicher notiert und vom Worker beantwortet. Ähnlich verhält es sich mit der Weiterleitung des Task-Statuses von SoccerGap über Celery zum Web Frontend. Dieses wird parallel zum asynchron arbeitenden Prozess in einer Datenbank vermerkt und steht für eine Abfrage über die Webserver-Schicht bereit.

SoccerGapPython

SoccerGapPython ist ein Wrapper für SoccerGap. Es stellt Funktionen von SoccerGap in Python bereit. SoccerGap kann auf diese Weise von Celery und SoccerGapScripts (Python) verwendet werden. Die Übersetzung macht es möglich, ressourcenintensive Aufgaben in C/C++ und den hier zur Verfügung stehenden Bibliotheken zu berechnen, während die umfassenden webspezifischen Werkzeuge von Python für den Webserver zum Einsatz kommen können.

SoccerGap, Metadaten

SoccerGap kennt zwei Arbeitsmodi: asynchron und synchron. Wie im Abschnitt zu Celery beschrieben, werden Bildverarbeitungsrouitinen in den Workern erledigt. Nachdem eine Aufgabe an einen Worker zugewiesen wurde, wird diese durch SoccerGapPython übersetzt und in SoccerGap asynchron bearbeitet. Die Ergebnisse, etwa Spielerpositionen oder Kameraparameter, werden als Metadaten in einer Datenbank hinterlegt. Bei der Abfrage dieser Daten kommt der zweite Modus zum Tragen. SoccerGap, genauso wie SoccerGapPython, stehen auch im Webserver zur Verfügung, um die Daten wieder aufzubereiten. Da hier nur Lesen und gegebenenfalls Formatieren notwendig ist, muss keine Arbeit über Celery verteilt werden und die Ergebnisse können durch Funktionen mit synchronem Ablauf sofort bereitgestellt werden.

SoccerGapScripts

Über SoccerGapScripts werden sowohl Aufträge an die Worker verteilt als auch Daten für den Anwender präsentiert. Einige SoccerGap-Funktionen werden ein weiteres Mal übersetzt, diesmal in JavaScript.

Eine zentrale Komponente von SoccerGapScripts ist die Wiedergabe von Videomaterial mit zusätzlichen Einblendungen (siehe Abbildung 2.4). Dabei wird das Video vom File-Server (Rohdaten) unverändert an den Anwender übertragen. Die Einblendungen werden synchronisiert in Echtzeit auf das Videoelement der Webseite gezeichnet. Hierfür musste ein neuer Videoplayer entwickelt werden. Da Video- und der Metadatenstrom über getrennte Puffer gestreamt werden müssen, ergeben sich Herausforderungen in der Synchronisation (etwa bei geringen Bandbreiten oder Sprüngen). Dies begründet sich insbesondere darin, dass Informationen über den Videodatenstrom nur bedingt vertrauenswürdig sind. Die zugehörige Veröffentlichung [Hoernig et al., 2014] fand unter Entwicklern von Browsern und Web-Anwendungen Zuspruch.

A paper presents some results of cross-browser testing of HTML5 video. This sort of work is very helpful since it draws attention away from browser vendor PR efforts and towards the sort of issues that really impact Web developers, with actionable data.

Having said that, it's also very interesting to compare how browsers fare on correctness benchmarks designed without a browser-vendor axe to grind. [...]

— O'Callahan [2014]

Es zeigte sich vor allem, dass ein zu unspezifisch formulierter Standard für HTML5 Video und fehlerhafte Implementationen in den Browsern eine vollständige HTML5-Lösung nicht möglich machen. Für gewisse Anwendungsfälle wurde somit ein Fallback (Flash) notwendig.



ABBILDUNG 2.4: Automatische Kameraparameter-Nachführung einer Mobiltelefon-Aufzeichnung innerhalb der Weblösung. Überlagerung eines HTML5-Videos mit JavaScript-Einblendungen auf Basis server-seitig berechneter Parameter.

Kapitel 3

Pixelweise Spielfeldanalyse

Dieses Kapitel basiert methodisch auf:

HOERNIG, Martin ; HERRMANN, Michael ; RADIG, Bernd:
Real-Time Segmentation Methods for Monocular Soccer Vi-
deos. In: *Pattern Recognition and Image Analysis* 25 (2015),
Nr. 2, S. 327-337

3.1 Einleitung

In diesem Kapitel wird die pixelweise Klassifikation von Einzelbildern bezüglich der Zugehörigkeit zur Spielfeldoberfläche besprochen. Unter gegebenen Voraussetzungen (etwa einer homogenen Rechteck-Spielfläche) werden die Grundebene und sich darauf evtl. befindliche Störungen (z. B. Spieler) bestimmt. Es werden zwei Aussagen getroffen: welche Fläche im Bild mutmaßlich vom Spielfeld eingenommen wird (die so genannte Hülle, Abschnitt 3.3) und welche Punkte des Bilds zum Spielfeld gehören (die Grasmasken, Abschnitt 3.4). Somit können letztlich für jeden Punkt des Eingangsbilds zwei Fragen beantwortet werden:

- Ist der Punkt innerhalb des Spielfelds?
- Gehört der Punkt zur Grasfläche?



Da es sich hierbei um die ersten Betrachtungen handelt und damit der Grundstein für nachfolgende Verfahren gelegt wird, stehen neben den Bildinformationen selbst keine weiteren Daten, etwa aus konsekutiven Bildern, bereit. Auf der einen Seite ist ein Verfahren, welches sich nur eines einzelnen Bilds bedient, selbstredend durch die Information dieses einzelnen Bilds eingeschränkt. Auf der anderen Seite ist es dadurch möglich, Szenen durch eine stichprobenartige Auswahl von Bildern einfacher bzw. schneller zu bewerten (dies wird in Kapitel 4 eine Rolle spielen). Ebenso brauchen Beleuchtungsveränderungen innerhalb einer Begegnung nicht berücksichtigt bzw. korrigiert werden und für eine robuste Messung über eine größere Anzahl von Bildern kann das Verfahren entsprechend erweitert werden (vgl. Abschnitt 3.3). Da es nicht gesichert ist, dass ein Spielfeld gewisse quantitative Eigenschaften im Bild erfüllt (Anzahl der zum Spielfeld gehörenden Pixel), muss ein Kontext somit durch den Algorithmus hervorgebracht werden.

Neben theoretischen Vorbetrachtungen werden in diesem Kapitel auch Anwendungen zur Grasmaskerung vorgestellt (Abschnitt 3.8). Augenmerk wird ebenso auf die Berechnung von Körperteilen von Spielern auf Grundlage der Kontur der Graspunkte gerichtet, da jene Kalkulation die Effektivität und Genauigkeit der Graserkennung verdeutlicht.



ABBILDUNG 3.1: Beispiel einer konvexen Hülle mit sechs Eckpunkten. Bildquelle: [MDR, 2013]

Das Beispiel einer berechneten Hülle ist in Abbildung 3.1 dargestellt. Dieses Hilfsmittel wird in den folgenden Kapiteln als Region of Interest (ROI) eine wichtige Rolle spielen und das Tracking des Spielfelds (bzw. der Kameraparameter) ermöglichen. Die Berechnung der Graspunkte wird insbesondere für die Erkennung von Spielern (Kapitel 7) eine Voraussetzung darstellen.

3.2 Clipping eines konvexen Vierecks in einem Rechteckfenster

Zur Gestalt des Spielfelds heißt es in den FIFA-Regularien [FIFA, 2014]:

Das Spielfeld ist rechteckig und wird mit Linien gekennzeichnet.

— DFB [2014a], FIFA [2014]

Das Spielfeld wird damit im Weiteren als Rechteck betrachtet (weiteres zum Spielfeld-Modell in Kapitel 6.4). Mit der Annahme, dass das Spielfeld durch eine projektive Kamera auf die Bildebene abgebildet wird, kann dies in homogenen Koordinaten als lineare Abbildung und schließlich als Homographie zwischen Spielfeldebene und Bildebene beschrieben werden. Mit der weiteren Annahme, dass sich kein Punkt des Spielfelds hinter der Kamera befindet, kann die Konvexität des abgebildeten Spielfelds sichergestellt werden.

Satz 3.1. Wenn das Rechteck-Spielfeld \mathbf{R} in \mathbb{R}^3 vollständig vor einer projektiven Kamera liegt, dann wird durch die projektive Abbildung der Kamera das Spielfeld konvex auf die Bildebene abgebildet.

Beweis 3.1. Siehe Theorem 21.4 in [Hartley u. Zisserman, 2003].

Anmerkung 1. Eine Einführung zur projektiven Geometrie und quasi-affinen Homographien wird in Kapitel 6 gegeben.

Mit der hinreichenden Bedingung aus Satz 3.1 kann die Gestalt des Spielfelds nach der Projektion (unter den genannten Bedingungen) damit auf ein konvexes Viereck \mathbf{R}' eingegrenzt werden. Das Viereck wird, wie in Abbildung 3.2 dargestellt, nun mit dem Bildrechteck (dem Fenster des Bilds) \mathbf{W} geschnitten. Dabei stellt sich die Frage nach der minimalen und maximalen Eckenanzahl des resultierenden konvexen Polygons. Sonderfälle, in welchen das Ergebnis auf eine Gerade verschwindet oder $\mathbf{W} \cap \mathbf{R}' = \emptyset$, werden dabei vernachlässigt. Daraus geht direkt die minimale Eckenanzahl von drei Ecken (Abbildung 3.2, C) hervor.

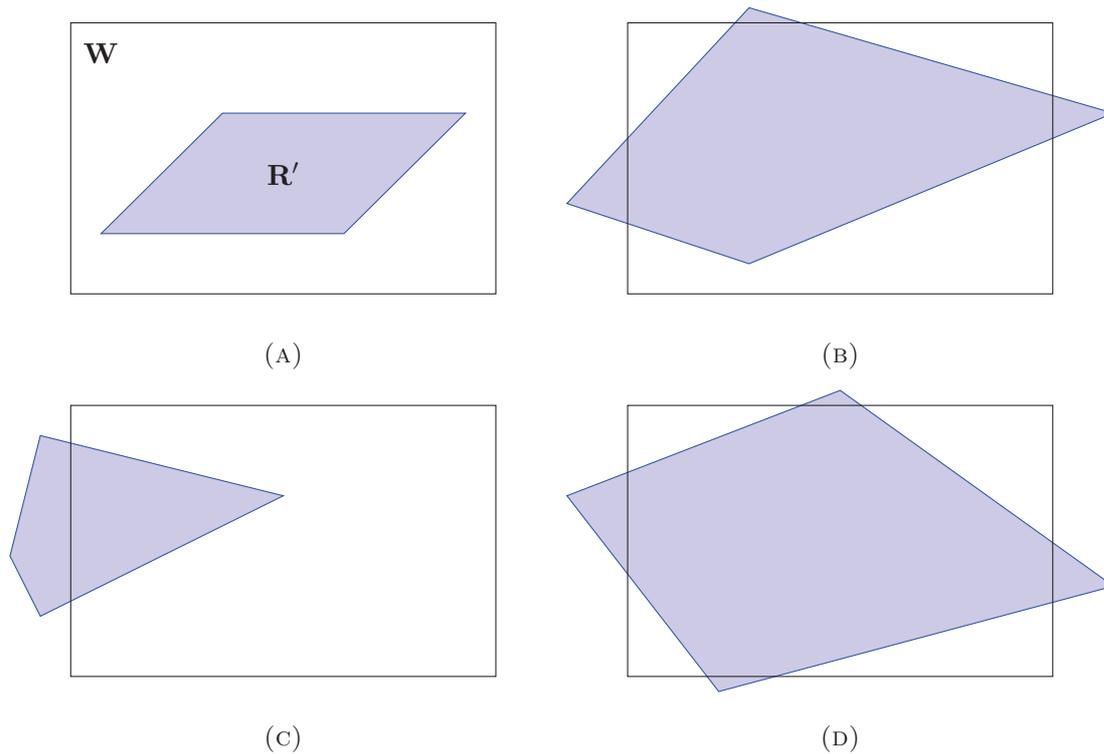


ABBILDUNG 3.2: Schneiden eines konvexen Vierecks \mathbf{R}' mit dem Bildrechteck \mathbf{W}

Für eine Bestimmung der maximalen Eckenanzahl wird der Sutherland-Hodgman-Algorithmus [Sutherland u. Hodgman, 1974] verwendet. Dieser durchläuft alle Eckpunkte von \mathbf{R}' . Liegt einer der Punkte dabei außerhalb von \mathbf{W} , so entstehen zwei neue Punkte an den Rändern von \mathbf{W} . Da \mathbf{R}' ein Viereck ist, kann so maximal ein Achteck entstehen. Eine Verzeichnung wird während der gesamten Betrachtung vernachlässigt. Die maximale Eckenanzahl beträgt acht (Abbildung 3.2, D).

Anmerkung 2. Für die Berechnung von Hüllen (konvexen Polygonen) werden nur endliche Anzahlen von Ecken in Betracht gezogen.

3.3 Berechnung der Spielfeldhülle

Wie bereits in der Einleitung vermerkt, wird die Berechnung der Grashülle als ROI einen zentralen Vorverarbeitungsschritt für weitere Verfahren darstellen. In diesem Kapitel geht sie beispielsweise der Berechnung der Grasmasken voraus. Da das Spielfeld in einem Bild nur einen verhältnismäßig kleinen Flächenanteil ausmachen kann, gestaltet sich die direkte Festlegung eines Verfahrens aus Histogrammen (etwa von Farbtönen) schwierig. Die FIFA-Regularien [FIFA, 2014] beschreiben die Oberfläche des Spielfelds mit den Worten:

Spiele können auf einer natürlichen oder künstlichen Unterlage ausgetragen werden. Maßgebend ist das jeweilige Wettbewerbsreglement. Kunstrasenfelder sind grün.

— DFB [2014a], FIFA [2014]

Da natürliche Unterlagen gemeinhin als Grasflächen verstanden werden können, wird für alle Unterlagen eine grüne Farbe gefolgert. In diesem Ansatz wird ein festes Intervall $[HueMin; HueMax]$ an Farbwerten (ein Hue-Intervall in HSV) für die Beschreibung der Spielfeldfarbe Verwendung finden. Dieses Intervall ist für die Einzelbildbetrachtung *a priori* zu trainieren. Liegt ein geeignetes Video vor, so kann $[HueMin; HueMax]$ auch robust aus dem Hue-Histogramm des gesamten Videos zuvor ermittelt werden.

Da achromatische Pixel ebenfalls in den Bereich dieser Farbtöne fallen können, wird für die Auswahl geeigneter grüner Pixel zuvor ein chromatisches Filter auf das Eingangsbild angewandt. [Dimopoulos u. Winkler, 2014] untersuchten, wie Menschen Wärme durch Farbtöne wahrnehmen, und testeten dabei verschiedene Wichtungen zur Modellierung der menschlichen Sensitivität gegenüber der Farbwahrnehmung in Abhängigkeit von Ent sättigung und Dunkelheit. Nach einer HSV-Konvertierung des Eingangsbilds erzielte dabei

$$w(\mathbf{x}) = s(\mathbf{x})v(\mathbf{x}) \in [0; 1] \quad (3.1)$$

das Resultat mit größter Übereinstimmung zur Kontrollgruppe. Dabei sind Sättigung s und Hellwert v in Abhängigkeit eines Pixels \mathbf{x} und im Bereich $[0; 1]$ anzugeben. Aus der Farbwichtung w , welche mit zunehmender Farbintensität (Farbwahrnehmung) wächst, kann über einen Schwellwert T_w die Menge der chromatischen Bildpunkte \mathbf{S} aus den Bildpunkten Ω bestimmt werden.

$$\mathbf{S} := \{\mathbf{x} \in \Omega \mid w(\mathbf{x}) > T_w\} \quad (3.2)$$

Unter Einbeziehung der trainierten Spielfeldfarbe kann die Menge \mathbf{N} der chromatischen Punkte in der Spielfeldfarbe notiert werden.

$$\mathbf{N} := \{\mathbf{x} \in \Omega \mid h(\mathbf{x}) \in [HueMin; HueMax]\} \cap \mathbf{S} \quad (3.3)$$

Der Farbwert (HSV) wird in diesem Fall mit der Funktion h ermittelt. Nun soll ein konvexes, maximal achteckiges Polygon gefunden werden, welches die größtmögliche Anzahl an Punkten, welche nicht zur Menge \mathbf{N} gehören, ausschließt und gleichzeitig möglichst viele Punkte aus \mathbf{N} einschließt. Sind alle Spielfeldpunkte innerhalb der Menge \mathbf{N} und keine Nicht-Spielfeldpunkte in \mathbf{N} , so entspricht dieses Polygon (nach Abschnitt 3.2) der

Spielfeldhülle. Wird mit \mathbf{C} eine Menge von Eckpunkten bezeichnet und ist Δ die symmetrische Differenz zweier Mengen, so kann dies als Optimierungsproblem beschrieben werden.

$$\#(\text{convexhull}(\mathbf{C}) \Delta \mathbf{N}) \xrightarrow{\mathbf{C}} \min!, \quad \text{mit } \#\eta(\mathbf{C}) \leq 8 \quad (3.4)$$

Dabei wird die konvexe Hülle auf den Eckpunkten \mathbf{C} (mit $\#\mathbf{C} \geq 1$) innerhalb des Bildraums Ω als Menge endlicher Konvexkombinationen definiert.

$$\text{convexhull}(\mathbf{C}) := \left\{ \sum_{i=1}^n w_i \mathbf{c}_i \in \Omega \mid n \in \mathbb{N}, \sum_{i=1}^n w_i = 1, \mathbf{c}_i \in \mathbf{C}, w_i \geq 0 \right\} \quad (3.5)$$

Die Hilfsgröße $\eta(\mathbf{C})$ beschreibt die Menge der Ecken, welche mindestens notwendig sind, um eine Hülle identisch zu $\text{convexhull}(\mathbf{C})$ zu erzeugen. Dies kann als

$$\begin{aligned} \mathbf{p}(\mathbf{C}) &:= \{ \mathbf{C}' \mid \text{convexhull}(\mathbf{C}) = \text{convexhull}(\mathbf{C}') \} \\ \eta(\mathbf{C}) &:= \{ \mathbf{C}' \in \mathbf{p}(\mathbf{C}) \mid \forall \mathbf{C}'' \in \mathbf{p}(\mathbf{C}) : \#\mathbf{C}' \leq \#\mathbf{C}'' \} \end{aligned} \quad (3.6)$$

definiert werden. Diese implizite Art der Definition macht Satz 3.2 erforderlich.

Satz 3.2. In einem kontinuierlichen Bildraum ist $\eta(\mathbf{C})$ wohldefiniert.

Beweis 3.2. Für jedes \mathbf{C} mit $\#\mathbf{C} \geq 1$ lassen sich konvexe Hüllen erzeugen, woraus neben der Eigenschaft $\mathbf{p}(\mathbf{C}) \neq \emptyset$ auch die Existenz von $\eta(\mathbf{C})$ folgt, da sich immer kleinste natürliche Zahlen finden lassen. Die Eindeutigkeit wird anhand einer Reduktion auf die minimale Eckenanzahl für eine Punktmenge \mathbf{C} besprochen. Da mit der Forderung nach Gleichheit der konvexen Hüllen in $\mathbf{p}(\mathbf{C})$ nur Punkte in \mathbf{C} entfallen können, welche sich durch Konvexkombinationen anderer Punkte in \mathbf{C} darstellen lassen (innere Punkte, Punkte auf Randstrecken) und die minimale Eckenanzahl sich damit aus der Anzahl an „äußeren“ Eckpunkten ergibt, ist mit der Eindeutigkeit auch die Wohldefiniertheit gezeigt. \square

In Anbetracht eines zu erwartenden Fehlers bei der Anpassung eines Polygons auf die Menge \mathbf{N} wird eine Kurzschreibweise für einen *symmetrischen Fehler der konvexen Hülle* (*SFH*) eingeführt.

$$\text{SFH}(\mathbf{C}, \mathbf{N}) := \#(\text{convexhull}(\mathbf{C}) \Delta \mathbf{N}) \quad (3.7)$$

Da diese Operation auch unter Berücksichtigung möglicher effizienter Implementierungen (siehe Abschnitt 3.9) aufwändig ist, wird die Menge zulässiger Ecken im Weiteren eingeschränkt, um eine echtzeitfähige Optimierung zu ermöglichen. Nach dem Auffüllen aller Löcher in \mathbf{N} und Bestimmung der Zusammenhangskomponenten $\mathbf{N}_{1:n}$, gilt es nun, zu jeder Komponente \mathbf{N}_i , mit $i \in \{1, \dots, n\}$, Ecken $\hat{\mathbf{C}}_i$ zu bestimmen. Diese Ecken werden in der Optimierungsheuristik als Kandidaten für die Ecken des Spielfeldpolygons dienen.

Die Suche nach Ecken wird über die Boundary Signature [Mustafa, 2001] durchgeführt. Dazu werden zunächst alle Punkte des Rands $\partial\mathbf{N}_i$ in Polarkoordinaten überführt, wobei der Schwerpunkt (\bar{x}, \bar{y}) von \mathbf{N}_i das Zentrum der Transformation bildet.

$$(x, y)^T \in \partial\mathbf{N}_i: \quad r(x, y) := \sqrt{(x - \bar{x})^2 + (y - \bar{y})^2} \quad (3.8)$$

$$\phi(x, y) := \text{atan2}(y - \bar{y}, x - \bar{x}) \quad (3.9)$$

Die Durchführung dieser Transformation für alle Punkte aus einem Rand $\partial\mathbf{N}_i$ führt zu einer Kurve wie in Abbildung 3.3 (A), der Boundary Signature.

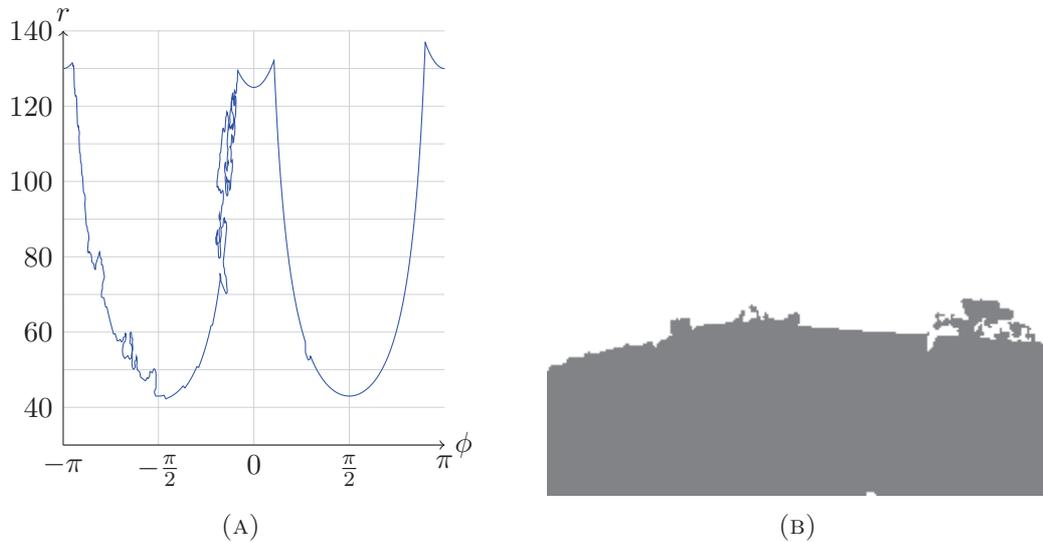


ABBILDUNG 3.3: Die Boundary Signature (A) zur Menge (B)

Satz 3.3. Ist $\mathbf{N}_i \subset \mathbb{R}^2$ konvex und $\#\eta(\mathbf{N}_i) \geq 3$, so ist $r \circ \phi^{-1}$ eine Funktion $\mathbb{R} \rightarrow \mathbb{R}$.

Beweis 3.3. $\mathbf{N}_i \neq \emptyset$ sei konvex. Für alle entsprechenden Mengen \mathbf{N}_i mit $\#\eta(\mathbf{N}_i) \geq 3$ ist der Schwerpunkt $(\bar{x}, \bar{y})^T$ ein Punkt der Menge $\mathbf{N}_i \setminus \partial\mathbf{N}_i$. Beginnend in $(\bar{x}, \bar{y})^T$ werden Strahlen durch jeden der Randpunkte $\partial\mathbf{N}_i$ bestimmt. Da $(\bar{x}, \bar{y})^T$ innerer Punkt ist, bedeutet dies zum einen, dass die Definitionslücke der atan2-Funktion nicht auftreten kann, und zum anderen, dass jeder Strahl den Rand $\partial\mathbf{N}_i$ nur in einem einzigen Punkt

schneiden kann, da sonst die Konvexitätsannahme direkt verletzt wäre. Dies heißt wiederum, dass sich jeder Randpunkt eindeutig einem Winkel zuordnen lässt, womit ϕ auch injektiv ist. Damit ist $r \circ \phi^{-1}$ eine Funktion.

□

Die Ausweitung von Satz 3.3 auf nicht-konvexe Mengen ist im Allgemeinen nicht richtig, wie Abbildung 3.3 zeigt. Dennoch stellt die gewonnene Erkenntnis die entscheidende Motivation zur Erkennung der Ecken einer Menge dar.

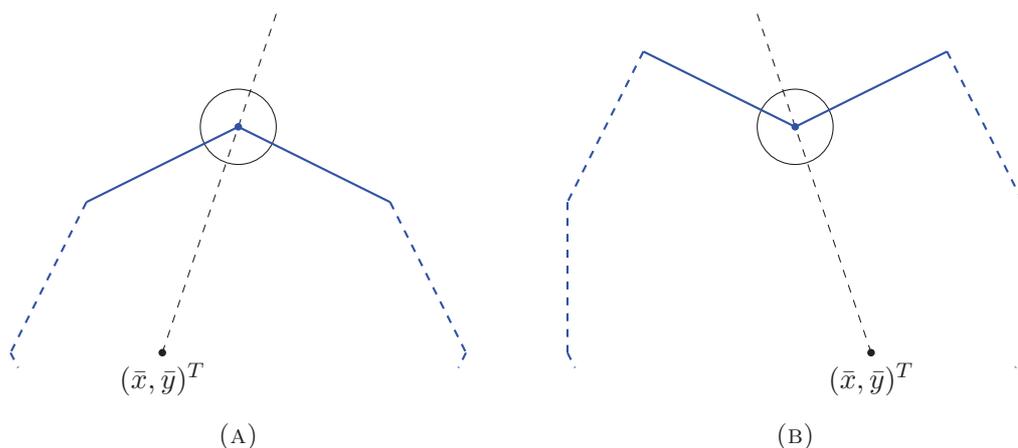


ABBILDUNG 3.4: Konvexe (A) und konkave (B) Eckpunkte werden bezüglich ihrer Innenwinkel unterschieden.

Eckpunkte lassen sich in die Kategorien konvex und konkav klassifizieren (siehe Abbildung 3.4). Konvexe Eckpunkte besitzen Innenwinkel von weniger als 180° , während bei Innenwinkeln von mehr als 180° konkave Eckpunkte vorliegen [Lee et al., 2011]. Liegt das Zentrum der Strahlen zu den Randpunkten, hier gewählt mit $(\bar{x}, \bar{y})^T$, geeignet innerhalb der Region, koinzidieren diese Kategorien mit lokalen Maxima und Minima der Boundary Signature (vgl. Abbildung 3.4). Da mit Satz 3.3 $r \circ \phi^{-1}$ für jede konvexe Ecke innerhalb einer Umgebung um jede Ecke eine Funktion ist, können konvexe Eckpunkte aus lokalen Maxima der Boundary Signature erkannt werden. Eben diese Maxima ergeben die Menge $\hat{\mathbf{C}}_i$.

Anmerkung 3. Mit dem vorgestellten Ansatz können konvexe Eckpunkte schnell gefunden werden. Zudem können Ecken, welche nicht entscheidend für die Gestalt sind, durch eine Approximation zur Boundary Signature leicht entfernt werden. Voraussetzung ist, dass der Beginn der Strahlen bzw. das Transformationszentrum $(\bar{x}, \bar{y})^T$ geeignet innerhalb der Menge \mathbf{N}_i liegt. Da in diesem Fall nach den konvexen Ecken des Spielfelds gesucht wird, ist es anzunehmen, dass in \mathbf{N}_i auf diese Weise wesentlich mehr ungeeignete Kandidaten für Ecken ausgeschlossen als sinnvolle Kandidaten nicht eingeschlossen werden.

Nachdem Kandidaten $\hat{\mathbf{C}} := \bigcup \hat{\mathbf{C}}_i$ für die Ecken der Hülle des Spielfelds berechnet sind, wird das eingeschränkte Problem zu Gleichung 3.4 gelöst.

$$\text{SFH}(\mathbf{C}, \mathbf{N}) \xrightarrow{\mathbf{C}} \min!, \quad \text{mit } \#\eta(\mathbf{C}) \leq 8 \wedge \mathbf{C} \subseteq \hat{\mathbf{C}} \quad (3.10)$$

Damit ist der Suchraum gegenüber einer Betrachtung von Kombinationen aus beliebigen Bildpunkten wesentlich verkleinert. Zur Lösung dieses Problems wird die folgende Greedy-Heuristik vorgeschlagen.

1. Auswahl einer zufälligen, bis zu acht-elementigen Teilmenge $\mathbf{C}_0 \subseteq \hat{\mathbf{C}}$.
2. Wiederholung der nachfolgenden Schritte für $i = 1, 2, \dots$
 - (a) $\mathbf{C}'_i := \eta(\mathbf{C}_{i-1})$.
 - (b) Für jeden Eckpunkt $\mathbf{c}' \in \mathbf{C}'_i$ wird der Eckpunkt in $\hat{\mathbf{C}}$ bestimmt, welcher bei einem Austausch die größte Verbesserung bezüglich der SFH der Menge erzielt.
$$\mathbf{c}' \in \mathbf{C}'_i: \arg \min(\mathbf{c} \in \hat{\mathbf{C}} \mapsto \text{SFH}(\mathbf{C}'_i \setminus \{\mathbf{c}'\} \cup \{\mathbf{c}\}, \mathbf{N})) \quad (3.11)$$
 - (c) Derjenige Eckpunkt $\mathbf{c}'_{opt} \in \mathbf{C}'_i$, für welchen (im gesamten Schritt i) die größte Verbesserung eingetreten ist, wird mitsamt seines Austausch-Eckpunkts $\mathbf{c}_{opt} \in \hat{\mathbf{C}}$ verwendet, um eine Verbesserung zu erzielen. $\mathbf{C}_i := \mathbf{C}'_i \setminus \{\mathbf{c}'_{opt}\} \cup \{\mathbf{c}_{opt}\}$.
 - (d) Der Algorithmus terminiert mit $\eta(\mathbf{C}_i)$, wenn $|\text{SFH}(\mathbf{C}_i, \mathbf{N}) - \text{SFH}(\mathbf{C}_{i-1}, \mathbf{N})|$ einen Schwellwert unterschreitet oder die Anzahl der Schritte eine Obergrenze erreicht.

In dieser Heuristik wird die Konvexität fortwährend durch die Funktion η sichergestellt. Dies garantiert zugleich, dass die zuvor festgelegte Zahl von acht Ecken nicht überschritten wird. Da der Algorithmus bei einer ungünstigen Startposition jedoch in frühen Schritten notwendige Eckpunkte verlieren kann, empfiehlt sich die mehrmalige Ausführung. Die beschriebene Iteration endet in der Praxis meist nach wenigen Schritten — ein solches Vorgehen kann deshalb vertreten werden.

Da ein Eckenaustausch auch mit identischen Ecken erfolgen kann, ist eine Verschlechterung nicht möglich. Somit muss es in jedem Schritt des Algorithmus hin zu einem lokalen Minimum eine Verbesserung des SFH geben oder den Abbruch.

3.4 Berechnung der Grasmask

Mit dem Wissen um die für ein Bild zur Verfügung stehende Hülle wird nun die Berechnung einer Grasmask thematisiert. Das Ziel ist es zunächst, Pixel binär zu klassifizieren, bevor in Abschnitt 3.5 auch Ergebnisse im Intervall $[0; 1]$ betrachtet werden. Hierbei wird der Fakt genutzt, dass sich das Spielgeschehen auf einem im Wesentlichen homogenen Grund abspielt. Wie in [Tong et al., 2004] dargelegt, eignet sich ein Helligkeitsbild selbst nicht für die Erkennung der Rasenfläche. Die abschnittsweise homogen gefärbten Rasenflächen besitzen jedoch eine hohe Korrelation der roten, grünen und blauen Farbkomponenten. Dies kann genutzt werden, um das Bild in dieser Region vereinfacht (einkanalig) zu betrachten und damit einhergehend den Aufwand zu reduzieren. Einen ersten Einblick hierzu bietet Abbildung 3.5.

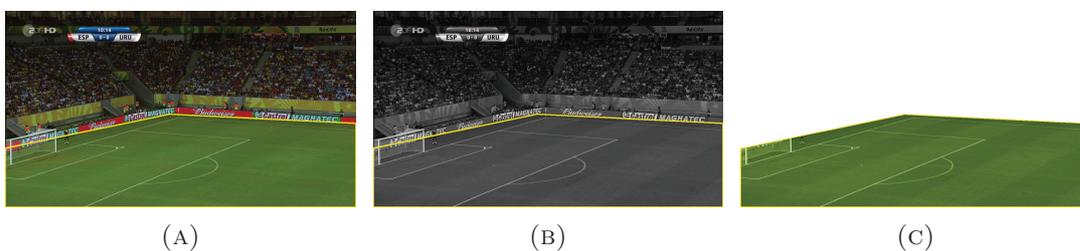


ABBILDUNG 3.5: Die Graspunkte innerhalb der Spielfeldhülle (C) können mit geringen Abweichungen zum Originalbild (A) aus einem einkanaligen Grauwertbild (B) durch eine Farbverschiebung wiederhergestellt werden. Bildquelle: [ZDF, 2013b]

Sei für ein Bild $\mathbf{I}: \Omega \rightarrow [0; 1]^3$, bestehend aus den drei Farbkanälen rot I_R , grün I_G und blau I_B ,

$$\forall \mathbf{x} \in \Omega: \mathbf{I}(\mathbf{x}) = (I_R(\mathbf{x}), I_G(\mathbf{x}), I_B(\mathbf{x}))^T \quad (3.12)$$

und das Helligkeitsbild $V: \Omega \rightarrow [0; 1]$,

$$V = \frac{1}{3}(I_R + I_G + I_B) \quad (3.13)$$

eine Menge $\mathbf{H} \subseteq \Omega$ homogener (Gras-)Pixel gegeben. Hierzu werden die empirischen Mittelwerte der Bilder \mathbf{I} und V berechnet.

$$\hat{\mu}_I = \frac{1}{\#\mathbf{H}} \sum_{\mathbf{x} \in \mathbf{H}} \mathbf{I}(\mathbf{x}) \quad (3.14)$$

$$\hat{\mu}_V = \frac{1}{\#\mathbf{H}} \sum_{\mathbf{x} \in \mathbf{H}} V(\mathbf{x}) \quad (3.15)$$

Mittels der Verschiebung $\boldsymbol{\tau} := \hat{\boldsymbol{\mu}}_I - \hat{\boldsymbol{\mu}}_V \mathbf{1}$, mit $\mathbf{1} = (1,1,1)^T$, wird nun ein Bild $\mathbf{V}_{\boldsymbol{\tau}}: \Omega \rightarrow [0; 1]^3$ in gleicher Weise wie in Abbildung 3.5 (C) erzeugt.

$$\mathbf{V}_{\boldsymbol{\tau}}(x) := V(x)\mathbf{1} + \boldsymbol{\tau} \quad (3.16)$$

Die lineare Abhängigkeit der Farbkanäle ist leicht einzusehen. Um schlussendlich den Fehler von $\mathbf{V}_{\boldsymbol{\tau}}$ gegenüber dem ursprünglichen Bild \mathbf{I} in der Menge \mathbf{H} zu bemessen, wird die korrigierte Stichproben-Kovarianzmatrix $\hat{\boldsymbol{\Sigma}}$ verwendet.

$$\hat{\boldsymbol{\Sigma}} := \text{Cov}_{\mathbf{x} \in \mathbf{H}} ((\mathbf{I}(x) - \hat{\boldsymbol{\mu}}_I) - (V(x)\mathbf{1} - \hat{\boldsymbol{\mu}}_V \mathbf{1})) \quad (3.17)$$

$$= \text{Cov}_{\mathbf{x} \in \mathbf{H}} (\mathbf{I}(x) - V(x)\mathbf{1} - \boldsymbol{\tau}) \quad (3.18)$$

Da, wie eingangs erwähnt und durch Abbildung 3.5 verdeutlicht, die Farbkanäle der Grasfläche in den allermeisten Fällen eine hohe Korrelation besitzen, können die Elemente außerhalb der Diagonalen in dieser Matrix bei der Berechnung vernachlässigt werden.

Die Berechnung der Menge \mathbf{H} homogener Punkte basiert auf den Punkten der konvexen Spielfeldhülle. Innerhalb dieser Menge werden alle Pixel bzgl. der Standardabweichungen in den drei Farbkanälen kategorisiert. Ein Pixel in der Hülle wird als homogen betrachtet, wenn das Maximum der Standardabweichungen aus allen Farbkanälen für eine Umgebung um dieses Pixel unterhalb des Durchschnitts aller solcher Umgebungen (wiederum mit kanalweisen Maxima) liegt. Pixel außerhalb der Hülle sind nicht Element von \mathbf{H} .

Aus dem normalisierten Histogramm $h(V(\mathbf{H}))_{1:n}$ mit n Einträgen entsteht die Indexmenge $\mathbf{B} \subseteq \mathbb{N}$ als kleinste Menge, für welche gilt:

$$\sum_{k \in \mathbf{B}} h(V(\mathbf{H}))_k > T_H \in [0; 1) \quad (3.19)$$

Der Schwellwert T_H bestimmt dabei die Anzahl $\#\mathbf{B}$ an Farben, welche notwendig sind, um die homogenen Punkte in der Hülle innerhalb des Helligkeitsbilds darzustellen. Ein Wert nah an 1 erscheint für T_H sinnvoll, um eine möglichst große Abdeckung jener Punkte sicherstellen zu können. 98% zeigt sich in empirischen Studien als erfolversprechend. Im Gegensatz zu ihrer Kardinalität ist die Menge \mathbf{B} durch die an Gleichung 3.19 gestellten Bedingungen im Allgemeinen nicht eindeutig festgelegt.

Während in diesem Ansatz die Anzahl der Farben des Eingangsbilds nicht reduziert wird (im Gegensatz etwa zu [Gedikli, 2009]), wird durch die Verschiebung des Helligkeitsbilds

eine effiziente Beschreibung der Grasfarben durch die Menge \mathbf{B} ermöglicht. Die Werte dieser Menge werden als Grautöne auf einer von schwarz nach weiß führenden Strecke in den RGB-Würfel übertragen und anschließend um $\boldsymbol{\tau}$ verschoben. Nach dieser Verschiebung in den Farbmittelpunkt des Eingangsbilds werden auftretende Fehler durch Ellipsoide um die erkannten Grasfarben modelliert. Die Halbachsen werden durch $\hat{\boldsymbol{\Sigma}}$ bestimmt. In Abbildung 3.6 wird dies beispielhaft an einem Bild mit zwei voneinander getrennten Grünschattierungen illustriert.

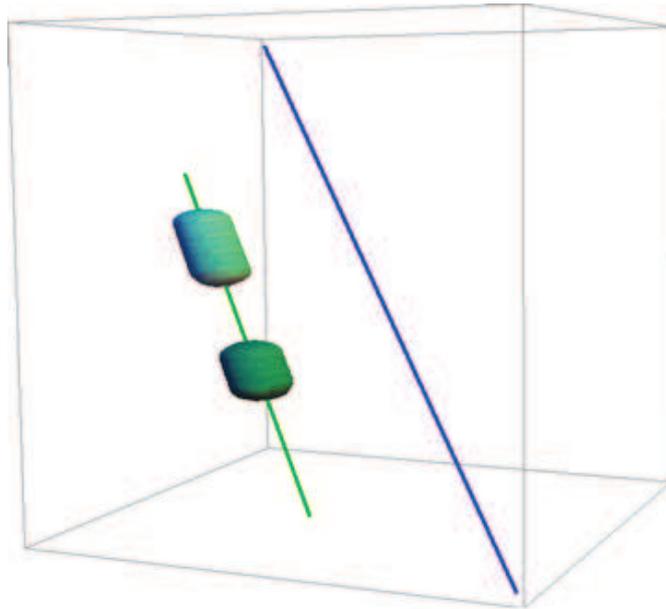


ABBILDUNG 3.6: Grasfarben innerhalb des RGB-Würfels. Nach einer $\boldsymbol{\tau}$ -Verschiebung (grüne Strecke) der von schwarz zu weiß verlaufenden Diagonalen (blaue Strecke) werden die Grasfarben zu Ellipsoiden extrudiert. Die Größe der Extrusion berücksichtigt den mit $\hat{\boldsymbol{\Sigma}}$ bemessenen Fehler der Annäherung durch einen Farbkanal.

Sei ein Pixel $\boldsymbol{x} \in \boldsymbol{\Omega}$ zur Klassifikation gegeben. Für dieses Pixel kann anhand seines Farbvektors $\boldsymbol{I}(\boldsymbol{x})$ geprüft werden, ob es sich innerhalb der Ellipsoide befindet und es sich somit um einen Graspunkt handelt.

$$\mathbf{1}^T(\boldsymbol{\tau} + \lambda \mathbf{1} - \boldsymbol{I}(\boldsymbol{x})) = 0 \quad \lambda \in \mathbb{R} \quad (3.20)$$

$$\mathbf{1}^T(\boldsymbol{\tau} + \lambda \mathbf{1}) - \mathbf{1}^T \boldsymbol{I}(\boldsymbol{x}) = 0 \quad (3.21)$$

$$3\lambda = \mathbf{1}^T \boldsymbol{I}(\boldsymbol{x}) - \mathbf{1}^T \boldsymbol{\tau} \quad (3.22)$$

$$y(\boldsymbol{I}(\boldsymbol{x})) := \frac{1}{3} (\mathbf{1}^T \boldsymbol{I}(\boldsymbol{x}) - \mathbf{1}^T \boldsymbol{\tau}) \quad (3.23)$$

Die mit y identifizierte Position auf der um $\boldsymbol{\tau}$ verschobenen Diagonalen wird weiter auf die Bereiche innerhalb der Indexmenge \mathbf{B} eingeschränkt.

$$y'(\mathbf{I}(\mathbf{x})) := \arg \min_{k \in B} |k - y(\mathbf{I}(\mathbf{x}))| \quad (3.24)$$

Zwischen der mit diesem Ergebnis assoziierten Farbe $y'(\mathbf{I}(\mathbf{x}))\mathbf{1} + \boldsymbol{\tau}$ und $\mathbf{I}(\mathbf{x})$ wird die Mahalanobis-Distanz $d_{\hat{\boldsymbol{\Sigma}}}$ berechnet, womit dem zuvor erwähnten Fehler Rechnung getragen wird.

$$d_{\hat{\boldsymbol{\Sigma}}}(\mathbf{I}(\mathbf{x}), y'(\mathbf{I}(\mathbf{x}))\mathbf{1} + \boldsymbol{\tau}) = \sqrt{(\mathbf{I}(\mathbf{x}) - \boldsymbol{\tau} - y'(\mathbf{I}(\mathbf{x}))\mathbf{1})^T \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{I}(\mathbf{x}) - \boldsymbol{\tau} - y'(\mathbf{I}(\mathbf{x}))\mathbf{1})} \quad (3.25)$$

Durch Verwendung eines Schwellwerts $T_b > 0$ für $d_{\boldsymbol{\Sigma}}$ kann die Grasmaske für das gesamte Bild, und somit auch für Punkte außerhalb der zuvor berechneten Hülle, bereitgestellt werden. Ein Beispiel einer auf diese Weise erzeugten binären Maske ist in Abbildung 3.7 dargestellt.

$$\mathbf{x} \in \text{Grasmaske}(\mathbf{I}) \Leftrightarrow d_{\hat{\boldsymbol{\Sigma}}}(\mathbf{I}(\mathbf{x}), y'(\mathbf{I}(\mathbf{x}))\mathbf{1} + \boldsymbol{\tau}) < T_b \quad (3.26)$$

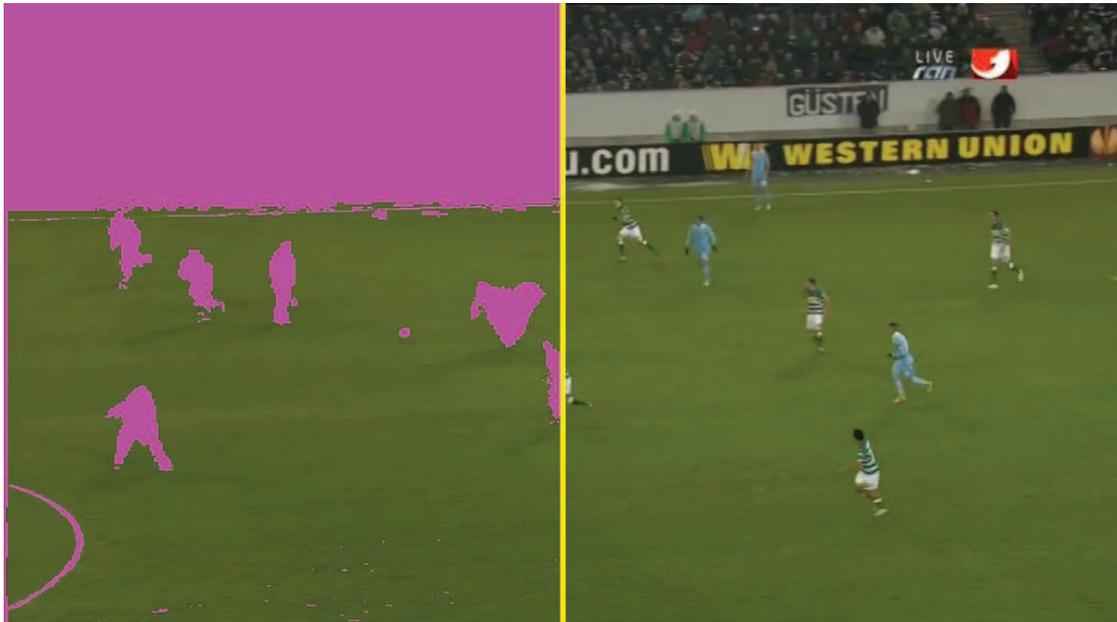


ABBILDUNG 3.7: Ergebnis der binären Grasmaskierung. Punkte außerhalb der Grasmasken sind magenta abgebildet. Bildquelle: [Kabel eins, 2013]

3.5 Ergänzungen zur Grasmaskerzeugung

Die im vorigen Abschnitt bereitgestellte binäre Grasmaskerzeugung hat durch die Wahl der Mahalanobis-Distanz eine Normierung erfahren, die die Abweichungen des Helligkeitsbilds gegenüber dem Eingangsbild berücksichtigt. Eine allgemeine Regel zur Festlegung eines geeigneten Schwellwerts lässt sich hieraus jedoch noch nicht ableiten. Da mit der Menge \mathbf{H} der homogenen Punkte jedoch bereits eine Stichprobe für die Distanzen vorliegt, wird ein Verfahren zur adaptiven Steuerung des Schwellwerts T_b aus den bisher berechneten Größen formuliert. Zuvor wird die Grasmaskerzeugung um kontinuierliche Wertebereiche erweitert, deren Parameter in Abschnitt 3.5.2 auch adaptiv bestimmt werden.

3.5.1 Kontinuierliche Wertebereiche



ABBILDUNG 3.8: Grasmaskerzeugung mit binärem Schwellwert (links) und kontinuierlichem Wertebereich (rechts)

Anstatt eine binäre Klassifikation durch einen einzigen Schwellwert wie in Gleichung 3.26 durchzuführen, kann es sinnvoll sein, basierend auf der Distanz $d_{\mathbf{Z}}$, mehrere Schwellwerte festzulegen. Während binäre Masken eindeutig klassifizieren, werden mit dem Ziel der Berücksichtigung von Farbüberlagerungen auch Unsicherheiten zwischen den Klassen berücksichtigt (vgl. Abbildung 3.8). Zu diesem Zweck wird T_a als unterer Schwellwert und T_b als oberer Schwellwert mit $0 < T_a < T_b$ festgelegt. Allen Distanzen $d_{\mathbf{Z}}$ kleiner T_a wird in der Grasmaskerzeugung mit kontinuierlichem Wertebereich $[0; 1]$ der Wert 0 zugeordnet, allen Distanzen größer T_b wird der Wert 1 zugeordnet. Anstatt die Distanzen zwischen T_a und T_b linear auf $[0; 1]$ abzubilden, wird die Fehlerfunktion

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (3.27)$$

verwendet. Transformiert auf die angegebenen Grenzen heißt dies:

$$\operatorname{erf}_{T_a, T_b}(x) := \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\sqrt{\pi} \frac{T_a + T_b - 2x}{T_a - T_b} \right) \quad (3.28)$$

Hierzu nun einige Anmerkungen:

- Die Schwellwerte T_a und T_b werden genauso wie die transformierte Fehlerfunktion erf_{T_a, T_b} auf den Werten der Distanz $d_{\hat{\Sigma}}(\mathbf{I}(\mathbf{x}), y'(\mathbf{I}(\mathbf{x}))\mathbf{1} + \tau)$ angewandt. Diese korrigiert die Metrik des Raums entsprechend der Matrix $\hat{\Sigma}$. In diesem Sinn sind T_a und T_b als Vielfache einer „gemittelten Standardabweichung“ zu betrachten.
- Die Verwendung der Fehlerfunktion erf begründet sich in der höheren Auflösung bzw. einer Kontrastausdehnung an den Rändern T_a und T_b . Weiterhin ist diese Funktion in vielen Programmiersprachen direkt verfügbar.
- Die subpixelgenaue Extraktion von Spielern erfordert qualitativ hochwertiges Eingangsmaterial.

3.5.2 Adaptive Schwellwerte

Bereits in Abschnitt 3.4 wurde diskutiert, dass sich die Mahalanobis-Distanz geometrisch in Form von extrudierten Ellipsoiden im RGB-Würfel interpretieren lässt. An dieser Stelle wird dieser Ansatz zusätzlich erweitert, indem jeder der Ellipsoide individuell extrudiert wird. Konkret heißt dies, dass für jeden Punkt der homogenen Menge \mathbf{H} markiert wird, mit welchem Element der Menge \mathbf{B} eine minimale Distanz $y'(\mathbf{I}(\mathbf{x}))$ (Gleichung 3.24) eintritt. Unter Einbeziehung des Schwellwerts $T_H \in [0; 1)$ (Gleichung 3.19) kann die Extrusionsgröße für jedes Element $k \in \mathbf{B}$ berechnet werden. Dazu wird zunächst die Menge \mathbf{A}_k aller dem Element k zugeordneten Punkte bestimmt.

$$\mathbf{x} \in \mathbf{A}_k \subseteq \mathbf{H} \Leftrightarrow y'(\mathbf{I}(\mathbf{x})) = k \in \mathbf{B} \quad (3.29)$$

Für die Bestimmung einer geeigneten Distanz für die mit k assoziierte Farbe wird das kleinste T_k bestimmt, welches die Bedingung

$$\sum_{\substack{k \in \mathbf{A}_k \\ d_{\hat{\Sigma}}(\mathbf{I}(\mathbf{x}), k\mathbf{1} + \tau) < T_k}} 1 > T_H * \#\mathbf{A}_k \quad (3.30)$$

erfüllt, und damit den Schwellwert beschreibt, welcher gerade groß genug ist, um einen Anteil größer T_H von A_k zu erfassen.

Es ist vorzuziehen, nicht jedes Element von \mathbf{B} einzeln zu betrachten, sondern benachbarte Werte zu Intervallen zu verbinden. \mathbf{B} geht aus einem Histogramm mit endlicher Klassenzahl hervor und die Elemente von \mathbf{B} besitzen eine natürliche Halbordnungsrelation, die dies erlaubt. Die Suche und Sortierung von Punkten muss so weniger oft

ausgeführt werden. Im Falle einer binären Maske wird der Schwellwert $T_b = T_k$ verwendet, für kontinuierliche Masken haben sich die Werte $T_a = 0,75T_k$ und $T_b = 2T_k$ als passend erwiesen.

3.6 Stand der Forschung

Die Bestimmung der Punkte einer Spielfläche in einem Bild besitzt Bedeutung für eine Vielzahl von sportspezifischen Verfahren in der Bildverarbeitung, hier im Sinn einer Schätzung des Hintergrunds. Eine Reihe von Verfahren nutzen vordefinierte Farbwerte und ordnen Punkte anhand von Schwellwerten zu. Zu nennen sind hier u. a. [Tong et al., 2004] und [Utsumi et al., 2002]. Andere wiederum bestimmen die Feldfarbe automatisch durch Ermittlung globaler Maxima in Histogrammen, wie etwa [Ekin et al., 2003, Seo et al., 1997, Yu u. Tian, 2003]. Diese Möglichkeit fand aufgrund seltener Szenarien, in welchen das Spielfeld im Bild kaum sichtbar ist, im vorgestellten Ansatz keine Berücksichtigung. ASpoGAMo [Gedikli, 2009] nutzt ein Gaussian Mixture Model (GMM) zur Farbraummodellierung in einem auf 16-Bit pro Farbkanal reduzierten Farbraum, kann damit jedoch nur Spielflächen mit einer einzigen Grasfarbe automatisch erkennen. Eine Gemeinsamkeit der Ansätze ist dabei, dass die Suche nach Graspunkten mit einer Schwellwertbildung oder, wie bei [Yu u. Tian, 2003], mit einem Region-Growing-Verfahren ihren Abschluss finden. Auf charakteristische Eigenschaften wie Konvexität wird keine Rücksicht genommen. Die oft zitierte und bildweise anwendbare Arbeit von [Ekin et al., 2003] wird in der Evaluation Anwendung finden.

Einen Ansatz auf Basis einer größeren Anzahl von Bildern verwenden sowohl [Figueroa et al., 2006] als auch [Schlipsing et al., 2013] und [De Marco et al., 2013]. Letztere verwenden eine Hintergrundsubtraktion, um eine Erkennung von Spielfeldobjekten im weiteren Verlauf zu ermöglichen. Dazu werden punktweise und mit exponentiellen Lernraten Erwartungswerte und Standardabweichungen geschätzt (und approximiert). Weicht ein Pixel um mehr als die doppelte Standardabweichung von seinem erwarteten Wert ab, so wird er als Vordergrund behandelt. [Schlipsing et al., 2013] verwenden zusätzlich ein ortsunabhängiges Modell, um Fehler, etwa durch Witterungseinflüsse, zu verringern. Der Ansatz von [Figueroa et al., 2006] basiert ebenso auf der Betrachtung von einzelnen Pixeln über die Zeit hinweg, um eine Aussage über den Bildhintergrund zu treffen. Hier werden morphologische Operationen auf dem 1D-Signal der Intensität in Abhängigkeit von der Zeit angewandt, um Abweichungen vom Hintergrundlevel zu identifizieren. Wie für die Ansätze von [De Marco et al., 2013, Schlipsing et al., 2013] findet auch hier eine statische Kamera Anwendung. Damit besteht ein klarer Nachteil zum vorgestellten Verfahren, welches dieser Einschränkung nicht unterliegt.

Einen deutlich anderen Weg beschreiten [Lu et al., 2013] im Kontext Basketball. Da in Aufzeichnungen dieser Sportart das Feld oft zu großen Teilen gezeigt wird und markante Werbebereiche enthält, ist es möglich, eine Kameraparameter-Nachführung (Kapitel 6), welche eine Abbildung zwischen Bild und Spielfeldmodell herstellt, auch ohne vorherige Farberkennung robust zu nutzen. Aufgrund einer geringen Anzahl sichtbarer Spielfeldlinien im Kontext Fußball wurde auf ein Vorgehen dieser Art verzichtet.

3.7 Evaluation

Für die Bereitstellung eines Testdatensatzes wurden aus sechs Bewegungen insgesamt 60 Bilder zufällig ausgewählt. Dabei wurden Bilder, welche das Spielfeld nicht in der Totalen zeigten, zuvor ausgeschlossen. Um auf möglichst vielfältigen Daten testen zu können, wurden TV-Aufzeichnungen von verschiedenen nationalen und internationalen Begegnungen auf unterschiedlichen Spielniveaus ausgewählt. Eine detaillierte Übersicht zu den verwendeten Aufzeichnungen ist Tabelle 3.1 zu entnehmen. Insbesondere ist darauf hinzuweisen, dass die Begegnung *Bayern* unter künstlicher Beleuchtung stattgefunden hat, während *S’Korea* und *Norwegen* in der Mittagssonne ausgetragen wurden. In der Begegnung *B’Faso* war der Rasen aufgrund starker Regenfälle am Vortag in ungewöhnlich schlechtem Zustand.

Kurzname	Begegnung	Liga / Pokal
Bayern	FC Bayern - FC Arsenal	UEFA Champions League 2013
B’Faso	Burkina Faso - Ghana	Africa Cup of Nations 2013
M’Gladbach	B. Mönchengladbach - Lazio Rom	UEFA Europa League 2013
S’Korea	Südkorea - Griechenland	FIFA WM 2010
Leipzig	RB Leipzig - SF Lotte	Regionalliga 2013
Norwegen	Norwegen - Äquatorialguinea	FIFA WM der Frauen 2011

TABELLE 3.1: Übersicht der Begegnungen, welche im Testdatensatz Berücksichtigung fanden

In jedem Bild des Testdatensatzes wurde manuell die Spielfeldhülle sowie die Menge der Graspunkte annotiert. Für die Durchführung kam das Bildbearbeitungsprogramm GIMP [Kimball et al., 2013] zum Einsatz. Mehreren annotierenden Personen wurden die in der Einleitung dargelegten Fragen als Aufgabenstellung erklärt. Um den Aufwand für die Mitarbeiter zu verringern, wurden vorgefertigte Masken, welche mit einer früheren Version des Algorithmus berechnet wurden, zur Bearbeitung angeboten.

Für eine Gegenüberstellung des vorgestellten Verfahrens mit einer Referenz wurde eine erneute Implementation der Graserkennung nach [Ekin et al., 2003] durchgeführt. Da dieser Ansatz eine manuelle Farbkalibrierung und Schwellwertauswahl notwendig macht, wurden diese Werte individuell für jede Begegnung trainiert. Aufgrund des Aufbaus ohne Berechnung einer Hülle muss weiterhin auf einen Vergleich in dieser Disziplin verzichtet werden.

Begegnung	Recall	Precision	F1-Score
Bayern	0,983	0,997	0,990
B'Faso	0,984	0,997	0,991
M'Gladbach	0,980	0,995	0,988
S'Korea	0,972	0,998	0,985
Leipzig	0,981	0,996	0,988
Norwegen	0,985	0,998	0,991

TABELLE 3.2: Punktweise Auswertung der Hüllendetektion in Relation zur Ground Truth

Die in Tabelle 3.2 dargestellten Ergebnisse zur Hüllenerkennung zeigen ein einheitliches Bild im gesamten Testdatensatz. Von 1 000 betrachteten Ground-Truth-Pixeln weichen im Mittel nur Einzelne ab. Fehler an den Rändern der Hülle fallen in Anbetracht der Kardinalität der Menge geringfügiger ins Gewicht.

Begegnung	adaptiver Schwellwert			trainierter Schwellwert			[Ekin et al., 2003]		
	R	P	F1	R	P	F1	R	P	F1
Bayern	0,968	0,991	0,979	0,961	0,996	0,978	0,970	0,994	0,982
B'Faso	0,983	0,985	0,984	0,978	0,996	0,987	0,960	0,967	0,964
M'Gladbach	0,980	0,985	0,983	0,978	0,986	0,982	0,986	0,976	0,981
S'Korea	0,958	0,985	0,971	0,974	0,975	0,974	0,767	0,911	0,833
Leipzig	0,980	0,957	0,969	0,967	0,991	0,979	0,997	0,811	0,895
Norwegen	0,992	0,987	0,989	0,994	0,951	0,972	0,527	0,989	0,688

TABELLE 3.3: Ergebnisse der Graserkennung des vorgestellten Verfahrens unter adaptiven und trainierten Schwellwerten sowie der Referenz [Ekin et al., 2003]. Gegenübergestellt werden (R)ecall, (P)recision und (F1)-Score. Die besten F1-Scores zu jeder Begegnung sind farblich hervorgehoben.

Für den Vergleich des vorgestellten Verfahrens mit Ground-Truth-Daten wurde auf binäre Maskenbildung zurückgegriffen. Hierbei kam sowohl die in Abschnitt 3.5.2 beschriebene adaptive Schwellwertbildung als auch ein zuvor trainierter, dann konstanter Schwellwert zur Anwendung. Die zu minimierende Zielfunktion während des Trainings

war der F1-Score des gesamten Testdatensatzes. Gemäß der in Tabelle 3.3 aufbereiteten Auswertung ergibt sich aus diesem Training jedoch kein entscheidender Vorteil gegenüber dem adaptiven Verfahren, welches für die Begegnungen *M'Gladbach* und *Norwegen* den besten F1-Score erreicht und sich für die Begegnungen *B'Faso*, *S'Korea* und *Leipzig* einem trainierten Schwellwert nur knapp geschlagen geben muss. Das Verfahren nach [Ekin et al., 2003] kann einzig in der Begegnung *Bayern* einen größeren F1-Score als das vorgestellte Verfahren erreichen. Aufgrund der künstlichen Beleuchtung besitzen die Farbwerte des Spielfelds im vorliegenden Video nur geringe Varianzen, was diesem Ansatz entgegenkommt. Die zuvor angesprochene Mittagssonne in den Begegnungen *B'Faso* und *Norwegen* bereiten diesem Verfahren jedoch Probleme, da die Spielfläche aufgrund von Rasenstreifen und Schattenwurf in vier Farb-Cluster zerfällt. Der vorgestellte Ansatz ist dennoch in der Lage, eine hohe Anzahl von Punkten korrekt zuzuordnen (siehe Abbildung 3.9) und die verschiedenfarbigen Bereiche individuell zu erkennen.

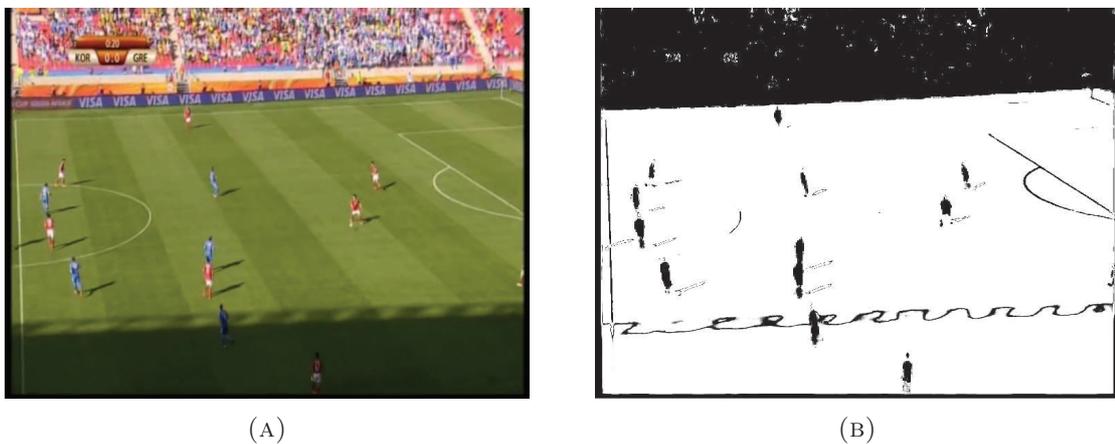


ABBILDUNG 3.9: Aufgrund von Rasenstreifen und Schattenwurf können vier separate Farbbereiche für den Rasen im Eingangsbild (A) der Begegnung *S'Korea* beobachtet werden. In der Grasmasken (B) sind diese Bereiche markiert (weiß).

Bildquelle: [ARD/SWR, 2010]

In weiteren Betrachtungen zeigt sich, dass der präsentierte Ansatz insbesondere bei überbelichteten Aufnahmen zu einem erhöhtem Fehler neigt. Bilder wie in Abbildung 3.10 (A) führen zuerst zu fehlerhaften Hüllen durch verfälschte Farbwerte innerhalb des chromatischen Bereichs und durch Einblendungen innerhalb des Spielfeldbereichs. In der nachfolgenden Berechnung der Grasmasken kann es weiterhin vorkommen, dass die berechneten Varianzen innerhalb der Matrix $\hat{\Sigma}$ (Gleichung 3.18), aufgrund der durch die Hülle vorgegebenen ungünstigen Stichprobe, deutlich zu groß geschätzt werden (Abbildung 3.10, B). Für entsprechende Fälle empfiehlt sich eine manuelle Wahl der Schwellwerte.



ABBILDUNG 3.10: Überbelichtete Eingangsbilder (A) stellen ein Problem für die Graserkennung (B) dar. Bildquelle: [Sport1, 2013a]

3.8 Anwendungen

Die Effektivität der Erkennung von Pixeln, welchen Spielfeldfarbe zugewiesen ist, wird anhand zweier Beispiele vorgestellt. In einer ersten Anwendung wird das Verfahren zur Berechnung der Grasmasken (und damit auch die Berechnung der Hülle) auf andere Sportarten angewandt. Die Abbildungen 3.11 und 3.12 zeigen die Ergebnisse für Hallenfußball und Handball, wobei bei Letzterem der Bereich $[HueMin; HueMax]$ der Feldfarbe speziell für diesen Fall angepasst wurde. Auch wenn Werbebanner im Spielfeld für nachfolgende Schritte, etwa bei einer Erkennung von Feldspielern, zu berücksichtigen sind, zeigt das Resultat, dass dieses Verfahren prinzipiell auch auf andere Feldsportarten anwendbar ist.

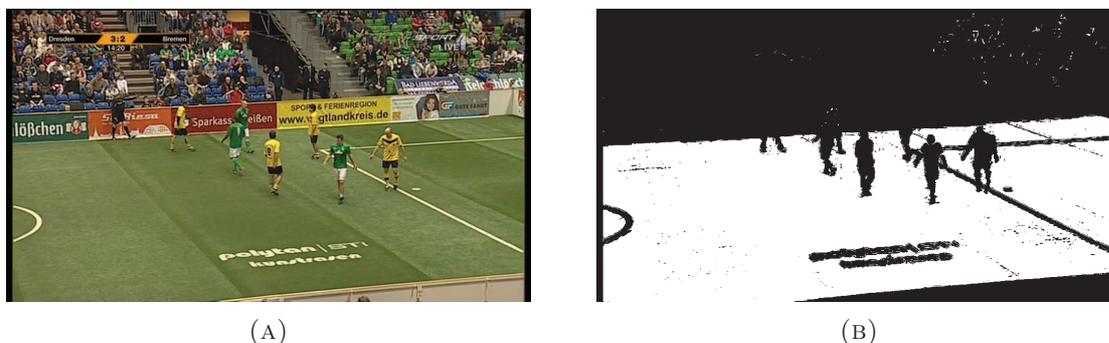


ABBILDUNG 3.11: (A) Hallenfußball, (B) Binäre Grasmasken. Bildquelle: [Sport1, 2013b]

Als zweite Anwendung wird die Klassifikation von Körperteilen auf Basis der hier vorgestellten Grasmasken betrachtet (siehe Abbildung 3.13). [Bigontina et al., 2014] verwenden monokulare Einzelbilder von Fußballspielern mit geringer Auflösung und untersuchen, welche Merkmale sich innerhalb eines Random Forests [Criminisi et al., 2012] am besten für die Bestimmung der Orientierung eines Spielers und die Bestimmung der Positionen seiner Körperteile (pixelweise) eignen. Wie sich zeigt, ist allein die vorgeschlagene Grasmasken für die Orientierung das Merkmal mit dem besten Klassifikationsergebnis. Im Fall der Klassifikation von Körperteilen erreicht eine Kombination von Merkmalen aus

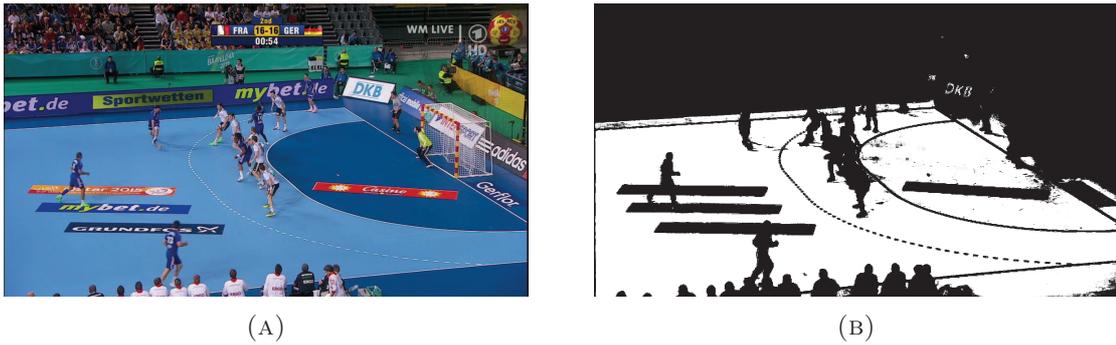


ABBILDUNG 3.12: (A) Handball, (B) Binäre „Gras“-maske mit angepasstem $[HueMin; HueMax]$ -Intervall innerhalb der Berechnung der Hülle. Bildquelle: [ARD/WDR, 2013b]

Grasmaske, Orientierung, relativer Pixelposition und Farbe das akkurateste Ergebnis mit 90,32% übereinstimmender Punkte zur Ground Truth bei 14 Körperteil-Klassen. Ausgehend von diesen Daten wird im Weiteren auch die Spielerpose berechnet.

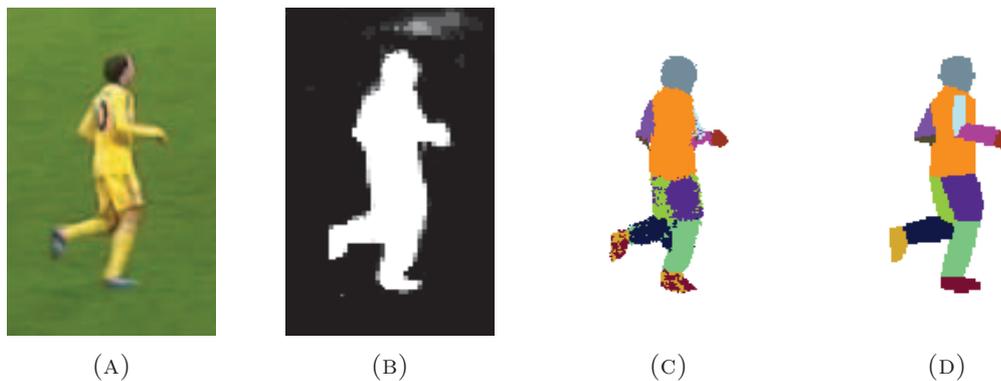


ABBILDUNG 3.13: Körperteilklassifikation. (A) Originalbild, (B) Grasmaske, (C) Klassifikation nach [Bigontina et al., 2014], (D) Ground Truth. Bildquelle: [Bigontina et al., 2014]

3.9 Diskussion und Ausblick

Sowohl für die berechneten Hüllen als auch die Grasmasken konnte in der Evaluation eine hohe Übereinstimmung mit dem manuell annotierten Ground-Truth-Datensatz nachgewiesen werden. Auch die F1-Scores der Referenz von [Ekin et al., 2003] wurden in den meisten Szenarien überboten, teils sogar deutlich, wenn die Umweltbedingungen anspruchsvoller wurden, etwa bei Sonneneinstrahlung. Einschränkend wurde aufgezeigt, dass auch das vorgestellte Verfahren in diesem Bereich an seine Grenzen stoßen kann, etwa bei überbelichteten Bildern. Die Differenzen zwischen den F1-Scores der automatischen und trainierten Schwellwertbildung fallen kaum ins Gewicht, was für die Qualität des adaptiven Schwellwerts spricht.

Damit die vorgestellten Verfahren in Echtzeit arbeiten können, wurden verschiedene Konzessionen gemacht. Beispielsweise wird die Menge \mathbf{N} (Gleichung 3.3), neben der Beachtung chromatischer Punkte, durch eine Teilmengenbildung in Abhängigkeit vom Farbwert gebildet. Dies erlaubt die Speicherung und Verarbeitung, insbesondere für die Bildung der symmetrischen Differenz, in Form einer Lauflängenkodierung. Alternativ wurde auch eine Schwellwertfunktion getestet, welche auf einen kontinuierlichen Wertebereich abbildet. Um die symmetrische Differenz SFH dennoch effizient berechnen zu können, wurde ein Verfahren auf Basis von Integralbildern implementiert, welches möglichst große Bildteile in möglichst wenige Rechtecke zerlegt — immer unter der Voraussetzung, dass ein konvexes Polygon (mit maximal acht Ecken) ausschließlich in Rechtecke zu zerlegen ist. Die SFH ist durch Integralbilder in den Rechtecken effizient zu berechnen. Da neben der initialen Bereitstellung der Integralbilder jedoch auch die Zerlegung in Rechtecke berechnet werden muss, wurde dieser Ansatz aufgrund nur geringer Vorteile der Genauigkeit erkannter Hüllen nicht weiter verfolgt.

In dem bisherigen Ansatz zur Berechnung der Grasmasken wird die Grauwert-Diagonale im RGB-Würfel entsprechend dem berechneten Versatz τ in den Bereich der Grasfarbe verschoben. Ist es der Fall, dass mehrere separate Grasfarben auftreten, hat diese Verschiebung den Nachteil, dass nicht berücksichtigt wird, falls diese eine andere Differenz haben als ein Vielfaches von $\mathbf{1}$, dem Richtungsvektor der Diagonalen. Dadurch kann es notwendig werden, dass eine Extrusion größere Ausmaße annimmt, als notwendig ist, und damit Pixel in die Grasmasken einbezogen werden, welche nicht berücksichtigt werden sollten. Eine Möglichkeit, diesem Problem zu begegnen, kann es sein, die beschriebene τ -verschobene Diagonale durch eine Kurve zu ersetzen. Diese könnte glatt zwischen den Punkten $\mathbf{0} = (0,0,0)^T$ und $\mathbf{1}$ verlaufen und auf ihrem Weg das Zentrum jedes Farbclusters im RGB-Würfel passieren. Die Cluster-Zentren könnten für eine schnelle Berechnung aus dem einfarbigen Bild bestimmt und anschließend im RGB-Würfel positioniert werden. Hierzu müssten alle Punkte, welche zu einem Cluster gehören, im Farbbild ein zweites Mal gemittelt werden. Eine individuelle Extrusion wäre dann nicht mehr von einer linearen Anordnung der Farben betroffen.

Kapitel 4

Erkennung und Klassifikation von Schnitten

Dieses Kapitel basiert methodisch u. a. auf:

HOERNIG, Martin ; HERRMANN, Michael ; RADIG, Bernd:
Multi Temporal Distance Images for Shot Detection in Soccer
Games. In: EUSIPCO 2014 (22nd European Signal Processing
Conference 2014). IEEE, 2013, S. 2420-2424

*First published in the Proceedings of the 22nd European Signal
Processing Conference (EUSIPCO 2014) in 2014, published by
EURASIP.*

4.1 Einleitung

Fernsehbilder werden praktisch bei jeder TV-Ausstrahlung mit verschiedenen Kameras aufgenommen und durch einen Regisseur zu einer Übertragung mit wechselnden Blickwinkeln arrangiert. Bei Fußballübertragungen werden so Handlungsabläufe aufbereitet und Spannungsbögen intensiviert. Hierzu dienen Rückblenden bedeutender Ereignisse, Blicke ins Publikum und auch Nahaufnahmen von Spielern. Ebenso ist außerhalb der Spielzeit mit Expertengesprächen oder Werbeunterbrechungen zu rechnen. Für eine automatische visuelle Analyse stellt sich zuerst die Frage, welche Bilder einer Aufzeichnung



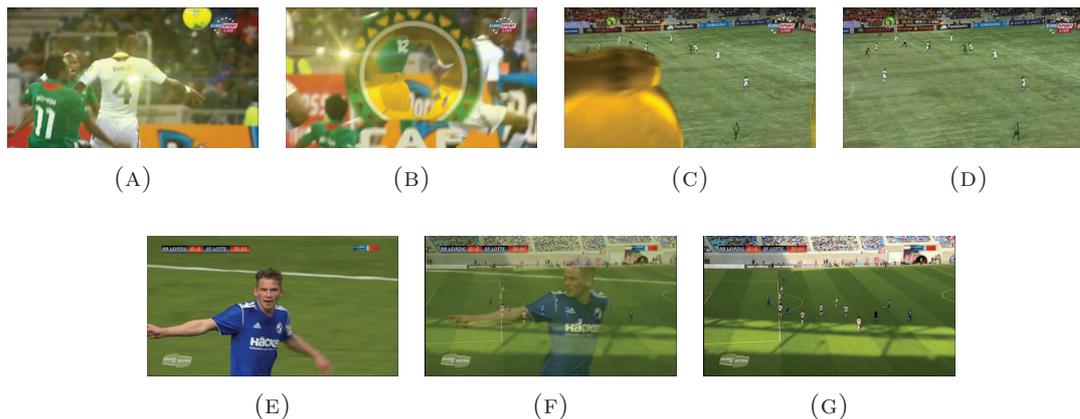


ABBILDUNG 4.1: Zwei weiche Übergänge in Fußballübertragungen. (A)-(D) Effekt-Übergang, (E)-(G) Überblendung. Bildquellen: [Eurosport, 2013, MDR, 2013]

verwertbare Informationen tragen. Nachdem Aufnahmen, welche das Spielfeld nicht zeigen, offensichtlich ausscheiden, sind auch Nahaufnahmen zu verwerfen, da Spieler in diesen zumeist nicht mit ausreichender Genauigkeit lokalisiert werden können. Relevanter Inhalt ist folglich den Bildern aus der Totalen zu entnehmen. Diese erfasst im Fußball die Hauptkamera, meist bezeichnet als *Kamera-1*. Beispiele für diese Sicht bieten Abbildung 4.1 (D) und (G). Je nach Bedeutung einer Begegnung kann der Kamera-1-Bereich durch mehrere Kameras besetzt oder die Totale durch weitere Blickwinkel erfasst werden (z. B. zwei Kamera-16, jeweils seitlich erhöht). In diesem Kapitel steht nur die Erkennung von Bildern der seitlichen Totalen (in aller Regel Kamera-1, darum kurz: Kamera-1) im Vordergrund. Aufnahmen von diesen Kameras stellen bei üblichen Ausstrahlungen den überwiegenden Anteil von Bildern aus der Totalen bereit. Torkameras, Flugkameras oder ähnlich ungewöhnliche Lösungen, welche ebenso Bilder aus der Totalen bieten können, sind nicht dafür bekannt, einen großen Anteil der Bilder in TV-Übertragungen bereitzustellen, sofern diese Kameratechniken vorhanden sind.

Eine Abfolge von Bildern, welche von der gleichen Kamera ohne Unterbrechung aufgenommen wurden, wird als *Einstellung* bezeichnet. Einstellungen werden durch *Schnitte* voneinander getrennt. Diese werden zuerst hinsichtlich ihrer Dauer in harte und weiche Schnitte unterteilt, wobei harte Schnitte den direkten, übergangslosen Wechsel zwischen zwei Einstellungen bedeuten. Im Gegensatz dazu gehen weich geschnittene Einstellungen allmählich ineinander über. Weiche Schnitte lassen sich weiterhin wie folgt klassifizieren:

- Wischblende (Wipe, siehe Abbildung 4.8)
- Überblendungen (Lap Dissolve, siehe Abbildung 4.1, E-G)
- Auf- und Abblenden, Überblendung üblicherweise zu schwarz (Fade)
- Effekt-Übergänge (z. B. Abbildung 4.1, A-D)

Effekt-Übergänge stellen eine Besonderheit für Schnitte bei Fußballübertragungen dar. Sie werden zumeist genutzt, um Rückblenden vom Live-Spielgeschehen unterscheidbar zu machen und sind in ihrer Erscheinung dem jeweiligen Wettbewerb entsprechend gestaltet. Da innerhalb der Schnitt-Erkennung auch Effekt-Übergänge zukünftiger Wettbewerbe automatisch erkannt werden sollen, gilt es ein Verfahren zu finden, welches unabhängig von Merkmalen in konkreten Effekten arbeiten kann. Mit dem zuvor formulierten Anspruch die Bilder der Kamera-1 in TV-Aufnahmen zu bestimmen, kann mit dem Wissen um Schnitte das Ziel für dieses Kapitel formuliert werden: die Erkennung der Schnitte zu Beginn und Ende aller Einstellungen der Kamera-1, die Nicht-Erkennung von Schnitten innerhalb dieser Einstellungen sowie die Klassifikation von Einstellungen in Kamera-1- oder Nicht-Kamera-1-Typ. Eine korrekte Schnittfindung innerhalb anderer Einstellungen wird nicht betrachtet.

4.2 Das System zur Klassifikation von Einstellungen

Das in diesem Kapitel vorgestellte Verfahren für die Erkennung von Kamera-1-Einstellungen basiert im ersten Schritt auf einer bildweisen Klassifikation von Einzelbildern innerhalb von Bildfolgen (-sequenzen). Jedes Bild in einer solchen Folge ist mit einer Zeitmarkierung versehen. Die Zeitmarkierungen wachsen mit größer werdendem Bildindex streng monoton und sind, wie bei TV-Übertragungen üblich, äquidistant. Auf der Grundlage der bildweisen Klassifikation wird für jedes Bild entschieden, ob unmittelbar vor oder nach diesem Bild ein harter Schnitt stattfindet oder ob besagtes Bild das (temporale) Zentrum eines weichen Schnitts ist.

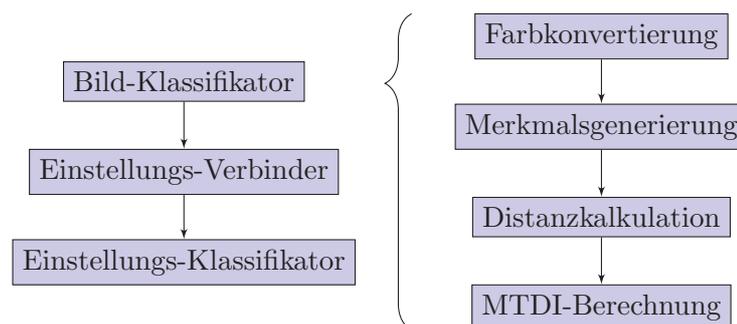


ABBILDUNG 4.2: Systemkomponenten zur Erkennung von Kamera-1-Einstellungen

Wie aus Abbildung 4.2 hervorgeht, setzt sich die Klassifikation von Einzelbildern aus einer Reihe von Subsystemen zusammen. Dies reflektiert insbesondere die Tatsache, dass für die Erkennung weicher Schnitte eine größere temporale Umgebung von Bildern zu berücksichtigen ist. Innerhalb der bildweisen Klassifikation sind folgende Teilschritte zu nennen: eine Farbkonvertierung in der Vorverarbeitung, eine Merkmalsgenerierung in

Einzelbildern (Abschnitt 4.3.1) und die Aufbereitung einer Distanz zwischen zwei Bildern (Abschnitt 4.3.2). Durch die Erweiterung vom paarweisen Vergleich zu einer vollständigen paarweisen Enumeration der Distanzen innerhalb einer Nachbarschaft entstehen Merkmalsmatrizen, welche weiche Schnitte in untersuchten Situationen aussagekräftig charakterisieren können (Abschnitt 4.3.3). Diese Matrizen werden als *Multi Temporal Distance Images (MTDIs)* bezeichnet, da jede ihrer „Achsen“ als Zeitachse interpretiert werden kann. Die Erkennung von harten Schnitten (Abschnitt 4.3.4) und weichen Schnitten (Abschnitt 4.3.5) kann anschließend auf MTDIs erfolgen.

Obwohl temporale Nachbarschaften, und damit vorausgehende und nachfolgende Bilder, betrachtet werden, geschieht die Klassifikation bis zu diesem Schritt bildweise. Einzelbilder werden anschließend im Einstellungs-Verbinder entsprechend den zuvor gewonnenen Schnittzeitpunkten zu Intervallen verknüpft (Abschnitt 4.4). Den abschließenden Schritt bildet die Anwendung des Einstellungs-Klassifikators (Abschnitt 4.5). Dieser beurteilt Einstellungen in einem Random Forest anhand von Gras- und Grashüllenmerkmalen.

Schwellwert	Bedeutung	Verwendete Größe [s]
T_s	Minimale Dauer einer Einstellung	2,0
T_t	Maximale Dauer eines weichen Schnitts	1,5

TABELLE 4.1: Quantitative Annahmen werden als Schwellwerte modelliert

Für das Verfahren wird eine Reihe von Annahmen getroffen. Tabelle 4.1 listet zwei Größen auf, welche quantitativ in das Verfahren eingehen. Beide Werte müssen nicht trainiert werden, da die minimale Dauer einer Einstellung für die Einstellungs-Klassifikation als Größe technisch bedingt ist (Abschnitt 4.5) und die maximale Dauer eines weichen Schnitts lediglich eine plausible obere Schranke darstellt. Neben diesen Schwellwerten werden weitere Annahmen für die Bild-Klassifikation getroffen:

- die szenisch aufbereitete Handlung wird dem Zuschauer vor allem über das Bildzentrum vermittelt,
- weder harte Schnitte noch weiche Schnitte in Form von Überblendungen werden zwischen Einstellungen mit nicht unterscheidbaren Bildinhalten gesetzt,
- die Übergangsfunktionen ($y(t) : [0; 1] \rightarrow [0; 1]$), welche eine Einstellung ($y = 0$) über einen gewissen Zeitraum ($0 \leq t \leq 1$) in eine andere Einstellung ($y = 1$) überführen, sind monoton wachsend und bezüglich der temporalen Mitte ($t = 0,5$) des Schnitts punktsymmetrisch, und
- die Bildsequenz stellt ein Fußballspiel in Vollbild dar, ohne Bild-in-Bild-Techniken.

Diese Annahmen reflektieren zum einen die Vorgabe an einen Kameramann, das Geschehen in der Bildmitte einzufangen, zum anderen das Regie-Prinzip, dem Zuschauer Einstellungswechsel klar kenntlich zu machen. Obwohl die Übergangsfunktionen, welche für den Wechsel zwischen Einstellungen verwendet werden, nicht bekannt sind, kann die Änderung der Bildinhalte über ein Zeitintervall beobachtet werden. Die Annahme der Symmetrie jener Funktionen lässt dann den Schluss auf den Mittelpunkt eines Schnitts zu. In der Praxis lässt sich die Plausibilität dieser Annahme sowohl für lineare als auch nicht-lineare Übergangsfunktionen feststellen [Fielding, 1985, Kapitel 7], [Nam u. Tewfik, 2000].

4.3 Erkennung von Schnitten

Die Bilder einer zu untersuchenden Bildsequenz mit paarweise identischen Bildgrößen werden für alle Berechnungen als Stapel betrachtet. Auch wenn eine Online-Lösung leicht realisierbar ist, werden endliche Bildfolgen im Sinn einer Stapelverarbeitung zugunsten einer einfacheren Formulierung des Verfahrens zur Erkennung von Schnitten bevorzugt.

4.3.1 Merkmalsgenerierung

Jedes Eingangsbild wird einer Farbraum-Konvertierung unterworfen. Die Verwendung von CIELAB hat sich an dieser Stelle als gewinnbringend herausgestellt — insbesondere da euklidische Distanzen in diesem Farbraum in Bezug auf die menschliche Wahrnehmung miteinander verglichen werden können (Wahrnehmungsbezogenheit). Durch den Farbraum wird angestrebt, dass Farben von optisch gleicher Differenz dies auch durch gleiche Abstände ausdrücken. Weiterhin kann davon profitiert werden, dass Histogramme eine größere Anzahl verschwindender Einträge enthalten, als dies bei RGB der Fall wäre, da der RGB-Farbraum, von welchem die Konvertierung ausgeht, kleiner ist (d. h. ein kleiner Teil der CIE-Normtafel wird abgedeckt).

Jedes einzelne Bild wird durch eine beliebig zu wählende, aber feste Zerlegung in n Teilbilder geteilt. Jedem dieser Teilbilder ist hierbei ein festes Gewicht $w_{1:n}$, mit $\sum_w = 1$ und $w_i \geq 0$, zuzuordnen. Um den Sachverhalt zu vereinfachen, wurden in der Implementation nur Zerlegungen in Rechtecke gleicher Größe berücksichtigt. Ein Beispiel ist in Abbildung 4.3 dargestellt. Mithilfe dieser Zerlegung wird die Forderung nach einer im Bildzentrum vermittelten Handlung umgesetzt. Bereiche, welchen eine größere Bedeutung beigemessen wird, werden in der folgenden Bemessung von Bilddistanzen stärker berücksichtigt. Die Bereiche sowie deren Gewichte sind *a priori* festzulegen.

1 ($\frac{1}{20}$)	2 ($\frac{1}{20}$)	3 ($\frac{1}{20}$)	4 ($\frac{1}{20}$)
5 ($\frac{1}{20}$)	6 ($\frac{2}{20}$)	7 ($\frac{2}{20}$)	8 ($\frac{1}{20}$)
9 ($\frac{1}{20}$)	10 ($\frac{2}{20}$)	11 ($\frac{2}{20}$)	12 ($\frac{1}{20}$)
13 ($\frac{1}{20}$)	14 ($\frac{1}{20}$)	15 ($\frac{1}{20}$)	16 ($\frac{1}{20}$)

ABBILDUNG 4.3: Zerlegung eines Bilds in Rechtecke. Die Zellen enthalten die zugeordneten Spaltenindizes der Matrix \mathbf{M}^i sowie, in Klammern, die Gewichte w_i .

Gegeben sei nun das i -te Bild innerhalb einer Bildsequenz. Nach Abschluss der Farbkonvertierung und Bestimmung einer Zerlegung werden Bildmerkmale extrahiert. Für jedes Teilbild von Bild $\#i$ wird ein normalisiertes 3D-Histogramm der Farbkomponenten berechnet. Als Auflösung bzw. Klassengröße des Histogramms empfehlen sich 3 Bit pro Kanal. Dies ergibt mit $(2^3)^3 = 512$ Klassen eine handhabbare Anzahl bei hinreichender Genauigkeit für den Vergleich von zwei Histogrammen. Die einzelnen Histogramme werden als 512-elementige Spaltenvektoren in eine Matrix eingetragen. Die Merkmalsmatrix \mathbf{M}^i kann damit unter Beachtung einer vorgegebenen Reihenfolge (siehe Abbildung 4.3) zusammengestellt werden. Da die Histogramme für jedes Teilbild einzeln normalisiert wurden, gilt für die Matrix $\|\mathbf{M}^i\|_1 = n$. Wie sich zeigen wird, besitzen Histogramme ausreichend Aussagekraft, um in Fußballspielen sowohl harte als auch weiche Schnitte erkennen zu können — vorausgesetzt Distanzen und Nachbarschaften werden geeignet für die Erkennung eingesetzt. Von gleicher Bedeutung wie die Erkennung der Schnitte zu Beginn und Ende von Kamera-1-Einstellungen ist der Ausschluss von falsch positiven Antworten in Kamera-1-Einstellungen. Da diese einen Spielfeldüberblick zeigen und die Histogramme auch bei schnellen Kamerabewegungen oder Veränderungen der Brennweite nicht im Verdacht stehen, drastische Änderungen in kurzen Zeiträumen zu durchlaufen, erscheint das Merkmal geeignet gewählt.

4.3.2 Kalkulation von Distanzen zwischen Bildern

Auf den für alle Bilder berechneten Merkmalsmatrizen kann eine Distanz definiert werden. Da die Histogramme die relativen Häufigkeiten von Farben in den jeweiligen Teilbildern darstellen, bietet sich die Verwendung eines stochastischen Distanzmaßes für diskrete Dichtefunktionen an, im Besonderen das χ^2 -Maß (für Weiteres siehe [Pele u. Werman, 2010]), welches symmetrisch ist und zudem im Ergebnis auf das Intervall $[0; 1]$ beschränkt ist. Es handelt sich hierbei jedoch um keine Distanz *per definitionem*, da sowohl das Subadditivitäts- als auch das Koinzidenzaxiom verletzt werden. Da diese

Punkte im Weiteren jedoch nicht ins Gewicht fallen, kann für die Distanz (der Begriff wird dennoch verwendet) zweier Merkmalsmatrizen M^i und M^j unter Berücksichtigung der Gewichte $w_{1:n}$ definiert werden:

$$d_{i,j} := \chi^2(\mathbf{M}^i, \mathbf{M}^j; w_{1:n}) := \frac{1}{2} \sum_{q=1}^n w_q \sum_{r=1}^{512} \frac{(\mathbf{M}_{r,q}^i - \mathbf{M}_{r,q}^j)^2}{\mathbf{M}_{r,q}^i + \mathbf{M}_{r,q}^j} \quad (4.1)$$

Summanden mit Wert $\frac{0}{0}$ entfallen. Es lässt sich leicht zeigen, dass auch $\chi^2(\mathbf{M}^i, \mathbf{M}^j; w_{1:n})$ symmetrisch und auf $[0; 1]$ beschränkt ist.

4.3.3 Berechnung von Distanzen für mehrere Zeitachsen

Bevor Bilder klassifiziert werden können, ist es sinnvoll, Distanzbildungen zwischen Bildern bzw. ihren Merkmalsmatrizen näher zu betrachten. Mit den bisher bereitgestellten Werkzeugen ist eine Erkennung von harten Schnitten leicht möglich. Dies kann beispielsweise auf einem Schwellwert zur Distanz aufeinanderfolgender Bilder basieren. Weiche Schnitte jedoch benötigen die Berücksichtigung einer größeren Anzahl konsekutiver Bilder.

Unter Verwendung des Schwellwerts für die maximale Transitionsdauer T_t wird die Fenstergröße $z := \lfloor \frac{1}{2} T_t * fps \rfloor$ eingeführt, fps gibt die Anzahl der Bilder pro Sekunde an. Eine geeignet gewählte Doppelindizierung über die Distanz χ^2 resultiert in der Matrix:

$$\hat{\mathbf{D}}(i) := \begin{pmatrix} d_{i-z, i-2z} & \cdots & d_{i-z, i-z} & \cdots & d_{i-z, i} \\ \vdots & & \vdots & & \vdots \\ d_{i, i-z} & \cdots & d_{i, i} & \cdots & d_{i, i+z} \\ \vdots & & \vdots & & \vdots \\ d_{i+z, i} & \cdots & d_{i+z, i+z} & \cdots & d_{i+z, i+2z} \end{pmatrix} \quad (4.2)$$

Diese $(2z+1) \times (2z+1)$ -Matrix wird als *Multi Temporal Distance Image (MTDI)* zum Bild $\#i$ bezeichnet. Das Zentrum mit Koordination $(z+1, z+1)$ ist ihr Referenzpunkt (mit Wert 0). Diese Darstellungsform bringt neben Vorteilen in der Erkennung von Schnitten auch erhebliche Einsparungen in Berechnungsaufwand mit sich. Ist $\hat{\mathbf{D}}(i)$ bekannt, sind für die Bestimmung von $\hat{\mathbf{D}}(i+1)$ unter Ausnutzen der Symmetrie der gewählten Distanzfunktion lediglich z Matrixeinträge neu zu berechnen und damit z Distanzen neu zu bestimmen.

Die Abbildungen 4.4 und 4.5 zeigen die Gestalt von MTDIs in zwei Szenarien. Die Eigenschaft des auf das Intervall $[0; 1]$ eingeschränkten Wertebereichs vererbt sich auch

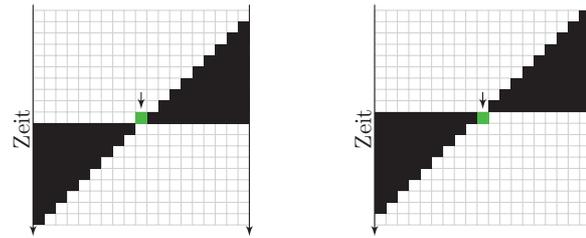


ABBILDUNG 4.4: In einem Raster dargestellte MTDI. Das MTDI des „cut-in“ (links) beschreibt die Situation einen Zeitschritt vor dem MTDI des „cut-out“ (rechts). Beide Abbildungen zeigen ideale Fälle.

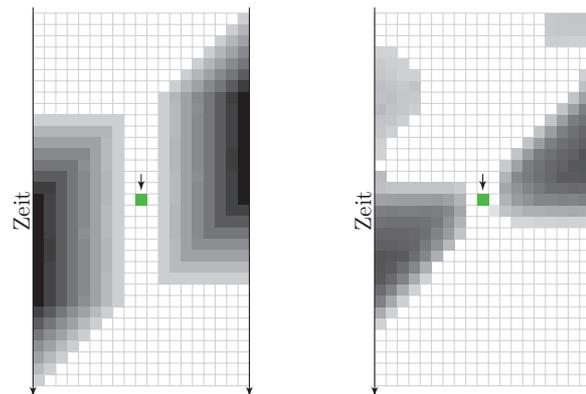


ABBILDUNG 4.5: MTDI für weiche Schnitte. Während das Laborbeispiel langsam und ohne Störung abläuft (links), zeigt das Szenario aus Abbildung 4.1 (E)-(G) einen schnellen Übergang mit weiteren Distanzeinträgen resultierend aus der vorhergehenden Nahaufnahme (rechts).

elementweise auf MTDI. Für die Betrachtung empfiehlt sich ein Bezug zum Referenzpunkt. Durch eine gedachte Verschiebung des Referenzpunkts entlang der Vertikalen können MTDI anderer Zeitpunkte bestimmt werden. Eine Verschiebung um eine Zeile nach oben resultiert im MTDI des vorherigen Bilds. Eine Verschiebung nach unten ergibt das MTDI des nächsten Bilds. Hierfür müssen die Zeilen der Matrix oben bzw. unten mit neuen Daten aufgefüllt werden, um eine Nachbarschaft der Größe z (in der Maximumnorm) um den Referenzpunkt im Bild zu berücksichtigen. Unter diesem Blickwinkel könnten alle berechneten Distanzen in ein einziges MTDI eingetragen werden, welches sich in der Länge entsprechend ausdehnen würde, jedoch noch immer die Breite (Spaltenanzahl) $2z + 1$ besäße. Für eine bessere Visualisierung wurde in Abbildung 4.5 auf eine vertikale Ausdehnung zurückgegriffen.

4.3.4 Erkennung harter Schnitte

Für die weitere Betrachtung ist es sinnvoll harte Schnitte in zwei Untertypen zu unterteilen: „cut-in“ und „cut-out“. Das letzte Bild einer Einstellung vor einem harten Schnitt

wird als „cut-in“ bezeichnet, das erste Bild nach einem Schnitt als „cut-out“. Ohne Beschränkung der Allgemeinheit wird im Folgenden nur der „cut-in“ betrachtet. Ausgehend von hinreichender Bilddistanz zwischen den Einstellungen resultiert ein harter Schnitt in einer doppelten Treppenstruktur analog zu Abbildung 4.4. Zur Erkennung derartiger Bilder findet ein Template Verwendung. An allen Stellen, die innerhalb des „cut-ins“ in Abbildung 4.4 (links) schwarz markiert sind, wird elementweise eine Schwellwertprüfung durchgeführt. Für die Erkennung eines harten Schnitts wird gefordert, dass alle durch das Template ausgewählten Matrixeinträge größer als ein *a priori* festzulegender Schwellwert B_c sind. Die Verwendung von MTDIs wirkt sich auch hier positiv durch die Berücksichtigung von Nachbarschaften aus. Schnitterkennungen über Einzelbilddifferenzen, wie sie häufig für die Erkennung von harten Schnitten Verwendung finden (etwa [Galambos, 2000]), benötigen zusätzliche Filter, um falsche Positivmeldungen z. B. durch Kamerablitzlichter zu korrigieren. In diesem Ansatz ist eine solche Korrektur nicht erforderlich.

4.3.5 Erkennung weicher Schnitte

Die Erkennung harter Schnitte kann aufgrund des abrupten Wechsels von Bildinhalten als relativ einfaches, gelöstes Problem angesehen werden. Weiche Schnitte stellen, speziell im Angesicht des Kontrastes von schnellen Kamerabewegungen und langsamen Übergängen, eine größere Herausforderung dar. Für die Bearbeitung dieses Sachverhalts werden vier Detektoren definiert. Diese prüfen ein MTDI auf einzelne Eigenschaften, welche einem weichen Schnitt unterstellt werden. Abbildung 4.5 zeigt MTDIs eines simulierten und eines realen Übergangs. Beide MTDIs wurden für jeweils das Bild berechnet, welches sich mittig in der Überblendung befindet und damit (punktweise) gleiche Farbanteile aus beiden angrenzenden Einstellungen besitzt. Das durch die Aufstellung der Detektoren verfolgte Ziel ist es, ausschließlich dieses Bild im Übergang als weichen Schnitt zu klassifizieren. Auf diese Weise sollen Mehrfacherkennungen für einzelne Ereignisse unterdrückt und Schnitte möglichst genau lokalisiert werden.

Für die Erkennung wird ausgenutzt, dass auch bei weichen Schnitten ein charakteristisches Muster in MTDIs erkennbar ist. Wie auch bei harten Schnitten, lässt sich eine treppenartige Struktur beobachten, welche im Gegensatz zu harten Schnitten im Idealfall punktsymmetrisch zum Referenzpunkt ist. Diese wird wiederum durch ein Template $\mathbf{T} = (t_{i,j})$ erkannt. Hierfür wird die obere Hälfte (Vergangenheit) des „cut-out“-Templates und die untere Hälfte (Zukunft) des „cut-in“-Templates verwendet. Auf der Zeile des Referenzpunkts seien alle Punkte außer dem Referenzpunkt besetzt ($= 1$). Nicht besetzte Punkte in \mathbf{T} haben den Wert 0. Damit ist auch das Template punktsymmetrisch. Mit dem bereits bekannten Template-Schwellwert B_c kann notiert werden:

$$p_1(\hat{\mathbf{D}}(i)) := \frac{\#\{(r, q) \mid \hat{d}_{r,q} * t_{r,q} \geq B_s\}}{\#\{(r, q) \mid t_{r,q} = 1\}} \quad (4.3)$$

Aufgrund der Symmetrie und des Wachstums der Treppenstruktur hin zum Schnittzentrum erscheint es sinnvoll anzunehmen, dass das Zentrum eines Schnitts die Zeile mit größter Zeilensumme innerhalb einer Nachbarschaft im MTDI ist. Im zweiten und dritten Detektor wird geprüft, wie groß der Wert der Zeilensumme des Referenzpunkts in Relation zu den Zeilensummen anderer Zeilen ist und ob die Zeilensummen von der Zeile des Referenzpunkts aus monoton fallend sind. Zur Realisierung des ersten Tests werden alle Zeilensummen aufsteigend sortiert. Die Position der Zeile des Referenzpunkts nach der Sortierung sei mit j bezeichnet. Dann kann für einen entsprechenden Detektor notiert werden:

$$p_2(\hat{\mathbf{D}}(i)) := \left(\frac{j}{2z+1} \right)^4 \quad (4.4)$$

Um eine stärkere Gewichtung von hohen Zeilensummen im MTDI zu gewährleisten, wurde der Exponent 4 gewählt. Eine Beurteilung der Monotonie geschieht auf Basis von Vergleichen der MTDI-Zeilensummen:

$$p_3(\hat{\mathbf{D}}(i)) := \frac{1}{2z} * \left(\# \left\{ r \mid 1 \leq r \leq z, \sum_q \hat{d}_{r,q} \leq \sum_q \hat{d}_{r+1,q} \right\} + \# \left\{ r \mid z+1 \leq r \leq 2z, \sum_q \hat{d}_{r,q} \geq \sum_q \hat{d}_{r+1,q} \right\} \right) \quad (4.5)$$

Die Punktsymmetrie der Übergangsfunktionen bedeutet, dass die Geschwindigkeitsfunktionen (erste Ableitung der Distanzen) des Übergangs symmetrisch (gerade) bzgl. der Geraden ($t = 0,5$) im Zentrum des Übergangs sind (vgl. Abschnitt 4.2). Um dies zu quantifizieren, wird ein Symmetrie-Maß basierend auf MTDI-Zeilen formuliert. Die Werte einer Zeile werden unter Verwendung des zentralen Differenzenquotienten numerisch differenziert. Die resultierende Folge $d_{1:2z-1}$ wird unter Auslassung des Mittelpunkts in einen linken (Index l) und einen rechten (Index r) Abschnitt unterteilt. Beide Wertelisten werden als Realisierungen einer im Intervall $[0; 1]$ äquidistant abgetasteten Funktion betrachtet. Mittels RANSAC (siehe Notation und Nomenklatur) wird für beide Teilfolgen eine Funktion $\phi(x \mid \theta \in \Theta)$ angepasst.

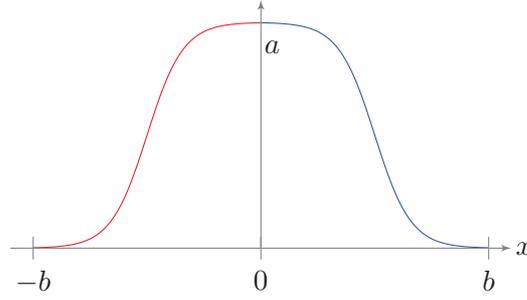


ABBILDUNG 4.6: Die Funktion $\phi(x) = \frac{a}{2} [\tanh(\pi(1 - 2bx)) + 1]$ mit den Parametern $\Theta = (a, b)$ in $[0; b]$ (blau) und ihr Spiegelbild $\phi(-x)$ in $[-b; 0]$ (rot). Diese Funktion wird für die Modellierung von Übergängen verwendet.

$$\theta_l = \text{RANSAC} \left(\phi, \Theta, d_{1:z-1}, \left(0, \frac{1}{z-2}, \dots, 1 \right) \right) \quad (4.6)$$

$$\theta_r = \text{RANSAC} \left(\phi, \Theta, d_{z+1:2z-1}, \left(1, \frac{z-3}{z-2}, \dots, 0 \right) \right) \quad (4.7)$$

Die tanh-Funktion zeigte in vielen Szenarien ein passendes Adaptionsvermögen. Der Parameterraum Θ beinhaltet Streckungsfaktoren in Richtung beider Achsen entsprechend Abbildung 4.6. Der Wahl anderer Modellfunktionen steht jedoch ebenso nichts im Weg. Zu nennen sind beispielsweise konstante Funktionen bzw. Rechteckfunktionen zur Modellierung linearer Übergangsfunktionen bei bekannter bzw. unbekannter Übergangslänge oder Gauß-Kerne. Der zentrale Gedanke ist, dass die Parameterschätzungen der linken sowie rechten Seite im Rahmen der Funktionsanpassung getauscht werden:

$$\phi_{1:z-1}^l := \phi \left(\left(0, \frac{1}{z-2}, \dots, 1 \right) \mid \theta_r \right) \quad (4.8)$$

$$\phi_{z+1:2z-1}^r := \phi \left(\left(1, \frac{z-3}{z-2}, \dots, 0 \right) \mid \theta_l \right) \quad (4.9)$$

Der daraus hervorgehende Fehler wird verwendet, um die Symmetrie zu bewerten. Der vierte Detektor beschreibt entsprechend die Abstände zwischen der Funktionsanpassung mit ausgetauschten Parametern und den tatsächlichen Werten:

$$p_4(d_{1:2z-1}, \phi_{1:z-1}^l, \phi_{z+1:2z-1}^r) := 1 - \left(\frac{1}{(z-1)} \sum_{q=1}^{z-1} |d_q - \phi_q^l|^{0,5} + \frac{1}{(z-1)} \sum_{q=z+1}^{2z-1} |d_q - \phi_q^r|^{0,5} \right) \quad (4.10)$$

Resultate von p_4 , welche kleiner als 0 ausfallen, werden 0 gesetzt. Der Tatsache, dass Funktionen bei vorliegender Symmetrie (nahe des Mittelpunkts des Übergangs) durch einen Parametertausch zielgerecht angepasst werden können, wird durch die Wichtung mit gebrochenem Exponenten 0,5 Rechnung getragen. Damit fallen die Werte dieses Detektors abseits des Mittelpunkts des Übergangs kleiner aus als im Schnitzzentrum.

Alle vier vorgestellten Detektoren für weiche Schnitte besitzen die Eigenschaft eines auf das Intervall $[0; 1]$ eingeschränkten Wertebereichs. Zusammenfassend kann notiert werden:

$$P(\text{weicher Schnitt in Bild } \#i) := \prod_{j=1}^4 p_j \tag{4.11}$$

Durch Anwendung eines Schwellwerts B_t auf den aus Gleichung 4.11 berechenbaren Wert, folgend als Schnitt-Wahrscheinlichkeit bezeichnet, wird für Einzelbilder die Klassifikation auf den Mittelpunkt eines weichen Schnitts durchgeführt. Ein Beispiel hierzu zeigt Abbildung 4.7.

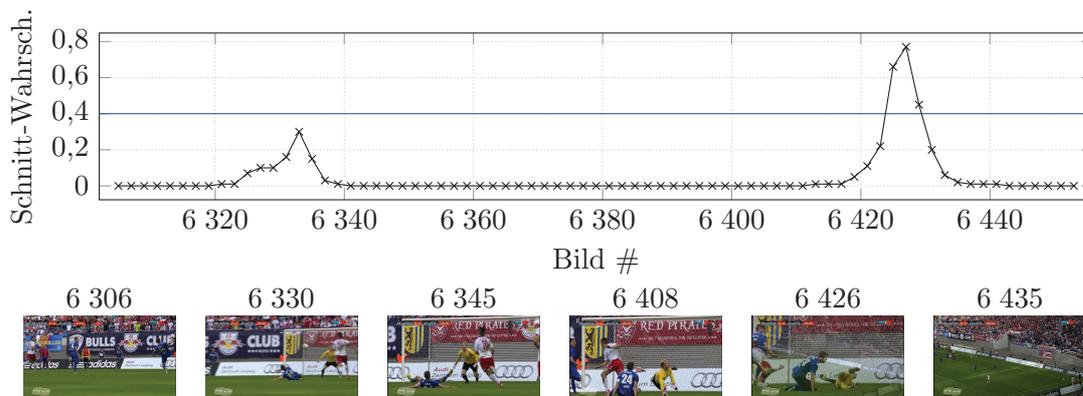


ABBILDUNG 4.7: Schnitt-Wahrscheinlichkeiten in einem Testvideo mit 25 Bildern pro Sekunde. Eine schnelle Bewegung in einer Nahaufnahme (um Bild #6 333) gefolgt von einer Überblendung (um Bild #6 427). Die Unterscheidung durch einen Schwellwert (blaue Gerade) ist möglich. Bildquelle: [MDR, 2013]

4.3.6 Behandlung von Wischblenden und Blitzlichtern

Wischblenden, wie in Abbildung 4.8 illustriert, sind in heutigen Übertragungen eher selten anzutreffen. Bezüglich der MTDI-Struktur besitzen sie, wie auch Blitzlichter, große Werte auf der von oben rechts nach unten links laufenden Diagonalen. Im Gegensatz

zu Blitzlicht-MTDIs sind jedoch auch die Nebendiagonalelemente aufgrund der längeren Dauer mit größeren Werten besetzt. Ausgenommen werden muss in jedem Fall der Referenzpunkt, welcher generell den Wert 0 annimmt.

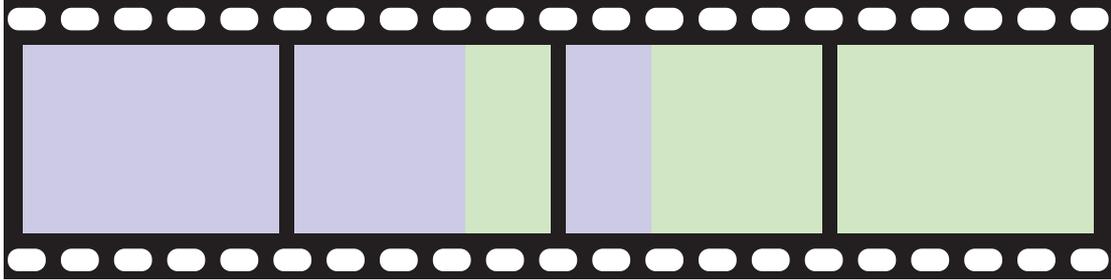


ABBILDUNG 4.8: Weicher Schnitt zwischen zwei Einstellungen (blau und grün) in Form einer Wischblende

Auf eine individuelle Erkennung wurde verzichtet, da die Wahrscheinlichkeit für Falsch-Positiv-Meldungen durch sich nah an der Kamera entlang bewegende Objekte weit größer ist als die Wahrscheinlichkeit des Einsatzes einer Wischblende im Fußball. Blitzlichter müssen nicht detektiert werden, da die Erkennung (harter) Schnitte mit der Einbeziehung einer größeren Nachbarschaft an Bildern robust gegenüber derartigen Störungen ist.

4.4 Berechnung von Schnitten aus Bildklassen

Bis zu diesem Zeitpunkt wurden lediglich Einzelbilder klassifiziert. Auf Grundlage der zeitlichen Markierungen der Bilder können die erkannten harten und weichen Schnitte sowie der Beginn- und Endzeitpunkt einer zu untersuchenden Bildsequenz auf einer Zeitachse zusammengetragen werden. Zuvor müssen jedoch die Grenzen weicher Schnitte geweitet werden, um ungünstiges Bildmaterial aus der Komposition mehrerer Einstellungen und Effekte auszuschließen. Dazu wird um jeden erkannten weichen Schnitt ein Bereich der Größe T_t mit besagtem Schnittpunkt im Zentrum als ungültig markiert. Im Fall überlappender ungültiger Bereiche werden diese individuell betrachtet. Nachdem somit neue Beginn- und Endpunkte hinzukommen, verbleiben zwischen diesen nur mehr harte Schnitte, welche den Zeitstrahl eindeutig in Einstellungen zerlegen. Alle daraus resultierenden Intervalle, welche größer als die kürzeste Einstellungsdauer T_s sind, werden für die weiteren Schritte als Einstellung in Betracht gezogen.

4.5 Schnitt-Klassifikation

Die Klassifikation von Einstellungen basiert auf Merkmalen in Einzelbildern. Da die Betrachtung jedes Bilds einer Einstellung eine große Menge für die Klassifikation überflüssiger Informationen generiert und Merkmalstupel ungleicher Länge erzeugen würde, wird ein Stichprobenverfahren vorgeschlagen. Für jedes Bild der Stichprobe werden gemäß Kapitel 3 Hülle \mathbf{H} und (binäre) Grasmasken \mathbf{G} sowie die Menge \mathbf{N} grüner Punkte (Gleichung 3.3) berechnet. Hieraus werden folgende Merkmale bestimmt:

- das Verhältnis aus der Anzahl $\#\text{Fill}(\mathbf{N}) \cap \mathbf{H}$ grüner Punkte innerhalb der Hülle zur Bildgröße,
- das Verhältnis aus der Anzahl $\#\text{Fill}(\mathbf{N}) \cap \mathbf{G}$ grüner Punkte in der Grasmasken zur Bildgröße,
- das Verhältnis der Anzahl $\#\mathbf{H}$ von Hüllpunkten zur Bildgröße,
- das Verhältnis der Anzahl $\#\mathbf{G}$ von Graspunkten zur Bildgröße,
- das Verhältnis der Anzahl $\#\mathbf{G}$ von Graspunkten zur Anzahl $\#\mathbf{H}$ von Hüllpunkten, so die Menge \mathbf{H} nicht leer ist, und
- die Anzahl $\#(\mathbf{H} \setminus \mathbf{G})$ der Hüllpunkte außerhalb der Grasmasken in Vielfachen der Hüllengröße $\#\mathbf{H}$, so die Menge \mathbf{H} nicht leer ist.

Für den Fall, dass die Menge \mathbf{H} leer ist, werden die letzten beiden Merkmale durch die Bildgröße ersetzt. Der Operator $\text{Fill}(\mathbf{N})$ bildet die Menge, welche zusätzlich zu den Elementen der Menge \mathbf{N} auch alle Elemente in Löchern von \mathbf{N} enthält. Die für jedes Bild erzeugten Merkmalsvektoren werden nach dem zuerst genannten Merkmal sortiert. Nach der Sortierung wird der erste, mittlere und letzte Vektor zu einem 21-elementigen Tupel zusammengesetzt. Dass dabei Ausreißer besonders ins Gewicht fallen, erscheint für die Klassifikation von Einstellungen der Totalen angebracht. Als Stichprobe wird ein Bild in jeder Sekunde eines Eingangsvideos verwendet. Aus der gewählten kürzesten Einstellungsdauer T_s von 2 s folgt damit, dass besagtes Tupel für jede gefundene Einstellung auch gebildet werden kann.

Als binärer Klassifikator für Kamera-1-Einstellungen wurde ein Random Forest [Criminisi et al., 2012] verwendet. Die Tiefe der Bäume wurde aufgrund der geringen Anzahl von Merkmalen nicht beschränkt. Im Besonderen ist es Aufgabe des Klassifikators in Szenen, welche Ähnlichkeiten zu Kamera-1-Aufnahmen aufweisen, eine sinnvolle Regel für eine Unterscheidung zu finden. Dafür werden im Sinn der zuvor eingeführten Merkmale vornehmlich die erkannte Größe des Spielfelds im Bild und die Größe von Störungen (Spielern) im Spielfeld berücksichtigt.

4.6 Stand der Forschung

Die Schnitterkennung in Videos von beliebigem Kontext gehört zu den Problemen, welche in der computergestützten Verarbeitung schon seit geraumer Zeit Beachtung gefunden haben. Ein Überblick früherer Ansätze kann [Boreczky u. Rowe, 1996] entnommen werden. In der seit 2003 jährlich stattfindenden unabhängigen Evaluation TRECVID mit angeschlossenem Workshop [Smeaton et al., 2006] werden Verfahren zur Schnitterkennung eingereicht und unter jährlich neu aufgelegten Testdatensätzen bewertet. Nach einer Bilanz im achten Jahr [Smeaton et al., 2010] wird auch hier festgehalten, dass weiche Schnitte das mit Abstand größere Problem im Vergleich zu harten Schnitten darstellten. Sich diesem Problem annehmend, formulieren [Amiri u. Fathy, 2010] eine Schnitterkennung auf Basis einer QR-Zerlegung. Ausgangspunkt hierbei ist, wie auch im vorgestellten Ansatz, eine Rechteckzerlegung des Eingangsbilds, aus welcher blockweise Farbhistogramme bestimmt werden, und die spaltenweise Zuordnung dieser in eine Merkmalsmatrix. Unter der Annahme, dass Bilder innerhalb einer Einstellung gleiche Informationen besitzen, können Bilder zwischen zwei Einstellungen durch eine QR-Zerlegung anhand höherer R-Werte (zugehörige Diagonalelemente der Matrix \mathbf{R}) identifiziert werden. Mit der Einschränkung, dass die Anzahl der Merkmale (Zeilen) in der Featurematrix größer oder gleich der Anzahl der Bilder (Spalten) ist, besteht für das Verfahren eine Einschränkung, welche eine gefensterte Ausführung erforderlich macht. [Bendale u. Talati, 2012] bestätigen grundlegend die Effektivität von Histogrammen, gerade auch in Kombination mit dem χ^2 -Maß. Die Idee, eine Untersuchung auf neu erstellten Bildern durchzuführen, verfolgen auch [Ngo et al., 2002]. Sie verwenden hierfür eine temporale und eine räumliche Achse, was als Verfolgung einer Bildspalte in der Zeit interpretiert werden kann. Das MTDI kann als Fortführung dieses Gedankens unter Verwendung von Histogrammen zur Erfassung des gesamten Bildinhalts interpretiert werden. Auch in dem Ansatz von [Ngo et al., 2002] werden verschiedene Detektoren beschrieben, welche aufgrund der räumlichen Achse in den Merkmalsbildern auf einen breiterem Fundament fußen müssen. So kommen Schätzungen des Bildhintergrunds ebenso wie eine Charakterisierung der Bewegungen im Bild zum Tragen. Eine Erkennung von Schnitten durch eine größere Diversität von Merkmalen betrachten auch [Baber et al., 2011] und [Petersohn, 2004] — die Ersteren durch SURF- und Entropy-Deskriptoren, die Letzteren auf Grundlage von Kanten-, Histogramm- und Farbdifferenz-Statistiken. Aus diesen Merkmalen wird zwischen harten und weichen Schnitten unterschieden. Für harte Schnitte ist eine Blitzlichterkennung integriert. Ein Kandidat für einen weichen Schnitt wird hinsichtlich Kantenenergie, Bildunterschied zu Bildern in der zeitlichen Umgebung des Kandidaten, Histogramm, Bilddistanzen zu den bewegungskorrigierten Bildern, Linearität und Gleichmäßigkeit des Übergangs klassifiziert, wodurch auch hier

der Symmetriegedanke bzgl. Übergängen verfolgt wird. Der Ansatz erzielte sehr gute Resultate bei TRECVID und wurde (siehe [Smeaton et al., 2006]) nach seiner Vorstellung für die Bereitstellung von Referenzdaten verwendet. Dieser kommerziell genutzte Ansatz wird in der Evaluation ebenso angewandt wie das fußballspezifische System von [Ekin et al., 2003]. Neben anderen Ansätzen, wie [Chen et al., 2003, Wan u. Xu, 2004, Xie et al., 2004], welche oft auch akustische Merkmale einsetzen oder die Highlight-Erkennung zum Ziel haben, sind die Ziele bei [Ekin et al., 2003] ähnlich den hier formulierten Forderungen hinsichtlich der Erkennung der Totalen. Basierend auf einer Farb- und Regionserkennung im Video werden Histogramme für die Erkennung von Schnitten verwendet. Die Schnitte werden verschiedenen Kategorien zugeordnet, um verwertbare Szenen zu klassifizieren. Im Gegensatz zu dem ebenfalls für Fußball spezifizierten System von [Siles Canales, 2014] sind im vorgestellten Ansatz keine Templates für die Erkennung von Effekt-Übergängen notwendig.

4.7 Evaluation

Die Evaluation erfolgt zweigeteilt. Zuerst wird die Qualität der Schnitterkennung in Kamera-1-Einstellungen bewertet, im Anschluss wird die Klassifikation auf den erkannten Einstellungen evaluiert. In diesem Rahmen kommt wiederum der Testdatensatz aus Kapitel 3 zum Tragen. Aufgrund ähnlicher Umweltbedingungen zwischen den WM-Begegnungen *S’Korea* und *Norwegen* wurde auf letzteren Datensatz verzichtet. Im Einzelnen werden damit die in Tabelle 4.2 aufgeführten Videos betrachtet.

Begegnung	Länge [m:s]	Einstellungen	Kamera-1-Einstellungen
Bayern	20:03	107	31
B’Faso	55:56	620	123
M’Gladbach	51:51	456	100
S’Korea	3:00	43	12
Leipzig	8:13	52	16
Σ	138:53	1 278	282

TABELLE 4.2: Übersicht der Videos im Testdatensatz

4.7.1 Evaluation der Schnitt-Erkennung

In den Testdaten wurden alle Einstellungen manuell gezählt. Für die Generierung der Ground Truth wurden insbesondere Bilder beachtet, welche durch die Kamera-1 aufgenommen wurden. Für alle derartigen Einstellungen wurden Schnittträger mit höchstmöglicher Exaktheit manuell bestimmt. Die Möglichkeit, ein Video in Einzelbildern zu

betrachten, war hierfür gegeben.

Neben dem vorgestellten Ansatz (mit den Parametern $B_c = 0,2$ und $B_t = 0,4$) wurden auch die Verfahren nach [Petersohn, 2004] und [Ekin et al., 2003] als Referenz herangezogen. Da für Letzteres keine Implementation verfügbar war, musste dies eigenständig nachgeholt werden. Die Schwellwerte des Systems wurden in Übereinstimmung mit den im Paper genannten Verfahrensweisen trainiert.

Da beide Referenzverfahren keine Modellparameter oder Resultate für die Schnittlänge bei weichen Schnitten bereitstellten, wurde ein Toleranzbereich verwendet. Eine in der Ground Truth verzeichnete Einstellung wird als korrekt erkannt betrachtet, wenn das Messsystem innerhalb jedes Toleranzbereichs zu Beginn und Ende der Einstellung mindestens einen Schnitt erkannt hat und wenn innerhalb der Einstellung (außerhalb der Toleranzbereiche) keine weiteren Schnitte erkannt wurden. Die Toleranzbereiche, zeitliche Intervalle um die Schnitte der Ground-Truth-Einstellung, wurden mit 1,5 s ausreichend groß gewählt und mittig zum Ground-Truth-Schnittzeitpunkt ausgerichtet. Das Tolerieren von Abweichungen hatte in keinem Fall einen negativen Effekt auf die Erkennungsrate eines der evaluierten Verfahren.

Begegnung	Vorgestellter Ansatz [%]	[Petersohn, 2004] [%]	[Ekin et al., 2003] [%]
Bayern	100,0	87,1	67,7
B'Faso	98,4	87,8	78,9
M'Gladbach	97,0	87,0	72,0
S'Korea	100,0	100,0	50,0
Leipzig	100,0	93,8	68,8
Σ	98,2	88,3	73,4

TABELLE 4.3: Anteil der Kamera-1-Einstellungen, für welche die Schnitte zu Beginn und Ende korrekt erkannt wurden und für welche keine Antworten innerhalb einer Einstellung vorlagen

Wie aus Tabelle 4.3 hervorgeht, erkennt der vorgestellte Ansatz 98,2% der Kamera-1-Einstellungen im Testdatensatz korrekt. Um auftretende Fehler unterscheiden zu können, wurde die Rate an unterbrochenen Einstellungen für die nicht erkannten Kamera-1-Einstellungen bestimmt. Für den vorgestellten Ansatz beträgt diese 0%, womit sich alle Fehler durch nicht erkannte Schnitte erklären lassen. Im konkreten Fall handelt es sich in den Videos *B'Faso* und *M'Gladbach* um weiche Schnitte zwischen ähnlichen Bildern.

Sowohl das Verfahren nach [Petersohn, 2004] als auch jenes nach [Ekin et al., 2003] erkennen Schnitte von Kamera-1-Einstellungen weniger häufig als der vorgestellte Ansatz. Mit Raten an unterbrochenen Einstellungen von 30% und 40% zeigten diese Verfahren eine hohe Anfälligkeit für falsche Detektionen während schnellen Kamerabewegungen.

4.7.2 Evaluation der Schnitt-Klassifikation

Für die Evaluation wurde die Ground Truth auf die Kamera-1-Einstellungen eingegrenzt, welche durch die Schnitt-Erkennung gefunden werden konnten. Die übrigen Kamera-1-Einstellungen sind durch den vorgestellten Ansatz nicht vollständig erkennbar, womit eine Klassifikation nicht berücksichtigt wurde. Dieser Fakt ist bei der Beurteilung der Klassifikation entsprechend zu beachten. Ergänzt wurde die Ground Truth durch weitere Einstellungen im Spiel, welche die Kamera-1 nicht zeigen. Eine Übersicht ist Tabelle 4.4 zu entnehmen.

Begegnung	Kamera-1-Einstellungen	Nicht-Kamera-1-Einstellungen
Bayern - Arsenal	31	19
B'Faso - Ghana	121	80
M'Gladbach	97	58
S'Korea	12	6
Leipzig	16	13
Σ	277	176

TABELLE 4.4: Übersicht der Ground Truth zur Einstellungs-Klassifikation

Für die Beurteilung der gewählten Merkmale wurden die Videos sowohl individuell betrachtet (Tabelle 4.5), das heißt dasselbe Video wird als Trainings- und als Testdatensatz verwendet, als auch gemeinsam (Tabelle 4.6), wobei dann ein Video den Testdatensatz darstellt, während die restlichen Videos die Trainingsmenge bilden. Im erstgenannten Fall wurde eine 10-fache Kreuzvalidierung verwendet. Abgesehen vom Ausreißer *S'Korea* zeigt sich, dass das individuelle Training wie erwartet von Vorteil sein kann. Der Fall dieser Begegnung kann mit einem zu kleinen Trainingsdatensatz erklärt werden. Andererseits verallgemeinert das Verfahren die Klassifikation in annehmbarer Weise. Ist das Testvideo im Trainingsdatensatz nicht vorhanden, konnten 430 von 453 Einstellungen korrekt klassifiziert werden (Tabelle 4.6).

Der Vollständigkeit halber wurde auch der gesamte Testdatensatz kreuzvalidiert. Das Ergebnis ist in Tabelle 4.7 dargestellt. Über diesen Datensatz unterschiedlichster Aufnahmen kann eine Rate von korrekten Klassifikationen von über 96 % erreicht werden. Beispiele für Klassifikationsfehler und -herausforderungen sind in Abbildung 4.9 zu finden.

Begegnung	TP	FP	TP-Rate	FP-Rate	Prc.	Rc.	F1	Klasse
Bayern	31	0	1	0	1	1	1	Kamera-1
	19	0	1	0	1	1	1	Rest
korrekte Klassifikationen: 100 %								
B'Faso	119	2	0,983	0,017	0,944	0,983	0,964	Kamera-1
	73	7	0,913	0,088	0,973	0,913	0,942	Rest
korrekte Klassifikationen: 95,52 %								
M'Gladbach	92	5	0,948	0,052	0,929	0,948	0,939	Kamera-1
	51	7	0,879	0,121	0,911	0,879	0,895	Rest
korrekte Klassifikationen: 92,26 %								
S'Korea	12	0	1	0	0,8	1	0,889	Kamera-1
	3	3	0,5	0,5	1	0,5	0,667	Rest
korrekte Klassifikationen: 83,33 %								
Leipzig	16	0	1	0	0,889	1	0,941	Kamera-1
	11	2	0,846	0,154	1	0,846	0,917	Rest
korrekte Klassifikationen: 93,10 %								

TABELLE 4.5: Resultate der Klassifikation bei individuellem Training der Begegnungen, jeweils 10-fach kreuzvalidiert. Neben der Betrachtung von richtig positiven (TP) und falsch positiven (FP) Ergebnissen, werden auch Precision, Recall und F1-Score angegeben.

Begegnung	TP	FP	TP-Rate	FP-Rate	Prc.	Rc.	F1	Klasse
Bayern	29	2	0,935	0,065	1	0,935	0,967	Kamera-1
	19	0	1	0	0,905	1	0,95	Rest
korrekte Klassifikationen: 96 %								
B'Faso	116	5	0,959	0,041	0,951	0,959	0,955	Kamera-1
	74	6	0,925	0,075	0,937	0,925	0,931	Rest
korrekte Klassifikationen: 94,53 %								
M'Gladbach	97	0	1	0	0,942	1	0,97	Kamera-1
	52	6	0,897	0,103	1	0,897	0,945	Rest
korrekte Klassifikationen: 96,13 %								
S'Korea	11	1	0,917	0,083	1	0,917	0,957	Kamera-1
	6	0	1	0	0,857	1	0,923	Rest
korrekte Klassifikationen: 94,44 %								
Leipzig	15	1	0,938	0,063	0,882	0,938	0,909	Kamera-1
	11	2	0,846	0,154	0,917	0,846	0,88	Rest
korrekte Klassifikationen: 89,66 %								

TABELLE 4.6: Resultate der Klassifikation bei einem Training durch die jeweils übrigen Begegnungen

4.8 Diskussion und Ausblick

Durch den vorgestellten Ansatz konnten im Testdatensatz 98,2 % der Kamera-1-Einstellungen korrekt erkannt werden. Damit übertraf der Ansatz zwei zur Referenz herangezogene Verfahren deutlich. Die Klassifikation von Einstellungen erreichte eine Rate

Begegnung	TP	FP	TP-Rate	FP-Rate	Prc.	Rc.	F1	Klasse
Σ	276	1	0,996	0,004	0,945	0,996	0,97	Kamera-1
	160	16	0,909	0,091	0,994	0,909	0,95	Rest
korrekte Klassifikationen: 96,25 %								

TABELLE 4.7: Resultate der Klassifikation bei Betrachtung aller Begegnungen ohne Unterscheidung des Videos, 10-fache Kreuzvalidierung



(A) Falsch Negativ



(B) Falsch Positiv



(C) Falsch Negativ



(D) Falsch Positiv

ABBILDUNG 4.9: Fehlerhafte Klassifikationen. Innerhalb einer Aufnahme können Einstellungen der Kamera-1 (A) und Einstellungen mit Nahaufnahmen (B) ähnliche Bildinhalte zeigen. Auch Einblendungen (C) oder Nahaufnahmen einzelner Spieler (D) können zu falschen Zuordnungen führen. Bildquellen: [Eurosport, 2013, Kabel eins, 2013]

korrekter Zuweisungen von 96,3%. Das Ziel, maßgebliche Bilder einer Begegnung extrahieren zu können, kann damit, zumindest innerhalb des Testdatensatzes, zum größten Teil als erreicht angesehen werden. Falsche Klassifikationen von Einstellungen gehen in erster Linie aus Situationen hervor, welche nicht zuletzt auch manuell eine Unterscheidung zwischen Totalen und Nahaufnahmen schwierig gestalten. Eine Möglichkeit der Verbesserung kann ein weiterer Schritt in der Vorverarbeitung sein, in welchem Einblendungen im Video maskiert und innerhalb der MTDI-Berechnung berücksichtigt (ausgelassen) werden. Außerhalb von Videos im Fußballkontext ist es zudem nicht unmöglich, dass Einstellungen, welche ähnliche Bilder im Sinn der Merkmale zeigen, als Kamera-1-Einstellungen klassifiziert werden. Vorstellbar wären Bilder des Tisches bei einem Snooker-Spiel oder Tiere auf einer Grasfläche. Durch Verwendung des Verfahrens und einer im Laufe der Zeit größer werdenden Trainingsmenge kann die Klassifikation an neue Gegebenheiten angepasst werden.

Kapitel 5

Erkennung von Geradensegmenten

5.1 Einleitung

Nachdem durch die Klassifikation von Kamera-1-Einstellungen aussagekräftige Videoausschnitte bestimmt werden können, wird es das Ziel in diesem Kapitel sein, ein Verfahren zur Extraktion von Geradensegmenten in Einzelbildern zu beschreiben. Dies ist ein Vorbereitungsschritt für die Kameraparameter-Nachführung, welche im nächsten Kapitel erläutert werden wird. Die dann zur Verfügung stehenden Geradensegmente in Einzelbildern werden genutzt, um Feldlinien dem Spielfeldmodell zuzuordnen zu können und damit eine bi-jektive Abbildung zwischen Bild und Modell herzustellen. Da die Extraktion der Geradensegmente kein fußballspezifisches Problem darstellt, wird in diesem Kapitel auf die Herausarbeitung eines sportspezifischen Kontextes verzichtet.

Die Erkennung von Geradensegmenten in Bildern ist ein seit geraumer Zeit untersuchtes Problem [Hough, 1962], welches weiterhin zu neuartigen Lösungsansätzen führt (siehe Abschnitt 5.6). Durch die Diskretisierung des Rasters in diskreten Bildern und die damit verbundene Pixel-Aliasierung stellt sich schon zu Anfang die Frage, wann Punktfolgen als gerade zu betrachten sind. Eine Übersetzung des differentialgeometrischen Krümmungsbegriffs auf diskrete Räume ist nicht trivial. Eine





ABBILDUNG 5.1: Eine als „gerade“ anzusehende Feldlinie. Der Versatz kann gegenüber der roten Geraden beobachtet werden. Bildquelle: [ARD/BR, 2012]

robuste Möglichkeit wurde zwar vorgestellt [Han u. Poston, 2001] und auch für Zwecke der Objekterkennung eingesetzt [Monroy et al., 2011], die Klassifizierung in gerade und nicht gerade Punktfolgen leidet hierbei jedoch unter dem gleichen Nachteil wie auch eine Vielzahl anderer Ansätze: Die Wahl der Modellschwellwerte ist unhandlich in dem Sinn, dass sie tief im Verfahren liegendes Verhalten beeinflussen, welches vom Anwender nicht direkt nachvollziehbar ist. Anderen Verfahren fehlt die gewünschte Treffsicherheit bei der Erkennung von leicht gekrümmten und kaum sichtbaren „Geraden“ (vgl. Abbildung 5.1).

Unter dem Aspekt der Erkennung von Geradensegmenten in einem Bild durch die Klassifikation von Punktfolgen (gerade, nicht gerade) werden folgende Teilziele festgehalten:

- Die Erkennung von Geradensegmenten sollte mindestens auf dem State-of-the-Art-Niveau sein.
- Der Algorithmus muss die Möglichkeit einer schnellen Ausführung bieten.
- Die Parameter sollen nachvollziehbar die Klassifizierung in gerade und nicht gerade Punktfolgen steuern.

Die einführenden Betrachtungen in diesem Kapitel sind zweigeteilt. Zuerst werden in formalen Überlegungen Punktfolgen in einer linearen Regression betrachtet und ein Maß für deren Geradheit wird abgeleitet (Abschnitt 5.2). In Abschnitt 5.3 wird anschließend erläutert, wie um Geraden im Raum Konfidenzbereiche berechnet werden. Diese Bereiche werden es sein, in welchen eine Punktfolge als gerade angesehen werden wird. Die Größe und Gestalt wird durch Parameter eines Hough-Raums charakterisiert. Auch die Zusammenführung beider Ansätze wird wiederum zweigeteilt beschrieben. Zuerst werden mathematische Grundlagen in Abschnitt 5.4 diskutiert, bevor ein geschlossener Algorithmus in Abschnitt 5.5 notiert werden kann, welcher es erlaubt, Punktfolgen im Bildraum mit den bekannten Hough-Parametern zu charakterisieren. Anstatt einzelne Punkte dabei in einen neuen Raum mittels einer Brute-Force-Methode transformieren zu müssen, wird durch den Konfidenzbereich eine zuvor aufgestellte Bewertungsgröße zur Extraktion für Geraden im Hough-Raum in den Bildraum übertragen. Die eigentliche Transformation kann entfallen.

5.2 Bestimmung der Geradheit von Punktfolgen

Für ein Binärbild I werden Punktfolgen $\mathbf{x}_{1:n} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ mit $\mathbf{x}_i \in \{1, \dots, \text{Breite}\} \times \{1, \dots, \text{Höhe}\}$ für $i \in \{1, \dots, n\}$ betrachtet. Für Punktfolgen wird ab sofort eine Nachbarschaftsbeziehung (8-Punkte-Nachbarschaft) `Neighbors` vorausgesetzt:

$$\mathbf{x}_{i+1} \in \text{Neighbors}(\mathbf{x}_i, I) \quad , \text{ für } i \in \{1, \dots, n-1\} \quad (5.1)$$

Eine Geradheit einer gegebenen Punktfolge wird durch lineare Regression bemessen. Dies setzt jedoch voraus, dass der Fall vertikaler Geraden zunächst ausgeschlossen wird. Zu diesem Zweck wird die Punktfolge, wie auch in [Lee et al., 2006], zunächst neu ausgerichtet. Die Streumatrix

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} \in \mathbb{R}^{2 \times 2} \quad (5.2)$$

der Punktfolge $\mathbf{x}_{1:n}$ und ihrem arithmetischen Mittel $\bar{\mathbf{x}}$ berechnet sich aus:

$$\mathbf{S} := \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (5.3)$$

Ihre Eigenwerte sind

$$\lambda_{1/2} = \frac{1}{2} \left(s_{11} + s_{22} \pm \sqrt{4 * s_{12}^2 + (s_{11} - s_{22})^2} \right) \quad (5.4)$$

mit $0 \leq \lambda_2 \leq \lambda_1$. Die zugehörigen normierten Eigenvektoren lauten dann

$$\mathbf{e}_i = \left(1 + \frac{(\lambda_i - s_{11})^2}{s_{12}^2} \right)^{-\frac{1}{2}} * \begin{pmatrix} 1 \\ s_{12}^{-1}(\lambda_i - s_{11}) \end{pmatrix} \quad (5.5)$$

Dies entspricht der Betrachtung von translationsinvarianten Momenten. Der erzeugte Vektor \mathbf{e}_1 besitzt in diesem Sinn die gleiche Ausrichtung wie die Punktfolge selbst. Der Vektor \mathbf{e}_2 ist rechtwinklig dazu. Es lässt sich eine sofortige Überführung der Punktfolge in die Hesse-Normalform in Abhängigkeit eines Vektors $\mathbf{x} \in \mathbb{R}^2$ (auf der Geraden) notieren.

$$\mathbf{x} \cdot \mathbf{e}_2 = \bar{\mathbf{x}} \cdot \mathbf{e}_2 \quad (5.6)$$

Der Winkel θ der Normalen ist leicht zu bestimmen.

$$\theta = \text{atan2}(\mathbf{e}_2) \quad (5.7)$$

Im Sonderfall, dass s_{12} verschwindet und in der Berechnung von \mathbf{e}_2 somit eine Division durch Null auftritt, kann die Berechnung auf

$$\theta = \text{atan2}(s_{11}, s_{22}) \quad (5.8)$$

gekürzt werden.

Die Punkte der Folge $\mathbf{x}_{1:n}$ werden zentriert und horizontal ausgerichtet in die Folge $\mathbf{x}'_{1:n}$ überführt:

$$\mathbf{x}'_i := \begin{pmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \end{pmatrix} (\mathbf{x}_i - \bar{\mathbf{x}}) \quad , \text{ für } i \in \{1, \dots, n\} \quad (5.9)$$

Im Rahmen einer linearen Regression werden die Gleichungen

$$x'_{2,i} = \beta_0 + \beta_1 x'_{1,i} + \epsilon_i \quad , \text{ für } i \in \{1, \dots, n\} \quad (5.10)$$

notiert. Dabei gilt $\beta_0, \beta_1 \in \mathbb{R}$ und die Fehler $\epsilon_{1:n} \sim \mathcal{N}(0, \sigma^2)$ werden als unabhängig normalverteilt angenommen. In Matrixnotation lässt sich dies zusammenfassen zu

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (5.11)$$

mit

$$\mathbf{y} = \begin{pmatrix} x'_{2,1} \\ \vdots \\ x'_{2,n} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x'_{1,1} \\ \vdots & \vdots \\ 1 & x'_{1,n} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad (5.12)$$

Die Modellparameter werden mittels Kleinste-Quadrate-Schätzer bestimmt. Diese Methode ist nicht robust gegenüber Ausreißern und dadurch dem Problem angemessen. Der Schätzer der Geradenparameter ist

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (5.13)$$

Das eigentliche Interesse liegt in der Anpassungsfähigkeit der Punkte an eine Gerade. Dies wird durch den Schätzer $\hat{\sigma}^2$ für die Varianz bestimmt.

$$\hat{\sigma}^2 = \frac{1}{n-2} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{1}{n-2} (\mathbf{y}^T \mathbf{y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y}) \quad (5.14)$$

Der Faktor $1/(n-2)$ resultiert aus einer Verminderung des Freiheitsgrads (im Vergleich zum Faktor $1/(n-1)$), da $\hat{\boldsymbol{\beta}}$ bereits Berücksichtigung in der Gleichung findet. Ein Messwert $\hat{\Psi}$ für die Geradheit kann nun festgehalten werden.

$$\hat{\Psi}(\mathbf{x}'_{1:n}) := \hat{\sigma}^2 \text{ für } \mathbf{x}'_{1:n} \quad (5.15)$$

5.3 Bestimmung eines adaptiven Schwellwerts

Dem im letzten Abschnitt berechneten Messwert der Geradheit $\hat{\Psi}$ gilt es einen geeigneten Test gegenüberzustellen. Eine anschauliche, einfache, aber auch problembehaftete Methode ist die bekannte Hough-Transformation [Hough, 1962] zur Geradendetektion. Hierbei werden die Punkte des Bildraums in den Hough-Raum (Akkumulationsraum) gesammelt. Da die Methode hinlänglich bekannt ist, wird auf eine Betrachtung neben Algorithmus 1 verzichtet.

Algorithmus 1: Hough-Transformation (vereinfacht)

Data: Binary input image $I : \{1, \dots, Width\} \times \{1, \dots, Height\} \rightarrow \{0, 1\}$

Result: Hough space H

$$H_{max} \leftarrow \sqrt{\frac{1}{4}Width^2 + \frac{1}{4}Height^2}$$

$$H_{min} \leftarrow -H_{max}$$

$$H[0, \dots, \pi][H_{min}, \dots, H_{max}] \leftarrow 0$$

foreach $i \in domain(I)$ **with** $I(i) \neq 0$ **do**

for $\alpha \leftarrow 0$ **to** π **do**

$$\quad d = i_1 \cos(\alpha) + i_2 \sin(\alpha)$$

/* d is distance to the origin of a line with angle α through i

*/

$$\quad H[\alpha][d] \leftarrow H[\alpha][d] + 1$$

Ohne auf alle Nachteile der Geradenerkennung durch die Hough-Transformation einzugehen, sei darauf verwiesen, dass die Erkennung geeigneter Punkte im Hough-Raum zur Extraktion von Geraden im Bildraum einen der wesentlichen Problempunkte darstellt [Leavers, 1993]. Ein Vorteil der Hough-Transformation ist zweifelsohne die Möglichkeit, sie durch einen leicht verständlichen Algorithmus zu implementieren, sowie eine einfache Parametrisierung des Hough-Raums selbst, da die Auflösungen für Winkel- und Distanzachse in der Matrix \mathbf{H} (Algorithmus 1) je nach erforderlicher Genauigkeit bzw. Geradheit gewählt werden können.

Ebendieser Vorteil wird für den Zweck einer adaptiven Schwellwertbestimmung ausgenutzt. Der Gedanke der Hough-Transformation wird damit umgekehrt: Anstatt Punkte in den Hough-Raum zu übertragen und diese anschließend zu beurteilen, wird im Hough-Raum der Begriff der Geradheit für alle Punktmengen einheitlich festgelegt. Hierzu wird eine Kerndichteschätzung im Hough-Raum durchgeführt, wodurch eine Charakterisierung der Geradheit von Punktfolgen festgehalten wird. Diese wird anschließend für konkrete Punktfolgen in Form eines Konfidenzbereichs in das Bild übertragen. Da die Geradheitsfestlegung lediglich über die Parametrisierung des Hough-Raums selbst erfolgt, ist der Gebrauch des Raums in der Anwendung nicht mehr erforderlich.

Wie bereits angedeutet, wird der Hough-Raum über die Winkel- und Distanzauflösung parametrisiert, welche im Folgenden mit (ϕ°, κ) angegeben werden. Die Winkelauflösung wird in Grad notiert. Hierbei sei auf die in Algorithmus 1 vorkommende Hesse-Normalform einer Geraden explizit hingewiesen:

$$d = x \cos \theta + y \sin \theta \quad (5.16)$$

Der Winkel der Geraden wird als Einheitsvektor in Richtung der Normalen angegeben und ist auf das Intervall $[0; \pi)$ festgelegt.

Definition 5.1. Ausgehend von einer gewählten Hough-Auflösung (ϕ°, κ) wird eine Punktfolge als *gerade* betrachtet, wenn sich zu allen Punkten dieser Folge paarweise eine gleiche Repräsentation im Hough-Raum (eine Zelle der Hough-Matrix) finden lässt.

Der Gedanke, Punktfolgen in den Hough-Raum einzutragen, wird durch Definition 5.1 umgekehrt. Ausgehend von einer durch eine Parametrisierung vorgegebenen Zellengröße wird ein Rechteckbereich im Hough-Raum vorgegeben. Statt die Auflösung zur Steuerung der Genauigkeit der Geradenrepräsentationen durch den Hough-Raum als untere Schranke zu verwenden, wird sie als obere Schranke für die Festlegung gerader Punktfolgen genutzt.

In [Fernandes u. Oliveira, 2008] wurde untersucht, wie Punktfolgen möglichst effizient in den Hough-Raum übertragen werden können. Dabei wurde der Fehler einer linearen Regression im Bildraum auf eine Kerndichteschätzung im Hough-Raum übertragen. Dies wird nachfolgend invers modelliert. Dabei werden die Bandbreiten so gewählt, dass unter der Dichte genau eine Zelle im Hough-Raum nicht verschwindet und die Definition auf diese Weise erfüllt wird. Für die Kerndichteschätzung wird zunächst der Epanechnikov-Kern K_E [Epanechnikov, 1969] verwendet.

$$K_E(x) := \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{x^2}{5}\right) & , \text{für } |x| \leq \sqrt{5} \\ 0 & , \text{sonst} \end{cases} \quad (5.17)$$

Er bietet drei zentrale Vorteile bei der Kerndichteschätzung:

- Er besitzt eine sehr ähnliche Gestalt zum Gauß-Kern; tatsächlich besitzen beide nahezu identische kanonische Bandbreiten. Bandbreiten zu beiden Kernen können leicht ausgetauscht werden.
- Sein Träger ist kompakt. Für jeden Summanden (Kern) der zu schätzenden Dichte genügt die Berücksichtigung einer beschränkten Menge.

- Er ist optimal in dem Sinn, dass er unter allen Kernen die mittlere quadratische Abweichung (MSE) des zugehörigen Kerndichteschätzers minimiert [Epanechnikov, 1969].

Die zweidimensionale Formulierung wird durch einen Produktkern, welcher einen rechteckigen Träger besitzt, gegeben. Der bivariate Schätzer, angepasst an die Bandbreiten h_θ und h_d , lautet dann

$$\hat{f}(\theta, d) = \frac{1}{nh_\theta h_d} \sum_{i=1}^n K_E\left(\frac{\theta - x_{\theta,i}}{h_\theta}\right) K_E\left(\frac{d - x_{d,i}}{h_d}\right) \quad (5.18)$$

für die n -elementige Punktfolge $\mathbf{x}_{1:n}$ mit $\mathbf{x}_i = (x_{\theta,i}, x_{d,i})^T$ für $i \in \{1, \dots, n\}$.

Die Bandbreiten sollen als Fehlerformulierung in einen anderen Raum übertragen werden. Hierfür wird der Gauß-Kern bevorzugt, da der Fehler als normalverteilt angenommen wird und die Bandbreiten damit als Standardabweichungen anzusehen sind. Die bisherigen Bandbreiten können jedoch durch das Konzept kanonischer Bandbreiten, in diesem Fall δ_E und δ_G , transformiert werden (siehe [Marron u. Nolan, 1988], dort unter Betrachtung eines abweichend definierten Epanechnikov-Kerns). Die Formeln lauten:

$$\sigma_\theta = h_\theta \frac{\delta_G}{\delta_E} \quad \text{und} \quad \sigma_d = h_d \frac{\delta_G}{\delta_E} \quad (5.19)$$

Mit Werten von $\delta_E = 0,7687$ und $\delta_G = 0,7764$ wird diese Skalierung aufgrund des geringen Einflusses auf die Ergebnisse im Weiteren übergangen. Hinsichtlich des nicht kompakten Trägers wurde der Gauß-Kern nicht für die Kerndichteschätzung auf einer Zelle im Hough-Raum selbst verwendet.

$$\sigma_\theta \approx h_\theta \quad \text{und} \quad \sigma_d \approx h_d \quad (5.20)$$

Die Funktion T bietet eine Möglichkeit zur Transformation von Geraden zwischen den Parametrisierungen. Dabei wird der Weg von der Hesse-Normalform (Gleichung 5.16) zur Hauptform ($y(x) = n + mx$) besprochen.

$$T: \begin{cases} [0; \pi) \times \mathbb{R} & \rightarrow \mathbb{R}^2 \\ (\theta, d) & \mapsto \left(\frac{d}{\sin \theta}, -\cot \theta\right)^T \end{cases} \quad (5.21)$$

Der Sonderfall einer vertikalen Geraden kann ausgeschlossen werden. Doch bevor die Begründung dargelegt wird, wird T verwendet, um die Bandbreiten in einer Fehlerrechnung in den Bildraum fortzupflanzen. Für den Fehler auf der Regressionsgeraden folgt:

$$\sigma_y^2(x) \approx \left(\frac{\partial y}{\partial \theta} \sigma_\theta \right)^2 + \left(\frac{\partial y}{\partial d} \sigma_d \right)^2 \quad (5.22)$$

$$= \sigma_\theta^2 (-d \cot \theta \csc \theta + x \csc^2 \theta)^2 + \sigma_d^2 \csc^2 \theta \quad (5.23)$$

Geht man nun davon aus, dass eine Punktfolge betrachtet wird, welche, wie in Abschnitt 5.2 beschrieben, ausgerichtet wurde, gilt zusätzlich $\theta = \frac{\pi}{2}$. Damit kann der Ausdruck weiter vereinfacht werden. Außerdem wird der zuvor angesprochene Fall vertikaler Geraden ausgeschlossen.

$$\sigma_y^2(x) \stackrel{\theta=\frac{\pi}{2}}{\approx} \sigma_\theta^2 (-d * 0 * 1 + x * 1)^2 + \sigma_d^2 * 1 \quad (5.24)$$

$$= \sigma_\theta^2 x^2 + \sigma_d^2 \quad (5.25)$$

Dass von x abhängige Resultat beschreibt dabei eine höhere Toleranz mit zunehmender Länge. Einsetzen der Bandbreiten unter Berücksichtigung des Definitionsbereichs (Gleichung 5.17)

$$\sigma_\theta \approx h_\theta = \frac{\phi\pi}{180\sqrt{5}} \quad \text{und} \quad \sigma_d \approx h_d = \frac{\kappa}{\sqrt{5}} \quad (5.26)$$

führt zu

$$\sigma_y^2(x) \approx \frac{\phi^2 \pi^2 x^2}{180^2 * 5} + \frac{\kappa^2}{5} \quad (5.27)$$

Hierbei ist jedoch die Varianz weiterhin von x abhängig. Unter der Annahme unabhängig und identisch normalverteilter Fehler der zugrundeliegenden Gleichung $y(x)$ kann das arithmetische Mittel der Varianzen genutzt werden, um eine Varianzaussage einer Punktfolge zu erhalten (das arithmetische Mittel von n unabhängig mit $\mathcal{N}(\mu, \sigma^2)$ verteilten Zufallsgrößen ist $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ verteilt).

$$\Psi_0(\mathbf{x}'_{1:n}) := \frac{\pi^2 \phi^2 \sum_{i=1}^n \mathbf{x}'_{1,i}{}^2}{180^2 * 5n} + \frac{\kappa^2}{5} \quad (5.28)$$

Damit ist eine Abhängigkeit ausschließlich von Elementen der Designmatrix \mathbf{X} ersichtlich. Dies korrespondiert direkt mit Gleichung 5.15 und wird in Fortführung des Gedankens zur Belegung einer einzigen Hough-Zelle als Schwellwert verwendet:

$$\mathbf{x}'_{1:n} \text{ gerade} \Leftrightarrow \hat{\Psi}(\mathbf{x}'_{1:n}) \leq \Psi_0(\mathbf{x}'_{1:n}) \quad (5.29)$$

Beispiel 1. Es sei eine Punktfolge $\mathbf{x}_{1:n}$ gegeben, für welche gilt:

$$\forall i \in \{1, \dots, n\}: \mathbf{x}_{2,i} = 0 \quad (5.30)$$

Damit ist die Punktfolge bereits horizontal ausgerichtet ($\hat{\Psi}(\mathbf{x}_{1:n}) = 0$) und der Schritt, diese zu transformieren, kann entfallen. Der Abstand benachbarter Punkte wird mit 1 vorausgesetzt. Gleichung 5.28 erlaubt die Berechnung eines Konfidenzbereichs um die gegebene Punktfolge entsprechend der gewählten Hough-Parametrisierung.

Gesucht wird die Umgebung der Punktfolge, in welcher sich gerade Punktfolgen befinden können. Unter den getroffenen Voraussetzungen ist dies durch $\Psi_0(\mathbf{x}_{1:n})$ vorgegeben. Die Betrachtung wird in Abhängigkeit der Folgenlänge durchgeführt, um die Auswirkung als Bereich abbilden zu können. Parametrisiert durch ϕ und κ wird er im Folgenden als Konfidenzbereich bezeichnet.

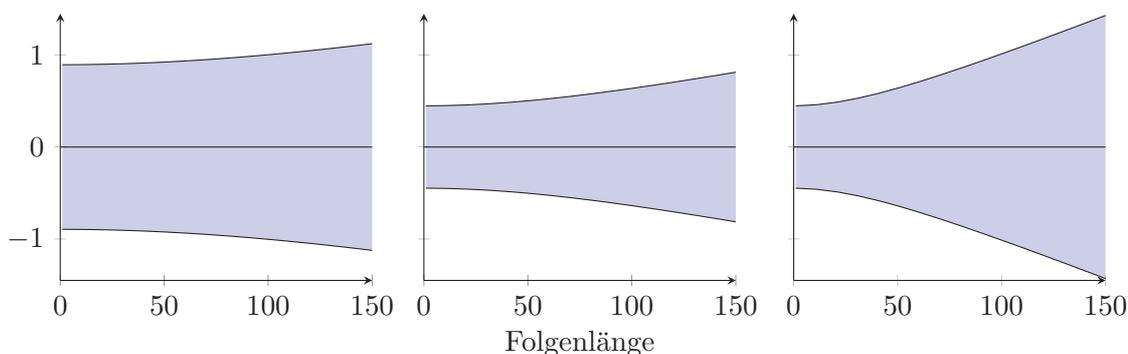


ABBILDUNG 5.2: Konfidenzbereiche um eine horizontale Punktfolge zu verschiedenen Parametern. Die Graphen bilden $\Psi_0^{\frac{1}{2}}$ und zur besseren Veranschaulichung auch $-\Psi_0^{\frac{1}{2}}$ in Abhängigkeit von der Punktfolgenlänge ab. Links: $\phi = 1, \kappa = 2$, mittig: $\phi = 1, \kappa = 1$, rechts: $\phi = 2, \kappa = 1$.

Je nach gewählten Hough-Parametern nimmt der Konfidenzbereich (Abbildung 5.2, blau) unterschiedliche Gestalten an. Aus Gründen einer besseren Darstellung wurde $\Psi_0^{\frac{1}{2}}$, die (größte erlaubte) Standardabweichung einer Punktfolge, sowohl positiv als auch negativ abgebildet. Eine Vergrößerung der Distanzauflösung κ (links) ist gegenüber der Standardparametrisierung (mittig) mit einer konstanten Vergrößerung der zulässigen Standardabweichungen verbunden. Damit werden unabhängig von der Länge der Punktfolge auch Punktfolgen mit stärkeren Abweichungen toleriert. Eine Erhöhung der Winkelauflösung ϕ (rechts) hingegen bewirkt eine Vergrößerung des Betrags des Anstiegs der Asymptote von $\Psi_0^{\frac{1}{2}}$, woraus eine größere Toleranz für Abweichungen mit zunehmender Folgenlänge einhergeht. Da auftretende Folgenlängen nicht zuletzt durch die Bildgröße beschränkt sind, können die Auswirkungen bei der Auswahl beider Werte in Anwendungen in der Regel auf endlich viele Werte eingegrenzt werden.

Wie bereits beschrieben, handelt es sich bei den in Abbildung 5.2 dargestellten Funktionen um die Standardabweichungen, welche eine Punktfolge entsprechender Länge nicht überschreiten darf (vgl. Gleichung 5.28, dort deren Quadrat, die Varianz), um als gerade angesehen zu werden. Standardabweichungen fungieren nicht als harte Schranken für einzelne Punkte. Lediglich für eine gesamte Punktfolge muss die Standardabweichung über ihre Hauptachse gering genug ausfallen, um sie als gerade charakterisieren zu können. Wird eine Punktfolge abschnittsweise auf ihre Geradheit bewertet, kann es somit vorkommen, dass Teilbereiche als gerade oder nicht gerade beurteilt werden, während die gesamte Punktfolge als gerade beurteilt wird.

5.4 Erkennung von gerader Punktfolgen

Bevor die Erkennung gerader Punktfolgen thematisiert wird, lohnt sich die Formulierung einer Hilfsfunktion für die Beurteilung der Orientierung einer Geraden.

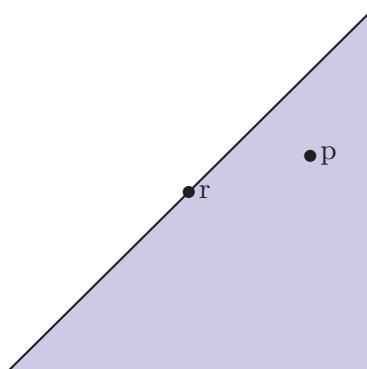


ABBILDUNG 5.3: Der schwarzen Gerade durch den Punkt \mathbf{r} ist eine Halbebene angeheftet, in welcher sich \mathbf{p} befindet.

Neben der Angabe in Hesse-Normalform und diversen anderen Möglichkeiten kann eine Gerade in der euklidischen Ebene auch durch einen Punkt der Gerade und ihren Winkel charakterisiert werden. Lässt man weiterhin zu, dass der Winkel sich nicht nur im Intervall $[0; \pi)$, sondern in $[0; 2\pi)$ bewegen kann, kann einer Geraden eine Orientierung zugeordnet werden. Hierzu wird folgende Testfunktion eingeführt:

$$\text{ProbePt}(\mathbf{p} \mid \theta, \mathbf{r}) := \text{sgn} \left(\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}^T (\mathbf{r} - \mathbf{p}) \right) \quad (5.31)$$

Hierbei wird, wie in Abbildung 5.3 schematisiert, an die zu testende Gerade eine Halbebene angeheftet. Die Seite, auf welcher dies geschieht, richtet sich dabei nach der Orientierung der Geraden. Abhängig von der Position des Testpunkts ergeben sich drei

Möglichkeiten:

$$\text{ProbePt}(\mathbf{p} \mid \theta, \mathbf{r}) = \begin{cases} 1, & \text{wenn } \mathbf{p} \text{ innerhalb der Halbebene ist,} \\ 0, & \text{wenn } \mathbf{p} \text{ auf dem Rand der Halbebene liegt, und} \\ -1, & \text{wenn sich } \mathbf{p} \text{ außerhalb der Halbebene befindet.} \end{cases} \quad (5.32)$$

Die Erkennung von geraden Punktfolgen wird mittels einer Greedy-Heuristik umgesetzt, da es mit einem exponentiellen Aufwand verbunden wäre, alle zusammenhängenden Kombinationen von Punkten in einem Bild auf ihre Geradheit zu überprüfen.

Gegeben sei für die weitere Betrachtung eine Punktfolge in einem Binärbild. Anhand der Nachbarn, welche sich an beiden Enden befinden, soll diese erweitert werden. Ziel ist, den Nachbarn auszuwählen, welcher die Punktfolge mit größtmöglicher Geradheit im Sinn von Gleichung 5.15 fortsetzt. Durch eine Wiederholung dieses Vorgehens können gerade Punktfolgen aus Saatpunkten berechnet werden. Abgebrochen wird, sobald die gesamte Punktfolge nicht mehr gerade ist (Gleichung 5.29). Um die Punktfolge sinnvoll fortzusetzen und nicht etwa den bereits beschrittenen Weg der Punktfolge erneut abzulaufen, wird die eingeführte Halbebene verwendet.

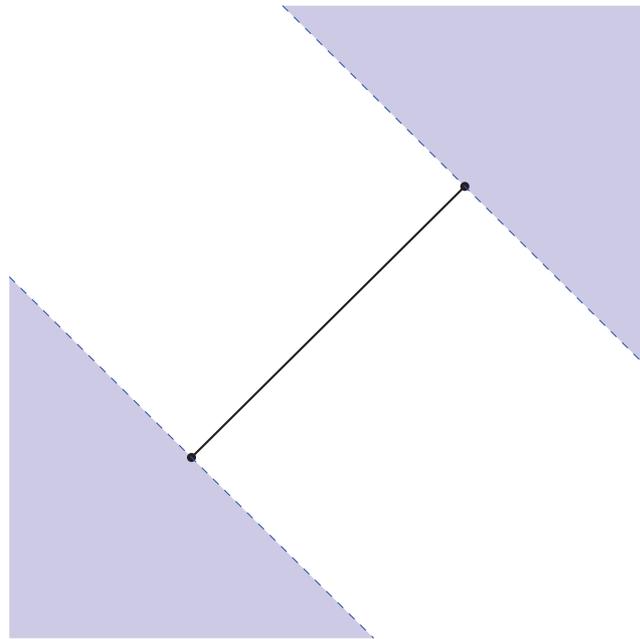


ABBILDUNG 5.4: An die Endpunkte des abgebildete Geradensegments wurden zwei Halbebenen mit parallel verlaufenden Rändern senkrecht zur Gerade positioniert. Nur der blaue Bereich kommt für neue Punkte in Frage.

Wie in den Gleichungen 5.3 bis 5.8 dargelegt, kann aus der Streumatrix einer Punktfolge $\mathbf{x}_{1:n}$ ihre Normale bestimmt werden. Sei deren Winkel wiederum mit θ bezeichnet und die Punktfolge weiterhin so ausgerichtet, dass $\text{ProbePt}(\mathbf{x}_1 \mid \theta, \mathbf{x}_n) < 0$ ist. Im Falle

$\text{ProbePt}(\mathbf{x}_1 \mid \theta, \mathbf{x}_n) > 0$ kann die Orientierung jener Punktfolgen durch Addition von π (modulo 2π) Berücksichtigung finden. Im Fall $\text{ProbePt}(\mathbf{x}_1 \mid \theta, \mathbf{x}_n) = 0$ stellt die Punktfolge kein Geradensegment dar. Nun können am Beginn und Ende der Punktfolge, wie in Abbildung 5.4 dargestellt, Halbebenen eingefügt werden. In Betracht für die Erweiterung der Punktfolge in einem Schritt kommen nur Punkte \mathbf{p} , für welche gilt:

$$\begin{aligned} \mathbf{p} \in \text{Neighbors}(\mathbf{x}_1, I) \cup \text{Neighbors}(\mathbf{x}_n, I): \\ \text{ProbePt}(\mathbf{p} \mid \theta, \mathbf{x}_n) \geq 0 \wedge \\ \text{ProbePt}(\mathbf{p} \mid \theta + \pi, \mathbf{x}_1) \geq 0 \end{aligned} \quad (5.33)$$

Unter diesen Punkten wird derjenige ausgewählt, welcher die Punktfolge mit minimalem $\hat{\Psi}$ fortsetzt. Stehen keine Punkte mehr zur Auswahl oder ist die Punktfolge nicht länger gerade, bricht das Verfahren ab.

Ausgangsbasis für die Suche können zufällig gewählte (Saat-)Punkte sein. Häufig entstehen Binärbilder jedoch durch Kantenberechnungen aus Farb- oder Graustufenbildern, sodass sich die lokalen Maxima aus Filteroperationen anbieten (und auch hier verwendet werden). Weiterhin ist anzumerken, dass die u. a. mit Gleichung 5.15 bereitgestellten Hilfsmittel bei sehr kurzen Punktfolgen schon aus algebraischen Gründen nicht anwendbar sind. Stehen durch die Kantenberechnung Kantenwinkel zur Verfügung, wird die folgende Strategie genutzt:

1. Auswahl eines Punktes im Binärbild, welcher mit einem lokalen Maximum im zugehörigen Kantenbild korrespondiert. Dieser Punkt ist der Saatpunkt für eine Punktfolge.
2. Erweiterung der Punktfolge in dem durch Gleichung 5.33 vorgegebenen Bereich mit dem Punkt, welcher die geringste Differenz zum Winkel der Punktfolge aufweist (ggf. mehrere).
3. Wiederholung von Schritt 2, bis die Punktfolge die notwendige Länge erreicht hat (mindestens drei Punkte).
4. Erweiterung der Punktfolge in dem durch Gleichung 5.33 vorgegebenen Bereich unter Minimierung von $\hat{\Psi}$ (Gleichung 5.15).
5. Wiederholung von Schritt 4, bis die Punktfolge Gleichung 5.29 verletzt.
6. Wiederholung beginnend mit Schritt 1, bis alle lokalen Maxima betrachtet wurden.

Da es im Allgemeinen wünschenswert ist, Saatpunkte nicht im Resultat zu berücksichtigen, wenn ausgehend von ihnen keine gerade Punktfolge berechnet werden konnte,

werden nur Punktfolgen gespeichert, welche bei Verletzung von Gleichung 5.29 eine Mindestlänge aufweisen.

5.5 Algorithmische Betrachtung

Das vorgestellte Verfahren wird mit den eingeführten Funktionen nachfolgend als Pseudo-Code notiert. Hierfür sei Q eine Transformation eines Punktes oder einer Punktfolge, ähnlich zu Gleichung 5.9. Q besitzt zwei Argumente: das erste Argument gilt es neu auszurichten, das zweite Argument bestimmt die Parameter der Transformation. Genauer wird von der Funktion Q das erste Argument mit derjenigen Transformation ausgerichtet, welche das zweite Argument horizontal und zentriert positioniert (unter Verwendung von arithmetischem Mittel und Eigenvektoren).

Die in Algorithmus 2 dargelegte Erkennung von geraden Punktfolgen stellt das Herzstück des Verfahrens dar. Mit der Erweiterung bzw. Fortsetzung von Punktfolgen werden die Schritte 4 und 5 der in Abschnitt 5.4 genannten Strategie umgesetzt. Wiederum werden gültige Punkte bestimmt (*candidates*), aus welchen dann der beste Punkt *bestpos* für eine Fortsetzung gesucht wird. Da dieser Quelltext in erster Linie der Erklärung dient, ergeben sich verschiedene Möglichkeiten der Optimierung:

- Infolge der mit Q einhergehenden Berechnung der Ausrichtung ist das Verfahren zur Erweiterung von Punktfolgen mit quadratischem Aufwand verbunden. Es lässt sich jedoch eine sequentielle Methode auf Basis von Givens Rotationen beschreiben, in welcher die Designmatrix zerlegt wird [Maindonald, 1984, Kapitel 4]. Für den Aufwand der Transformation in jedem Schritt kann damit eine obere Schranke gefunden werden. Der Aufwand für die Fortsetzung von Punktfolgen ist dann linear (jeweils zur Folgenlänge).
- Die Anzahl der Saatpunkte kann reduziert werden. Es müssen keine Saatpunkte betrachtet werden, welche bereits in Punktfolgen vorhanden sind.
- Wie bei [Grompone von Gioi et al., 2012] kann eine Verkleinerung des Eingangsbilds nicht nur zur Glättung, sondern auch für eine verbesserte Erkennung genutzt werden. Da sich die Reduktion der Bildgröße quadratisch auswirkt ([Grompone von Gioi et al., 2012] verwenden eine 80 %-Skalierung), verkleinert dies auch den Suchraum deutlich.

Algorithmus 2: Erweiterung von geraden Punktfolgen**Data:** Binary input image $I : \{1, \dots, Width\} \times \{1, \dots, Height\} \rightarrow \{0, 1\}$ Straight point list $line$ **Result:** Straight point list $line$ **repeat**

```

    /* reorientate the line list */
     $\theta \leftarrow$  [normal of line, Equation 5.3 to 5.8]
     $signum \leftarrow$  ProbePt( $line.front \mid \theta, line.back$ )
    if  $signum > 0$  then
        |  $signum \leftarrow signum + \pi$ 
    /* candidates, back of line */
     $candidates_b \leftarrow$  Neighbors( $line.back, I$ )
    forall the  $c \in candidates_b$  do
        |  $signum \leftarrow$  ProbePt( $c \mid \theta, line.back$ )
        | if  $signum < 0$  then
            | |  $candidates_b \leftarrow candidates_b \setminus \{c\}$ 
    /* candidates, front of line */
     $\theta \leftarrow \theta + \pi$ 
     $candidates_f \leftarrow$  Neighbors( $line.front, I$ )
    forall the  $c \in candidates_f$  do
        |  $signum \leftarrow$  ProbePt( $c \mid \theta, line.front$ )
        | if  $signum < 0$  then
            | |  $candidates_f \leftarrow candidates_f \setminus \{c\}$ 
     $candidates \leftarrow candidates_b \cup candidates_f$ 
    /* which one continues the line best? */
     $line_t \leftarrow Q(line \mid line)$ 
     $bestval \leftarrow \infty, bestpos \leftarrow \emptyset$ 
    forall the  $c \in candidates$  do
        |  $c_t \leftarrow Q(c \mid line)$ 
        |  $s = \hat{\Psi}(line_t \cup \{c_t\})$ 
        | if  $s < bestval$  then
            | |  $bestval \leftarrow s, bestpos \leftarrow c$ 
    if  $bestval = \infty$  then
        | break
     $line \leftarrow line \cup \{bestpos\}$ 

```

until $\hat{\Psi}(Q(line \mid line)) \leq \Psi_0(Q(line \mid line))$

5.6 Stand der Forschung

Da die Hough-Transformation in der Geradenerkennung grundlegende Vorteile, wie Robustheit gegenüber Störungen, Aliasierungsfehlern und Lücken, mit sich bringt, zielt eine Reihe von Verfahren darauf ab, die angesprochenen Nachteile abzumildern [Leavers, 1993]. [Shi et al., 2010] stellen einen Ansatz unter Verwendung der Radon Transformation vor und zeigen, dass sich im Akkumulationsraum Geraden mit geringerer

Falsch-Positiv-Rate erkennen lassen. Sie erklären jedoch auch, dass der Ansatz damit das Problem der Erkennung geeigneter Kantenkandidaten im Akkumulationsraum von der Hough-Transformation erbt. Ebenfalls zu erwähnen ist die randomisierte Hough-Transformation [Xu u. Oja, 1993], bei welcher der Bildraum nicht sukzessiv betrachtet wird, sondern zufällig Punkte im Bild ausgewählt werden. In jedem Schritt wird aus zwei im Bild gesetzten Pixeln eine Geradenhypothese formuliert, welche in den Hough-Raum eingetragen wird. Damit kann das Verfahren als akkumulierender RANSAC betrachtet werden. Speziell mit dem Problem des Rechenaufwands beschäftigen sich [Fernandes u. Oliveira, 2008]. Wie im vorgestellten Ansatz wird hier eine Kerndichteschätzung im Hough-Raum durchgeführt. Der ausschlaggebende Unterschied ist jedoch, dass, nach einer Suche nach annähernd kollinearen Punktfolgen im Bildraum, Fehlerabschätzung und -fortpflanzung für die Akkumulierung im Hough-Raum genutzt werden. Im Hough-Raum reduziert sich die Anzahl der Zellen, welche in einem Schritt des Algorithmus verändert werden müssen, im Vergleich zum Brute-Force-Ansatz der Hough-Detektion deutlich. Ebenso ist die Anzahl der Iterationen geringer, da nur Geradenkandidaten in den Hough-Raum eingetragen werden. Mit der Vorauswahl zusammenhängender Punktfolgen und der nachfolgenden Bestimmung von Geraden im Hough-Raum stellt sich das Problem der Auswahl geeigneter Geraden(-segmente) bzw. der Parameter zur Auswahl jedoch doppelt.

Auch hierarchische Ansätze zur Hough-Transformation sind zu finden. [Princen et al., 1990] verwenden eine Bildpyramide, in welcher das Bild bis zur untersten Stufe in kleiner werdende Teilbilder zerlegt wird. Nur auf der untersten Stufe wird die Hough-Transformation für die einzelnen Teilbilder angewandt. Höhere Stufen ergeben sich durch Gruppierung von benachbarten Geradensegmenten zu längeren Geradensegmenten. In Kombination mit Random Forests, so genannten Hough Forests, hat sich die Hough-Transformation auch im Bereich der Objekt- [Gall u. Lempitsky, 2013] und Personenerkennung [Barinova et al., 2012] als effektiv erwiesen.

Die Erkennung von Geradensegmenten betrachten auch [Lee et al., 2006]. Inspiriert von den Veröffentlichungen [Guru et al., 2004, Nagabhushan et al., 2005, Shekar et al., 2006] leiten sie ein Maß für die Geradheit von Punktfolgen aus der Hauptkomponentenanalyse ab. Dieses ist λ_2 (Gleichung 5.4), der kleinere Eigenwert der Streumatrix. Entsprechend ist dieser Wert eine Schätzung für die Streuung der Punktmenge senkrecht zur Hauptachse und aus einem kleinen Wert kann eine gerade Punktfolge geschlossen werden. Obwohl ein Schwellwert in Abhängigkeit von der Punktfolgenlänge notwendig ist, konnte leider keine Herleitung beschrieben werden. [Schmid u. Zisserman, 1997] verwenden einen Wurm im binären Kantenbild, um „Ecken“ in Punktfolgen zu lokalisieren. Anschließend werden Abschnitte zwischen Ecken an ein Geradenmodell angepasst. Unterschreitet der Fehler dabei einen (gering gewählten) Schwellwert, gilt ein Geradensegment als gefunden.

Einen Ansatz auf Grundlage der Gestalt-Theorie [Desolneux et al., 2007, 2000] verfolgen [Von Gioi et al., 2012]. Auf Basis des Verfahrens von [Burns et al., 1986] werden Punkte im Bild zu Punktfolgen verbunden. Motiviert durch das Helmholtz-Prinzip, welches besagt, dass keine Strukturerkennung im Rauschen erfolgen soll, wird ein *A-contrario*-Ansatz beschrieben, welcher Punktfolgen auf die Wahrscheinlichkeit ihres zufälligen Auftretens hin bewertet. Eine derartige Modellierung erlaubt die Schwellwertbildung über der Anzahl falscher Alarme im gesamten Bild. Obwohl die Autoren das Verfahren als parameterfrei bezeichnen, ist anzumerken, dass aufgrund von vordefinierten Verfahrensparemtern, etwa dem Schwellwert für die Klassifizierung von zwei Winkeln als gleich ($22,5^\circ$) in der Bildung der Punktfolgen, dies auch durch vorgegebene Werte erzielt wird. Dieses State-of-the-Art-Verfahren wird zum Vergleich in der Evaluation herangezogen. In Weiterführung dieses Verfahrens stellen [Akinlar u. Topal, 2011] einen Ansatz vor, welcher Punktfolgen auf Grundlage des Helmholtz-Prinzips validiert. Die Kandidatensuche selbst basiert auf Grundlage der hier ebenso verwendeten Anpassung durch einen Kleinste-Quadrate-Schätzer. Da längere Punktfolgen im Rauschen immer unwahrscheinlich sind, erübrigt sich für diese der Helmholtz-Validierungsschritt. Letztlich stellt sich auch hier die Frage nach der Modellierung eines geeigneten Schwellwerts für den Fehler der linearen Anpassung (Gleichung 5.28). Die Geschwindigkeitsvorteile des Verfahrens von [Akinlar u. Topal, 2011], etwa vom Faktor 10, sind im vorgestellten Verfahren aufgrund der gemeinsamen Strategie der linearen Anpassung, gegenüber der auf Zufall basierenden Bewertung von [Von Gioi et al., 2012], ebenso realisierbar.

5.7 Evaluation

Im ersten Teil der Evaluation wird die Auswirkung der Parameter betrachtet. Im zweiten Teil folgt ein Vergleich des vorgestellten Verfahrens mit dem Verfahren nach [Von Gioi et al., 2012]. Abbildung 5.5 zeigt das Ergebnis der Detektion von Geradensegmenten nach dem vorgestellten Verfahren unter zwei verschiedenen Parametersätzen. Durch die Wahl eines größeren Parameters κ (Abbildung 5.5, C) bleiben gegenüber dem binären Kantenbild (Abbildung 5.5, B) kurze Punktfolgen erhalten, da hierfür höhere Abweichungen toleriert werden. Gut ist dies vor allem an den Punkten im linken Bildteil erkennbar, welche aus dem Blattwerk resultieren. Bei Verminderung des Parameters κ (Abbildung 5.5, D) entfällt eine Vielzahl an Punkten, während die geradlinige Gebäudestruktur korrekt und der Erwartung entsprechend wiedergegeben wird.

Der Vergleich mit dem Verfahren nach [Von Gioi et al., 2012] wird zuerst anhand zweier Bilder aus dessen Veröffentlichung (Abbildung 5.6, A und D) sowie der Kombination aus



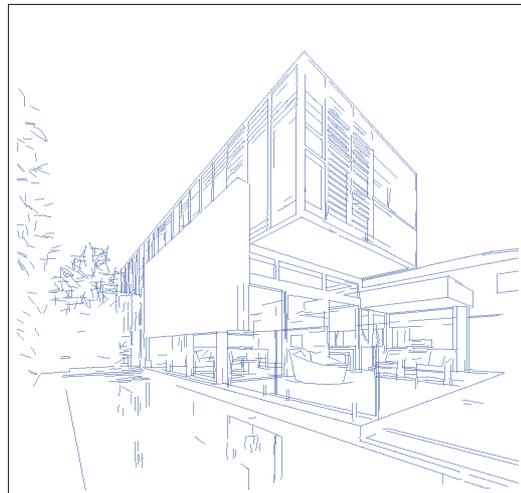
(A) Originalbild.
Bildquelle: [Swalwell u. Clarke, 2012]



(B) Binäres Kantenbild (Sobelfilter),
60 060 Punkte



(C) Detektion von Geradensegmenten
mit $\phi = 1$ und $\kappa = 3$, 45 560 Punkte



(D) $\phi = 1$ und $\kappa = 1$,
36 044 Punkte

ABBILDUNG 5.5: Einfluss des Parameters κ auf den vorgestellten Ansatz

Rauschen und einer Strukturformel (Abbildung 5.6, G) geführt. Für die Verfahrensparameter wird im Folgenden einheitlich ($\phi = 1, \kappa = 1$) festgelegt.

Im Beispiel *Chairs* sind nur geringfügige Unterschiede zwischen beiden Ergebnissen feststellbar. Mit dem vorgestellten Ansatz können im Hintergrund geringfügig bessere Ergebnisse erzielt werden, im Vordergrund werden wenige Kanten abweichend gehandhabt. Durch den vorgestellten Ansatz ist es hingegen möglich, die gerade Treppenstruktur am Gebäude im Vordergrund im Beispiel *Le Piree* zu detektieren. Ebenso werden die geraden Kanten der Gebäudefront zu großen Teilen korrekt wiedergegeben — Strukturen, die dem Verfahren nach [Von Gioi et al., 2012] zum Teil verborgen bleiben. Beide Algorithmen erkennen Geradensegmente in anisotropen Bereichen, zum Beispiel dem Wasser — eine Eigenschaft, für die der vorgestellte Ansatz möglicherweise größere Anfälligkeit

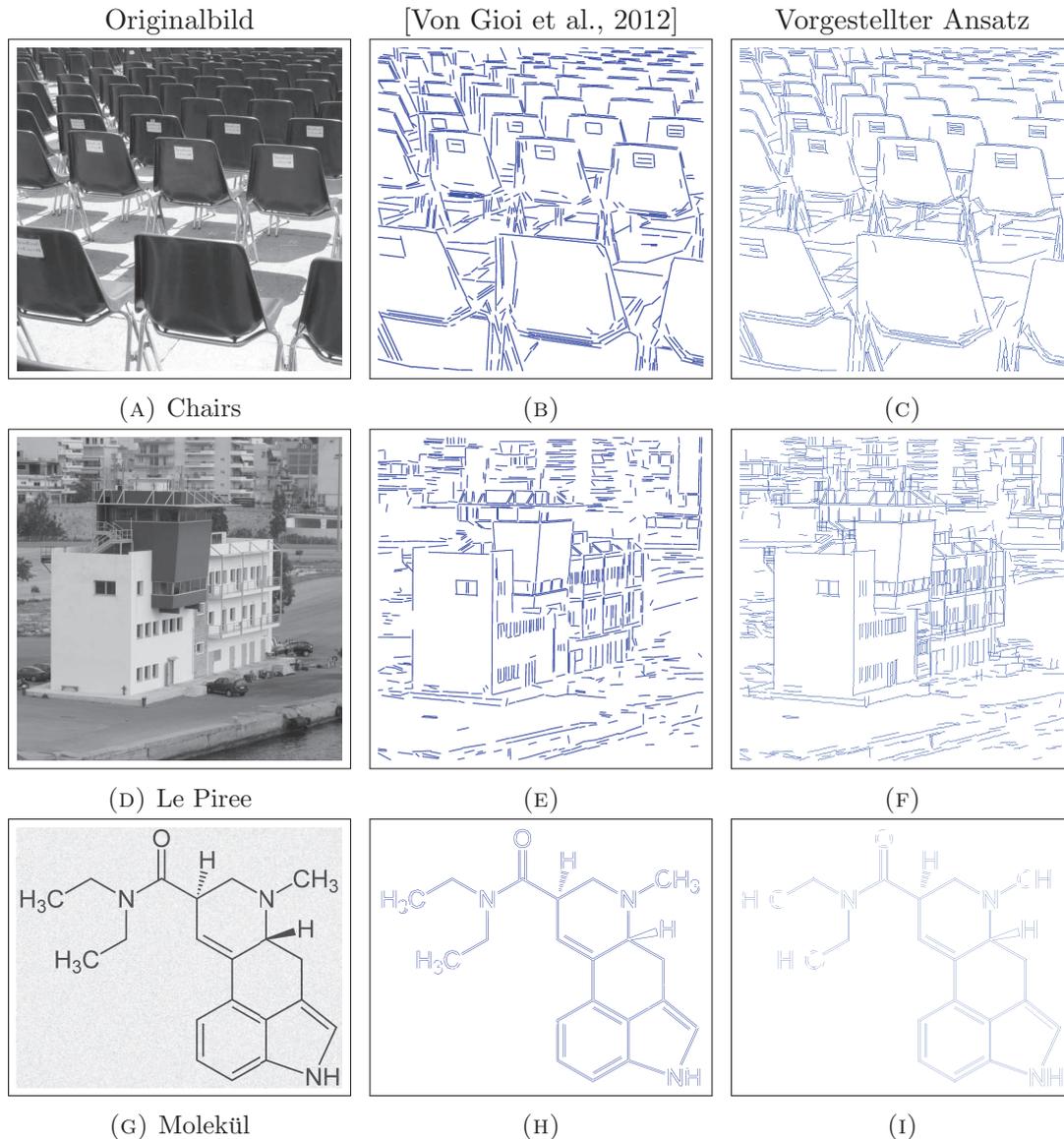
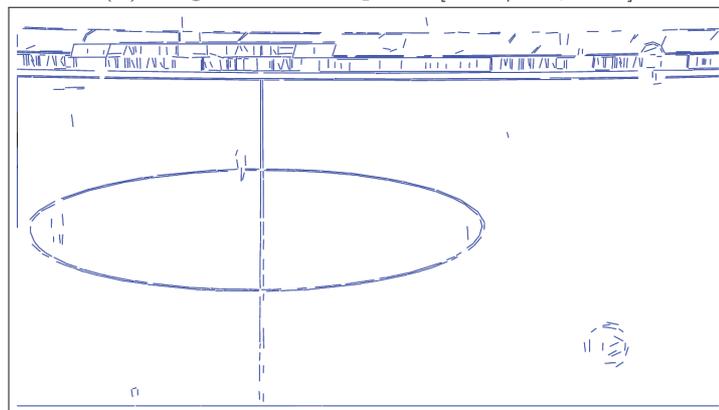


ABBILDUNG 5.6: Vergleich des Verfahrens nach [Von Gioi et al., 2012] mit dem vorgestellten Ansatz. Im Bild erkannte gerade Punktfolgen werden für die Darstellung der Ergebnisse des vorgestellten Ansatzes an ihren ursprünglichen Positionen (pixelweise) wiedergegeben. Dem Verfahren von [Von Gioi et al., 2012] nach werden sie als Strecken gezeichnet. Diese sind in den Bildern (B), (E) und (H) dicker dargestellt.

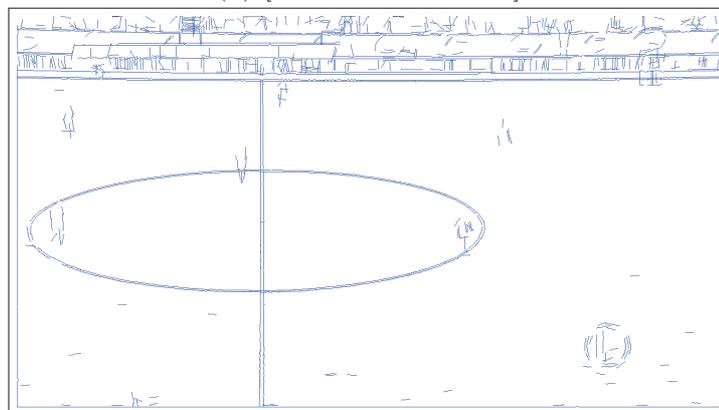
besitzt, da keine Kantenrichtungen berücksichtigt werden. Die Überlagerung durch ein Rauschen in Beispiel *Molekül* stellt in dieser Hinsicht jedoch kein Problem für eines der beiden Verfahren dar. Der vorgestellte Ansatz entfernt die Zahl „3“ sowie die Teile der Buchstaben „C“ und „O“ mit größtem „Krümmungsradius“ aus dem Bild, während das Referenzverfahren nahezu alle Kanten des Originalbilds beibehält. Je nach Wahl des Parameters κ kann auch dieses Verhalten erreicht werden oder die genannten Buchstaben können komplett entfernt werden.



(A) Originalbild. Bildquelle: [ARD/BR, 2012]



(B) [Von Gioi et al., 2012]



(C) Vorgestellter Ansatz

ABBILDUNG 5.7: Vergleich der Geradensegment-Erkennung im Kontext Fußball

Abschließend wird der Vergleich beider Ansätze im Fußballkontext illustriert. Abbildung 5.7 zeigt die erkannten Geradensegmente beider Verfahren. Dem vorgestellten Ansatz ist anzulasten, dass mehr Geradenstücke in anisotropen Bereichen, hier im Publikum und im Rasen, gefunden werden. Dies gründet u. a. in der Tatsache, dass für die vorangestellte Kantendetektion ein niedriger Schwellwert verwendet wurde. Auf der anderen Seite wird ein höherer Anteil der Feldlinien erkannt. Die Mittellinie, welche aufgrund des Rasens eine Störung aufweist, wird durch das gewählte Modell zufriedenstellend beschrieben.

Ebendiese treffsichere Geradenerkennung des Spielfelds wird im nächsten Kapitel in ein robustes Verfahren zur Kamerakalibrierung eingehen.

5.8 Diskussion und Ausblick

In diesem Kapitel wurde ein neuer Ansatz zur Erkennung von Geradensegmenten vorgestellt. Den Schwerpunkt bildete die formale Herleitung eines Konfidenzbereichs, in welchem Punktfolgen als gerade angesehen werden. Die Größe dieses Bereichs wird über die bekannten Parameter der Hough-Transformation zur Festlegung der Auflösung des Hough-Raums gesteuert. So wie jeder andere betrachtete Ansatz ist damit auch das vorgestellte Verfahren letztlich von der Wahl von Schwellwerten abhängig und muss sich dieser Kritik stellen. Unter der Annahme, dass die Hough-Transformation unter Bildverarbeitung Bekanntheit genießt, ist diese Kritik jedoch teilweise zu entkräften. Vor der Ausführung des Algorithmus muss festgelegt werden, bis zu welchen Abweichungen man Punktfolgen als gerade anerkennt. Sind die Eingangsdaten unbekannt oder variieren sie in ihrer Skalierung, kann dies einen Nachteil darstellen. Ausgehend von Saatpunkten können gerade Punktfolgen mit linearem Aufwand in einem Bild gefunden werden. Der vorgestellte Ansatz gehört damit zu den schnellsten Verfahren zur Detektion von Geradensegmenten. Eine Implementation unterstrich die Qualität oberhalb des State-of-the-Art-Niveaus. Obwohl hierbei mit quadratischem Aufwand berechnet wurde, war die Bearbeitungszeit vergleichbar mit der des Referenzsystems.

Im Rahmen einer folgenden Veröffentlichung des vorgestellten Ansatzes ist angedacht, eine Implementation mit linearer Laufzeit zur Verfügung zu stellen. Neben einem zusätzlichen Augenmerk auf Anisotropie soll das Verfahren dabei in erster Linie um die Erkennung von geraden Punktfolgen mit Subpixelgenauigkeit erweitert werden. Im Rahmen einer Optimierung ist der Ansatz für eine verteilte Berechnung prädestiniert. Die Möglichkeit der parallelen Ausführung ist angedacht. Ebenso kann in diesem Rahmen die Einschränkung auf binäre Kantenbilder fallen. Da diese üblicherweise aus Bildern mit einem größerem Wertebereich hervorgehen, werden letztendlich weitere Parameter im Rahmen der Kantendetektion und (Hysteresis-)Schwellwertbildung verborgen. Dies ist eine Gemeinsamkeit der betrachteten Ansätze, welche einer gesonderten Betrachtung bedarf.

Kapitel 6

Kameraparameter-Nachführung

6.1 Einleitung

Die Berechnung von Weltkoordinaten zu Koordinaten von Bildpunkten erlaubt die Zuordnung eines festen Bezugssystems zu Bildobjekten unabhängig von der jeweiligen Kameraposition und -ausrichtung, von welcher das Bild erzeugt wurde. Einer solchen Zuordnung kommt in der Verarbeitung von Aufnahmen von Fußballspielen durch bewegte Kameras eine entscheidende Bedeutung zu, da im (orthonormalen) Weltkoordinatensystem Spielfeldobjekte mit den gebräuchlichen physikalischen Größen vermessen werden können. Da das Spielfeld selbst als planar angesehen werden kann (mehr dazu in Abschnitt 6.4), wird für ein Bild eine Abbildung zwischen zwei Ebenen, der Bildebene und der Spielfeldebene (dem Spielfeldmodell, dem Weltkoordinatensystem) gesucht. Statt einer Kamerakalibrierung hinsichtlich äußerer und innerer Parameter einer projektiven Kamera wird das Problem auf die Bestimmung einer Homographie (Rektifizierung) reduziert. Mit einer gegebenen Bildfolge (aufgenommen aus der Kamera-1-Position) erweitert sich die Problemstellung auf die Berechnung der Bild-Spielfeld-Homographien für alle enthaltenen Bilder. Die nachzuführenden Kameraparameter werden durch Homographien ausgedrückt. Unter Berücksichtigung der



Ähnlichkeit aufeinanderfolgender Bilder werden Bild-zu-Bild-Homographien in planaren Spielfeldbereichen der Bilder eingesetzt, um eine Bild-Spielfeld-Zuordnung auch dort zu ermöglichen, wo dies die natürlichen Landmarken im Bild allein nicht zulassen.

In diesem Kapitel wird nach Aufbereitung wichtiger Grundlagen zu Homographien (Abschnitt 6.2) die Klasse der quasi-affinen Homographien in Abschnitt 6.3 eingeführt. Wie erklärt werden wird, stellen jene ein entscheidendes Werkzeug für die robuste Bestimmung relevanter Homographien dar. In diesem Rahmen wird ein Verfahren zur Feststellung der Eigenschaft der Quasi-Affinität vorgeschlagen und im gesamten Kapitel angewandt.

Nach der Einführung des Spielfeldmodells (Abschnitt 6.4) wird die Kalibrierung einer Kamera (d. h. die Bestimmung von Kameraparametern in Form einer Homographie) durch einen modifizierten *Iterative-Closest-Point- (ICP)* Algorithmus vorgestellt, welcher sowohl bezüglich des festgelegten Fehlerterms als auch bezüglich der Verknüpfungen zwischen Bildpunkten und Modellpunkten symmetrisch ausgeführt wird. Dieses in Abschnitt 6.5 beschriebene Vorgehen stellt ein nicht-lineares Optimierungsverfahren dar, welches im Moment des Abbruchs das Auffinden eines globalen Minimums nicht sicherstellen kann. In Kombination mit einem planaren Tracking-Verfahren für Bilder, für welche keine automatische Bild-Modell-Zuordnung möglich ist (etwa in solchen, in denen nicht genügend Spielfeldlinien sichtbar sind) wird der modifizierte ICP-Algorithmus in Abschnitt 6.6 eingesetzt, um ein Verfahren zur Bestimmung der Homographien in langen Bildsequenzen zu konstruieren. Da hierbei eine hinreichend genaue Näherung für den Anfangszustand des modifizierten ICP-Algorithmus durch das jeweils vorige Bild anzunehmen ist, kann die Einschränkung auf eine Suche in Richtung lokaler Minima toleriert werden. Die Thematik der Bestimmung einer Homographie ohne Wissen um eine Näherung wird in Abschnitt 6.7 betrachtet. Letztlich werden innerhalb der Evaluation (Abschnitt 6.8) synthetische Daten, reale Begegnungen und das Beispiel eines Fußgängerüberwegs näher untersucht. Neben der Robustheit werden auch die Einschränkungen des Verfahrens erörtert.

6.2 Grundlagen zur projektiven Geometrie

Die gängige Literatur bietet bereits gute Übersichten zur projektiven Geometrie [Beutelspacher u. Rosenbaum, 1992, Fischer, 2001, Hartley u. Zisserman, 2003], sodass jene an dieser Stelle nicht vollständig wiederholt wird. Einige Erkenntnisse werden jedoch gekürzt zusammengefasst, um ein Vokabular für dieses Kapitel zu schaffen.

Definition 6.1. Der reell-projektive Raum $\mathbb{P}^n(\mathbb{R})$ ist ein Quotientenraum des \mathbb{R}^{n+1} mit kanonischer Quotiententopologie, auf einer Äquivalenzrelation \sim . Es gilt:

$$\mathbb{P}^n(\mathbb{R}) := (\mathbb{R}^{n+1} \setminus \{\mathbf{0}\}) / \sim \quad (6.1)$$

$$(x_1, \dots, x_{n+1})^T \sim (y_1, \dots, y_{n+1})^T \quad \Leftrightarrow \quad \exists \lambda \in \mathbb{R} \setminus \{0\}: (x_1, \dots, x_{n+1})^T = \lambda(y_1, \dots, y_{n+1})^T \quad (6.2)$$

$\mathbb{P}^n(\mathbb{R})$ kann als Raum aller Geraden bzw. eindimensionaler Unterräume in \mathbb{R}^{n+1} , welche den Ursprung passieren, angesehen werden. Da im Weiteren immer von reell-projektiven Räumen ausgegangen wird, kann verkürzt $\mathbb{P}^n := \mathbb{P}^n(\mathbb{R})$ geschrieben werden. Obschon die Elemente von \mathbb{P}^n Äquivalenzklassen sind, werden sie entsprechend der kanonischen Quotiententopologie $(\mathbb{R}^{n+1} \setminus \{0\}) \rightarrow (\mathbb{R}^{n+1} \setminus \{0\}) / \sim$ der üblichen Schreibweise folgend mit Vektoren ausgedrückt. Dabei sind Vielfache gemäß Gleichung 6.2 äquivalent. Dies wird im Weiteren durch ein \propto -Zeichen kenntlich gemacht. Die Koordinaten der Punkte des Raums heißen homogene Koordinaten. Ist für einen Punkt die letzte Komponente $\neq 0$, korrespondiert er zu einem finiten Punkt in \mathbb{R}^n . Die übrigen Punkte werden als Punkte im Unendlichen bezeichnet. Vorausgesetzt wird fortan $n \geq 2$.

Sowohl Geraden als auch Punkte der Ebene können als Elemente von \mathbb{P}^2 dargestellt werden. Eine Geradengleichung $Ax + By + C = 0$ besitzt die Darstellung $(A, B, C)^T \in \mathbb{P}^2$ und es gilt $x^T(A, B, C) = 0$ für alle Punkte $x \in \mathbb{P}^2$, welche auf der Geraden liegen. Nicht-identische Geraden mit gleichem Anstieg schneiden sich in Punkten $(x, y, 0) \in \mathbb{P}^2$ im Unendlichen.

Definition 6.2. Eine Projektivität ist eine Abbildung $h: \mathbb{P}^n \rightarrow \mathbb{P}^n$, welche von einem Vektorraumautomorphismus $\mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$ induziert wird.

Eine Projektivität wird auch als Homographie oder projektive Transformation bezeichnet. Aufgrund der Eigenschaft der Erhaltung von Geraden ist auch der Begriff der Kollineation gebräuchlich. Mit der Komposition als Verknüpfungselement bilden die Projektivitäten auf einem projektiven Raum \mathbb{P}^n eine Gruppe. Der Automorphismus $\mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$ induziert eine Bijektion $\mathbb{P}^n \rightarrow \mathbb{P}^n$. Damit gilt:

Satz 6.1. Eine Abbildung $h: \mathbb{P}^n \rightarrow \mathbb{P}^n$ ist genau dann eine Projektivität, wenn eine nicht-singuläre reelle $(n+1) \times (n+1)$ -Matrix \mathbf{H} existiert, sodass für jeden Punkt $\mathbf{x} \in \mathbb{P}^n$ die Gleichung $h(\mathbf{x}) = \mathbf{H}\mathbf{x}$ erfüllt wird.

Für einen Beweis sei an dieser Stelle auf die Literatur verwiesen [Gruenberg u. Weir, 1977]. Aufgrund der Eigenschaft der Matrix \mathbf{H} , die Homographie vollständig zu beschreiben, wird auch diese selbst als Homographie bezeichnet. Auch für sie gilt die aus der

Äquivalenzrelation rührende Tatsache, dass Vielfache als äquivalent zu betrachten sind und damit zur gleichen Transformation führen. Die Anzahl der Freiheitsgrade ist mit $(n + 1)^2 - 1$ bestimmt.

Eine Untergruppe bilden affine Homographien. Eine ihrer Eigenschaften ist es, die Menge der Punkte im Unendlichen zu fixieren (jedoch nicht punktweise). Eine affine Homographie ist eine nicht-singuläre Matrix der Form:

$$\mathbf{H} = \begin{pmatrix} h_{1,1} & \cdots & \cdots & h_{1,n+1} \\ \vdots & \ddots & \ddots & \vdots \\ h_{n,1} & \cdots & \cdots & h_{n,n+1} \\ 0 & \cdots & 0 & 1 \end{pmatrix} \quad (6.3)$$

Dadurch sinkt die Anzahl der Freiheitsgrade auf $n(n + 1)$.

Für eine Homographie \mathbf{H} und zwei (homogene) Punkte $\mathbf{x}, \mathbf{y} \in \mathbb{P}^n$ sei $\mathbf{y} \propto \mathbf{H}\mathbf{x}$. Dann gilt für $\lambda \in \mathbb{R} \setminus \{0\}$:

$$\begin{aligned} y_1 \lambda &= h_{1,1}x_1 + \cdots + h_{1,n+1}x_{n+1} \\ &\vdots \\ y_{n+1} \lambda &= h_{n+1,1}x_1 + \cdots + h_{n+1,n+1}x_{n+1} \end{aligned} \quad (6.4)$$

Sei nun weiter $n = 2$ und $x_3 = 1$, so lautet die Transformation in inhomogene Koordinaten $(x', y')^T$:

$$x' = \frac{y_1}{y_3} = \frac{h_{1,1}x_1 + h_{1,2}x_2 + h_{1,3}}{h_{3,1}x_1 + h_{3,2}x_2 + h_{3,3}} \quad (6.5)$$

$$y' = \frac{y_2}{y_3} = \frac{h_{2,1}x_1 + h_{2,2}x_2 + h_{2,3}}{h_{3,1}x_1 + h_{3,2}x_2 + h_{3,3}} \quad (6.6)$$

Auf der rechten Seite der Gleichungen können mit $(x_1, x_2)^T$ inhomogene Koordinaten direkt verwendet werden. Die Wirkung einer Homographie \mathbf{H} auf ein Bild kann durch die Neuordnung von Punkten beobachtet werden. Es handelt sich um ein bijektives Warping, wodurch Bilder von einem anderen Blickwinkel betrachtet werden können. Da üblicherweise endliche, diskrete Bilder wie in Abbildung 6.1 verwendet werden, geht die Bijektivität im endlichen Bildraster im Allgemeinen verloren. Die Aliasierung kann durch geeignete Verfahren behandelt werden. Neben der Nearest-Neighbor-Methode empfehlen sich u. a. bilineare und bikubische Interpolation [Gonzalez u. Woods, 2002].



ABBILDUNG 6.1: Rektifizierung. Korrektur der perspektivischen Verzerrung der Spielfeldebene durch eine Homographie, welche die Gerade (in Bild A), in welcher sich parallele Spielfeldlinien schneiden, auf die Gerade $l_\infty = (0,0,1)^T$, die Gerade im Unendlichen, abbildet (B).

6.3 Quasi-affine Homographien

Werden Punkte aus der konvexen Hülle eines Bildobjekts durch eine Homographie auf Punkte im Unendlichen abgebildet, tritt der Fall ein, dass die Punkte der konvexen Hülle nach der Transformation keine konvexe Menge bilden. Dieser Fall ist in Abbildung 6.2 dargestellt und in einem überwiegenden Teil der Anwendungen unerwünscht, da eine derartige Transformation aus keinem (Kamera-)Blickwinkel erreicht werden kann. Oftmals gehen derartige Homographien aus einer zufälligen Auswahl von Bezugspunkten zweier Bilder hervor und müssen ausgeschlossen werden. Zu diesem Zweck wird im Folgenden die Klasse der quasi-affinen Homographien betrachtet. Diese lassen sich zwischen affinen und projektiven Homographien hinsichtlich der mit ihnen verbundenen Freiheiten eingruppiert. Die Verhinderung von Fällen wie in Abbildung 6.2 (B) ist der Leitgedanke dieses Abschnitts.

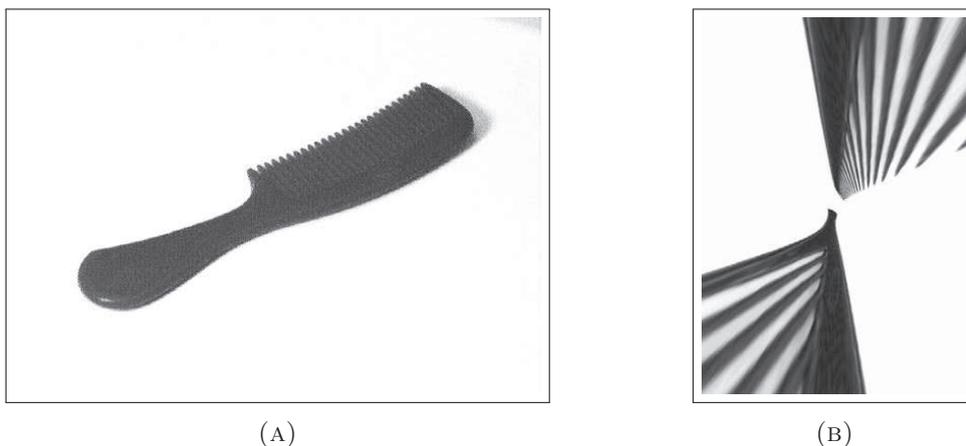


ABBILDUNG 6.2: Das Bild (A) eines Kamms und das Resultat (B) der Anwendung einer projektiven Transformation auf das Bild. Die projektive Transformation erhält die konvexe Hülle der Punkte des Kamms nicht. Quelle: [Hartley u. Zisserman, 2003]

Definition 6.3. Sei \mathbf{H} eine Homographie $\mathbb{P}^n \rightarrow \mathbb{P}^n$, \mathbf{E}_∞ die Hyperebene, welche die Punkte im Unendlichen beinhaltet (die letzte homogene Koordinate ist 0) und $\emptyset \neq \mathbf{B} \subset \mathbb{R}^n$, dann ist \mathbf{H} bezüglich \mathbf{B} quasi-affin, falls

$$\mathbf{H}(\text{hom} \circ \text{convexhull})(\mathbf{B}) \cap \mathbf{E}_\infty = \emptyset \quad (6.7)$$

Der Operator hom dient der Einbettung der (finiten) Punkte der Menge \mathbf{B} in \mathbb{P}^n . Entsprechend werden diese „homogenisiert“:

$$\mathbf{x} = (x_1, \dots, x_n)^T \mapsto (\alpha x_1, \dots, \alpha x_n, \alpha)^T, \alpha \in \mathbb{R} \setminus \{0\} \quad (6.8)$$

Anmerkung 4. Die Multiplikation zwischen Matrizen und Mengen erfolgt elementweise: $\mathbf{H}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} = \{\mathbf{H}\mathbf{x}_1, \mathbf{H}\mathbf{x}_2, \dots, \mathbf{H}\mathbf{x}_n\}$

Die konvexe Hülle wird in diesem Kapitel wie folgt definiert:

$$\text{convexhull}(\mathbf{B}) := \bigcap_{\substack{\mathbf{B} \subset \mathbf{X} \subset \mathbb{R}^n \\ \mathbf{X} \text{ konvex}}} \mathbf{X} \quad (6.9)$$

Auch wenn die Hintereinanderausführung von quasi-affinen Homographien bezüglich einer identischen Menge \mathbf{B} wieder quasi-affin ist, bilden quasi-affine Homographien keine Gruppe — ein entscheidender Unterschied sowohl zu affinen als auch projektiven Homographien. Die Anzahl der Freiheitsgrade entspricht dem projektiven Fall. Tatsächlich weicht die gegebene Definition von Hartleys Festlegung ab, ist jedoch äquivalent zu seiner Wahl [Hartley u. Zisserman, 2003, Definition 21.3], wie Satz 6.2 zeigt.

Satz 6.2. \mathbf{H} ist genau dann eine quasi-affine Homographie bezüglich einer Menge \mathbf{B} , wenn die konvexe Hülle von \mathbf{B} auf eine konvexe Menge abgebildet wird.

Beweis 6.2. Siehe Theorem 21.2 in [Hartley u. Zisserman, 2003].

Ferner ist $\mathbf{H}(\text{hom}(\mathbf{B}))$ konvex und \mathbf{H}^{-1} quasi-affin bzgl. $\mathbf{H}(\text{hom}(\mathbf{B}))$.

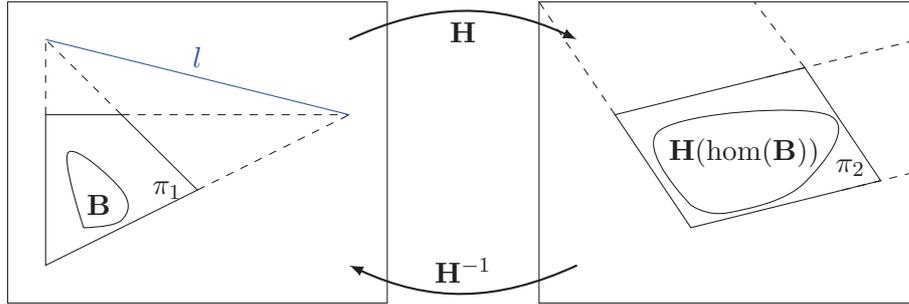


ABBILDUNG 6.3: Die Homographie \mathbf{H} bildet die finite Gerade l von der euklidischen Ebene π_1 auf l_∞ auf der Ebene π_2 ab. Ist \mathbf{H} quasi-affin bzgl. der konvexen Menge \mathbf{B} , so ist auch $\mathbf{H}(\text{hom}(\mathbf{B}))$ konvex.

Für den Fall $n = 2$ sei \mathbf{H} nun eine quasi-affine Homographie bezüglich der konvexen Menge $\mathbf{B} \in \mathbb{R}^2$, welche die Ebene π_1 auf π_2 abbildet. Dabei soll eine (finite) Gerade $l = (l_1, l_2, l_3)^T$ mit $l_3 \neq 0$, in welcher sich parallele Geraden in π_1 schneiden, gemäß Abbildung 6.3, auf die Gerade $l_\infty = (0,0,1)^T$ im Unendlichen transformiert werden. Die Homographie \mathbf{H} ist damit bis auf eine affine Homographie \mathbf{H}_A eindeutig bestimmt (vgl. Gleichung 2.19 in [Hartley u. Zisserman, 2003]).

$$\mathbf{H} = \mathbf{H}_A \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{pmatrix} \quad (6.10)$$

Für die Charakterisierung der Quasi-Affinität ist es ausreichend, die Schnittmenge der konvexen Menge \mathbf{B} mit der (finiten) Geraden l zu betrachten.

Satz 6.3. Eine Homographie $\mathbf{H}: \mathbb{P}^2 \rightarrow \mathbb{P}^2$ ist genau dann quasi-affin bezüglich einer Menge $\mathbf{B} \subset \mathbb{R}^2$, wenn:

$$(\text{hom} \circ \text{convexhull})(\mathbf{B}) \cap \{\mathbf{x} \mid \mathbf{x}^T \mathbf{H}^T l_\infty = 0\} = \emptyset \quad (6.11)$$

Beweis 6.3. Die Richtigkeit dieser Aussage kann durch elementweise Transformation mit \mathbf{H} vor der Schnittbildung gezeigt werden:

$$\begin{aligned} \emptyset &= (\text{hom} \circ \text{convexhull})(\mathbf{B}) \cap \{\mathbf{x} \mid \mathbf{x}^T \mathbf{H}^T l_\infty = 0\} = \\ &= \mathbf{H}(\text{hom} \circ \text{convexhull})(\mathbf{B}) \cap \mathbf{H}\{\mathbf{x} \mid \mathbf{x}^T \mathbf{H}^T l_\infty = 0\} = \\ &= \mathbf{H}(\text{hom} \circ \text{convexhull})(\mathbf{B}) \cap \{\mathbf{H}\mathbf{x} \mid \mathbf{x}^T \mathbf{H}^T l_\infty = 0\} = \\ &= \mathbf{H}(\text{hom} \circ \text{convexhull})(\mathbf{B}) \cap \{\mathbf{x} \mid \mathbf{x}^T l_\infty = 0\} = \\ &= \mathbf{H}(\text{hom} \circ \text{convexhull})(\mathbf{B}) \cap \mathbf{E}_\infty \end{aligned} \quad (6.12)$$

Die Charakterisierung der Quasi-Affinität kann gemäß Satz 6.3, durch Betrachtung der Geraden $\mathbf{H}^T \mathbf{l}_\infty$ vorgenommen werden. Entsprechend Abbildung 6.3 ist dies die (finite) Gerade

$$\mathbf{H}^T \mathbf{l}_\infty = \left(\begin{pmatrix} 1 & 0 & l_1 \\ 0 & 1 & l_2 \\ 0 & 0 & l_3 \end{pmatrix} \mathbf{H}_A^T \right) \mathbf{l}_\infty \propto \begin{pmatrix} 1 & 0 & l_1 \\ 0 & 1 & l_2 \\ 0 & 0 & l_3 \end{pmatrix} \begin{pmatrix} h_{11}^A & h_{21}^A & 0 \\ h_{12}^A & h_{22}^A & 0 \\ h_{13}^A & h_{23}^A & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \mathbf{l} \quad (6.13)$$

Statt die Erhaltung der Konvexität einer Menge zu testen, um die Eigenschaft der Quasi-Affinität für eine Homographie zu überprüfen, genügt es die Lage der Punkte besagter Menge relativ zur Geraden \mathbf{l} zu überprüfen. Schneidet die Gerade \mathbf{l} die konvexe Hülle der Menge \mathbf{B} nicht, ist \mathbf{H} quasi-affin bzgl. \mathbf{B} . Da eine affine Homographie finite Punkte nicht auf Punkte im Unendlichen abbildet, ist diese immer quasi-affin.

Anmerkung 5. In [Hartley, 1999, Proof Theorem 7] wird bereits von der Möglichkeit Gebrauch gemacht, alle Punkte einer Menge auf einer Seite der Geraden $\mathbf{l} = \mathbf{H}^T \mathbf{l}_\infty$ zu verorten, ohne dies für die Charakterisierung der Quasi-Affinität zu nutzen (vgl. [Hartley, 1999, Theorem 8]).

Im Kontext Fußball bedeutet dies, dass jede Homographie, deren Horizontgerade \mathbf{l} innerhalb der konvexen Spielfeldhülle (Kapitel 3.3) liegt, als eine nicht-quasi-affine Lösung ausgeschlossen werden muss. Damit gestaltet sich die Charakterisierung deutlich einfacher als durch andere Ansätze (siehe Abschnitt 6.9).

6.4 Das Spielfeld-Modell

Die Vermessung von Spielfeldobjekten wird innerhalb der Spielfeldebene durchgeführt. Abstände und Winkel sind den FIFA-Regularien [FIFA, 2014] entnommen. Als Ursprung wurde der Spielfeldmittelpunkt bestimmt. Obwohl für das Spielfeld keine feste Breite und Länge vorgeschrieben sind, wird der überwiegende Teil der höherklassigen Begegnungen auf den Standardmaßen eines 105 m langen und 68 m breiten Felds ausgetragen. Diese Maße werden auch im Modell verwendet, sind jedoch für Begegnungen auf anderen Spielfeldern anpassbar.

Anmerkung 6. Wie dem vorigen Absatz zu entnehmen ist, wird das Spielfeld als planar betrachtet. Die Annahme eines planaren Spielfelds im Bild führt zwangsweise zu Fehlern in der Zuordnung des Spielfeldmodells, welche sich aufgrund unbekannter Umgebungsparameter nur bedingt beheben lassen. Dies begründet sich zum einen in der

Verzeichnung optischer Systeme, zum anderen in der Tatsache, dass Spielfelder infolge von Bewässerungsproblematiken (Drainagesysteme, Rasenheizungen) oft nicht eben gebaut werden.

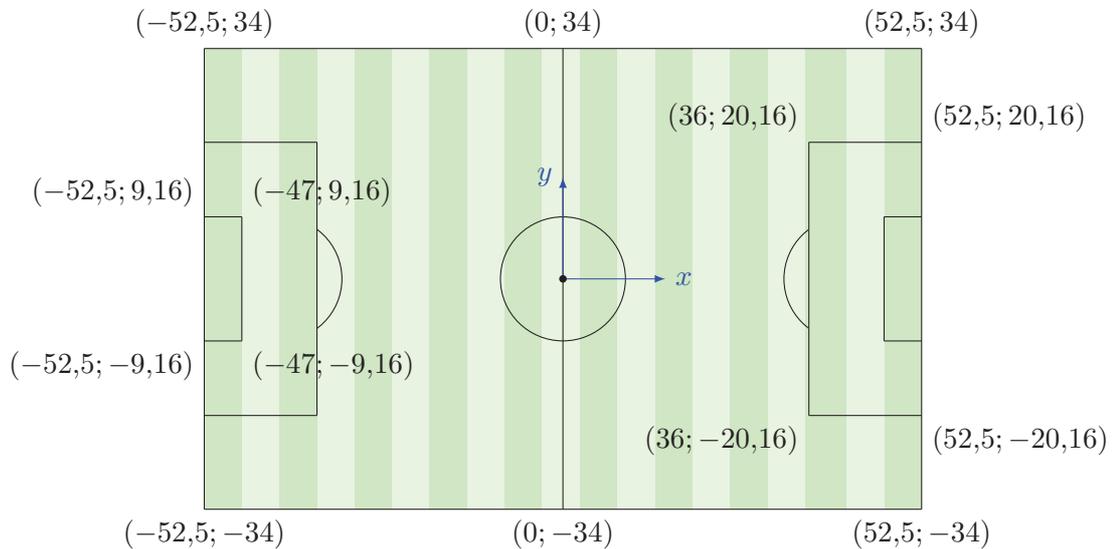


ABBILDUNG 6.4: Koordinaten der Eckpunkte im Spielfeldmodell. Auf die Angabe homogener Koordinaten wurde für eine bessere Darstellung verzichtet. Ebenso wurde jeweils nur ein Tor- und Strafraum eingetragen. Die weiteren Punkte ergeben sich aus der Symmetrie des Spielfelds.

Aus den in Abbildung 6.4 dargestellten Punkten können die Gleichungen der Geraden im Spielfeld bestimmt werden. In homogenen Koordinaten dargestellt lauten diese:

Horizontale Spielfeldgeraden:

$$(0; 1; 34)^T$$

$$(0; 1; 20,16)^T$$

$$(0; 1; 9,16)^T$$

$$(0; 1; -9,16)^T$$

$$(0; 1; -20,16)^T$$

$$(0; 1; -34)^T$$

Vertikale Spielfeldgeraden:

$$(1; 0; -52,5)^T$$

$$(1; 0; -47)^T$$

$$(1; 0; -36)^T$$

$$(1; 0; 0)^T$$

$$(1; 0; 36)^T$$

$$(1; 0; 47)^T$$

$$(1; 0; 52,5)^T$$

Um die Geradensegmente des Spielfelds zu bestimmen, werden für die Geraden Nebenbedingungen anhand der gegebenen Start- und Endpunkte formuliert. Für Punkte $\boldsymbol{x} \in \mathbb{P}^2$ auf der Geraden $\boldsymbol{l} = (1; 0; -52,5)^T$ (d. h. $\boldsymbol{x}^T \boldsymbol{l} = 0$), der äußeren Spielfeldbegrenzung, lautet diese etwa $-34x_3 \leq x_2 \leq 34x_3 \wedge x_3 \neq 0$.

Der Mittelkreis und die Kreissegmente am Torraum werden gesondert betrachtet. Ihre Mittelpunkte sind:

$$\begin{aligned} \mathbf{c}_0 &= (0; 0) \\ \mathbf{c}_l &= (-41,5; 0) \\ \mathbf{c}_r &= (41,5; 0) \end{aligned}$$

Der Radius r beträgt jeweils 9,15 m . Die (inhomogene) Kreisgleichung

$$(x - c_1)^2 + (y - c_2)^2 = r^2 \quad (6.14)$$

kann durch die Ersetzungen $x \mapsto \frac{x_1}{x_3}$ und $y \mapsto \frac{x_2}{x_3}$ in die homogene Form überführt werden.

$$x_1^2 + x_2^2 - 2c_1x_1x_3 - 2c_2x_2x_3 + (c_1^2 + c_2^2 - r^2)x_3^2 = 0 \quad (6.15)$$

Dies ist äquivalent mit der Matrixnotation $\mathbf{x}^T \mathbf{C} \mathbf{x} = 0$ für die symmetrische Matrix

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & -c_1 \\ 0 & 1 & -c_2 \\ -c_1 & -c_2 & c_1^2 + c_2^2 - r^2 \end{pmatrix} \quad (6.16)$$

Für die Kreisbögen sind folgende Nebenbedingungen relevant:

linker Kreisbogen:

$$x_1 \geq -36x_3 \wedge x_3 \neq 0$$

rechter Kreisbogen:

$$x_1 \leq 36x_3 \wedge x_3 \neq 0$$

Werden die Kreise (Kegelschnitte) und Geraden unter Beachtung der Nebenbedingungen zusammengetragen, entsteht das vollständige Modell des Spielfelds analog zu Abbildung 6.4. Im Modell sind sowohl Längenmaße als auch Winkel mit den FIFA-Regularien vereinbar. In diesem Zusammenhang wird der Begriff der Weltkoordinaten verwendet.

Das Spielfeld ist rechteckig und wird mit Linien gekennzeichnet. Die Linien gehören zu den Räumen, die sie begrenzen. [...] Alle Linien sind gleich breit. Ihre Breite beträgt höchstens 12 cm.

— DFB [2014a], FIFA [2014]

Die Modellpunkte werden im Weiteren verwendet, um das Spielfeld im Bild zu erkennen. Zwar ist mit einer Ungenauigkeit zu rechnen, da im Modell für kein Geradensegment

eine Fläche berücksichtigt wurde und die Geradensegmente auch nicht mittig auf einer im Spielfeld markierten Feldlinie (unbekannter Breite) ausgerichtet wurden, jedoch unterscheidet sich die Linienbreite bereits auf einzelnen Spielfeldern teils so deutlich, dass eine Berücksichtigung hier nicht maßgeblich erscheint.

6.5 Automatische Zuordnung zwischen Bild und Spielfeldmodell

Nachdem in Kapitel 5 ein Verfahren zur Erkennung von Geradensegmenten in einem Bild vorgestellt wurde und in Abschnitt 6.4 das (in erster Instanz aus Geradensegmenten bestehende) Spielfeldmodell eingeführt wurde, werden beide Aspekte nun in der Kamera-Kalibrierung zusammengeführt. Das Ziel ist es, die im Bild erkannten Strukturen durch eine quasi-affine Homographie möglichst passgenau auf das Modell des Spielfelds zu transformieren. Sowohl aus dem Bild als auch aus dem Modell bestimmte Landmarken werden hierfür betrachtet. Das Verfahren gliedert sich in die in Abbildung 6.5 dargestellten Schritte. Im letzten Schritt, dem modifizierten *Iterative Closest Point-* (ICP) Algorithmus, wird die Homographie bereitgestellt, welche die Transformation des Modells auf Ankerpunkte im Bild herstellt.

Der ICP-Algorithmus [Besl u. McKay, 1992] zur Registrierung zweier Punktmengen (Punktwolken) besteht essenziell aus folgenden Schritten:

1. Bestimmung des Punkts mit kürzester Distanz aus der Zielpunktmenge für jeden Punkt der Eingangspunktmenge. Für jeden Punkt der Eingangspunktmenge wird dadurch eine Referenz auf einen Punkt der Zielpunktmenge bestimmt.
2. Ausführung der Rotation und Translation auf der Eingangspunktmenge, welche die Eingangspunkte bestmöglich an ihre Referenzen (aus dem letzten Schritt) anpassen.
3. Wiederholung, beginnend bei Schritt 1, bis die Summe der Distanzen zwischen den Punkten der Eingangspunktmenge und ihren Referenzen hinreichend klein ist.

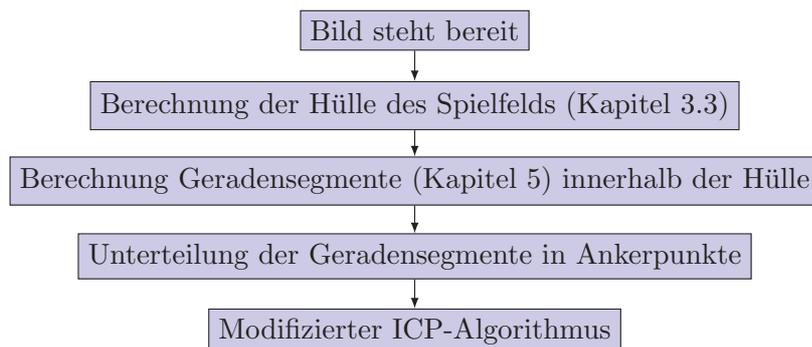


ABBILDUNG 6.5: Verarbeitungsschritte hin zur Kamera-Kalibrierung durch einen modifizierten ICP-Algorithmus

Ein Beispiel für die Ausführung des ICP-Algorithmus auf zwei Punktmengen kann in Abbildung 6.6 betrachtet werden. Dieses Beispiel zeigt jedoch einen Sonderfall in welchem Eingangs- und Zielpunktmenge deckungsgleich abgebildet werden können. Da im Falle von Fußballübertragungen auch mit Änderungen in der Skalierung der Objekte (Änderungen der Brennweite) zu rechnen ist und das Spielfeld nur sehr selten komplett im Bild ist, wird das ICP-Verfahren modifiziert.

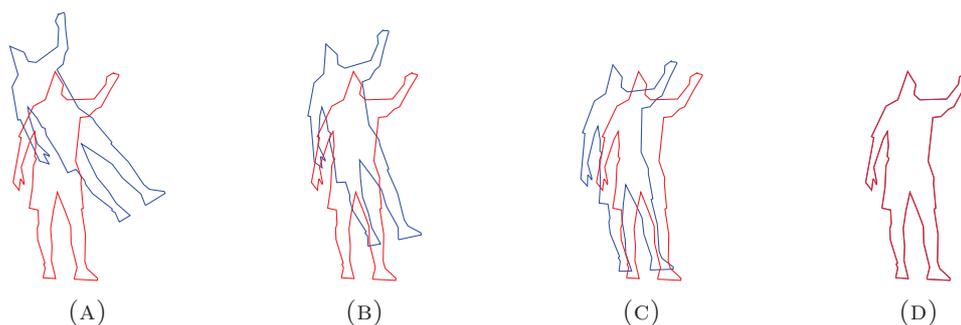


ABBILDUNG 6.6: Iterative Registrierung zweier Punktmengen (rot, blau) durch Translation und Rotation der blauen Punktmenge.

Für den Start des Verfahrens wird von einer Homographie \mathbf{H}_0 ausgegangen, welche die Bildpunkte mit einem Fehler auf das Modell überträgt (Näherungslösung). Den Fehler dieser Näherung gilt es in Schritten durch die Homographien $\mathbf{H}_1, \mathbf{H}_2, \dots$ zu minimieren. Die Zielmenge geht aus den gefundenen Geradensegmenten hervor. Diese werden jedoch nicht direkt verwendet, sondern in kurze Abschnitte zerlegt, deren Mittelpunkte, bezeichnet als Ankerpunkte, die Zielmenge bilden. Auf diese Weise lässt sich sowohl der Aufwand als auch die Genauigkeit über die Länge der Abschnitte steuern. Die Eingangsmenge bildet ein durch \mathbf{H}_{i-1}^{-1} transformiertes Spielfeldmodell (Abschnitt 6.4). Es werden in jedem Schritt des ICP-Algorithmus die kürzesten Distanzen im Bildraum zwischen

Ankerpunkten und Transformationen des Modells, einschließlich der Kegelschnitte, gesucht. Die dabei gefundenen Zuordnungen von Punktpaaren bilden die erste Liste von Referenzen.

Dieser Aufbau optimiert jedoch nur gemäß der Evidenz. Es wird somit nicht bestraft, falls Linien des Spielfeldmodells an Positionen im Bild auftreten, für welche keine Geradensegmente gefunden wurden. Um auch diesen Fall zu behandeln, wird eine zweite Liste von Referenzen zwischen der Eingangsmenge (dem Spielfeldmodell) und der Zielmenge (den Geradensegmenten) erzeugt. Alle nach der Transformation \mathbf{H}_{i-1}^{-1} innerhalb der Spielfeldhülle sichtbaren Modellfeldlinien werden im Bildraum gerastert und ebenso wie die Geradensegmente in Abschnitte zerlegt. Daraus werden wiederum Ankerpunkte bestimmt. In diesem Fall werden Referenzen durch eine Suche nach Punkten mit jeweils kürzester Distanz von Ankerpunkten zu Punkten in der Menge der gefundenen Geradensegmente hergestellt. Eine Übersicht der sich daraus ergebenden Abhängigkeiten ist in Abbildung 6.7 dargestellt. Da der Suchraum im Fall der Suche nach Geradensegmenten unabhängig vom Iterationsschritt $\#i$ ist, erscheint eine Vorverarbeitung der Daten in vielen Fällen sinnvoll. Unter Verwendung von *KD-Trees* [Friedman et al., 1977] wird der durchschnittliche Aufwand für eine Punktsuche auf $\mathcal{O}(\log n)$ reduziert. Da nicht davon ausgegangen werden kann, dass die Suche nach Geradensegmenten auch die Kegelschnitte erkennt, werden diese bei der Transformation des Spielfeldmodells nicht berücksichtigt, deren Nicht-Auffinden im Bild wird also nicht bestraft.

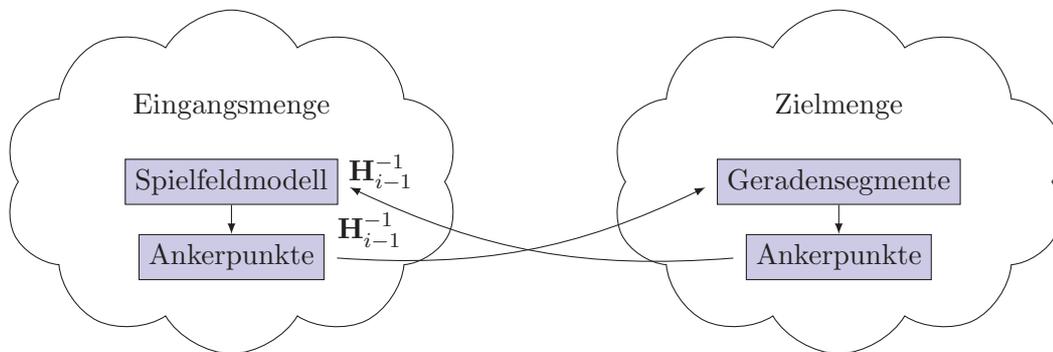


ABBILDUNG 6.7: Referenzen werden jeweils ausgehend von Ankerpunkten, entweder des Spielfeldmodells (Eingangsmenge) oder der Geradensegmente (Zielmenge), gebildet. Da Referenzen zwischen gefundenen Geradensegmenten und dem Spielfeldmodell ausschließlich im Bildraum bestimmt werden, sind alle Koordinaten des Spielfeldmodells in den Bildraum durch die Homographie \mathbf{H}_{i-1}^{-1} zu transformieren.

In jedem Schritt des Verfahrens entstehen zwei Listen von Referenzen im Bildraum. Diese lassen sich derart verknüpfen, dass zwei Folgen, $\mathbf{x}_{1..m}$ im Bildraum und $\mathbf{x}'_{1..m}$ im Spielfeldmodell, notiert werden können, wobei \mathbf{x}_j mit \mathbf{x}'_j korrespondiert für $j \in \{1, \dots, m\}$ mit $m \in \mathbb{N} \setminus \{0\}$. Dies wird dadurch erreicht, dass die Punkte der Eingangsmenge durch \mathbf{H}_{i-1} zurück in den Modellraum übertragen werden. Ob eine Referenz ausgehend von

einem Ankerpunkt des Spielfeldmodells oder der Geradensegmente bestimmt wurde, spielt an dieser Stelle keine Rolle mehr. Der Gesamtfehler der Homographie \mathbf{H}_i wird symmetrisch im Bild- und Modellraum durch die Summe der euklidischen Distanzen aller Korrespondenzen bestimmt:

$$\sum_{j=1}^m (d(\mathbf{x}_j, \mathbf{H}_i^{-1} \mathbf{x}'_j) + d(\mathbf{x}'_j, \mathbf{H}_i \mathbf{x}_j)) \quad (6.17)$$

Hierbei wird die euklidische Distanz zweier homogener Punkte durch $d(\mathbf{a}, \mathbf{b})$ berechnet. Der Fall, dass diese nicht berechenbar ist, wird im Anschluss behandelt. Im allgemeinen Fall einer nicht-affinen Homographie \mathbf{H}_i ist die Minimierung nach dem Fehlerterm aus Gleichung 6.17 nicht-linear auszuführen. Hierfür wurde der Levenberg-Marquardt-Algorithmus [Lourakis, 2005] verwendet. Unter dem Aspekt, Ausreißer in der Optimierung nicht zu berücksichtigen, wird zuvor ein *Least-Median-Of-Squares*- (*LMedS*) Filter [Rousseeuw, 1984] auf die Daten angewandt. Hierfür wird das einfachere Optimierungsproblem

$$\text{median}_j d^2(\mathbf{x}'_j, \mathbf{H}_i \mathbf{x}_j) \longrightarrow \min! \quad (6.18)$$

gelöst, wodurch der kleinste Medianwert des Residuums für den gesamten Datensatz (linear) bestimmt wird. Auf Grundlage des Medians als robusten Schätzer können nahezu 50 % Ausreißer toleriert werden. Erkannte Ausreißer werden im weiteren Verlauf des Schritts $\#i$ nicht beachtet. Weiterhin werden Buckets, wie durch [Zhang et al., 1995] vorgeschlagen, eingesetzt. Dabei handelt es sich um eine Monte-Carlo-Technik, durch welche die Wahrscheinlichkeit auf eine adäquate räumliche Verteilung der Korrespondenzen erhöht wird. Dafür werden Testpunkte für die Bestimmung einer Homographie bevorzugt so ausgewählt, dass sie jeweils aus verschiedenen Bildregionen (Buckets) stammen. Die Implementation der nicht-linearen Optimierung und des LMedS-Filters wurden durch eine angepasste Version der Bibliothek *homest* [Lourakis, 2006] realisiert.

Anmerkung 7. Die Berechnung von Homographien durch Punktkorrespondenzen wird durch den *Direct-Linear-Transformation*- (*DLT*) Algorithmus [Hartley u. Zisserman, 2003] realisiert. Punktfolgen sind dabei gemäß Hartleys Vorgaben [Hartley, 1997a] zu normalisieren, um einen algebraischen Fehler zu vermeiden, welcher aus einer unausgewogenen Skalierung von Komponenten der Vektoren der Folgen $\mathbf{x}_{1:m}$ und $\mathbf{x}'_{1:m}$ resultiert.

Das an dieser Stelle beschriebene Verfahren optimiert eine Näherung zwischen einer diskreten Punktwolke und einem kontinuierlichen Modell. Da Ausreißer in jedem Schritt neu zugewiesen werden, ist keine Konvergenz gesichert. Es ist möglich, dass das Verfahren zwischen zwei Lösungen pendelt. Neben einem (unteren) Schwellwert für die Differenz des Gesamtfehlers zweier aufeinanderfolgender Schritte muss das Verfahren damit auch abbrechen, falls die Gesamtzahl der Iterationen eine Obergrenze erreicht. Ein weiteres

Kriterium für den Abbruch ist die Quasi-Affinität der Homographie \mathbf{H}_i (und daraus folgend die Berechenbarkeit von $d(a, b)$). Ist diese nicht mehr gegeben, bricht das Verfahren mit der Rückgabe einer entsprechenden Meldung ab. Wird das Verfahren auf eine andere Art beendet, wird die zuletzt berechnete Matrix \mathbf{H}_i zurückgegeben. Diese ist sowohl im Sinn der gebildeten Referenzen, vom Modell zum Bild sowie vom Bild zum Modell, als auch im Sinn der Fehlerfunktion, welche im Modell und im Bild betrachtet wird, symmetrisch. Da zuerst genannte Referenzen genauso wie die Ausreißersuche dabei nur in jeweils einem Raum Geltung finden, besteht keine Gefahr, dass verschiedene Distanzmaße in den unterschiedlichen Koordinatensystemen die Ergebnisse verfälschen bzw. der Fehler in einem Raum den Fehler in einem anderen Raum skalierungsbedingt überwiegt. Ähnliches gilt für die Berechnung der Ankerpunkte, welche aus dem Spielfeldmodell und den Geradensegmenten auf dieselbe Art gebildet werden.

6.6 Quasi-affine Kameraparameter-Nachführung

Nicht für jedes Bild besteht die Möglichkeit, eine Kamera-Kalibrierung durch den ICP-Algorithmus sinnvoll durchführen zu können, da hierfür eine ausreichend große Anzahl von Feldlinien (Landmarken) im Bild sichtbar sein muss. Welcher Anteil im Einzelfall notwendig ist bzw. welchen Fehler eine Homographie-Berechnung (bzgl. Ground-Truth-Daten) aufweist, hängt neben der Ebenheit des Spielfelds auch von der dreidimensionalen Anordnung der im Bild sichtbaren Spielfeldlinien ab. Bildern, in welchen die sichtbaren Feldlinien für das ICP-Verfahren nicht ausreichend sind (Details unten), müssen auf eine andere Weise Homographien zugewiesen werden. Ausgehend von einer Homographie \mathbf{H}_0 , welche die Transformation des Spielfelds von Bild- in Modellkoordinaten für ein erstes Bild beschreibt, werden weitere Homographien ${}^T\mathbf{H}_1, {}^T\mathbf{H}_2, \dots$ gesucht, welche ebendiese Transformationen in den Folgebildern darstellen. Der Index beschreibt in diesem Fall die Bildnummer.

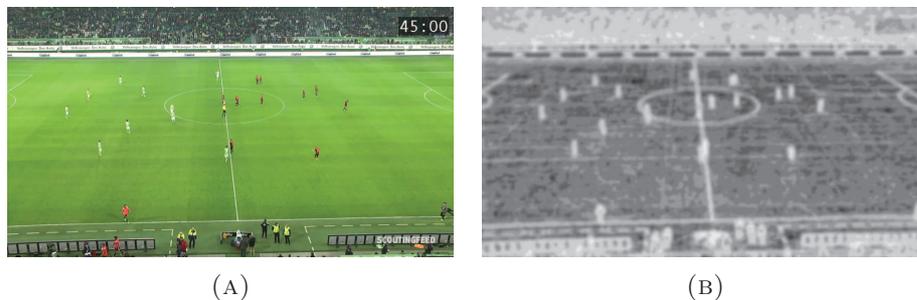


ABBILDUNG 6.8: (A) Bild eines Spielfelds und (B) das Resultat der Berechnung der Entropie in Rechteckfenstern zu jedem Bildpunkt. Zunehmende Helligkeit ist mit einem ansteigendem Wert der Entropie zu interpretieren.

Für die Berechnung der Homographien wird das Verfahren von [Prince et al., 2002] angewandt. Gegeben seien eine Menge von Merkmalspunkten (Modelllandmarken) zu einem Bild $\#i$ und eine Menge von Merkmalspunkten zu dem nachfolgenden Bild $\#(i+1)$. Wiederholt werden zufällig jeweils vier Korrespondenzen von der ersten zur zweiten Menge gebildet. Aus den daraus berechenbaren Test-Homographien (DLT-Algorithmus, [Hartley u. Zisserman, 2003]) kann die Anzahl der Korrespondenzen beider Mengen ermittelt werden, welche im Rahmen dieser Transformation als Inlier angesehen werden. Eine der Test-Homographien mit der größten Anzahl an Inliern wird ausgewählt und alle Inlier werden verwendet, um die endgültige Homographie zwischen beiden korrespondierenden Bildern zu bestimmen. Zusätzlich zum Ansatz nach [Prince et al., 2002] werden nicht-quasi-affine Homographien aus der weiteren Betrachtung ausgeschlossen. Das Verfahren gleicht weitestgehend RANSAC und ist robust gegenüber Ausreißern.

Als Merkmals- (Saat-)punkte des Verfahrens werden lokale Maxima aus dem zugehörigen Entropiebild (Abbildung 6.8) innerhalb der Grasmasken (Kapitel 3.4) bestimmt. Diese können somit als Landmarken des Grasmodells interpretiert werden. Zu jeder Umgebung eines Saatpunkts wird ein Template erstellt. Ähnlich dem Bucket-Verfahren [Zhang et al., 1995] wird die Grasmasken des Spielfelds in ein Rechteck-Gitter eingeteilt. In jeder Zelle wird unabhängig eine vordefinierte Anzahl der größten lokalen Maxima ausgewählt, um eine räumliche Verteilung der Punkte in der Grasmasken zu gewährleisten und die Relevanz der Homographien damit auf den größtmöglichen planaren Bildanteil auszudehnen.

Die Korrespondenzen werden durch ein Template Matching (Summe der quadrierten Differenzen, normiert) unter Berücksichtigung aller Farbkanäle gebildet. Ausgehend von einem Template um den Punkt \mathbf{x} in Bild $\#i$ wird der Inhalt des Templates in Bild $\#(i+1)$ gesucht. Das Argument \mathbf{x}' eines globalen Minimums der Matching-Kostenfunktion wird ausgewählt, zu welchem wiederum ein Template erzeugt und mit gleichem Verfahren in Bild $\#i$ gesucht wird. Dieses Vorgehen wird von [Hartley, 1997b] vorgeschlagen. Nur wenn ein globales Minimum in unmittelbarer Nähe zum Punkt \mathbf{x} eintritt, wird eine Korrespondenz zwischen den Punkten \mathbf{x} und \mathbf{x}' angenommen. Diese Ausreißersuche hat eine geringer werdende Anzahl der nachgeführten Templates zur Folge, da Templates, für welche keine Korrespondenz gefunden werden konnte, nicht weiter betrachtet werden. Das aus Bild $\#(i+1)$ erstellte Template wird verworfen. Sinkt die Anzahl von verbleibenden Tracks — z. B. durch einen Kameraschwenk — unter einen zuvor festgelegten Schwellwert, werden neue Saatpunkte in einem Entropiebild gesucht und anschließend neue Templates auf die zuvor beschriebene Art und Weise erzeugt. Bereits vorhandene Templates werden weiterhin nachgeführt, da aufgrund bereits erfolgter Nachführungen eine erhöhte Wahrscheinlichkeit für den Erfolg weiterer vermutet wird.

Sind die Homographien zwischen den Bildern $\mathbf{H}_{0 \rightarrow 1}, \mathbf{H}_{1 \rightarrow 2}, \dots$ berechnet, können auch die gesuchten Bild-Modell-Homographien ${}^T\mathbf{H}_1, {}^T\mathbf{H}_2, \dots$ des Template Trackers bestimmt werden.

$$\begin{aligned} {}^T\mathbf{H}_1 &= \mathbf{H}_0\mathbf{H}_{0 \rightarrow 1}^{-1} \\ {}^T\mathbf{H}_2 &= \mathbf{H}_0\mathbf{H}_{0 \rightarrow 1}^{-1}\mathbf{H}_{1 \rightarrow 2}^{-1} \\ &\dots \end{aligned} \quad (6.19)$$

Ist eine Homographie für ein erstes Bild gegeben, werden der soeben beschriebene Template Tracker und das ICP-Verfahren (Abschnitt 6.5) entsprechend Abbildung 6.9 zusammengeführt.

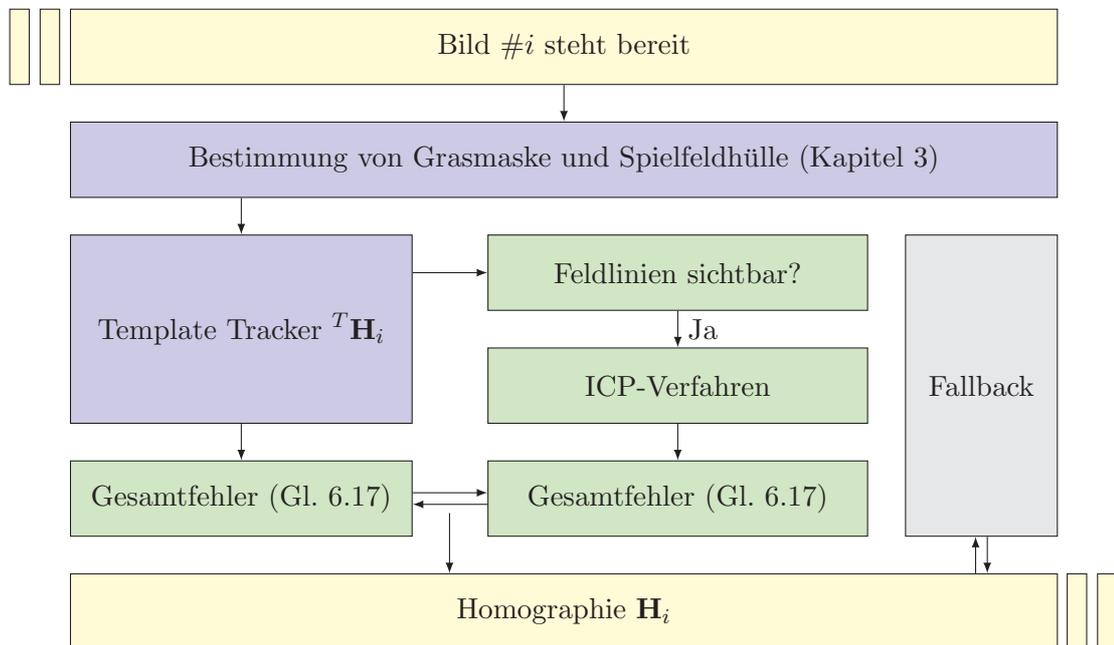


ABBILDUNG 6.9: Aufbau der quasi-affinen Kameraparameter-Nachführung. Die Komponenten des ICP-Subsystems sind grün hinterlegt.

Für jedes Bild werden die Templates in der Rasenfläche verfolgt. Sind ausreichend Spielfeldlinien sichtbar, wird zusätzlich das ICP-Verfahren ausgeführt. Für diese Bilder werden letztlich zwei Homographien berechnet. Beide Verfahren stehen dann im Wettbewerb zueinander. Durch die Ausführung des ICP-Algorithmus, insbesondere durch die Erkennung von Geradensegmenten innerhalb der Spielfeldhülle, ist es möglich, den Gesamtfehler (Gleichung 6.17) bezüglich der Spielfeldlinien im Bild zu berechnen. Dieser Fehler kann ebenso für die Homographie des Template Trackers berechnet werden, womit die Auswahl des geeignetsten Ergebnisses über einen Vergleich erfolgt. Als Näherungslösung (Argument) für den ersten Schritt des ICP-Verfahrens wird die Lösung ${}^T\mathbf{H}_i$ des Template Trackers verwendet. Um Rechenaufwand zu reduzieren, wurde eine zusätzliche

Beschränkung eingeführt, um das ICP-Verfahren nicht häufiger als fünfmal je Videosekunde auszuführen. Die Gesamtfehler-Bewertung entfällt, falls keine Feldlinien sichtbar sind und der Template Tracker die Homographie allein festlegt.

Im Fall des ICP-Verfahrens kann durch Rückgabe einer Meldung eine nicht-quasi-affine Homographie erkannt werden. Für den Template Tracker besteht die Möglichkeit, dass nicht genügend Saatpunkte oder Korrespondenzen gefunden werden. Für den Fall, dass beide Verfahren versagen, ist ein Fallback integriert. Dann wird, für einen begrenzten Zeitrahmen, die zuletzt ermittelte Homographie auch für die nächsten Bilder weiterverwendet.

Die Sichtbarkeit der Spielfeldlinien wird regelbasiert geprüft. Es ist leicht einzusehen, dass aus gewissen Bildausschnitten eine Einschränkung auf endlich viele Bild-Modell-Homographien selbst unter dem Aspekt einer utopischen Fehlerminimierung nicht möglich ist. Diese Fälle und solche, in welchen die Spielfeldlinien im Bild ungünstig verteilt sind, werden mit den nachfolgenden Regeln geprüft. Die ausschlaggebende Länge ist 16,5 m, die kürzeste Linienlänge im Spielfeld unter Vernachlässigung der 5,5 m langen Begrenzungslinien des Torraums. Ein solches Geradensegment wird als (theoretisch) sichtbar betrachtet, wenn sich mindestens 9 m einer Feldlinie im Bild befinden.

Ausreichend Feldlinien sind im Bild sichtbar \Leftrightarrow

Mindestens drei vertikale Feldlinien sind im Bild sichtbar \wedge

Mindestens eine der Kombinationen horizontaler Feldlinien ist im Bild sichtbar:

$$(ot \wedge lt \wedge lb) \vee (ot \wedge lt \wedge ob) \vee (ot \wedge lb \wedge ob) \vee (lt \wedge lb \wedge ob) \vee \\ (ot \wedge rt \wedge rb) \vee (ot \wedge rt \wedge ob) \vee (ot \wedge rb \wedge ob) \vee (rt \wedge rb \wedge ob)$$

wobei die Abkürzungen im Einzelnen die nachfolgenden Feldlinien beschreiben:

- ot ... obere Spielfeldbegrenzung
- ob ... untere Spielfeldbegrenzung
- rt ... obere, rechte Strafraumbegrenzung
- rb ... untere, rechte Strafraumbegrenzung
- lt ... obere, linke Strafraumbegrenzung
- lb ... untere, linke Strafraumbegrenzung

Da kein definitives Wissen um die tatsächlich im Bild sichtbaren Feldlinien vorliegt, wird die Sichtbarkeit aus der Homographie ${}^T\mathbf{H}_i$ des Template Trackers und den Abmessungen der Eingangsbilder geschätzt. Dazu werden die Modellfeldlinien über die Homographie ${}^T\mathbf{H}_i^{-1}$ in ein Testbild übertragen und dort vermessen. Sind genügend Modellfeldpunkte unter den angegebenen Bedingungen im Bild sichtbar, wird das Spielfeld als geeignet für die Ausführung des ICP-Algorithmus erachtet.

6.7 Kombination der Startposition

Die Suche nach einer Start-Homographie für die quasi-affine Kameraparameter-Nachführung wird am Beispiel eines Strafraumbilds (Abbildung 6.10, A) erklärt.

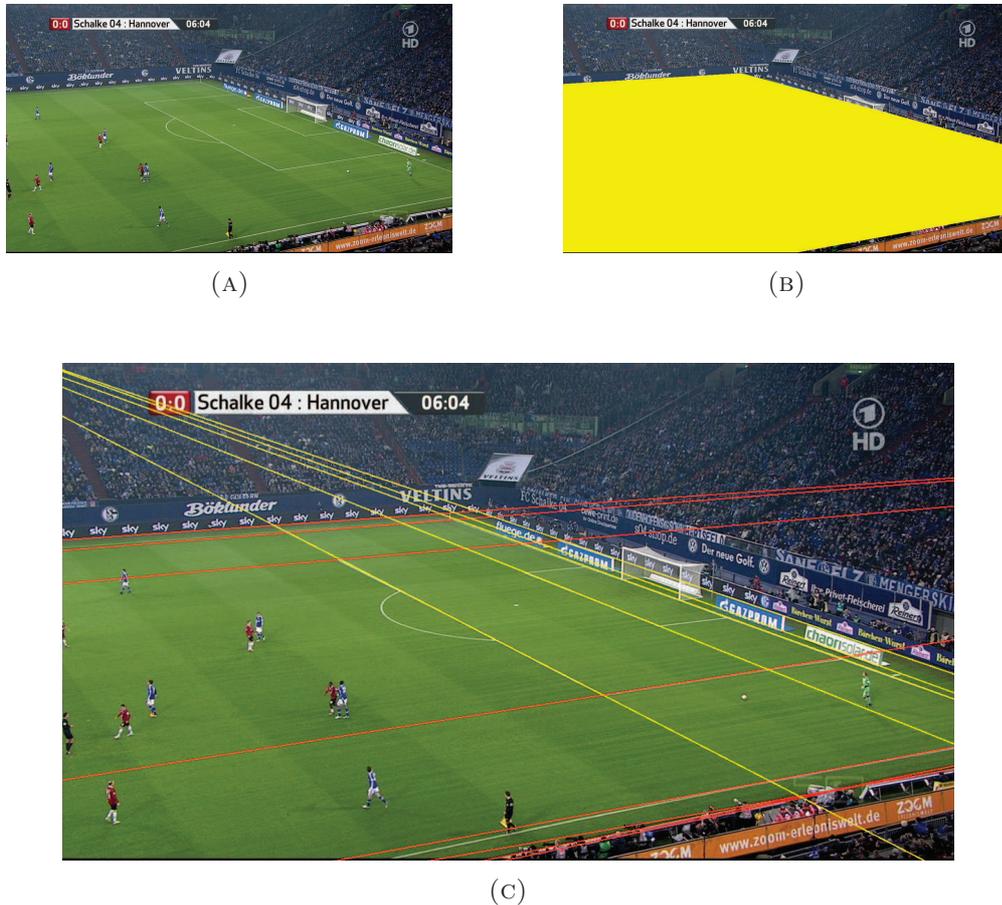


ABBILDUNG 6.10: Berechnung eines globalen Optimums in einer Strafraumsituation (A) unter Betrachtung des Hüllbereichs (B, gelb). Im Bild erkannte Geradensegmente wurden zu Geraden fusioniert und nach Winkel klassifiziert (C).
Bildquelle: [ARD/WDR, 2013a]

Geradensegmente (Kapitel 5) innerhalb der Spielfeldhülle (Kapitel 3.3, Abbildung 6.10, B) stellen, wie auch im ICP-Ansatz, die verfahrenstechnische Grundlage für die Bereitstellung der Startposition dar. Den im Bild erkannten geraden Punktfolgen werden durch lineare Regression Geraden zugeordnet, welche anschließend in einen Hough-Raum eingetragen werden. Jede Gerade wird entsprechend der Anzahl der Punkte innerhalb der Folge gewichtet. Nach einer anschließenden gaußschen Filterung des Hough-Raums werden die Geraden abhängig von ihren Winkeln klassifiziert (zwei Klassen). Dazu wird der Hough-Raum zeilenweise (für jeden Winkel) aufsummiert, um anschließend ein normiertes Histogramm der Winkel des Hough-Raums bestimmen zu können. Mittels eines *Expectation-Maximization-* (EM) Verfahrens [Russell u. Norvig, 1995] wird ein *Gaussian*

Mixture Model (GMM) [Reynolds, 2009] bestehend aus zwei Normalverteilungen auf diese Daten angepasst. Je nach Winkeldistanz zu den damit geschätzten Erwartungswerten werden Geraden aus dem Hough-Raum extrahiert und zu einer Menge (Klasse) hinzugefügt (Abbildung 6.10, C). Für anschließende Betrachtungen wird ausgehend von der Vermutung einer mittig an der Seitenauslinie positionierten Kamera eine Hypothese auf der Grundlage der geschätzten Erwartungswerte über die Position des Strafraums (rechts / links) überprüft. Im vorliegenden Beispiel kann die Existenz eines rechten Strafraums beobachtet und gemessen werden.

Aus beiden Klassen von Geraden (horizontal oder vertikal im Bild, vgl. Abbildung 6.10, C) werden alle Teilmengen mit einer zuvor festgelegten Kardinalität bestimmt. Allen Kombinationen aus jeweils einer Teilmenge horizontaler und vertikaler Geraden werden gleichmächtige Kombinationen aus Teilmengen horizontaler und vertikaler Geraden des Modells der rechten oder linken Spielfeldhälfte zugeordnet. Aus jeder der Zuordnungen wird durch einen geradenbasierten DLT-Algorithmus [Hartley u. Zisserman, 2003] eine Homographie erzeugt. Die Enumeration der Kombinationen muss sowohl im Bild als auch im Modell durchgeführt werden.

Durch Auswahl übereinstimmender Geraden-Reihenfolgen in Bild und Modell (mit jeder Kombination ohne Berücksichtigung der Reihenfolge der Auswahl) können alle anderen Reihenfolgen (Permutationen) der Geraden unberücksichtigt bleiben. Dazu werden sowohl die Positionen der Geraden im Bild als auch im Modell intelligent miteinander verbunden. Nach der Auswahl einer Kombination im Bild werden die Geraden mit der Reihenfolge ihres Auftretens relativ zur Spielfeldebene verknüpft. Dies basiert auf ihrer Lage relativ zu den Geraden der jeweils anderen Klasse. Ähnliches findet statt bzgl. der Kombinationen der Geraden des Spielfeldmodells: diese werden in Abhängigkeit von der Position des Strafraums in vorgegebener Weise sortiert. Eine zufällige Bestimmung der Reihenfolge innerhalb der Kombinationen ist damit überflüssig und eine große Zahl von Transformationen (welche zum Teil nicht quasi-affin wären) kann im Voraus ausgeschlossen werden.

Ist ein DLT-Ergebnis quasi-affin, wird die erzeugte Homographie hinsichtlich der Summe der Quadrate der euklidischen Abstände zwischen erkannten Geradensegmentenpunkten und Feldlinien (jeweils im Modell) bewertet. Das beste Ergebnis innerhalb dieser Bewertung, im Sinn der niedrigsten Summe, wird als Start-Homographie herangezogen. Beispiele hierzu sind Abbildung 6.11 zu entnehmen. Für die Kardinalitäten der Teilmengen wurden Elemente der Menge $\{2,3\}$ zugelassen, da so zum einen eine Zuordnung von vier Geraden die Bildung einer Homographie erlaubt und zum anderen die Rechenzeit vertretbar bleibt.

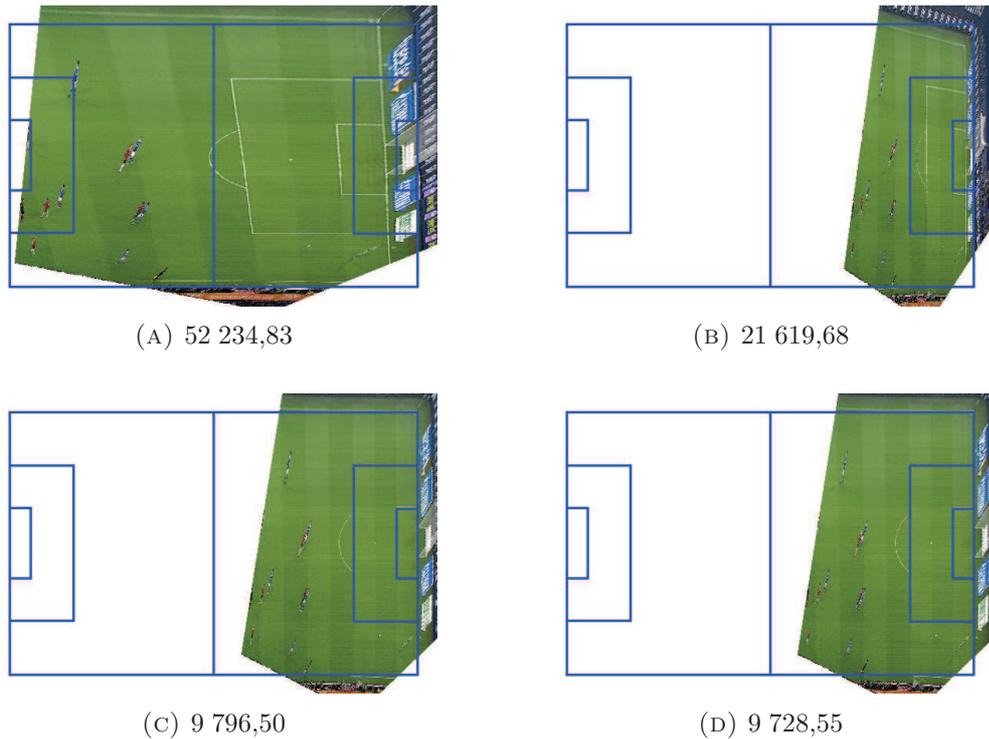


ABBILDUNG 6.11: Summe der quadrierten Distanzen (Scores) von erkannten Geradensegmenten zu Feldlinien des Modells bzgl. verschiedener Bild-Modell-Zuordnungen. Unter Verwendung einer Lookup-Tabelle für das Spielfeldmodell kann eine effiziente Berechnung dieses Fehlers erfolgen.

6.8 Evaluation

Die Evaluation erfolgt dreigeteilt. Zuerst wird die vorgestellte quasi-affine Kameraparameter-Nachführung auf eine simulierte Kamerabewegung angewandt. Hierbei wird eine virtuelle Kamera ein reales Fußballspiel aufzeichnen. Die Position der Kamera ist fest, ihre Ausrichtung und Brennweite werden in jedem Bild neu bestimmt. Aus den bekannten Kameraparametern kann für jeden Spielfeldpunkt der Fehler der Kameraparameter-Nachführung exakt bestimmt werden. Somit ist für diesen Fall eine Ground-Truth-Evaluation möglich.

Im zweiten Teil werden vollständige Halbzeiten von realen Kamerabewegungen mit anspruchsvollen Einstellungen nachverfolgt. Auch wenn für einzelne Bilder eine optische Einschätzung der Kameraparameter-Nachführung erfolgt, werden in diesem Abschnitt keine quantitativen Aussagen getroffen. Im Mittelpunkt steht die Robustheit des Verfahrens als entscheidendes Evaluationskriterium. Eine Möglichkeit, dennoch quantifizierbare Messwerte zu den berechneten Homographien zu erhalten, wird in Kapitel 7 besprochen.

Im letzten Teil wird die Kameraparameter-Nachführung in der Szenerie eines Fußgängerüberwegs untersucht. Das Spielfeldmodell wird an dieser Stelle von einem Modell

eines Fußgängerüberwegs ersetzt. Neben einer Auswertung des vorgestellten Verfahrens werden die Beispiele auch seine Einschränkungen festhalten.

6.8.1 Evaluation auf Basis einer simulierten Kamerabewegung

Die Bereitstellung einer Ground Truth, also einer Kamerabewegung, für welche alle Kameraparameter für jedes Bild bekannt sind, stellt für reale Begegnungen kein triviales Problem dar. Zu diesem Zweck wird eine nicht-statische Kamera simuliert. Um einen Realitätsbezug zu erhalten, wird die Kamera eine Begegnung filmen, welche tatsächlich ausgespielt wurde.

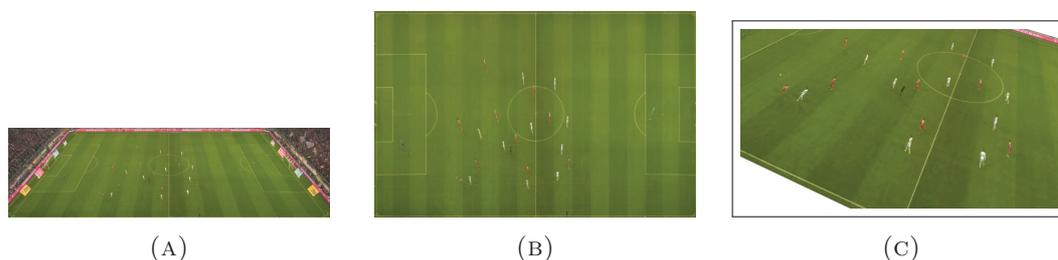


ABBILDUNG 6.12: Der Panorama Feed (A) wird auf Weltkoordinaten übertragen (B) und anschließend von einer virtuellen Kamera gefilmt (C).

Ausgehend von einer statischen Kamera (siehe Abbildung 6.12, A), mit welcher die Testbegegnung aufgezeichnet wurde, wurde zuerst die Homographie vom Bild zum Spielfeldmodell berechnet. Sei sie an dieser Stelle mit \mathbf{H}_s bezeichnet. Auf das in Weltkoordinaten beschriebene Modell (Abbildung 6.12, B) wird eine projektive Kamera gerichtet. Hierfür wird die übliche Notation $\mathbf{x} \propto \mathbf{P}\mathbf{X}$ mit

$$\mathbf{P} = \mathbf{KR} \left[\mathbf{I} \mid -\tilde{\mathbf{C}} \right] \quad (6.20)$$

$\mathbf{x} \in \mathbb{P}^2$ sowie $\mathbf{X} \in \mathbb{P}^3$ verwendet. Die internen Kameraparameter werden in der oberen Dreiecksmatrix \mathbf{K} abgelegt. Mit dem Hauptpunkt (Principal Point) im Mittelpunkt $(\frac{w}{2}, \frac{h}{2})$ der Bildebene, quadratischen Pixeln und der Brennweite f vereinfacht sich jene auf

$$\mathbf{K} = \begin{pmatrix} f & 0 & \frac{w}{2} \\ 0 & f & \frac{h}{2} \\ 0 & 0 & 1 \end{pmatrix} \quad (6.21)$$

Die Kameraposition $\tilde{\mathbf{C}}$ (die Tilde weist an dieser Stelle auf die Verwendung des Weltkoordinatensystems hin, in welchem $\tilde{\mathbf{C}}$ festzulegen ist) wird an einem erhöhten Punkt

abseits des Spielfelds festgelegt. Hierfür wurden die Koordinaten $(0; -45; 25)^T$ [m] verwendet, welche der tatsächlichen Position einer TV-Kamera (Kamera-1, siehe Kapitel 1.2 und 4.1) entsprechen könnten (Abbildung 6.12, C).

Die Bestimmung der weiteren externen Kameraparameter, d. h. der Rotationsmatrix \mathbf{R} und der Brennweite f , erfolgt bildweise abhängig von der Position der Spieler. Zuerst lohnt sich jedoch ein weiterer Blick auf die Matrix \mathbf{P} . Dem Modellaufbau ist zu entnehmen, dass alle Bildpunkte ihren Ursprung in der Spielfeldebene haben. Die dritten Komponenten aller in Frage kommenden Vektoren $\mathbf{X} \propto (x, y, 0, 1)^T$ verschwinden somit. Folgend kann die dritte Spalte von \mathbf{P} gestrichen werden.

$$\mathbf{x} \propto \underbrace{\begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \end{pmatrix}}_{=\mathbf{P}} \underbrace{\begin{pmatrix} x \\ y \\ 0 \\ 1 \end{pmatrix}}_{=\mathbf{X}} = \underbrace{\begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,4} \end{pmatrix}}_{=\mathbf{H}} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (6.22)$$

Die Transformation vom Eingangsbild (Abbildung 6.12, A) hin zu der festgelegten Kamerasicht (Abbildung 6.12, C) kann durch die Homographie $\mathbf{H}\mathbf{H}_s$ beschrieben werden. Da sowohl Brennweite als auch die Rotationsmatrix bildweise bestimmt werden, werden die Homographien zusätzlich mit einem Index entsprechend ihrer Bildnummer notiert: $\mathbf{G}_i := \mathbf{H}_i\mathbf{H}_s$. Die Transformation $\mathbf{x} \propto \mathbf{G}_i\mathbf{x}'$ erzeugt die Sicht der virtuellen Kamera aus den Punkten \mathbf{x}' des Panorama Feeds.

Die Bestimmung der Rotation und Brennweite erfolgt in Abhängigkeit von den Spielerpositionen. Die Positionen aller Feldspieler im Welt- bzw. Modellkoordinatensystem werden für jedes Bild als bekannt vorausgesetzt (mehr hierzu in Kapitel 7). Für ein beliebiges Bild $\#i$ kann der Schwerpunkt aller Feldspieler leicht bestimmt werden. Sei dieser $(\bar{x}, \bar{y}, 0) \in \mathbb{R}^3$. Auf diesen Punkt soll die Kamera gerichtet werden. Gesucht wird eine Rotationsmatrix, welche die Sicht entlang $\mathbf{e}_3 = (0, 0, 1)^T$ für den Fall $\mathbf{R} = \mathbf{I}$ auf eine Sicht entlang des Vektors $(\bar{x}, \bar{y}, 0)^T - \tilde{\mathbf{C}}$ korrigiert. Hierfür lässt sich das Kreuzprodukt beider Vektoren als Drehachse verwenden. Seien

$$\mathbf{a} := \frac{\left((\bar{x}, \bar{y}, 0)^T - \tilde{\mathbf{C}} \right) \times \mathbf{e}_3}{\left\| \left((\bar{x}, \bar{y}, 0)^T - \tilde{\mathbf{C}} \right) \times \mathbf{e}_3 \right\|_2} \quad (6.23)$$

$$s := \sin \angle \left((\bar{x}, \bar{y}, 0)^T - \tilde{\mathbf{C}}, \mathbf{e}_3 \right) \quad (6.24)$$

$$c := \cos \angle \left((\bar{x}, \bar{y}, 0)^T - \tilde{\mathbf{C}}, \mathbf{e}_3 \right) \quad (6.25)$$

und $[\mathbf{a}]_{\times}$ für $\mathbf{a} \in \mathbb{R}^3$ die schiefsymmetrische Kreuzproduktmatrix

$$[\mathbf{a}]_{\times} := \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix} \quad (6.26)$$

dann kann eine Rotationsmatrix \mathbf{R} mithilfe der Rodrigues-Formel in Matrizennotation [Faugeras, 1993] bestimmt werden.

$$\mathbf{R} = c\mathbf{I} + s[\mathbf{a}]_{\times} + (1 - c)\mathbf{a}\mathbf{a}^T \quad (6.27)$$

Bei diesem Vorgehen ist jedoch anzumerken, dass auch der Rollwinkel der Kamera in die Bewegung mit einbezogen wird. In TV-Aufzeichnungen von Fußballspielen ist dies in der Regel nicht der Fall.

Anmerkung 8. Das Modellkoordinatensystem, und damit das Spielfeldmodell, ist in das Weltkoordinatensystem eingebettet.

Für die Bestimmung der Brennweite werden vom Kamerazentrum ausgehende Strahlen auf jede Spielerposition ausgerichtet. Für jedes Bild wird die größte Brennweite ausgewählt, bei welcher kein Strahl die Bildebene außerhalb des Bilds (dem achsenparallelen Rechteckbereich zwischen den Punkten $(0,0)$ und (w, h) in der Bildebene) schneidet. Dazu wird für jeden Strahl individuell die größte Brennweite bestimmt, unter welcher ein Spieler noch im Bild sichtbar ist. Das Minimum der berechneten Werte ist die gesuchte Brennweite, unter welcher alle Spieler sichtbar sind.

Homographien des Tracking-Verfahrens werden an dieser Stelle mit $\hat{\mathbf{H}}_i$ bezeichnet. Diese können den Ground-Truth-Homographien \mathbf{G}_i gegenübergestellt werden. Als Fehler wird die *durchschnittliche Distanz (DD)* für jeden Punkt des Spielfelds Ω (in Modellkoordinaten) definiert.

$$DD_i := \frac{1}{\int_{\Omega} d\mathbf{x}} \int_{\Omega} d(\mathbf{x}, \hat{\mathbf{H}}_i \mathbf{G}_i \mathbf{H}_s^{-1} \mathbf{x}) d\mathbf{x} \quad (6.28)$$

$$= \frac{1}{\int_{\Omega} d\mathbf{x}} \int_{\Omega} d(\mathbf{x}, \hat{\mathbf{H}}_i \mathbf{H}_i \mathbf{x}) d\mathbf{x} \quad (6.29)$$

Die Integration wird durch Abtastungen des Spielfelds numerisch approximiert. Dabei wurde eine Auflösung von 1 m verwendet. Durch eine graphische Auswertung der erzeugten Werte können Fehlerkarten wie in Abbildung 6.13 erzeugt werden.

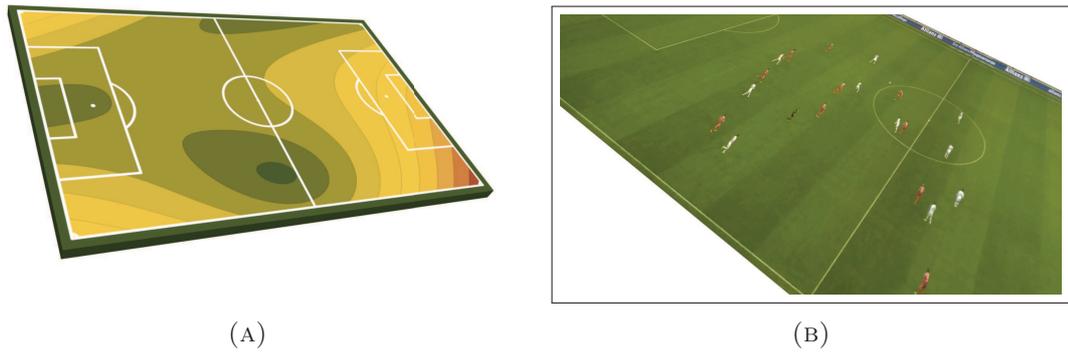


ABBILDUNG 6.13: Die von der Kameraparameter-Nachführung verursachten Fehler können für jeden Punkt des Spielfelds bestimmt werden (A). Hohe Abweichungen sind rot, geringe grün dargestellt. Festzustellen ist, dass Fehler in den sichtbaren Bereichen im Kamerabild (B) kleiner ausfallen, da diese Bereiche im ICP-Verfahren Berücksichtigung finden. Die DD beträgt 0,19 m.

Als Testdatensatz dient eine Begegnung, welche 2014 in der Allianz Arena ausgetragen wurde. Die Homographie \mathbf{H}_s wurde durch den DLT-Algorithmus unter Zuhilfenahme manuell ermittelter Korrespondenzen bestimmt. Die durchschnittlichen Distanzen für 3 000 Bilder können in Abbildung 6.14 betrachtet werden. Die Bildwiederholrate beträgt 25 *fps*.

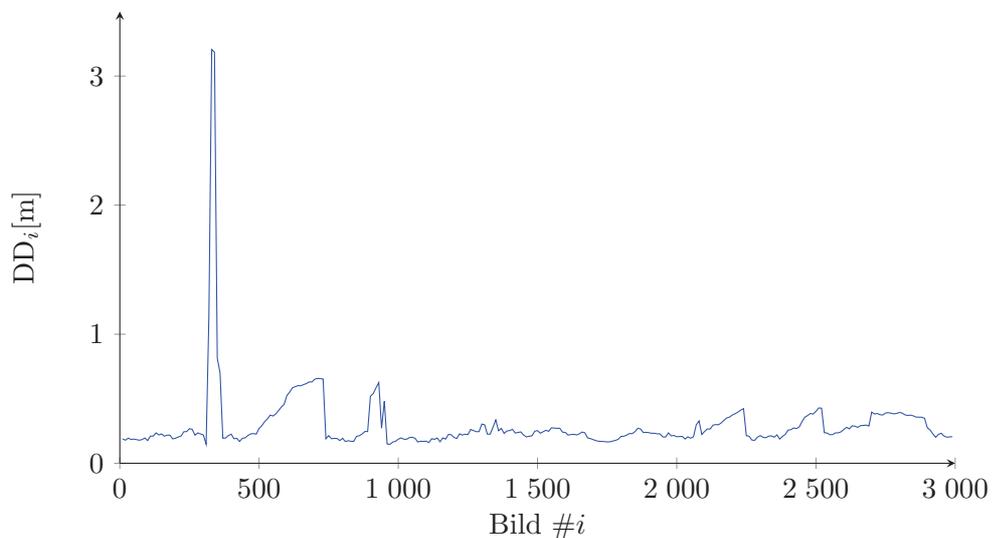


ABBILDUNG 6.14: Visualisierung des bildweisen Fehlers der Kameraparameter-Nachführung gemäß Gleichung 6.29

In dieser Evaluation kann eine Reihe von Beobachtungen notiert werden. Parallel zu Bild #320 tritt ein Verlust des Trackings ein, welcher zu einer Anwendung des Fallbacks (Abschnitt 6.6) führt. Dieser begründet sich in einem relativ kleinen Spielfeldausschnitt im Bild bei zugleich großem Rollwinkel (Abbildung 6.15). Da die Saatpunkte des Template Trackers aus einem achsenparallelen Rechteckgitter zu bestimmen sind, verblieben

in diesem Szenario unter den gewählten Schwellwerten nicht genügend Lösungen, um besagten Tracker weiter zu nutzen. Nachdem der Spielfeldausschnitt in den folgenden Bildern wieder an Größe zunimmt, kann die Nachführung mit einer ICP-Kalibrierung (automatisch) fortgesetzt werden. Da Drehungen in der Roll-Achse einen ungewöhnlichen Fall darstellen, wird dieses Problem im Weiteren vernachlässigt. Die größten übrigen lokalen Maxima in Abbildung 6.14 erklären sich durch längere oder starke Vergrößerungen des Spielfelds und damit durch Bildfolgen, in welchen aufgrund der Regeln über die Sichtbarkeit von Feldlinien im Allgemeinen die Ausführung des ICP-Verfahrens unterdrückt wird. Nach einer ICP-Kalibrierung ist im Verlauf der Funktion in Abbildung 6.14 zumeist eine abrupte Abnahme der DD festzustellen.



ABBILDUNG 6.15: Beginn einer Tracking-Abweichung in Bild #320 durch zu großen Rollwinkel

Nicht überraschend kann festgehalten werden, dass die Möglichkeit einer regelmäßigen Ausführung des ICP-Verfahrens für die Güte der Homographie-Bestimmung entscheidend ist. Innerhalb der simulierten Kamerabewegungen konnte in den meisten Bildern eine durchschnittliche Distanz zu den (tatsächlichen) Modellpunkten unterhalb von 0,25 m beobachtet werden. In Anbetracht einer Breite von Feldlinien von bis zu 0,12 m ist dies ein durchaus akzeptabler Wert.

6.8.2 Evaluation auf Basis realer Kamerabegegnungen

Als Evaluationsszenario entscheidend sind reale Begegnungen und Kamerabewegungen, welche durch einen Kameramann zum Zeitpunkt der Austragung vorgenommen wurden. Wie eingangs erwähnt, steht in diesem Zusammenhang keine Ground Truth zur Verfügung, jedoch werden in Kapitel 7 Homographien gemeinsam mit einer Spieler-Erkennung und -Nachführung genutzt, um die Position von Spielfeldobjekten in Weltkoordinaten zu bestimmen. Für diese Daten kann eine Ground Truth bereitgestellt werden, sodass eine gemeinsame Bewertung von Spieler- und Kameraparameter-Nachführung möglich ist.

Da Fragen nach der Genauigkeit der berechneten Homographien vorerst durch menschliche Beobachter zu bewerten sind, stehen in diesem Abschnitt Fragen im Mittelpunkt, welche Funktionsweisen und Einschränkungen des vorgestellten Verfahrens in seinem spezifizierten Anwendungsszenario charakterisieren sollen: Wie häufig muss auf die Fallback-Lösung zurückgegriffen werden? Wie oft entspricht eine akzeptierte Lösung des ICP-Verfahrens einem lokalen Minimum, welches nicht mit einer sinnvollen Bild-Modell-Zuordnung vereinbar ist? Und letztlich, wie oft muss das Verfahren nach einem manuellen Eingriff aufgrund eines derartigen Fehlers neu gestartet werden? Als Testdatensatz zur Anwendung kommen Scouting Feeds zu den in Tabelle 6.1 dargelegten Halbzeiten von vier Begegnungen der ersten Bundesliga. In Scouting Feeds sind zumeist alle Feldspieler im Bild sichtbar, jedoch können auch, bei Eintreten festgelegter Ereignisse (z. B. Torerfolg), einzelne Spieler in Nahaufnahme gezeigt werden. Schnitte kommen nicht vor.

Begegnung (Video)	Spielzeit	Halbzeit	Dauer [m:s]	Einzelbilder
FC Bayern - Schalke 04	2013/14	2	45:04	135 200
FC Bayern - Bayer Leverkusen	2013/14	1	44:53	134 650
Hertha BSC - BVB	2013/14	1	45:16	135 800
Augsburg - Paderborn	2014/15	1	45:01	135 050
Σ			180:14	540 700

TABELLE 6.1: Scouting Feeds, welche als Testdatensatz herangezogen wurden

Das Bildmaterial selbst liegt in 1280×720 mit 50 fps vor und wurde mit handelsüblichen Kleinkameras aufgezeichnet (siehe Kapitel 1.2). Für jedes Video wurde das erste Bild anhand markanter Spielfeldpunkte manuell kalibriert. Um die Bewertung zu erleichtern, wurde das Spielfeldmodell in das Kamerabild eingeblendet. Für den Fall, dass die Kameraparameter-Nachführung neu gestartet werden musste, d. h. dass es im weiteren Verlauf des Videos zu keiner weiteren Überdeckung zwischen Modell- und Spielfeldlinien kam, wurde an der nächstmöglichen Stelle nach dem Verlust eine neue manuelle Kalibrierung durchgeführt. Vorausgesetzt wurde auch hierfür eine ausreichende Sichtbarkeit des Spielfelds. Eine detaillierte visuelle Auswertung zu allen Begegnungen ist in Anhang B vorzufinden.

Auswertung FC Bayern - Schalke 04

Die Parameter der Kamerabewegung der gesamten Halbzeit konnten vollständig und damit ohne die Notwendigkeit eines manuellen Eingriffs berechnet werden. Auf die Fallback-Lösung wurde nicht zurückgegriffen. In der 64. Spielminute jedoch läuft eine Vielzahl

von Spielern der Gastmannschaft nach einem ausgeführten Eckball zurück in ihre Spielfeldhälfte (Abbildung 6.16, A-B). Die Bewegung wird begleitet durch eine Änderung der Brennweite, um die Feldspieler, welche nun in einem größeren Abschnitt des Spielfelds verteilt sind, innerhalb des Bilds darstellen zu können. Dies erfordert auch eine neue Bestimmung von Saatpunkten im Template Tracker. Aufgrund künstlicher Beleuchtung stellen Spielerschatten einen nicht zu vernachlässigenden Anteil der Rasenfläche im homogenen Rasen der Allianz Arena dar. Diese Bereiche werden aufgrund erhöhter Entropie für Saatpunkte vornehmlich aufgegriffen. Werden sie nicht sofort durch das Template Tracking für die weitere Verarbeitung ausgeschlossen, wie in diesem Fall, führt die Nachführung dieser nicht-ortsfesten Landmarken zu entscheidenden Abweichungen. Der ICP-Algorithmus optimiert im Fall ausreichender Sichtbarkeit von Feldlinien nun in Richtung eines ungeeigneten lokalen Minimums (Abbildung 6.16, C). Die Abweichung der Homographie wird zum Ende der 67. Spielminute in Richtung des globalen Minimums korrigiert (Abbildung 6.16, D).

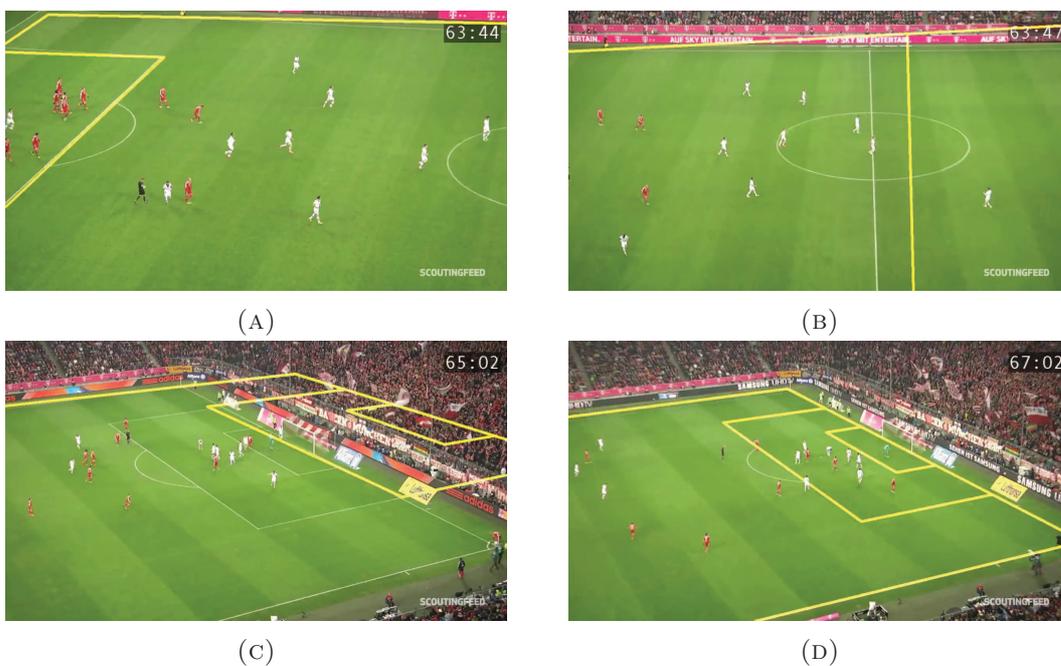


ABBILDUNG 6.16: Eine fehlerhafte Nachführung in der 64. Spielminute (A, B) führt zu einer ICP-Kalibrierung in Richtung eines ungeeigneten lokalen Minimums (C). Dies wird gegen Ende der 67. Spielminute selbstständig korrigiert (D).

Auswertung FC Bayern - Bayer Leverkusen

Obwohl das vorliegende Video ebenso wie die Aufzeichnung FC Bayern - Schalke 04 in der Allianz Arena angefertigt wurde und beide Aufnahmen die gleiche durchschnittliche Datenrate von 5 MBit/s aufweisen, sind in diesem Testszenario größere Abweichungen der Homographien festzustellen. Anhand der bereitstehenden Informationen ist nicht

möglich festzustellen, ob beide Aufnahmen mit einem ähnlichen Camcorder angefertigt wurden. Zu beobachten ist jedoch eine geringere Qualität in den MPEG-Rohdaten, welche sich in Blockartefakten äußert. Das ICP-Verfahren kann diese Abweichungen nur ausgleichen, falls (temporale) Abstände zwischen zwei ICP-Kalibrierungen ausreichend klein sind. Aus diesem Grund musste das gesamte Tracking-Verfahren insgesamt zweimal neu gestartet werden.

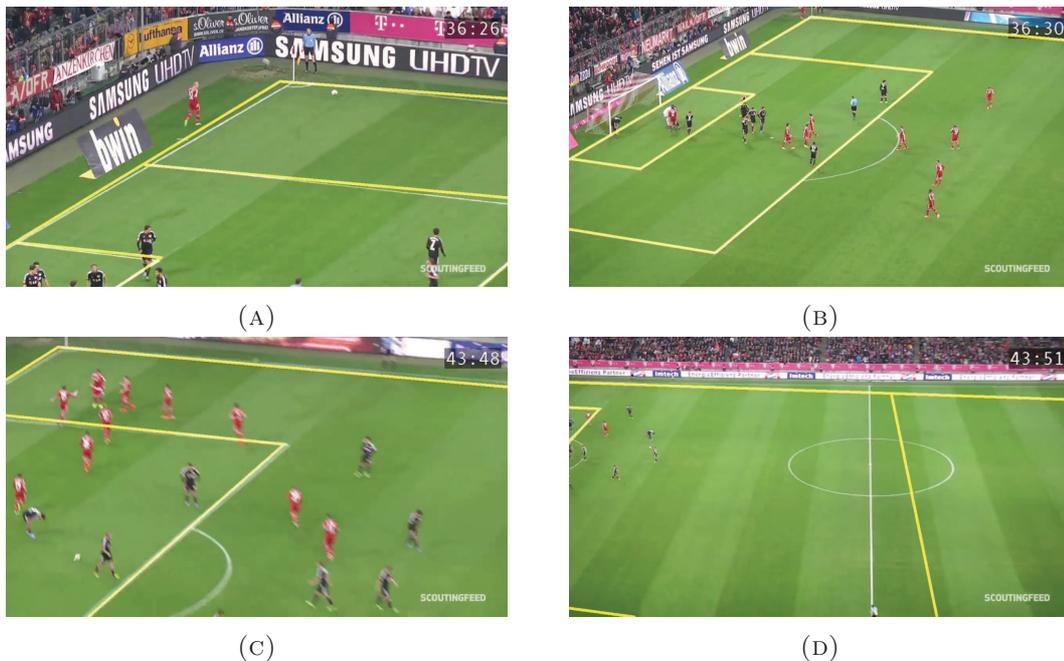


ABBILDUNG 6.17: Begründet durch die automatische Kalibrierung sind Abweichungen infolge von Skalierungen in aller Regel zu vernachlässigen (A, B). Ist keine Neuausrichtung möglich, stellt Bewegungsunschärfe eine zusätzliche Herausforderung dar (C) und kann zu Abweichungen führen (D).

Ungeachtet der Fähigkeit des ICP-Verfahrens, eine passende Zuordnung zwischen Bild und Modell zu finden (Beispiele hierzu in den Abbildung 6.17 B und C), kann an dieser Stelle eine entscheidende Einschränkung notiert werden: Der Template Tracker sucht nach Bild-zu-Bild-Zuordnungen in der Menge der Graspunkte, welche u. a. aufgrund ihrer Homogenität für diese Menge ausgewählt werden (Kapitel 3.4). Im Rahmen einer Videokompression können die gleichen Bereiche aufgrund der Annahme einer geringen Relevanz für das menschliche Auge verstärkt für die Einsparung von Speicherplatz in Betracht gezogen werden. So kann beobachtet werden, wie einzelne Blöcke zeitversetzt nachrücken, wodurch die Nachführung des Hintergrunds zusätzlich erschwert wird. Für das Erreichen einer möglichst hohen Übereinstimmung zwischen Modell- und Spielfeldlinien innerhalb des Template Trackers ist eine ausreichende Datenrate bzw. eine genügend genaue Abbildung der Kamerarohdaten in den Graspunkten erforderlich.

Auswertung Hertha BSC - Borussia Dortmund

Innerhalb des Testdatensatzes ist die Begegnung zwischen Hertha BSC und Borussia Dortmund im Berliner Olympiastadion das einzige Spiel, welches in einem Mehrzweckstadion ausgetragen wurde, in welchem sich neben dem Spielfeld auch eine 400 m Laufbahn befindet. Dies beeinflusst auch die Positionierung der Scouting-Kamera, welche das Spielgeschehen aus größerer Entfernung und mit größeren Änderungen der Brennweite als in den übrigen Begegnungen abbildet. Wie in den übrigen Videos ist auch hier ein Rückgriff auf die Fallback-Lösung nicht von Nöten. Die Kameraparameter-Nachführung ist in einem großen Anteil der Bilder dazu in der Lage, das Modell nahezu deckungsgleich in das Bild zu übertragen. Ein manueller Eingriff ist nicht erforderlich. Beispiele für die Zuordnung zwischen Bild und Modell sind Abbildung 6.18 zu entnehmen.

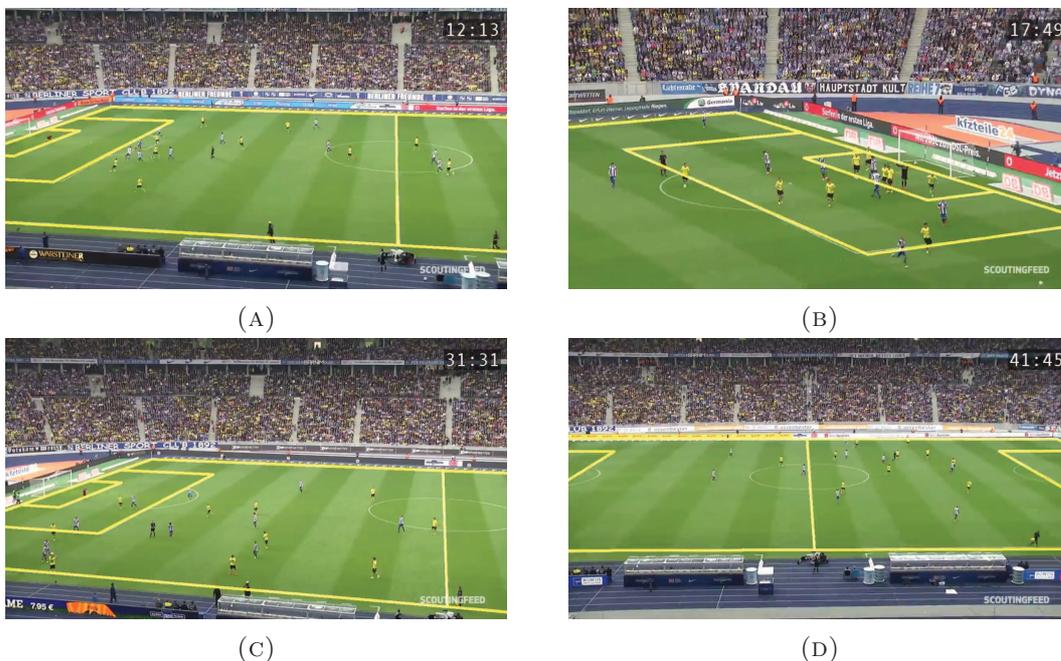


ABBILDUNG 6.18: Der Großteil der Homographien, die aus dem Tracking hervorgehen, ergeben eine hohe Übereinstimmung mit den Spielfeldlinien.

Auswertung FC Augsburg - Paderborn

Die Begegnung in der SGL Arena wurde in der 25. Spielminute nach einem Fehler manuell neu gestartet. In diesem Fall wurde die Spielfeldhülle (Kapitel 3.3) nicht korrekt bestimmt, wodurch der ICP-Algorithmus auf einer deutlich zu kleinen Menge an Geradensegmenten ausgeführt wird (Abbildung 6.19, A).

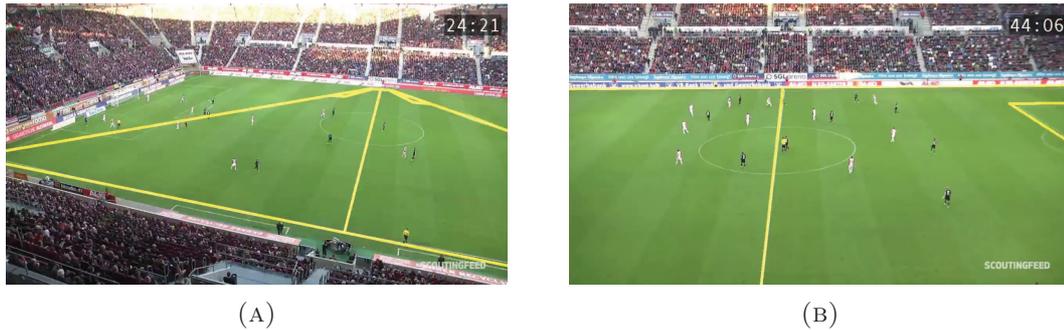


ABBILDUNG 6.19: Aufgrund einer besseren Bewertung in Gleichung 6.17 wird eine ungeeignete Spielfeldzuordnung angenommen (A). Dies begründet sich in einer falsch berechneten Spielfeldhülle. Dieser Fehler trat in den übrigen Bildern (B) nicht auf.

Aus diesem Szenario lässt sich eine weitere Einschränkung ableiten: Die Menge der sichtbaren Spielfeldlinien wird theoretisch anhand einer Homographie bestimmt. Ist die Diskrepanz zwischen dieser Menge und den tatsächlich bestimmten Geradensegmenten zu groß, führt dies leicht zu einer falschen Bild-Modell-Zuordnung. Neben einer ungenügenden Hülle und einer nicht ausreichend genauen Homographie für die Bestimmung der sichtbaren Spielfeldlinien kann ein weiterer Grund für einen derartigen Fehler auch eine schlechte Bildqualität sein. Für die Ausführung des ICP-Verfahrens ist es erforderlich, dass Spielfeldmarkierungen, welche durch die Kamera auf den Bildausschnitt abgebildet werden, im Bild sichtbar sind. Auch im Kontext des ICP-Verfahrens ist somit eine Güteschranke für Eingangsbilder durch das Verfahren vorgegeben.

6.8.3 Evaluation am Beispiel eines Fußgängerüberwegs

Im dritten Teil der Evaluation wird die Anwendbarkeit des vorgestellten Verfahrens in der Szenerie eines Fußgängerüberwegs untersucht. An diesen sind die bisher verwendete Hüllen- und Graserkennung sowie das Spielfeldmodell anzupassen. Erstere wurden durch eine schwellwertbasierte Farbauswahl in Verbindung mit morphologischen Operatoren ersetzt. Das neue Modell wurde speziell für die Szenerie bestimmt.

Fußgängerüberwege werden mit 0,50 m breiten Strichen markiert, die parallel zur Fahrtrichtung der Fahrzeuge liegen und untereinander Abstände von 0,50 m haben [...].

— Bundesministerium für Verkehr [1993]

Im konkreten Fall handelt es sich um eine 5,30 m Breite Ausführung in Garching b. München, welche, aufgebracht auf einer an den Rändern abfallenden Fahrbahn, bereits

gewisse Alterungserscheinungen aufweist (Abbildung 6.20, A). Unter Vernachlässigung dieser Faktoren wurde das in Abbildung 6.20 (B) dargestellte Modell festgelegt.

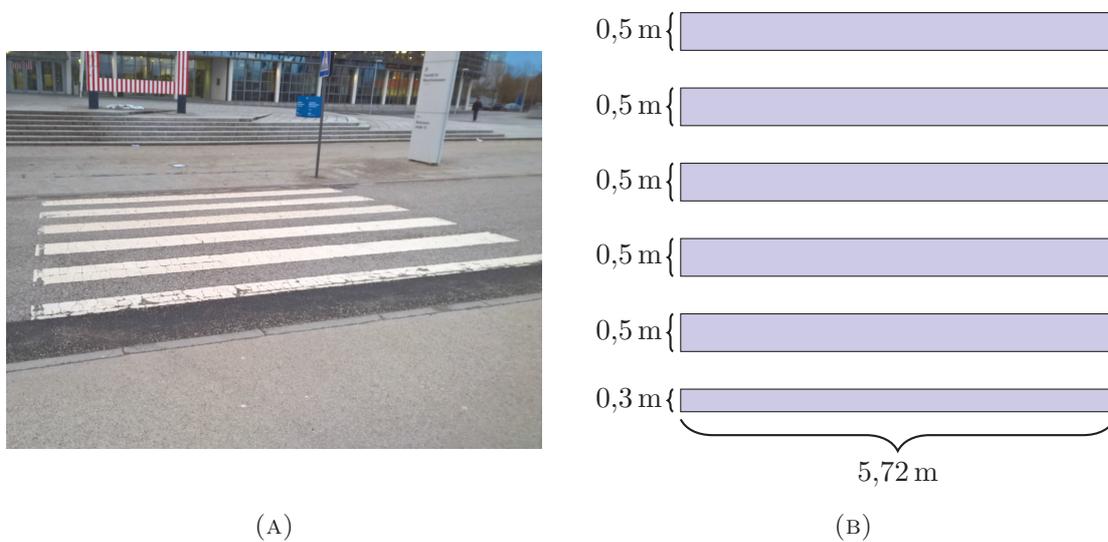


ABBILDUNG 6.20: Der Fußgängerüberweg des Testszenarios in einer Nahaufnahme (A) sowie das verwendete Modell (B). Abstände zwischen den Streifen des Fußgängerüberwegs werden im Modell mit 0,5 m vorgegeben.

Als Testdatensatz dient eine kurze Videosequenz, welche durch eine handelsübliche Mobiltelefon-Kamera aufgezeichnet wurde. Die Aufzeichnung wurde am Tag, ohne Stativ, bei einer Entfernung von ca. 20 m aus einer Höhe von ca. 6 m angefertigt. Die Kamera war dabei auf den Fußgängerüberweg ausgerichtet. In einer 70 s langen Videosequenz wurde dieser von einem Verkehrsschild, vier Personenkraftwagen, vier Passanten und zwei Fahrradfahrern (teilweise) verdeckt.

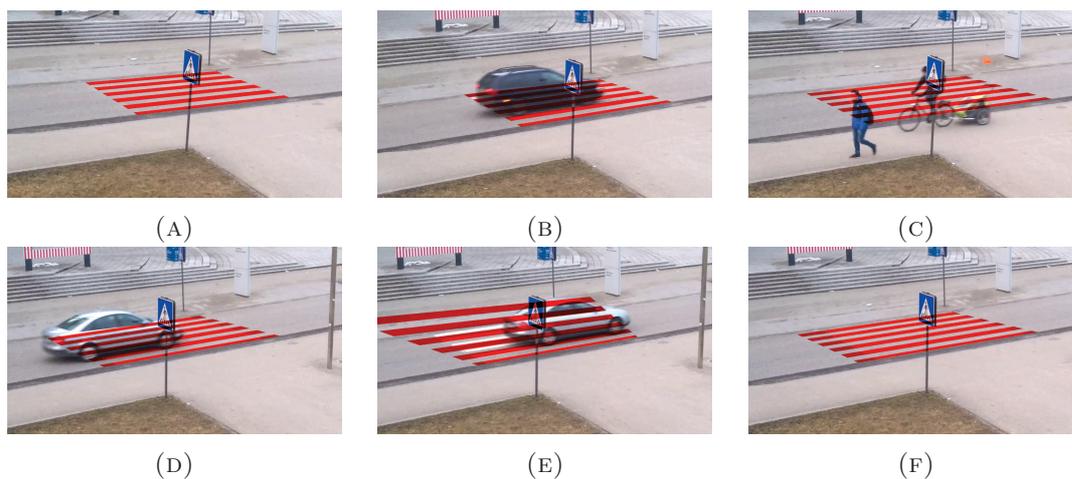


ABBILDUNG 6.21: Das ins Bild transformierte Modell überlagert den Fußgängerüberweg (rot). Verschiedene Verdeckungen des Fußgängerüberwegs werden beispielhaft gezeigt.

Innerhalb der Videosequenz wird eine sinnvolle Bild-Modell-Zuordnung ein einziges Mal verloren: in einem Moment, in welchem ein silbernes Fahrzeug den Fußgängerüberweg verdeckt (Abbildung 6.21, E). Dies geht auf Reflektionen auf der Oberfläche des Fahrzeugs zurück, welche im Bild zu falschen Zuordnungen innerhalb der (simplen) schwellwertbasierten „Gras-“ und Hüllenerkennung führen. Teile des Fahrzeugs werden ebenso als Fußgängerüberweg erkannt. Nachdem das Fahrzeug den Bildbereich verlässt, kann eine sinnvolle Zuordnung automatisch wiederhergestellt werden (Abbildung 6.21, F). Die weiteren genannten Verdeckungen führten nicht zu optisch bemerkbaren Abweichungen (Beispiele: Abbildung 6.21, A, B und C).

In einer Vielzahl der Bilder können vertikalen Modelllinien keine Korrespondenzen im Bild zugewiesen werden, da keine passenden Geradensegmente aufgrund der schlechten Beschaffenheit des Fußgängerüberwegs gefunden werden. In Bildern, in welchen Geradensegmente gefunden werden, ist deren Anteil im Sinn des robusten LMedS-Filters zu klein, um für eine Zuordnung berücksichtigt zu werden. Diese fehlende Evidenz für die Kalibrierung entlang der horizontalen Modellachse führt zu einem Fehler, welcher mit zunehmender Bildnummer ansteigt (Abbildung 6.21, A-F). Punkte entlang der vertikalen Modellachse werden entsprechend optischer Einschätzung mit hoher Genauigkeit zugeordnet.

6.9 Stand der Forschung

Das in diesem Kapitel vorgestellte Verfahren zur Kameraparameter-Nachführung macht an verschiedenen Stellen Gebrauch von der Theorie zu quasi-affinen Homographien, um ungeeignete Lösungen auszuschließen. Auch in der Literatur finden sich hierzu Ansätze. Auf der Basis von [Hartley u. Zisserman, 2003, Hartley, 1999] beschreiben [Moisan et al., 2012] mithilfe der Jacobi-Matrix eine Methode, welche irreguläre Homographien ausgehend von der Transformation inhomogener Punkte erkennt. In diesem Ansatz zur Erkennung „wilder Homographien“ wird letztlich eine Bedingung für die Erkennung ausgehend von der Umkehrabbildung \mathbf{H}^{-1} formuliert. Durch die Voraussetzung einer positiven Funktionaldeterminante und der damit einhergehenden Forderung nach der (lokalen) Erhaltung der Orientierung besitzt dieser Ansatz ein beschränktes Anwendungsfeld. Können Spiegelungen des Koordinatensystems auftreten — beispielsweise von Bild und Spielfeldmodell — entspricht dieses Verfahren nicht mehr der üblichen Definition quasi-affiner Homographien. Der Einschränkung zu gespiegelten Koordinaten unterliegt auch der von [Monnin et al., 2010] gewählte Ansatz. In der vorgestellten Methode werden die Orientierungen ausgewählter Punkte vor und nach der Transformation betrachtet. Im Kontrast hierzu stellen [Hess u. Fern, 2007] eine Methode mit größerem Abstand

zu Hartleys Definition vor. Sie verwenden zufällige Störungen in den Punktzuordnungen, um die Stabilität einer Homographie zu bewerten. Keiner der betrachteten Ansätze lässt sich ähnlich einfach und entsprechend der gebräuchlichen Definition [Hartley u. Zisserman, 2003] verwenden wie die vorgestellte Methode.

Die Bestimmung von Kameraparametern unter Verwendung von Homographien ist in der Bildverarbeitung ein gebräuchlicher Ansatz. Besondere Bedeutung kommt dabei dem Bild $\omega = \mathbf{K}^{-T}\mathbf{K}^{-1}$ des absoluten Kegelschnitts (vgl. Gleichung 6.20) und dessen Dual $\omega^* = \mathbf{K}\mathbf{K}^T$ zu. Solange die Kamera sich nicht ändert, bleiben die Bilder der Kegelschnitte (auf der Ebene im Unendlichen) ebenso unverändert. Eine schwenkende Kamera mit sich ändernden intrinsischen Parametern bei konstanter Position untersuchen [de Agapito et al., 1998]. Sie verwenden ω^* , um Einschränkungen aufzustellen und aus diesen die Kameraparameter abzuleiten. Einen ähnlichen Weg unter Verwendung von ω verwenden [Sturm u. Maybank, 1999], wobei sie aufzeigen, dass bei einer Kamera-Kalibrierung unter Verwendung einer einzigen Kalibrierebene nicht alle Unbekannten aufgelöst werden können und Fehler im Mittel größer als bei einer Standard-Kalibrierung [Faugeras u. Toscani, 1987] ausfallen. Basierend auf einer Geradenerkennung und einer Fluchtpunktsuche, welche jeweils unter Anwendung von LMedS-Schätzern umgesetzt wurden, bestimmen auch [Tang et al., 2007] die internen Kameraparameter (die Brennweite) durch ω^* . Auf die Anfälligkeit der Schätzung gegenüber Störungen wird hingewiesen (für einen tieferen Hintergrund zur Fluchtpunktsuche siehe [Förstner, 2010]). Aus mehreren Homographien wird analog zu Gleichung 6.19 auf den Zusammenhang einer Bildfolge geschlossen. Ebenfalls zu erwähnen ist die Modulus-Restriktion: Mit besagter Konstanz des Kegelschnitts ω^* bei gleichbleibenden internen Kameraparametern gilt $\omega^* \propto \mathbf{H}_\infty \omega^* \mathbf{H}_\infty^T$ für eine Homographie \mathbf{H}_∞ , welche die Ebene im Unendlichen von einem Bild in ein weiteres überführt. \mathbf{K} kann aus ω^* durch eine Cholesky-Zerlegung berechnet werden. Für die Bestimmung von \mathbf{H}_∞ zeigen [Pollefeys u. Van Gool, 1999], dass die Eigenwerte dieser Matrix allesamt den gleichen Betrag (Modulus) besitzen müssen. Diesen Gedanken fortsetzend stellen [Chandraker et al., 2007] ein schrittweises Verfahren vor, wobei wiederum für ω^* die Modulus-Restriktion formuliert wird. Eine algebraische Kostenfunktion wird im zweiten Schritt global minimiert. Eine automatische Parameterbestimmung anhand einzelner Homographien unter Berücksichtigung variierender interner Kameraparameter untersuchen [Johansson u. Cipolla, 2002] am Beispiel von Häuserwänden unter Einbeziehung dabei bekannter Modelle. Eine objektive Betrachtung der entstehenden Fehler konnte nicht durchgeführt werden. Anwendungen entsprechender Verfahren auf berechnete Homographien im vorliegenden Fußballkontext zeigten neben der Anwendbarkeit auch die Anfälligkeit für Störungen. Da im Verlauf der weiteren Kapitel ausschließlich die Positionen bodennaher Objekte vom Bild in Weltkoordinaten überführt werden, wird auf den Umweg eines vollständigen Kameramodells verzichtet. In einer Studie wurde

das Vorgehen untersucht, gewisse Kameraparameter analog zu [Gedikli, 2009] einzeln zu korrigieren. Dabei stellte sich heraus, dass eine gemeinsame Optimierung der Parameter (z. B. acht Parameter einer Homographie) i. Allg. zu besseren Resultaten führt. Aus diesem Grund wird die schrittweise Bestimmung von Kameraparametern, etwa über die Erkennung von Horizontgeraden, nicht weiter verfolgt.

Der in Abschnitt 6.7 besprochenen globalen Suche einer Bild-Modell-Zuordnung im FußballszENARIO widmen sich auch [Alemán-Flores et al., 2014]. Sie verwenden morphologische Operationen und die Hough-Transformation zur Erkennung der Spielfeldlinien. Diesen werden Modelllinien zugeordnet, woraus durch den DLT-Algorithmus Homographien berechnet werden. Wie im hier vorgestellten Verfahren werden diese bezüglich der Distanzen zu erkannten Geradensegmenten bewertet. Aus dem Argument des Minimums aller Bewertungen werden schließlich die Kameraparameter bestimmt. [Bu et al., 2011] bestimmen ebenfalls Bild-Modell-Korrespondenzen, um eine Homographie mit dem DLT-Algorithmus zu bestimmen. Die Ausgangsbasis hierfür bilden Schnittpunkte aus erkannten Geradensegmenten. Damit einhergehende Abweichungen sind in den Resultaten zu beobachten. Auf quasi-affine Homographien oder Einschränkungen zur Auswahl von Zuordnungen wird in beiden Ansätzen nicht eingegangen. Instabile Lösungen in diesem Sinn können demzufolge nicht ausgeschlossen werden. Auf Punktkorrespondenzen gründet ebenfalls der Ansatz von [Farin et al., 2005], in welchem sowohl Geradensegmente bzgl. ihrer Orientierung (horizontal, vertikal) klassifiziert werden als auch eine (grundlegende) Überprüfung für berechnete Homographien berücksichtigt wird. Es werden zwei mögliche Verfahrensweisen vorgestellt: Entweder die Endpunkte der Geradensegmente oder die Schnittpunkte „orthogonaler“ Geradensegmente werden für die Bildung der Korrespondenzen verwendet.

Einen Ansatz zur Kameraparameter-Nachführung in Bildsequenzen im Fußballkontext betrachten [Hayet u. Piater, 2007]. Wie im vorgestellten Ansatz werden dabei sowohl ein Verfahren der bildweisen Punktverfolgung als auch der Modellzuordnung miteinander verbunden. Die Bild-zu-Bild-Korrespondenzen werden durch einen KLT-Tracker [Tomasi u. Kanade, 1991] berechnet. Nachgeführte Feldlinien werden bei Sichtbarkeit in einem einzigen DLT-Schritt der Bildevidenz zugeordnet. Die Besonderheit in diesem Ansatz liegt in der Modellierung von Unsicherheiten in den Bild-zu-Bild-Homographien. Im Kontext Eishockey verwenden auch [Okuma et al., 2004] einen KLT-Tracker in Kombination mit einer Bild-Modell-Korrektur. Durch die Begrenzung auf einen einzigen DLT-Schritt ist auch dieses Verfahren abhängig von einer kaum zu erreichenden Genauigkeit bei der Auswahl der Bild-Modell-Korrespondenzen, was mit größer werdender Bildnummer letztlich zu einem signifikanten Drift führt [Gupta et al., 2011]. Die Evaluationen zu den jeweiligen Ansätzen wurden in Größenordnungen unterhalb von 2 000 Bildern durchgeführt. Einen linearen Ansatz verfolgen [Farin et al., 2003]. Ist \mathbf{H}_i die

Bild-Modell-Homographie zu Bild $\#i$, so lautet die lineare Schätzung für Bild $\#(i + 1)$ unter Kenntnis der Homographie \mathbf{H}_{i-1} zu Bild $\#(i - 1)$: $\hat{\mathbf{H}}_{i+1} = \mathbf{H}_i \mathbf{H}_{i-1}^{-1} \mathbf{H}_i$. Ausgehend von dieser Schätzung wird ein nicht-linearer Projektionsfehler optimiert. Eine robuste Zuordnung durch ein ICP-Verfahren verwenden [Lu et al., 2013]. Im Kontext Basketball mit relativ kleinen Spielfeldern und ausreichend markanten Spielfeldlinien und Werbe­flächen in der Spielfeldebene konnte dieses Verfahren für jedes der untersuchten Bilder eingesetzt werden. Innerhalb von Szenarien, die eine Lokalisierung nicht eindeutig gestalten, betrachten [Hess u. Fern, 2007] die Möglichkeit der simultanen Kartierung des Spielfelds. Wie sie erwähnen, ist auch in diesem interessanten Verfahren die Abhängigkeit von Ausreißern nicht ohne Bedeutung.

6.10 Diskussion und Ausblick

In diesem Kapitel wurde ein robustes Verfahren zur Kameraparameter-Nachführung auf der Grundlage planarer Objekte mit wenigen unterscheidbaren Landmarken beschrieben. Die Kameraparameter wurden dabei in Form von Homographien notiert. Die Evaluation zeigte, dass auch in langen Bildsequenzen (in der Summe drei Stunden) die Kameraparameter zuverlässig nachgeführt werden konnten. Da die in Abschnitt 6.9 genannten Verfahren eine Auswertung auf wenige Videosekunden beschränken, kann im Gegensatz zu anderen Verfahren eine Eignung für lange Bildsequenzen festgestellt werden. Ermöglicht wird dies, neben der robusten Extraktion von Geradensegmenten (Kapitel 5), auch durch die geeignete Auswahl von Homographien durch die Bestimmung ihrer Quasi-Affinität.

Die Nachführung der Kameraparameter wird in einer Abwägung zwischen einer Neu­zuordnung des Spielfeldmodells (Abschnitt 6.5) und einer Nachführung von Modell­landmarken (Abschnitt 6.6) entschieden. Insbesondere das vorgestellte modifizierte ICP-Verfahren trägt entscheidend zur Robustheit des Gesamtverfahrens bei. Durch eine automatische Erkennung des Verlusts der Bild-Modell-Zuordnung und einen damit einhergehenden Wechsel von lokaler Suche (ICP-Verfahren) zu globaler Suche (Abschnitt 6.7) könnten auftretende Fehler automatisch behoben werden. In Letzterer ist eine Normalisierung der Geraden vor der Anwendung des DLT-Algorithmus entsprechend Hartleys Anmerkungen zur numerischen Stabilität [Hartley, 1997a] eine sinnvolle Ergänzung. Möglichkeiten hierzu werden von [Zeng et al., 2008] und [Dubrofsky u. Woodham, 2008] betrachtet.

Das Template Matching zur Herstellung der Modelllandmarken (Punktkorrespondenzen) stellt die primäre Quelle von Abweichungen dar. Um die Allgemeingültigkeit des Verfahrens hinsichtlich der Videoformate sicherzustellen, fand die Bewegungsvorhersage

innerhalb gängiger Videokompressionen keine Berücksichtigung. Für die Herstellung von Bild-zu-Bild-Homographien sollte jedoch von der Bestimmung aus einem „dichten“ Vektorfeld, wie dem Ergebnis der Bewegungsvorhersage, abgesehen werden. Erste Studien zeigten, dass Störungen in einem solchen Vektorfeld einen entscheidenden Einfluss auf die Homographie, insbesondere auf einen Drift, haben. Vielmehr empfiehlt es sich auch in diesem Zusammenhang, geeignete Saatpunkte auszuwählen [Förstner et al., 2009].

Eine Homographie mit acht Freiheitsgraden bot in der überwältigenden Mehrheit der Bilder die Möglichkeit einer optisch kaum verbesserungsbedürftigen Zuordnung zwischen Bild und Modell. Da aus der Kalibrierung über eine einzige Ebene keine vollständige Matrix der internen Kameraparameter abgeleitet werden kann und eine Vielzahl von Bildern wiederum nicht ausreichend Merkmale für eine Berechnung der Verzeichnung ermöglicht, wurde auf eine Korrektur eben jener verzichtet.

Weitere Möglichkeiten der Verbesserung bestehen in einer nicht-linearen stochastischen Filterung der Homographien oder in einer Kartierung des Spielfelds zum Zeitpunkt der Kameraparameter-Nachführung. Auch wenn Fußball-Spielfelder meist nur wenige markante Punkte enthalten, kann es durch Letztere ermöglicht werden, zusätzliche Merkmale (wie Streifen im Rasen) in der Berechnung zu berücksichtigen. Erstere wird im Rahmen eines erweiterten Kalman-Filters von [Caballero et al., 2007] untersucht. Neben der Möglichkeit, Sprünge des ICP-Verfahrens in ungünstigen Lagen zu vermeiden (stochastisches Filter), kann ein entsprechendes Verfahren verwendet werden, um den aus einer Kalibrierung hervorgehenden Versatz (siehe Abbildung 6.14) durch eine nachträgliche Verarbeitung (stochastische Glättung) zu reduzieren.

Kapitel 7

Spieler- und Ballbehandlung

Dieses Kapitel basiert methodisch u. a. auf:

HERRMANN, Michael ; HOERNIG, Martin ; RADIG, Bernd: Online Multi-player Tracking in Monocular Soccer Videos. In: *AASRI Procedia* 8 (2014), S. 30-37

7.1 Einleitung

In diesem Kapitel wird nach der Kameraparameter-Nachführung die Erkennung von Spielfeldobjekten thematisiert. Dies wird sowohl monokular geschehen, und damit unter den gleichen Voraussetzungen wie im vorigen Kapitel, als auch im Kontext im Markt etablierter *Content Information-Provider- (CIP)* Positionsdaten-Erfassungssysteme erfolgen, welche eine Vielzahl von Kameras einsetzen. Da die Kameraparameter-Nachführung eine Transformation erkannter Spieler aus monokularen Bildern in Weltkoordinaten erlaubt, kann im Anschluss ein Vergleich beider Systeme durchgeführt werden. Dies kann auf der einen Seite Anwendung finden, um die zuvor eingeführte Kameraparameter-Nachführung hinsichtlich ihrer Genauigkeit quantitativ beurteilen zu können. Auf der anderen Seite werden auch Fehler in den CIP-Positionserkennungs- und Nachführungssystemen untersucht, erkannt und behoben.



Am Markt etablierte Systeme verwenden statische und kalibrierte Multikameralösungen und verfolgen die Spieler semi-automatisch. Das heißt, dass in Spielfeldnähe agierende Scouts die Möglichkeit haben, in die Ausgabe der berechneten Daten einzugreifen, bevor diese weitergeleitet werden. Dieser Puffer, der üblicherweise wenige Sekunden beträgt, soll genutzt werden, um Spielerverwechslungen, Abbrüche in der Nachführung von Spielern, Schiedsrichtern oder dem Ball sowie Sprünge in den Trajektorien korrigieren zu können. Solange ein Puffer noch nicht an einen anderen Dienstleister weitergeleitet ist, zum Beispiel für eine Auswertung von Laufwegen im Live-TV-Bild, wird eine entsprechende Änderung auch rückwärts innerhalb des Puffers durchgeführt. Fehler, die nach diesem Verfahren verblieben sind, werden im Anschluss an die Begegnung aus den Livedaten manuell entfernt. Für die Betrachtung dieses Kapitels werden Live-Daten betrachtet, da im nächsten Kapitel Konstrukte (Ballbesitz und Spielereignisse) eingeführt werden, für welche eine zeitnahe Verfügbarkeit bedeutsam ist.

7.2 Monokulare Spieler- und Ball-Nachführung

Die Erkennung und Nachführung von Objekten auf dem Spielfeld wird auf beliebigem Bildmaterial betrachtet. Hierzu müssen zunächst keine Kameraparameter vorliegen. Als Objekte werden sowohl die Feldspieler als auch der Ball und der Schiedsrichter und ggf. seine Assistenten betrachtet. Da es ohne Kenntnis um die Position des Spielfelds schwierig ist, Personen außerhalb der Spielfeldränder auszuschließen, werden auch Trainer oder Spieler außerhalb des Spielfelds in die Betrachtungen eingeschlossen. Unter nachträglicher Einbeziehung der Spielfeldinformationen, welche durch das Verfahren aus Kapitel 6 berechnet werden, können diese in einem nachgestelltem Verarbeitungsschritt ausgeschlossen werden.

7.2.1 Mehrspieler-Nachführung

Die Nachführung der Spieler erfolgt auf Grundlage stochastischer Filter. Dabei werden die durch das Bild gegebenen Evidenzen mit Kalman-Prädiktionen abgeglichen. Für jedes erkannte Objekt wird hierbei ein individuelles Filter verwendet. Neue und verloren gegangene Objekte (etwa durch Eintreten oder Verlassen des Bildbereichs) werden durch deterministische Algorithmen erfasst. Ein rechenintensives Tracking durch Objekterkennungen wird umgangen. Das Zustandsmodell schließt die Positionen und Geschwindigkeiten der Objekte ein (ein vergleichbares Modell wird in Abschnitt 7.5 beschrieben).

Sei ein neues Bild (Abbildung 7.1, A) einer Sequenz zusammenhängender Bilder gegeben. Durch die bisherigen Positionen und Geschwindigkeiten (in vergangenen Bildern)

der jetzt im Bild auftretenden Objekte (Personen) kann eine Annahme über die Position jedes Objekts im neuen Bild getroffen werden. In einem Vorverarbeitungsschritt wird weiterhin die binäre Grasmasken (Kapitel 3.4) zur Erkennung der Vordergrundregion (der Menge der Nicht-Graspixel) verwendet (Abbildung 7.1, B). Objekte im Bild werden durch achsenparallele Rechtecke (Rahmen) beschrieben. Durch das gegebene Bild können die folgenden Maßzahlen in Abhängigkeit der Rahmenpositionen bestimmt werden. Jede Bewertung bezieht sich auf jeweils eine Zusammenhangskomponente der Vordergrundregion, welche einen, keinen oder mehrere Spieler enthalten kann:

1. *Region-based evidence*. Der Grad an Abdeckung des Vordergrunds, welcher durch die Objektrahmen erreicht wird.
2. *Over-coverage feature*. Bestrafung von Graspixeln innerhalb der Rahmen.
3. *Overlap feature*. Außer in sehr seltenen Fällen überlappen zwei unterschiedliche Spieler einander im Bild nicht völlig, entsprechend werden einander überlappende Rahmen bestraft.
4. *Color-based evidence*. In einem Training auf dem ersten Bild werden die Farben beider Teams, der Schiedsrichter und Torwärtler bestimmt. Jedem Objekt werden die Farben einer dieser Klassen zugewiesen. Der Abstand der nun gemessenen Farbwerte innerhalb der Rahmen zu den trainierten Werten wird in diesem Maß wiedergegeben.
5. *HOG-based evidence*. Erkennung von Menschen durch Histogramme orientierter Gradienten [Dalal u. Triggs, 2005] innerhalb der Rahmen.
6. *Gating*. Distanz der Rahmen zur Kalman-Prädiktion.

Wird das Gating zusätzlich auf den maximalen Wert von 1 beschränkt, können alle sechs Maßzahlen bei geeigneter Formulierung im Intervall $[0; 1]$ ausgedrückt werden. In einer Konvexkombination werden sie unter geeigneten Gewichten, welche *a priori* zu trainieren sind, linear kombiniert. Hieraus ergeben sich Konfidenzen zu einzelnen Rahmenpositionen entsprechend des Bildinhalts (Abbildung 7.1, C). Damit können die Positionen aller Objekte gleichzeitig optimiert (Abbildung 7.1, D) werden.

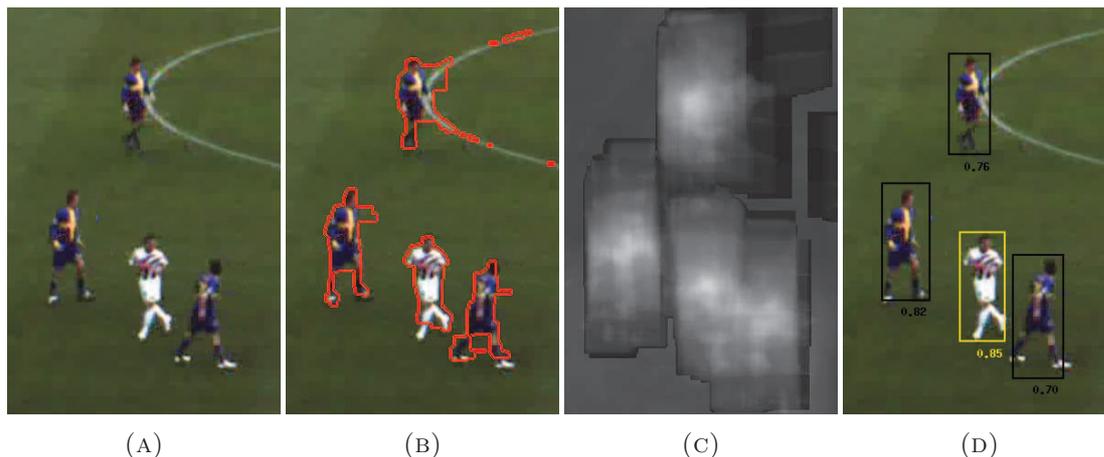


ABBILDUNG 7.1: Erkennung von Spielern in einer Konfidenzkarte. Originalbild (A), Segmentierung (B), Konfidenzkarte (C), Darstellung erkannter Spieler (D).
Bildquelle: [Herrmann et al., 2014]

7.2.2 Ballerkennung

Im Gegensatz zu anderen Objekten, welche sich am Boden oder bodennah befinden, stellt die Erkennung des Balls eine besondere Herausforderung dar. Eine exakte Positionsbestimmung in Weltkoordinaten ohne weitere Informationen gestaltet sich in vielen Situationen schwierig. Da Bälle oft „am Fuß“ eines Spielers anzutreffen sind und die Höhe zu diesen Zeitpunkten annähernd als „am Boden“ angesehen werden kann, erscheint eine Ballerkennung beginnend in diesen Situationen als zielführend. Eine solche Erkennung über eine Formulierung zur Spielernähe untersuchten bereits [Wang et al., 2014]. Perspektivisch kann es neben der Ballerkennung ein weiteres Ziel sein, die Genauigkeit der erkannten Spielerpositionen durch das Wissen um den Ball zu erhöhen oder die Positionen von Spielern, welche nicht getrackt werden, zu schätzen [Kim et al., 2010].

Für eine erste Untersuchung im Rahmen einer Bachelor-Arbeit [Hopper, 2015] zur monokularen Ballerkennung wurde in 531 Bildern die Ballposition (in Bildkoordinaten) markiert.

Merkmal	Haar-like	LBP	HOG
Kaskaden	13	17	17
Trefferquote in %	61,58	80,23	66,10
durchschnittliche Falsch-Positiv-Treffer	370	206	230

TABELLE 7.1: Vergleich der Ergebnisse zur Balldetektion. Anwendung fanden Kaskaden aus Klassifikatoren unter Verwendung von Haar-like Features, Local Binary Patterns (LBPs) und HOGs. Quelle: [Hopper, 2015]

Die Klassifikation wurde in Kaskaden durchgeführt [Lienhart u. Maydt, 2002, Viola u. Jones, 2001]. Neben den dafür gebräuchlichen Haar-like Features wurden ebenso Local Binary Patterns [Ojala et al., 2002] sowie die bereits zur Personenerkennung im vorigen Abschnitt eingesetzten HOGs verwendet. Die Anzahl der Kaskaden wurde je nach Merkmal sinnvoll gewählt.

In den durchgeführten Tests zeigten die Local Binary Patterns die beste Performance und übertrafen kaskadiert klassifizierte Haar-like Features sowie HOGs bzgl. Trefferquote und Anzahl falsch positiver Messungen. In Anbetracht der Verschiedenartigkeit der optischen Repräsentationen von Bällen (hervorgerufen u. a. durch Bewegungsunschärfe, unterschiedliche Ball-Designs) ist die Trefferquote von über 80 % vertretbar, jedoch erkauft durch eine Vielzahl falscher Detektionen, vor allem im Publikum. Durch die Einschränkung der Erkennung auf die Spielfeldhülle (Kapitel 3.3) lässt sich die Anzahl der Falsch-Positiv-Treffer um 60 % reduzieren. Da die Verwendung von LBPs zu Fehldetektionen in der Rasenfläche führen kann, ist es sinnvoll, Graspixel (Kapitel 3.4) ebenfalls auszuschließen. Ein entsprechendes Ergebnis kann in Abbildung 7.2 betrachtet werden.



ABBILDUNG 7.2: Ballerkennung innerhalb der Spielfeldhülle unter Ausschluss der Rasenfläche. Bildquelle: [Hopper, 2015, ZDF, 2013a]

Natürlich bleiben nach den gewählten Suchregionen Bälle im Bereich des Publikums verborgen, mit geeigneten temporalen Filtern erscheint es jedoch möglich, die Ballposition abschnittsweise zu ermitteln.

7.2.3 Ausblick

Nachdem Kameraparameter für jedes Bild berechenbar sind, ist es möglich, ein beliebiges (Kamera-1-)Eingangsvideo auf eine statische Kamera zu überführen. Möglichkeiten ergeben sich etwa durch eine neue virtuelle Kamera (wie in Kapitel 6.8.1 gezeigt) oder direkt durch die Übertragung in Weltkoordinaten (siehe hierzu Abbildung 6.12, B). Auch wenn keine vollständige Abdeckung des Spielfeldbereichs sichergestellt werden kann, können Spielfeldobjekte durch die statische Kamera leichter erkannt werden. [Figuroa et al., 2006] nutzen morphologische Operationen auf zeitlichen Veränderungen (1D-Signalen) für fixe Pixel, um Objekte zu erkennen. In [Xu et al., 2004] wird dies durch exponentielle Lernraten erweitert, um zusätzlich eine größere Robustheit, gerade bei wechselnden Umgebungsbedingungen, zu erzielen. Der Vordergrund ergibt sich jeweils aus signifikanten Unterschieden zur Hintergrundschätzung. Neben Schwierigkeiten durch Objektverzerrungen in Abhängigkeit von der Feldposition (bei einem Warping) ist bei einer Erkennung aufgrund einer Hintergrundschätzung auch der durch die ICP-Kalibrierung verursachte Fehler zu berücksichtigen (Kapitel 6.8.1). Bei einem zukünftigen Erkennungsverfahren, welches auf den berechneten Kameraparametern fußt, muss der Versatz, der durch eine Kalibrierung hervorgerufen wird, in der Hintergrundschätzung berücksichtigt werden. Für einen entsprechenden Ansatz stehen jedoch alle Transformationen aus der Kameraparameter-Nachführung bereit.

7.3 Spieler-Nachführung in CIP-Systemen

Im zweiten Teil dieses Kapitels werden Fehler in einem marktüblichen Fußball-Objekt-Nachführungssystem untersucht, bevor dieses System im Anschluss als Datengrundlage für eine Auswertung der in Kapitel 6 und Abschnitt 7.2.1 vorgestellten Kameraparameter- und Spieler-Tracking-Systeme verwendet wird. Da auch durch dieses System keine Höheninformationen zur Verfügung gestellt werden, werden alle Positionsangaben als geordnete Paare betrachtet. Für einen Zeitpunkt t wird die Position des Spielers $\#i$ im Weiteren mit $\mathbf{x}_t^i \in \mathbb{R}^2$, für den Ball mit $\mathbf{b}_t \in \mathbb{R}^2$ bezeichnet. Die Abtastungen werden bezüglich ihres temporalen Auftretens als äquidistant betrachtet (z. B. 25 *fps*). Bei n Abtastungen (Bildern) bilden die Folgen $\mathbf{x}_{1:n}^i$ und $\mathbf{b}_{1:n}$ $n * fps^{-1}$ Sekunden ab. Neben den Positionsdaten wird ebenso ein Running Flag zu jedem Bild zur Verfügung gestellt. Dieses gibt Auskunft über den aktuellen Spielzustand, d. h. ob in diesem gerade das Spielgeschehen läuft, oder ob es unterbrochen ist (etwa durch ein gefallenes Tor, Abseits oder einen Ball außerhalb des Spielfelds).

7.3.1 Analyse eines exemplarischen Datensatzes

Für die Beurteilung der Qualität des Trackings von Spielern und Ball wurde eine Begegnung zweier Erstligamannschaften gewählt. Die durch Opta [Perform, 2015] bereitgestellten Spielstatistiken können Tabelle 7.2 entnommen werden.

	Heimmannschaft	Gastmannschaft
Ballbesitz	75,8 %	24,2 %
Torschüsse	22	10
Gewonnene Zweikämpfe	52,1 %	47,9 %
Fouls	8	12
Ecken	14	4

TABELLE 7.2: Spielstatistik, bereitgestellt durch Opta [Perform, 2015]

Durch einen Fachkundigen wurden in diesem Spiel Fehler sowohl in den Spielerpositionen als auch im Running Flag markiert. Soweit dies möglich war, wurde dabei mit bildweiser Genauigkeit gearbeitet. Neben den Positions- und Running-Flag-Daten war hierfür auch eine Aufzeichnung des Spiels in Form des Panorama Feeds gegeben. Da die Daten naturgemäß in zwei Spielhälften geteilt sind, wurde diese Teilung für die Evaluation beibehalten. Neben dem sich daraus ergebenden Ermüdungstest der Scouts kann auf diese Weise eine mögliche Korrelation zwischen den erzielten Toren und den auftretenden Fehlern überprüft werden: So kann man annehmen, dass Spielervertauschungen nach gefallen Toren besonders häufig auftreten und damit in der ersten Hälfte verstärkt vertreten sind.

7.3.1.1 Fehlerbetrachtung des Running Flags

Running Flags werden in der weiteren Betrachtung als Zeiträume angesehen, in welchen das Flag gesetzt (*wahr*) ist. Eine Übersicht für die gewählte Begegnung ist in Tabelle 7.3 vorzufinden. Ein Running-Flag-Intervall ist ein Intervall auf der diskreten Zeitachse, wobei zu jedem Zeitpunkt innerhalb des Intervalls die Running Flags *wahr* sind und es kein Running-Flag-Intervall gibt, welches besagtes Intervall als echte Teilmenge umfassen würde. Die Größe wird anhand der Anzahl enthaltener Running Flags bemessen.

	Erste Hälfte	Zweite Hälfte
Bilder	67 500	67 807
Running-Flag-Intervalle	46	43
Flag-Wechsel	92	86

TABELLE 7.3: Übersicht zu den Running-Flag-Daten innerhalb der betrachteten Begegnung. Angegeben werden die absoluten Häufigkeiten.

Für die Bemessung der zeitlichen Übereinstimmungen mit den tatsächlichen Spielunterbrechungen werden Beginn und Ende der Running-Flag-Intervalle herangezogen. Eine Änderung des Flags (und damit ein Beginn oder ein Ende des Intervalls) wird als Flag-Wechsel bezeichnet. Letztendlich sind es diese Wechsel, die von Scouts am Spielfeldrand manuell annotiert werden. Gemäß Tabelle 7.3 tritt dieses Ereignis in den jeweiligen Halbzeiten etwa gleich häufig auf. Für die Beurteilung der Genauigkeit der Wechsel (und damit des Running Flags) wurden alle Zeitpunkte für einen Wechsel ein zweites Mal manuell annotiert. Hat ein Wechsel (Intervallbeginn oder -ende) ein Gegenüber in der Ground Truth, ist also auch dort ein entsprechender Wechsel verzeichnet, wird er als gefunden bewertet. Um Ungenauigkeiten zu tolerieren, werden die Untersuchungen im Weiteren gefenstert durchgeführt, d. h. in Abhängigkeit von einer gegebenen Suchfenstergröße wird in der Messung (den Daten des Nachführungssystems) nach dem Wechsel gesucht.

Anmerkung 9. Angaben zur Fenstergröße sind in beide Richtungen des Bezugspunkts x zu interpretieren. Ist von einer Fenstergröße von 0,6 s die Rede, so umfasst das Fenster den Bereich $[x - 0,6; x + 0,6]$ der in Sekunden skalierten Zeitachse. Dies ist entsprechend mit $\pm 0,6$ s zu lesen.

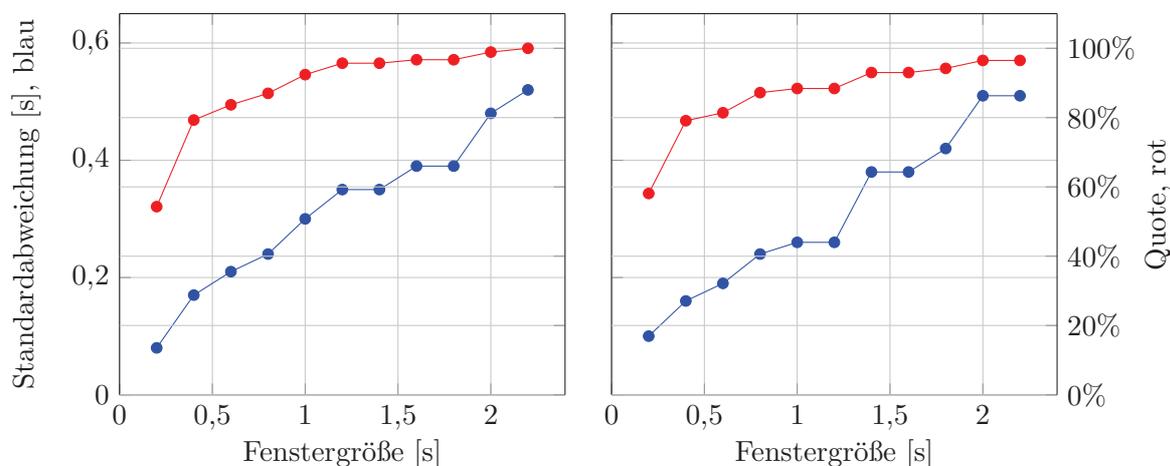


ABBILDUNG 7.3: Quote (rot) gefundener Wechsel und Standardabweichung (blau) der jeweils gefundenen Wechsel in Abhängigkeit von der Fenstergröße. Resultate für erste (links) und zweite (rechts) Halbzeit sind einzeln aufgeführt.

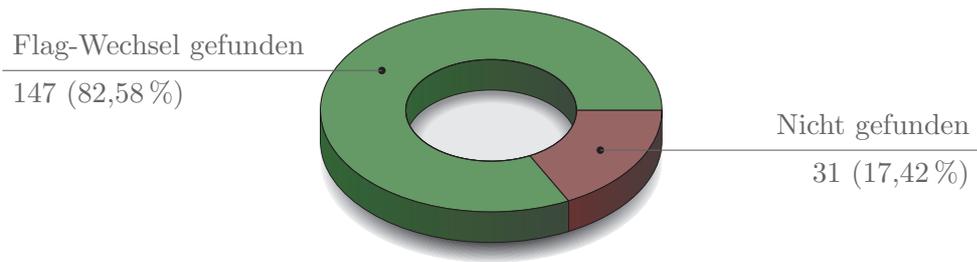


ABBILDUNG 7.4: Innerhalb eines Fensters von 0,6 s gefundene Running-Flag-Wechsel

Den in Abbildung 7.3 dargestellten Graphen lässt sich entnehmen, dass bei einer Fenstergröße von 0,2 s die Hälfte der Wechsel gefunden werden. Nahezu kein Wechsel hat eine Abweichung von mehr als den abgebildeten 2,2 s. Für die Wechsel innerhalb dieser Fenstergröße liegt damit eine Standardabweichung von etwa 0,5 s für beide Halbzeiten vor. Die empirischen Mittelwerte sind, mit Werten von $-0,15$ s und $-0,12$ s (in beiden Fällen für die gefundenen Ereignisse aus dem 2,2 s Fenster) wie erwartet klein. Die Fehler der Scouts mitteln sich im Schnitt um die tatsächlichen Zeitpunkte. Da ebenso kein systematischer Fehler für Zeitpunkte des Running-Flag-Intervallbeginns und -endes feststellbar war, bleibt die Nettospielzeit nahezu unberührt.

Unter Verwendung der maximal aufgetretenen Standardabweichung als Fenstergröße (0,6 s) können 147 Wechsel gefunden werden (siehe Abbildung 7.4). Eine Übersicht aller aufgetretenen Abweichungen ist in Abbildung 7.5 wiedergegeben. Da an dieser Stelle keine Fenstergröße vorgegeben ist, sind auch Ausreißer enthalten.

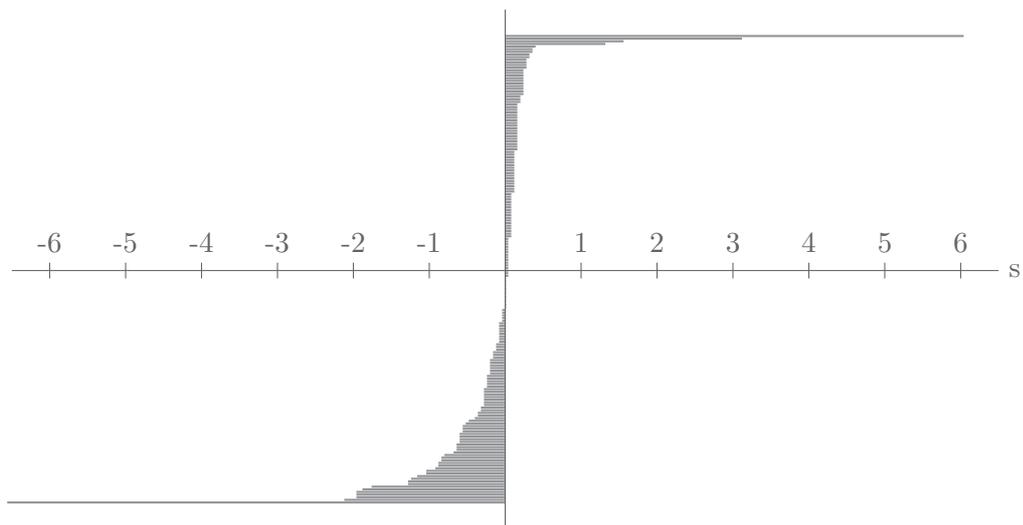


ABBILDUNG 7.5: Vollständige Liste der Abweichungen der Running-Flag-Wechsel relativ zur Ground Truth. Jede Abweichung ist als horizontale, bei 0 beginnende Strecke dargestellt.

7.3.1.2 Fehlerbetrachtung der Spieler- und Ball-Nachführung

Unabhängig von dem durch das Running Flag gegebenen Spielzustand stehen zu jedem Zeitpunkt innerhalb der Begegnung für jeden Spieler auf dem Feld und den Ball (teilweise auch für den Schiedsrichter und die Linienrichter) Weltkoordinaten zur Verfügung. Diese wurden für eine visuelle Untersuchung innerhalb des Panorama Feeds sichtbar gemacht. Dabei auftretende Fehler wurden mit jeweils genau einem der folgenden Typen klassifiziert:

1. Die Ballposition wird nicht korrekt nachgeführt (dieser Fehlertyp wurde auf Running-Flag-Intervalle eingeschränkt).
2. Mindestens ein Spieler wird durch das System nicht nachgeführt.
3. Mindestens zwei Spieleridentifikationen (IDs) werden vertauscht.

Entsprechend dieser Festlegung können für ein Bild auch mehrere Fehlertypen gleichzeitig auftreten. Wird zum Beispiel ein Trainer an der Seitenlinie statt einem Spieler nachgeführt, also keine zwei Spieler vertauscht, entspricht dies einem Fehler vom Typ 2. Während der Annotation von Fehlern wurden dem System gewisse Zugeständnisse gemacht, da es technisch unrealistisch ist, Spieler und Ball bei Verdeckung beliebig genau unterscheiden zu können. Insbesondere bei Fehlern vom Typ 1 wurden Markierungen nur eingefügt, wenn die tatsächliche Position eine Abweichung von mehreren Metern von der annotierten Position aufwies.

Erste Hälfte						
Fehlertyp	# Intervalle	Durchs.- länge [s]	Min.- länge [s]	Max.- länge [s]	Std.- abw. [s]	Ges.- länge [s]
1.	115	4,97	0,08	47,8	7,70	571,52
2.	33	12,47	0,4	107,12	20,83	411,52
3.	6	5,32	1,36	17,84	6,30	31,92
U	116	7,32	0,08	189,76	19,32	849,2
Zweite Hälfte						
1.	83	3,90	0,16	81,64	10,07	323,32
2.	19	4,39	0,2	18,36	4,78	83,44
3.	15	13,77	0,4	86,84	27,62	206,6
U	95	6,10	0,16	94,88	15,18	579,12

TABELLE 7.4: Fehlerauswertung gemäß den zuvor eingeführten Klassen. Eine Vereinigung (U) über alle Fehlertypen ist in den jeweils letzten Zeilen gegeben.

Die Übersicht zu den klassifizierten Fehlertypen kann Tabelle 7.4 entnommen werden. Hierbei sei darauf hingewiesen, dass sich die letzte Zeile aus einer Vereinigung (\cup) der Fehlerintervalle ergibt und damit kürzer ausfällt als die Summe aller Einzelfehler. Tatsächlich können auch Unterschiede zwischen beiden Halbzeiten ausgemacht werden. Während der erste Fehlertyp in beiden Halbzeiten ähnliche Charakteristika zeigt, sind die Fehlertypen 2 und 3 nicht gleich verteilt. Entgegen der eingangs formulierten Vermutung sind Spielervertauschungen in der ersten Hälfte nicht häufiger anzutreffen.

Da gerade die Ballposition für weitergehende Konstrukte von Bedeutung sein wird, ist die Länge von Fehlern vom Typ 1 nicht zu unterschätzen. In Abbildung 7.6 ist die Anzahl fehlerbehafteter Bilder in Relation zur Gesamtzahl der Bilder (Bruttozeit) abgebildet. An dieser Stelle ist ersichtlich, dass auch Anbieter, welche die Möglichkeit besitzen, eine größere Anzahl statische Kameras zu verwenden, keine perfekten Daten liefern können. Eine Ermüdung der Scouts kann hingegen in der ausgewählten Begegnung weitestgehend ausgeschlossen werden.

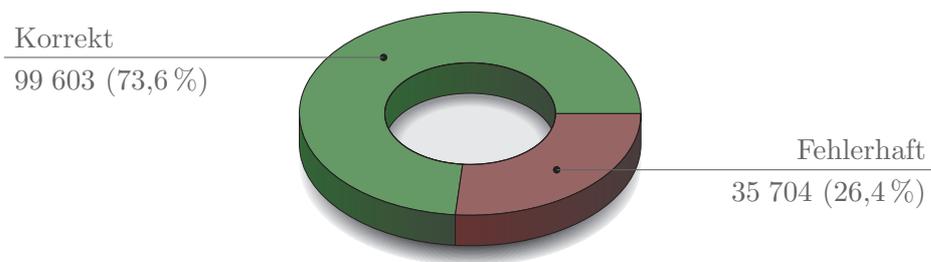


ABBILDUNG 7.6: Verhältnis der Bilder, welche fehlerbehaftete Tracking-Informationen tragen, zur Bruttospielzeit. Bilder mit multiplen Fehlern wurden einfach gezählt.

7.4 Evaluation des monokularen Systems

Auf Basis bereitstehender Positionsdaten kann ebenso die Mehrspieler-Nachführung (Abschnitt 7.2.1) evaluiert werden. Hierzu wurden wiederum die Videos aus Kapitel 6.8.2 verwendet, für welche bereits die Kameraparameter bestimmt wurden. Die Spieler-Nachführung wird in diesem Sinn gemeinsam mit dem Algorithmus zur Kameraparameter-Nachführung ausgewertet, da eine getrennte quantitative Betrachtung aufgrund nicht vorhandener Ground-Truth-Kameraparameter ausgeschlossen werden muss.

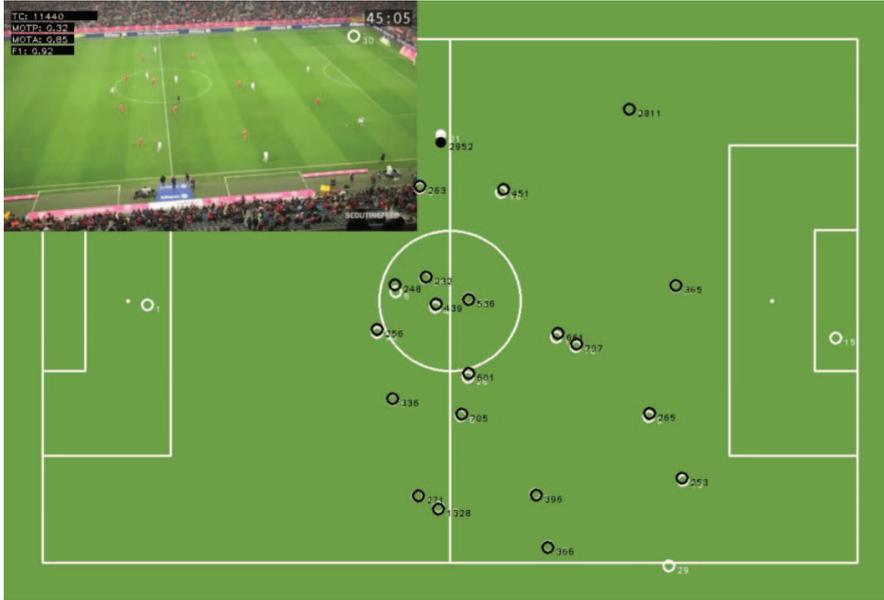


ABBILDUNG 7.7: Abweichungen zwischen den Tracking-Systemen innerhalb eines Bilds. Weiße Kreise stellen Ground-Truth-Positionen dar, schwarze Kreise erkannte Spieler im Rahmen des vorgestellten Systems. Sind die Kreise nicht ausgefüllt, wurden Korrespondenzen beider Systeme innerhalb eines 1 m-Radius gefunden oder es handelt sich um Schiedsrichter oder Torwächter. Quelle: Michael Herrmann

Für die Auswertung wurden die Positionen getrackter Spieler in Weltkoordinaten mit den Positionen der Spieler in den Ground-Truth-Daten verglichen. Unterschreitet der Abstand zwischen Erkennungen der zwei Systeme einen Schwellwert, wird eine Korrespondenz hergestellt (Abbildung 7.7). Die Torwächter beider Mannschaften werden ebenso wie der Schiedsrichter und Linienrichter als optionale Objekte betrachtet. Eine Erkennung wird nicht als Fehler gewertet, eine Nichterkennung ebenso. Dies begründet sich darin, dass die Torhüter in den Scouting Feeds oft nicht im Bild sichtbar sind und die Positionen der Schiedsrichter nicht zu allen Begegnungen in der Ground Truth zur Verfügung stehen. Ausgehend von der Ground Truth werden aussagekräftige und gebräuchliche Maßzahlen in Abhängigkeit von Suchradien um die gegebenen Spielerpositionen verwendet. Neben dem F1-Score, welcher die Precision- und Recall-Werte harmonisch mittelt, werden die *Multiple Object Tracking Accuracy (MOTA)* sowie die *Multiple Object Tracking Precision (MOTP)* genutzt. Damit können Abweichungen beider Tracking-Systeme über ein Video hinweg berechnet werden.

$$\text{MOTA} := 1 - \frac{\sum_t (\text{FN}_t + \text{FP}_t + \text{IDSW}_t)}{\sum_t \text{GT}_t} \quad (7.1)$$

$$\text{MOTP} := \sum_t \frac{\sum_i d_{t,i}}{c_t} \quad (7.2)$$

Die Anzahl FN_t an falsch negativen, FP_t an falsch positiven und GT_t an Ground-Truth-Objekten wird in jedem Bild $\#t$ für die Berechnung der MOTA genauso berücksichtigt wie die Anzahl $IDSW_t$ der Wechsel der Tracker IDs im Bezug auf die Ground-Truth-Objekte. Für die Berechnung dieser Anzahl wird das folgende Verfahren verwendet:

1. Berechnung eines bipartiten Matchings (in einem ungerichteten Graph) zwischen den Objekten (Knoten) der Ground Truth und der Messung (Knoten) im ersten Bild, wobei Assoziationen (Kanten) nur erlaubt sind, falls ein zuvor festgelegter Abstand nicht überschritten wird.
2. Wiederherstellung der bestehenden Zuordnungen in einem neuen Bild, solange der festgelegte Abstand nicht überschritten wird.
3. Matching der verbleibenden Objekte im neuen Bild (das Erhalten einer Zuordnung besitzt damit Priorität vor einem Matching mit geringerem Abstand).
4. Wiederholung, beginnend von Schritt 2.

Insgesamt finden auf diese Weise drei mögliche Fehlerquellen in der MOTA Berücksichtigung. Es gilt $MOTA \leq 1$, wobei negative Werte auftreten können. Die MOTP hingegen gibt die durchschnittliche Abweichung zwischen wahr positiven Antworten und ihren korrespondierenden Ground-Truth-Objekten in Metern an. Die Abweichung eines Spielers $\#i$ zum Zeitpunkt t ist mit $d_{t,i}$ gegeben, die Anzahl der Matches innerhalb des Bilds $\#t$ mit c_t . Alle Distanzen werden euklidisch bestimmt.

Weitere Details zu beiden Größen können in [Leal-Taixé et al., 2015] nachgeschlagen werden.

Abbildung 7.8 gibt einen ausführlichen Überblick über die Abweichungen zwischen beiden Systemen für verschiedene Suchfenster. Bei der Wahl eines 3 m-Fensters wird schließlich ein F1-Score nahe 1 erreicht, wobei die MOTP mit etwa 0,7 m auf eine hohe Genauigkeit des monokularen Systems hinweist. Bei einer Fenstergröße von 2 m liegt der geringste F1-Score aller vier Videos bei 83 %, erreicht im Video FC Augsburg - Paderborn. Dies teilt sich auf in 81 % Recall und 85 % Precision. Diesen hohen und ausgeglichenen Anteilen zwischen ausgewählten und relevanten Entitäten steht mit 0,7 m eine geringe Abweichung der erkannten Objekte gegenüber. Im Fall des 2 m-Schwellwerts ist dies zugleich die höchste durchschnittliche Abweichung. Da die MOTP für die Qualität der Kameraparameter-Nachführung von besonderer Bedeutung ist, werden die Werte in Tabelle 7.5 gesondert wiedergegeben.

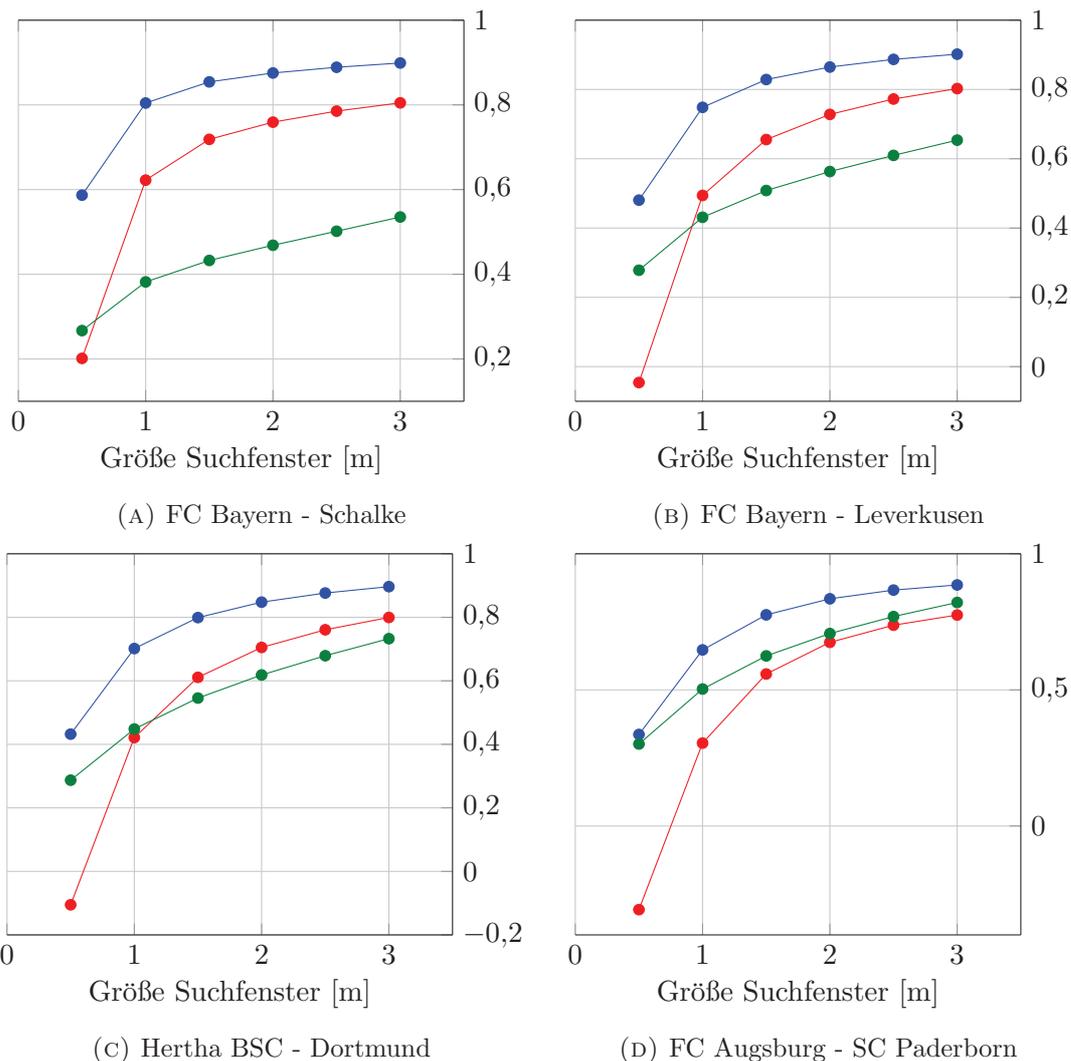


ABBILDUNG 7.8: Abweichungen zwischen den Messungen und der Ground Truth in den Testvideos, jeweils betrachtet in Weltkoordinaten. Dargestellt werden F1-Score (blau), MOTA (rot) und MOTP (grün, in Meter)

Video \ Radius [m]	0,5	1,0	1,5	2,0	2,5	3,0
FC Bayern - Schalke	0,2669	0,3819	0,4326	0,4685	0,5016	0,5352
FC Bayern - Leverkusen	0,2783	0,4311	0,5083	0,5631	0,6097	0,6537
Hertha BSC - Dortmund	0,2871	0,4482	0,5457	0,6184	0,6790	0,7327
FC Augsburg - SC Paderborn	0,3012	0,5032	0,6248	0,7070	0,7694	0,8211

TABELLE 7.5: Durchschnittlicher Abstand (MOTP) in Meter

Tabelle 7.5 ermöglicht eine quantitative Beurteilung der Kameraparameter-Nachführung. Die Aussagekraft dieser Tabelle schränkt sich dadurch ein, dass die MOTP nur

wahr positive Erkennungen innerhalb des Suchfensters berücksichtigt. Innerhalb dieses Szenarios werden etwa im Video FC Bayern - Schalke mit 0,54 m MOTP bei 3 m-Schwellwert sehr geringe Abweichungen erreicht. Der Worst Case tritt im Video FC Augsburg - SC Paderborn mit einer Abweichung von 0,82 m ein. Auch bei Abweichungen dieser Größenordnung disqualifiziert sich das System nicht für übliche Anwendungen wie die Erfassung der Spielerlaufeistung, da gewisse Abweichungen auch in kommerziell eingesetzten Systemen bekannt sind und toleriert werden.

Einschränkend sollte auch die MOTA betrachtet werden: Für den 2 m-Radius bewegt sich diese zwischen 67 % und 76 %. Dies lässt sich zum einen mit Objekten erklären, welche durch das monokulare System nicht oder falsch getrackt wurden oder deren IDs vertauscht wurden. Ebenso kann dies auf eine ungenaue Kameraparameter-Nachführung zurückgeführt werden, wodurch sich zumeist alle Spielerpositionen außerhalb ihrer Suchfenster verschieben. Die bereitstehenden Ground-Truth-Daten können aus den in Abschnitt 7.3 genannten Gründen nicht im engeren Sinn als Ground Truth betrachtet werden. Fehler in diesen Daten können ebenso nicht ausgeschlossen werden.

7.5 Stochastische Glättung

Die Fehler, welche die CIP-Positionsdaten betreffen, wurden thematisiert. Eine erste Behandlung umfasst die im Folgenden dargelegten zwei Schritte.

Da Sprünge in der Position sehr häufig in Verbindung mit jedem der drei Fehlertypen stehen, werden sie durch einen Schwellwert erkannt: Überschreiten ein Mensch oder der Ball die physikalisch mögliche Geschwindigkeit entsprechend ihrer Objektklasse, wird die Folge der Objektpositionen an dieser Stelle durchtrennt. Entstehen dadurch sehr kurze Teilfolgen von Objektpositionen, werden diese verworfen. Die übrigen Folgen werden im zweiten Schritt weiter betrachtet.

Sei eine Folge $\mathbf{x}_{1:n}$ ohne Sprünge gegeben. Es wird nun davon ausgegangen, dass die Vektoren \mathbf{x}_k , $k \in \{1, \dots, n\}$, eine Störung durch weißes Gaußsches Rauschen $\mathbf{r}_k \sim \mathcal{N}(0, \mathbf{R})$ beinhalten.

$$x_{1,k} := y_{1,k} + r_{1,k} \tag{7.3}$$

$$x_{2,k} := y_{2,k} + r_{2,k} \tag{7.4}$$

Um Position und Geschwindigkeit in einem linearen System zu modellieren, werden die Zustandsvektoren \mathbf{y}_k als Quadrupel jener Größen notiert. Mit $\Delta t = fps^{-1}$ und dem

zeitlich unkorrelierten Prozessrauschen $\mathbf{q}_{k-1} \sim \mathcal{N}(0, \mathbf{Q})$ kann das dynamische Modell

$$\begin{pmatrix} y_{1,k} \\ y_{2,k} \\ y_{3,k} \\ y_{4,k} \end{pmatrix} := \underbrace{\begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{=: \mathbf{A}} \mathbf{y}_{k-1} + \mathbf{q}_{k-1} \quad (7.5)$$

notiert werden. Für die Messungen $\mathbf{x}_{1:n}$ gilt dann:

$$\mathbf{x}_k = \mathbf{H}\mathbf{y}_k + \mathbf{r}_k, \quad \text{mit } \mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad (7.6)$$

Innerhalb dieses linearen Modells werden auftretende Beschleunigungen durch den Rauschterm \mathbf{q}_{k-1} beschrieben. Während mit \mathbf{r}_k der Fehler des Messsystems modelliert wird, wird durch \mathbf{q}_{k-1} der Modellfehler beschrieben.

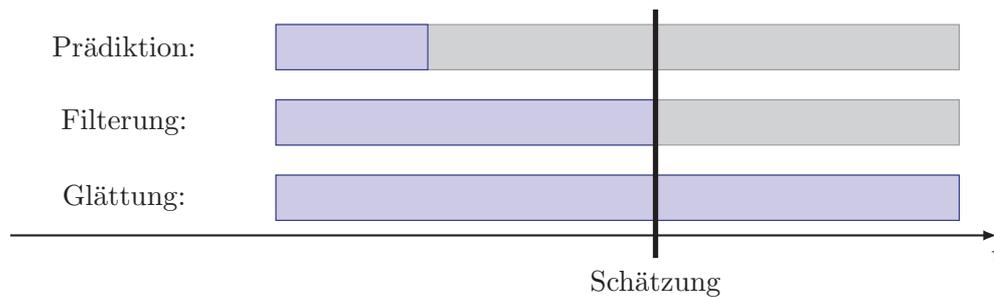


ABBILDUNG 7.9: Basierend auf den Zeitfenstern (blau), für welche Messungen zur Verfügung stehen, können stochastische Zustandsschätzer in Prädiktion, Filterung und Glättung eingeteilt werden, je nachdem ob die Schätzung für einen zukünftigen, aktuellen oder vergangenen Zeitpunkt erfolgen soll. Vgl.: [Särkkä, 2013, Figure 1.7]

Im Gegensatz zu einem Kalman-Filter [Kalman, 1960], bei welchem keine Erkenntnis über zukünftige Positionen vorliegt, ist dies bei Positionsdaten gegeben (vgl. Abbildung 7.9). Statt auf eine Filterung kann somit auf eine stochastische Glättung zurückgegriffen werden. Die optimale Lösung kann aus dem Satz von Bayes auf den gesamten Stapel $\mathbf{x}_{1:n}$ übertragen werden und ist als Rauch-Tung-Striebel-Filter bekannt. Hierbei wird der Stapel in einem Vorwärts- und Rückwärtsschritt insgesamt zweimal durchlaufen. Für die Vorwärts-Rekursion werden berechnet:

Prädiktion:

$$\text{Mittelwert: } \mathbf{m}_k^- = \mathbf{A}\mathbf{m}_{k-1} \quad (7.7)$$

$$\text{Kovarianz: } \mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q} \quad (7.8)$$

Korrektur:

$$\text{Residualkovarianz: } \mathbf{S}_k = \mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R} \quad (7.9)$$

$$\text{Kalman-Matrix: } \mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}^T\mathbf{S}_k^{-1} \quad (7.10)$$

$$\text{Mittelwert: } \mathbf{m}_k = \mathbf{m}_k^- + \mathbf{K}_k(\mathbf{x}_k - \mathbf{H}\mathbf{m}_k^-) \quad (7.11)$$

$$\text{Kovarianz: } \mathbf{P}_k = \mathbf{P}_k^- + \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T \quad (7.12)$$

Die Gleichungen der anschließenden Rückwärts-Rekursion lauten:

Prädiktion:

$$\text{Mittelwert: } \mathbf{m}_{k+1}^- = \mathbf{A}\mathbf{m}_k \quad (7.13)$$

$$\text{Kovarianz: } \mathbf{P}_{k+1}^- = \mathbf{A}\mathbf{P}_k\mathbf{A}^T + \mathbf{Q} \quad (7.14)$$

Korrektur:

$$\text{Kalman-Matrix: } \mathbf{G}_k = \mathbf{P}_k\mathbf{A}(\mathbf{P}_{k+1}^-)^{-1} \quad (7.15)$$

$$\text{Mittelwert: } \mathbf{m}_k^s = \mathbf{m}_k + \mathbf{G}_k(\mathbf{m}_{k+1}^s - \mathbf{m}_{k+1}^-) \quad (7.16)$$

$$\text{Kovarianz: } \mathbf{P}_k^s = \mathbf{P}_k + \mathbf{G}_k(\mathbf{P}_{k+1}^s - \mathbf{P}_{k+1}^-)\mathbf{G}_k^T \quad (7.17)$$

Die Herleitung dieser Gleichungen kann in [Särkkä, 2013] Kapitel 3, 4 und 8 nachvollzogen werden. Für die *a priori* festzulegenden Kovarianzmatrizen gilt:

$$\mathbf{Q} = \begin{pmatrix} \frac{q\Delta t^3}{3} & 0 & \frac{q\Delta t^2}{2} & 0 \\ 0 & \frac{q\Delta t^3}{3} & 0 & \frac{q\Delta t^2}{2} \\ \frac{q\Delta t^2}{2} & 0 & q\Delta t & 0 \\ 0 & \frac{q\Delta t^2}{2} & 0 & q\Delta t \end{pmatrix} \quad \mathbf{R} = \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \end{pmatrix} \quad (7.18)$$

Die Varianz r^2 des Messrauschens wird genauso wie die spektrale Dichte q des Prozessrauschens unabhängig von den Koordinatenachsen betrachtet. Die Werte wurden im Weiteren entsprechend physikalisch sinnvoller Rahmenbedingungen gewählt. Stochastisch geglättete Größen werden in den folgenden Termen durch einen Unterstrich gekennzeichnet, Ableitungen (der Positionsdaten) — durch Punkte.

$$\begin{aligned} \underline{x}_{1,k} &:= m_{1,k}^s & \dot{\underline{x}}_{1,k} &:= m_{3,k}^s \\ \underline{x}_{2,k} &:= m_{2,k}^s & \dot{\underline{x}}_{2,k} &:= m_{4,k}^s \end{aligned} \quad (7.19)$$

Wird auch die zweite Ableitung benötigt, wird diese durch den zentralen Differenzenquotienten bereitgestellt.

$$\ddot{\mathbf{x}}_k := \frac{1}{2\Delta t} (\dot{\mathbf{x}}_{k+1} - \dot{\mathbf{x}}_{k-1}) \quad (7.20)$$

Generell sind für alle Gleichungen in diesem Abschnitt die Ränder $k = 1$ und $k = n$ gesondert zu betrachten.

7.6 Diskussion und Ausblick

Nachdem in diesem Kapitel Verfahren zur Ball- und insbesondere Spieler-Nachführung in monokularen Daten eingeführt wurden, konnte das vollständige System aus Kameraparameter- und Spieler-Tracking mit einem CIP-System verglichen werden. Hinsichtlich der Distanzen zwischen den Messungen beider Systeme für mutmaßlich einen Spieler konnte dabei gezeigt werden, dass diese für die gewählten Testvideos innerhalb praxistauglicher Größenordnungen liegen. Mit einer Worst-Case-Durchschnittsdistanz von 0,82 m kann von einer hohen Genauigkeit für ein monokulares System ausgegangen werden. Andererseits wird auch eine Reihe von Objekten nicht erkannt (MOTA zwischen 64 % und 76 % im 2 m-Radius). Obwohl gezeigt wurde, dass auch die Daten eines kommerziellen Anbieters nicht frei von Fehlern sind, kann dies nicht als einzige Fehlerursache angesehen werden. Neben Fehlern im Spieler-Tracking sind auch die bereits angesprochenen Fehler in der Kameraparameter-Nachführung zu berücksichtigen. Eine Ungenauigkeit oder ungeeignete Kalibrierung in diesem Verfahren führt in aller Regel zu einem Verlust der Nachführung einer großen Anzahl von Spielern (in Weltkoordinaten) oder einem Wechsel von Tracker IDs.

Neben dem bereits vorgestellten Rauch-Tung-Striebel-Filter und der Sprungerkennung in Trajektorien sind weitere Verfahren für die Fehlererkennung und -bereinigung der CIP-Daten denkbar. Für die aus der Sprungerkennung resultierenden Abschnitte gilt es, eine Klassifikation zu konstruieren, welche es ermöglicht, für Trajektorien beliebiger Länge zu entscheiden, ob besagte Punktfolge aus einem Fehler hervorgegangen ist oder die tatsächliche Bewegung eines Spielers abbildet. Aus dem Wissen um die Position innerhalb der Formation einer Mannschaft und die Bewegung der anderen Spieler ließe sich ein geeigneter Test formulieren.

Kapitel 8

Ballbesitz und Spielereignisse

Dieses Kapitel basiert methodisch u. a. auf:

HOERNIG, Martin ; LINK, Daniel ; HERRMANN, Michael ; RADIG, Bernd ; LAMES, Martin: Detection of Individual Ball Possession in Soccer. In: *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)* 392. Springer, 2016, S. 103-107

8.1 Einleitung

Die in Weltkoordinaten zur Verfügung stehenden Positionsdaten werden in diesem Kapitel genutzt, um Ballbesitzinformationen, Statistiken, visuelle Spielauswertungen und Spielereignisdaten zu generieren. Zuerst werden neue individuelle und mannschaftsbezogene Ballbesitzmodelle eingeführt, welche sich im Grad der Ballkontrolle differenzieren. Darauf aufbauend werden Algorithmen aufgestellt, welche die Ballbesitzmodelle funktional umsetzen. Auf diese Weise werden, soweit bekannt, zum ersten Mal im Rahmen großer Datenerhebungen spielerindividuelle Ballbesitzinformationen bereitgestellt. Die manuelle Erhebung des spielerbezogenen Ballbesitzes wird aufgrund zu hoher Kosten durch *Content Information Provider (CIPs)* nicht durchgeführt. Anzumerken ist, dass innerhalb von CIP-Tracking-Systemen durchaus auch der Ballbesitz einzelnen Spielern zugewiesen werden kann. Da Objekte



bei großer Nähe oft schwer zu unterscheiden sind, wird mit dem Wissen um die Ballführung eines einzelnen Spielers bei gleichzeitiger Nähe beider Objekte die Position des Balls an die des besagten Spielers angeglichen, um falsche Antworten zu unterdrücken. Die gewonnenen Erkenntnisse um den Ballbesitz werden im Anschluss evaluiert und für eine beispielhafte Begegnung auf Spieler- und Mannschaftsebene ausgewertet. Die Ergebnisse werden grafisch durch Heatmaps unterstrichen.

Spielereignisse wie Pässe, Einwürfe oder Zweikämpfe werden aus den erhobenen Ballbesitzinformationen bestimmt. Sie gehen nach ihrer Erhebung in ein Spieljournal ein, also in eine Beschreibung einer Begegnung durch die Geschehnisse, welche sich in selbiger zugetragen haben. Auch wenn es hierbei prinzipiell keine Rolle spielt, von welchem der Tracking-Systeme die Daten stammen (von CIP-Fußball-Tracking-Systemen oder den in Kapitel 6 und 7.2.1 beschriebenen monokularen Trackern), wird in den Anwendungen auf ein CIP-System zurückgegriffen. Bei dieser Wahl sind verlässlichere Ballpositionsdaten verfügbar, welche sowohl für den Ballbesitz als auch für die darauf beruhenden Spielereignisse von entscheidender Bedeutung sind.

8.2 Stand der Forschung

Da CIP-Systeme zur Positionserfassung erst in den letzten Jahren weite Verbreitung in den Top-Ligen gefunden haben, stehen frühere Veröffentlichungen zur semantischen Fußballanalyse üblicherweise in direkter Verbindung mit Verfahren der Bildverarbeitung. Hier werden entweder Positionsdaten für eine weiterführende Auswertung aus Bildinhalten abgeleitet [Yu et al., 2005] oder Bildinhalte werden direkt mit Modellen verglichen, um eine Bedeutung mit den Bildinhalten assoziieren zu können [Assfalg et al., 2003, Ekin et al., 2003, Luo et al., 2003]. Das Ziel hierbei ist es, bedeutende Spielereignisse wie Tore oder Verwarnungen in einem Video erkennen zu können. Einen umgekehrten Ansatz hierzu verfolgen [Wang et al., 2014]. Hier steht nicht die Erkennung einer taktischen Struktur im Vordergrund, sondern durch Berechnungen auf Spieler- und Ballpositionen aus einem Spielertracking wird die fehleranfällige Bestimmung der Ballposition eingeschränkt. Ausgangsbasis dabei ist ein an Ball-Spieler-Distanzen gebundener Ballbesitz. In einer frühen Publikation auf Basis von Positionsdaten beschreiben [Tovinkere u. Qian, 2001] die Zerlegung eines Spiels in seine relevanten Ereignisse. Dieser Ansatz gliedert das Spiel in ein hierarchisches Entity-Relationship-Model. Hierauf aufbauend werden Kausalketten konstruiert. Voraussetzung für ein gefallenes Tor ist etwa, dass ein Schuss, Strafstoß, Freistoß, Eckstoß oder dergleichen erfolgte. Auf ein Tor folgt wiederum ein Anstoß. Die Datensätze zur Evaluation umfassen jedoch nur wenige Sekunden. Mit breiterer Verfügbarkeit von Positionsdaten erhöhte sich die Zahl der Publikationen gerade

in den letzten Jahren. In [Kang et al., 2006] wird ähnlich zu dem im Weiteren vorgestellten Verfahren ein Ballbesitz bestimmt. Die Grundlage bilden hier wie im Folgenden Ballberührungen, welche durch eine Unterschreitung eines Schwellwerts für den Abstand zwischen einem Spieler und dem Ball definiert sind. Darauf aufbauend werden Ballbesitz und Handlungsalternativen von Spielern formuliert. Die Verfahren unterscheiden sich jedoch bereits im Ansatz, da besagte Publikation mit der Trajektorienanalyse schließt, während im Folgenden ein sportpraktischer Hintergrund sowie Grade der Ballkontrolle und Anwendungen des Ballbesitzes zentrale Eckpfeiler bilden. [Kang et al., 2006] untersuchen unter allgemeineren Aspekten Spielertrajektorien auf simulierten Daten.

Eine Erkennung von taktischen Strukturen verfolgen sowohl [Kim et al., 2011] als auch [Grunz et al., 2012]. Beide haben das Ziel, Aussagen über Spieler aus den Positionen von kleinen Spielergruppen zu gewinnen. Während bei [Kim et al., 2011] hierfür morphologische Maße für Spielerketten definiert und berechnet werden, nutzen [Grunz et al., 2012] selbstorganisierende Karten, um das Verhalten von kleinen Spielergruppen in trainierten Situationen, etwa einer Spieleröffnung, zu klassifizieren. Ein Verfahren, welches das Rollenverhalten einzelner Spieler und gesamter Mannschaften adaptiv handhaben kann, stellen [Bialkowski et al., 2014] vor. Der vorgestellte Ansatz zur Formationserkennung erlaubt es über die gesamte Spieldauer, Mannschaften automatisch Formationen zuzuordnen. Dazu wird die Entropie einer Menge von Spielerrollenverteilungen minimiert. Eine Untersuchung zu Auffälligkeiten in Spielerpositionen führen [Moura et al., 2012] durch. Auf zehn Datensätzen der ersten brasilianischen Liga werden die von den Spielern abgedeckte Fläche des Spielfelds, Abstände zwischen den Spielern und Geschwindigkeiten von Spielern innerhalb ihrer Formationen betrachtet.

Insgesamt lässt sich feststellen, dass bislang keine Arbeiten zur robusten Erkennung des individuellen Ballbesitzes auf Basis realer Positionsdaten vorliegen.

8.3 Einführung spielerbezogener Ballbesitzmodelle

Der Ballbesitz ist aus sportwissenschaftlicher Sicht ein relevanter und der am häufigsten untersuchte Leistungsindikator im Fußball [Mackenzie u. Cushion, 2013]. In den bisherigen Veröffentlichungen in diesem Kontext, etwa bei [Jones et al., 2004] oder [Hughes u. Franks, 2005], wird jedoch ausschließlich auf mannschaftsbezogenen Ballbesitz Bezug genommen, welcher auch durch die am Markt etablierten CIPs bereitgestellt wird. Dieser wird durch Scouts am Spielfeldrand während des Spiels manuell erhoben. Ein entsprechend größerer Personalaufwand und die damit einhergehenden Kosten behindern die Erfassung von Daten auf Spielerebene. Ein direkter Vorteil aus der automatischen Erhebung besteht in der inhärenten Kombination mit Positionsdaten, wodurch zusätzlich

zu den Zeitpunkten auch die Positionen der ballführenden Spieler erfasst werden. Sich hieraus ergebende Möglichkeiten werden in den Abschnitten 8.3.7 und 8.5 vorgestellt. Da einvernehmliche Definitionen nur für Mannschaftsballbesitz existieren, bleibt die Frage zu beantworten, wann für einen Spieler der Ballbesitz beginnt und wann dieser endet.



ABBILDUNG 8.1: Verschiedene Grade an Ballkontrolle. Während der Ball in der Abwehrreihe zirkuliert, bestehen verschiedene Handlungsalternativen (A), während eines Kopfballduells im Mittelfeld bestehen diese zumeist nicht (B).

In Abbildung 8.1 werden zwei Situationen gegenübergestellt, welche den für diesen Abschnitt maßgeblichen Begriff einführen: die Ballkontrolle. Während beim klassischen mannschaftsbasierten Verständnis für Ballbesitz zu jedem Zeitpunkt, in dem das Spiel nicht unterbrochen ist, einer Mannschaft der Ballbesitz zugeordnet wird, wird dies im Weiteren nicht nur auf Spielerebene konkretisiert — es werden auch Modelle definiert, welche keine Zuordnung des Ballbesitzes zulassen und damit einem Spieler nur dann Ballbesitz zuweisen, wenn dieser die Bahn des Ball tatsächlich beeinflussen oder (bewusst) zwischen verschiedenen Handlungsalternativen wählen kann (siehe Abbildung 8.1).

Der Ballbesitz wird unabhängig von der Art der Definition nur erhoben, während der Ball im Spiel ist (d. h. während das Spiel läuft). Innerhalb dieser zeitlichen Grenzen gilt es den Ballbesitz bzw. die Grenzen für Beginn- und Endzeitpunkte des Ballbesitzes zu bestimmen. Hierfür werden die in Kapitel 7.5 eingeführten stochastisch geglätteten Größen verwendet. Damit ist die Ballposition durch die Folge $\underline{b}_{1:n}$ bestimmt und die Positionen der auf dem Feld vertretenen Spieler werden durch $\underline{x}_{1:n}^i$ notiert. Auch hierbei werden Ball- und Spielerpositionen wiederum in \mathbb{R}^2 ausgedrückt. Die Folge der Spielerpositionen ist fortlaufend mit Spieleridentifikationen indiziert und die n Bilder entstammen einer äquidistanten Abtastung (mit *fps* Bildern pro Sekunde).

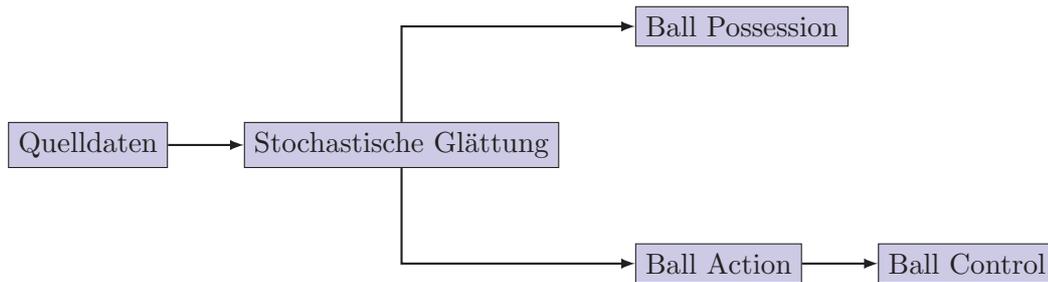


ABBILDUNG 8.2: Verarbeitungsschritte zur Erkennung der drei Ballbesitztypen

In den nächsten Abschnitten werden verschiedene Ballbesitzmodelle eingeführt und algorithmisch auf Grundlage der Positionsdaten umgesetzt. Anschließend werden die Algorithmen evaluiert und Anwendungen auf den Ballbesitzdaten vorgestellt. Abbildung 8.2 zeigt den Weg dorthin: Zuerst wird die Ball Possession eingeführt, welche keine Zeitpunkte ohne Ballbesitz kennt und damit das einfachste Modell darstellt. Als zweiter Pfad wird die Ball Action eine Möglichkeit bieten, den Ballbesitz auf Intervalle einzuschränken, in welchen eine physikalische Interaktion möglich ist. Spieler können hier den Ballbesitz also auch verlieren, indem sie sich den Ball zu weit vorlegen oder ihn flanken. Hierauf aufbauend wird die Ball Control definiert, in welcher Situationen wie in Abbildung 8.1 unterschieden werden. Die Unterscheidung auf Grundlage einer Distanz kann für diesen Ballbesitz nicht aussagekräftig sein, da der Spieler auch im Stande sein muss, bewusst die Bahn des Balls zu kontrollieren.

Die Algorithmen zu den Ballbesitzmodellen werden zumeist auf Berechnungen von Spieler-Ball-Distanzen beruhen und *a priori* trainierte Schwellwerte verwenden. Als Distanz d kommt die euklidische Metrik zur Anwendung:

$$d(\mathbf{a}, \mathbf{b}) := \|\mathbf{a} - \mathbf{b}\|_2 \quad (8.1)$$

8.3.1 Ball Possession

Das Ball-Possession-Modell verfolgt den Gedanken, zu jedem Zeitpunkt den Ballbesitz einem Spieler zuweisen zu können, und stimmt demzufolge mit den gängigen Modellen der CIPs überein. Dabei bleibt ein Spieler so lange in Ballbesitz, bis ein neuer Spieler „am Ball“ ist. Im Fall einer Flanke heißt dies beispielsweise, dass der Flankengeber während der Flugphase des Balls in Ballbesitz bleibt. Damit dient dieses Modell auch als Zerlegung der Balltrajektorie in Spieler- und Mannschaftsanteile.

Definition 8.1. Die *Individual Ball Possession (IBP)* eines Spielers beginnt, sobald dieser Spieler erstmals (nach einer IBP eines anderen Spielers oder einer Spielunterbrechung) eine Aktion mit dem Ball durchführen kann. Sie endet, sobald eine IBP für einen anderen Spieler beginnt.

Um den Zeitpunkt der Interaktion mit dem Ball zu bestimmen, werden zwei Methoden vorgestellt. Zuerst wird ein einfacher Distanztest Verwendung finden. Ein Spieler $\#i$ erhält Ballbesitz, sobald die Distanz zwischen Spieler und Ball unter einen gewissen Schwellwert fällt und kein anderer Spieler näher am Ball ist (siehe Abbildung 8.3 A).

$$d(\underline{\mathbf{x}}_t^i, \underline{\mathbf{b}}_t) < T_p \quad \wedge \quad \forall j \neq i: d(\underline{\mathbf{x}}_t^j, \underline{\mathbf{b}}_t) > d(\underline{\mathbf{x}}_t^i, \underline{\mathbf{b}}_t) \quad (8.2)$$

Der Wert T_p beschreibt den Bereich, bis zu welchem eine Interaktion mit dem Ball möglich ist, und muss *a priori* trainiert werden (siehe Abschnitt 8.3.6.1). Diese Methode wird im Weiteren als *Naïve Physical* bezeichnet.

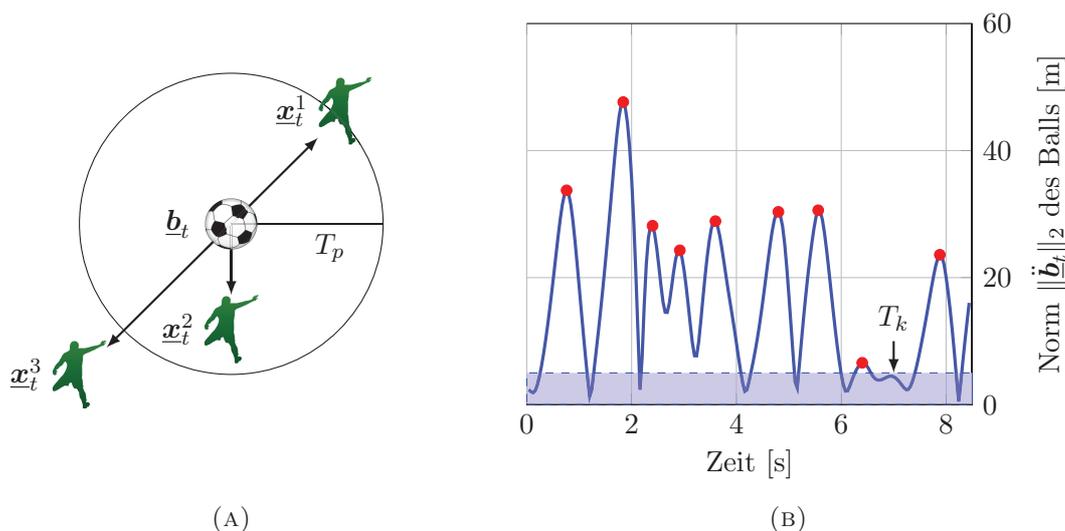


ABBILDUNG 8.3: Bestimmung des Beginns einer Ballbesitzphase (IBP). Im Naïve-Physical-Verfahren beginnt der Ballbesitz, sobald ein Ball-Spieler-Abstand den Schwellwert T_p unterschreitet und kein anderer Spieler näher am Ball ist (A). Für die Kick Detection muss zusätzlich ein lokales Maximum (rot dargestellt) in der Norm der Ballbeschleunigung vorliegen (B).

Wie eingangs festgelegt, wird auch für die Ballposition keine Höheninformation vorausgesetzt. Falls der Ball, etwa im Fall einer Flanke, die Position eines Spielers in großer Höhe kreuzt, so wird dies im Naïve-Physical-Verfahren nicht unterschieden und der Spieler erlangt dennoch Ballbesitz. Dieses Problem wird durch die Einbeziehung einer *Kick Detection* bearbeitet. Hierbei werden Ballbesitzwechsel nur zu Zeitpunkten in Betracht gezogen, für welche ein lokales Maximum in der Norm der Ballbeschleunigungen vorliegt (siehe Abbildung 8.3, B).

$$\text{lokales Maximum zum Zeitpunkt } t \Leftrightarrow \|\ddot{\mathbf{b}}_{t-1}\|_2 \leq \|\ddot{\mathbf{b}}_t\|_2 \geq \|\ddot{\mathbf{b}}_{t+1}\|_2 \quad (8.3)$$

Ein Schwellwert T_k geringer Größe (genutzt wurden 4 ms^{-2}) wird zusätzlich verwendet, um Antworten bei geringen Geschwindigkeitsveränderungen zu unterdrücken. Ballbeschleunigungen im Fußball können durchaus $3\,000 \text{ ms}^{-2}$ erreichen [Shinkai et al., 2009]. Nur wenn das Kriterium aus Gleichung (8.3) erfüllt ist und *gleichzeitig* die Kriterien des Naïve-Physical-Verfahrens erfüllt werden, erlangt ein neuer Spieler Ballbesitz.

8.3.2 Ball Action

Von dem grundlegenden Ballbesitz grenzt sich die Ball Action dadurch ab, dass hier auch Zeiträume im laufenden Spiel existieren können, in welchen keinem Spieler Ballbesitz zugeordnet ist. Ein Spieler bleibt nur so lange in Ballbesitz, wie er in der Lage ist, taktische Aktionen mit dem Ball durchzuführen. Diese Zeiträume besitzen eine hohe Relevanz, da sie innerhalb der Nettospielzeit diejenigen Zeitpunkte zusammenfassen, in welchen tatsächlich ein Einfluss auf den Ball ausgeübt werden kann.

Definition 8.2. Die *Individual Ball Action (IBA)* eines Spielers beginnt, sobald dieser Spieler eine Aktion mit dem Ball durchführen kann und unmittelbar zuvor keine IBA hatte. Sie endet, sobald für den Spieler keine Aktion mit dem Ball mehr möglich ist.

Die Identifizierung der Zeitpunkte für den Beginn einer IBA kann analog zur IBP erfolgen und wird übernommen. Da in diesem Modell der Ballbesitz jedoch nicht ausschließlich mit einer Spielunterbrechung oder dem Ballbesitz eines anderen Spielers enden muss, gilt es an dieser Stelle einen geeigneten Test zu formulieren, um für jeden Zeitpunkt überprüfen zu können, ob der sich in Ballbesitz befindliche Spieler weiterhin eine Aktion mit dem Ball durchführen kann.

Wieder werden zwei Ansätze vorgestellt. In beiden werden die aktuellen Positionen und Geschwindigkeiten verwendet, um eine Aussage über zukünftige Positionen zu treffen. Die Schätzung zukünftiger Positionen wird genutzt, um zu betrachten, ob sich dem ballführenden Spieler auch in naher Zukunft noch immer die Möglichkeit der Ballinteraktion bieten wird. Ein Spieler könnte sich den Ball etwa im Dribbeln vorlegen und dabei in Ballbesitz bleiben. Der Zeitraum der Vorhersage wird mit 1 s festgelegt. Weiteres hierzu kann in Abschnitt 8.5 vorgefunden werden.

Über Positions- und Geschwindigkeitsvektoren werden lineare Vorhersagen für die Einschätzung der nahen Zukunft extrapoliert. Die Bewegungsgeschwindigkeit und -richtung geht direkt in die Bestimmung des Ballbesitzes ein. Im ersten Verfahren zur Bestimmung der Endpunkte eines Ballbesitzes wird die Betrachtung zuerst auf den Ball eingegrenzt.

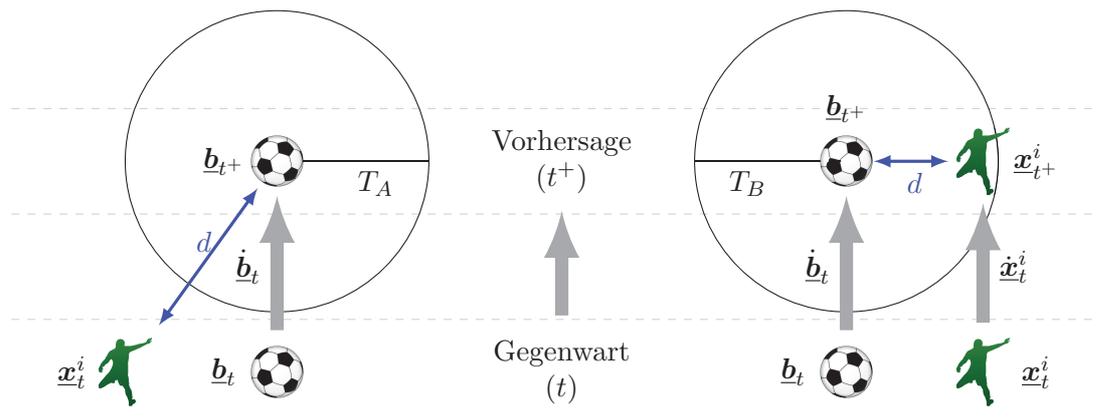


ABBILDUNG 8.4: Test auf Ballinteraktionsvermögen des Ballführers \underline{x}_t^i auf Grundlage der Positionen \underline{b}_t und \underline{x}_t^i und Geschwindigkeiten $\dot{\underline{b}}_t$ und $\dot{\underline{x}}_t^i$. Während in der Ball Prediction nur die Ballposition vorhergesagt wird (links), wird in der Player Prediction auch die Spielerposition geschätzt (rechts).

Voraussagen werden im Folgenden mit einem hochgestellten + an der Position des Zeitindex versehen.

$$\underline{b}_{t+} := \underline{b}_t + \dot{\underline{b}}_t * [1 \text{ s}] \quad (8.4)$$

Der Ballbesitz für den ballführenden Spieler $\#i$ endet (neben den bisher genannten Bedingungen: Spielunterbrechungen und Ballbesitz eines anderen Spielers) auch, wenn

$$d(\underline{x}_t^i, \underline{b}_{t+}) > T_a \quad (8.5)$$

erfüllt ist (siehe Abbildung 8.4, links). Das heißt, wenn die zukünftige Ballposition sich zu weit von der jetzigen Spielerposition entfernt. Mit T_a wird auch an dieser Stelle wieder ein Schwellwert eingeführt, welchen es *a priori* festzulegen gilt. Dieses Modell der Ball Prediction wird durch Hinzunahme der Vorhersage der Spielerposition

$$\underline{x}_{t+}^i := \underline{x}_t^i + \dot{\underline{x}}_t^i * [1 \text{ s}] \quad (8.6)$$

erweitert (siehe Abbildung 8.4, rechts). Damit kann die Bedingung aus Gleichung (8.5) erweitert werden zu

$$d(\underline{x}_{t+}^i, \underline{b}_{t+}) > T_b \quad (8.7)$$

Diese Bedingung korrespondiert entsprechend Abbildung 8.4 zu der Vorhersage von Spieler und Ball und sieht damit das Ende des Ballbesitzes vor, sobald der vorhergesagte Abstand beider Objekte zu groß wird. Dieser Ansatz wird als *Player Prediction* bezeichnet.

8.3.3 Ball Control

Bereits mit dem Übergang von IBP zu IBA wurde eine Aussage über eine Möglichkeit zur Ballkontrolle (-aktion) in die Formulierung einbezogen. Eine weitere Differenzierung innerhalb der IBA erfolgt durch die *Individual Ball Control*. Entscheidendes Kriterium ist, dass ein Spieler die Möglichkeit hat, innerhalb seines (IBA-)Aktionsintervalls den Ball sicher zu kontrollieren, und ihm eine bewusste Entscheidung zwischen mehreren Handlungsalternativen offen steht. In Situationen, wie in Abbildung 8.1 (B) angedeutet, in welcher der Ball durch ein Zuspiel des Torwarts ins Mittelfeld geflankt wurde und der Ballaktionszeitraum auf ein Kopfballduell beschränkt ist, ist dies klar nicht der Fall. Nach einer erfolgten Analyse kann das Spiel zudem weiter zerlegt werden, insbesondere können mannschaftstaktische Episoden der Ballkontrolle im Spiel erkannt werden (mehr zur Erkennung auf mannschaftstaktischer Ebene in Abschnitt 8.3.5).

Definition 8.3. Hat ein Spieler innerhalb einer IBA die Möglichkeit der Entscheidung zwischen verschiedenen Handlungsalternativen, so besitzt der Spieler für den Zeitraum der IBA *Individual Ball Control (IBC)*.

Da die IBC eine Binnendifferenzierung zur IBA darstellt, werden keine weiteren Intervallgrenzen berechnet. Hierfür können Kombinationen aus Naïve Physical und Kick Detection für die Berechnung des Beginns sowie Ball und Player Prediction für die Berechnung des Endzeitpunkts genutzt werden. Als zentrale Frage stellt sich, ob in einem vorliegenden Intervall Ballkontrolle ausgeübt wurde. Die Entscheidung darüber kann in gewissen Fällen Informationen voraussetzen, welche aus Positionsdaten nicht entnommen werden können. Für eine Annäherung an die übrigen Fälle wurde ein Bayes'sches Netz als binärer Klassifikator trainiert (Training in Abschnitt 8.3.6.3). Für die Klassifikation fanden die folgenden Merkmale im Bezug auf ein IBA-Intervall Berücksichtigung:

- Länge des Intervalls
- Durchschnittliche Ballgeschwindigkeit und -beschleunigung (jeweils als Vektor und euklidische Vektornorm)
- Varianz der Ballgeschwindigkeit und -beschleunigung (jeweils als Vektor und euklidische Vektornorm)
- Durchschnittliche Distanz zwischen Ball und ballführendem Spieler
- Anzahl an Gegenspielern innerhalb gewisser Distanzen (0,5 m, 1,0 m, ..., 5,0 m), jeweils im Durchschnitt und summiert über die Zeit

8.3.4 Berechnung der Ballbesitz-Zeiträume

Aus den klassifizierten Zeitpunkten kann sowohl die IBP als auch die IBA eindeutig berechnet werden. Technisch erfolgt die Erkennung von Zustandsübergängen im Ballbesitz auch in den Spielunterbrechungen. Erst nach der Berechnung der Ballbesitzintervalle werden die Phasen ausgeschlossen, in denen der Ball nicht im Spiel ist. Dies erlaubt eine größere Toleranz gegenüber Ungenauigkeiten bzgl. des Running Flag (siehe dazu auch Kapitel 7.3). Im umgekehrten Fall müsste mit Fehlern gerechnet werden, welche dadurch entstehen, dass sich das Running Flag erst kurz nach der tatsächlichen Rückkehr ins laufende Spiel ändert, etwa nach einer getretenen Ecke. Die Kick Detection würde dann keine Antwort mehr liefern und dem ausführenden Spieler könnte kein Ballbesitz mehr zugeordnet werden.

Ein IBP-Intervall beginnt mit den gefundenen Zeitpunkten des Naïve-Physical-Verfahrens bzw. der Kick-Detection. Als Endzeitpunkt ergibt sich der nächste gefundene IBP-Beginn oder eine Spielunterbrechung. IBA-Intervalle werden auf gleiche Weise oder aus Paaren von Beginn- und Endpunkten (unter Berücksichtigung von Spielunterbrechungen) berechnet. Alle IBP-, IBA- und IBC-Intervalle sind mit der Identifikation des ballführenden Spielers markiert.

8.3.5 Mannschaftsballbesitz

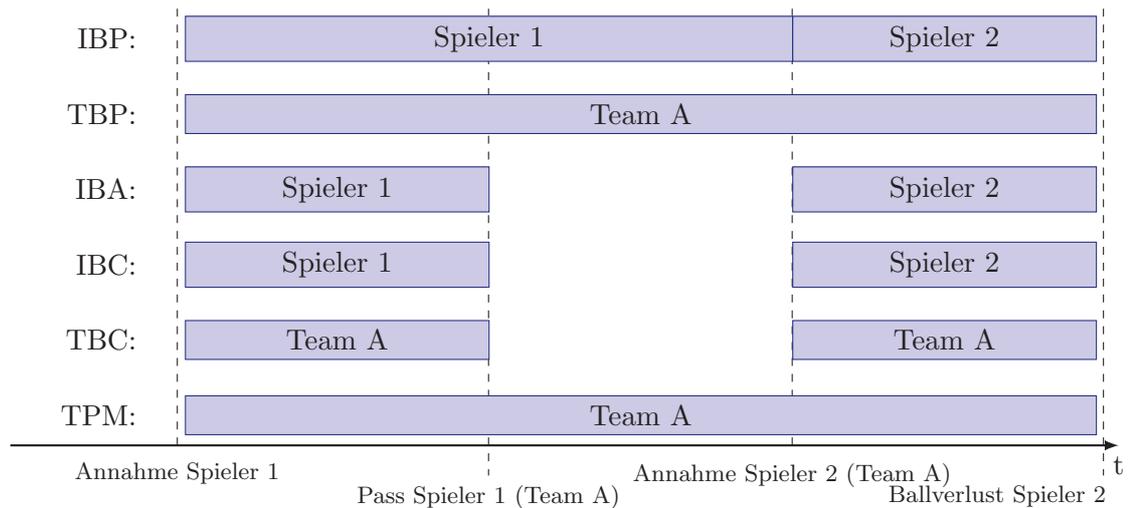
Bis zu diesem Zeitpunkt wurden nur individualtaktische Ballbesitzmodelle eingeführt. Durch eine Interpretation aus mannschaftstaktischer Sicht, mathematisch gesehen durch die Vereinigung von Intervallen, können Aussagen auch auf dieser Ebene getroffen werden. Dies entspricht auch der Ebene, welche durch die CIPs bereitgestellt wird.

Definition 8.4. Die *Team Ball Possession (TBP)* einer Mannschaft beginnt, sobald für einen ihrer Spieler erstmals (nach IBP eines gegnerischen Spielers oder Spielunterbrechung) eine IBP beginnt. Sie endet mit der ersten nachfolgenden IBP für einen Spieler der anderen Mannschaft.

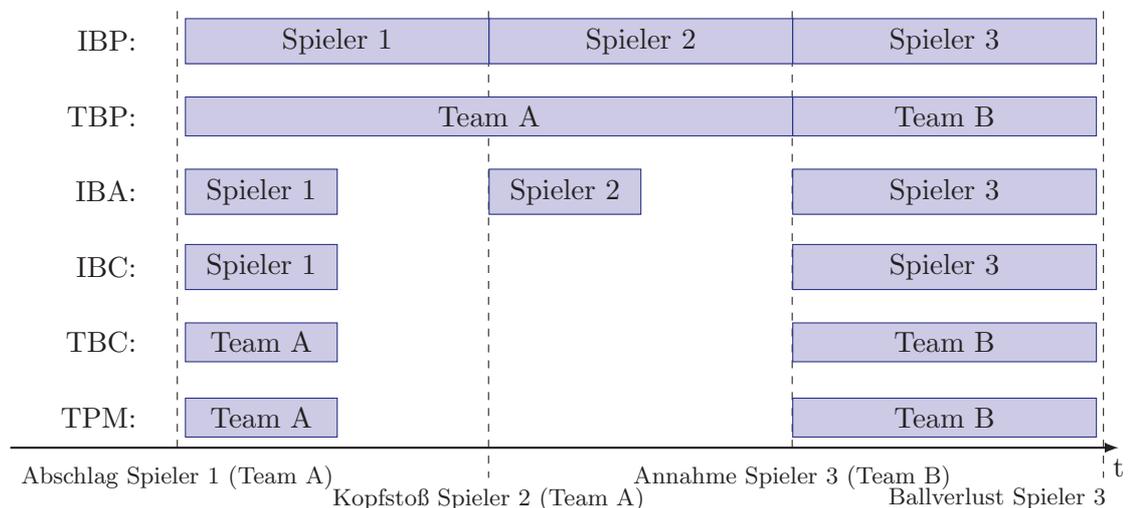
Definition 8.5. Die *Team Ball Control (TBC)* für eine Mannschaft beginnt, sobald für einen Spieler dieser Mannschaft eine IBC beginnt und endet sobald diese IBC endet.

Definition 8.6. Das *Team Playmaking (TPM)* für eine Mannschaft beginnt, sobald für einen Spieler dieser Mannschaft IBC beginnt und diese Mannschaft unmittelbar zuvor keine IBC hatte. Sie endet mit der letzten IBC vor der nächsten IBA eines Spielers der gegnerischen Mannschaft.

Diese Definitionen werden an zwei exemplarischen Situationen näher erklärt.



Situation 1: Annahme des Balls durch Spieler 1, Pass zu Spieler 2, Ballverlust durch Spieler 2



Situation 2: Abstoß Spieler 1, Verlängerung Spieler 2 (Kopfball), Ballverlust zu Spieler 3

ABBILDUNG 8.5: Zwei exemplarische Situationen werden bezüglich der aufgestellten Ballbesitzmodelle ausgewertet. Bildquelle: Daniel Link, Martin Hoernig

In der ersten Situation (Abbildung 8.5, oben) nimmt Spieler 1 den Ball an und führt diesen mit Ballkontrolle, bevor er ihn in einem Passspiel zu Spieler 2 der gleichen Mannschaft abspielt. Damit besteht für den gesamten Zeitraum TPM. Die TBC hingegen umfasst den Zeitraum des Abspiels nicht, da kein Spieler in diesem Zeitraum Ballkontrolle besitzt. Entsprechend ist die TBP und TPM identisch. Die zweite Situation stellt das Konzept der Ballkontrolle heraus. Ein abstoßender Torwart (Spieler 1) hat Ballkontrolle. Sein Abstoß wird jedoch nicht von seinem Mitspieler (Spieler 2) erfolgreich angenommen, welcher im Mittelfeld den Ball mit dem Kopf zu verlängern versucht. Sowohl TBC als auch TPM enden direkt mit dem Ende der IBC von Spieler 1. Nach dem

erfolgten Kopfball erhält das gegnerische Team Ballkontrolle durch Spieler 3, womit auch die TBP von Team A endet.

Die TBP entspricht damit weitestgehend dem eher klassischen Ballbesitzverständnis zwischen zwei Mannschaften und wird aus der Ball Possession bestimmt. Für die Berechnung werden IBP-Spieler-Intervalle miteinander vereinigt, falls die Spieler paarweise zur gleichen Mannschaft gehören, bis nur noch zwei Mengen vorhanden sind. Die IBC hingegen wird, entsprechend Abbildung 8.6, mannschaftstaktisch in zwei Konstrukte überführt. Die TBC stellt die Vereinigung der IBC-Phasen aller Spieler einer Mannschaft dar, während das Team Playmaking (TPM) die Zeiträume umfasst, in welchen eine Mannschaft den Ball unter ihrer Kontrolle halten kann. Sie geht entsprechend dem fachlichen Modell aus zwei neuen Intervallfolgen (eine für jede Mannschaft) hervor. Der Beginn eines solchen Intervalls ist der Beginn der ersten IBC eines Spielers einer Mannschaft nach einer Pause oder dem Ballbesitz einer anderen Mannschaft. Das Ende eben dieser Folge ist das Ende der letzten IBC der zugehörigen TBP-Phase.

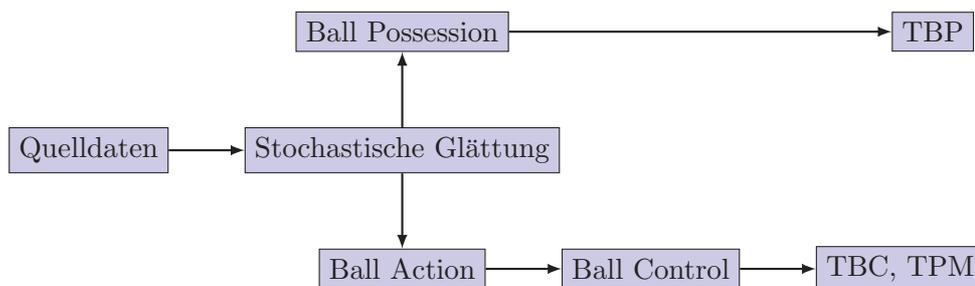


ABBILDUNG 8.6: Verarbeitungsschritte zur Erkennung des Mannschaftsballbesitzes

8.3.6 Evaluation und Training

Als Trainings- und Testmenge wurde wiederum die bereits in Kapitel 7.3 besprochene Begegnung zweier Erstligamannschaften (Heim und Gast) verwendet. Für dieses Spiel wurden IBA- und IBC-Referenzdaten durch einen Sportwissenschaftler manuell unter Verwendung einer Aufzeichnung des Spiels erhoben. Für die Annotation stand ausreichend Zeit und das Panorama-Video (mit Einblendung des bildgenauen Timecodes) zur Verfügung. Insgesamt wurden 1 369 IBA-Phasen annotiert, wobei in 1 264 dieser Phasen Ballkontrolle vorlag. Die IBP wurde im Nachgang aus der IBA berechnet. Eine gesonderte Annotation (und Evaluation) der mannschaftsspezifischen Konstrukte geschah nicht, da diese Daten aus den spielerindividuellen Daten berechenbar sind.

8.3.6.1 Bestimmung der Modellparameter

Durch die Annotation eines Trainingsdatensatzes werden die Modellschwellwerte T_p , T_k , T_a und T_b individuell kalibriert. Eine vollständige Gridsuche [Powell, 1998] zur Bestimmung des globalen Optimums ist an dieser Stelle ausreichend. Dafür wurden die Eingangsdaten auf die fehlerfreien Bereiche (siehe Kapitel 7.3) eingegrenzt.

Da die Schwellwerte in starker Korrelation zur gewählten Datenaufbereitung und der Qualität des Tracking-Systems stehen (bei Unterschreitung gewisser realer Distanzen sind die Ball- und Spielerpositionen nicht mehr unterscheidbar), werden sie an dieser Stelle nicht wiedergegeben.

8.3.6.2 Evaluation von Ball Possession und Ball Action

Die Beurteilung der Erkennungsqualität der IBP- und IBA-Verfahren im Kontext der Ground Truth geschieht über die Wechsel des Ballbesitzes. Letztlich erkennen alle vorgestellten Verfahren Zeitpunkte der Wechsel im Ballbesitzverhalten, unabhängig davon, ob Intervallbeginn- oder -endzeitpunkte betrachtet werden. Hierbei wird auch der Wechsel auf einen hypothetischen Platzhalter betrachtet, falls ein Spieler Ballbesitz verliert und keinem anderen Spieler Ballbesitz zugewiesen wird. Die Beurteilung der Güte kann nun über Precision und Recall erfolgen. Weiterhin wird ein Timeline-Maß eingeführt. Hierbei werden die Markierungen in der Ground Truth und in den IBP- bzw. IBA-Messungen für jeden Zeitpunkt der Zeitachse miteinander verglichen. Die Quote aus korrekten Zeitpunkten zur Anzahl an Zeitpunkten innerhalb der Nettospielzeit wird unter der Bezeichnung Timeline notiert. Um gewisse Toleranzen zu berücksichtigen, werden für Auswertungen verschiedene Fenstergrößen berücksichtigt (Abbildung 8.7).

Gemäß Abbildung 8.7 zeigt die Ball Prediction für Fenstergrößen ab 0,4s geringfügig bessere F1-Scores als die Player Prediction. Da mit geringeren Fenstergrößen auch die Erkennungsraten relativ stark sinken, erscheint die zusätzliche Vorhersage der Spielerposition nicht nutzbringend. Ab einer Fenstergröße von 0,6s treten keine deutlichen Verbesserungen innerhalb einzelner Verfahren auf. Für diese Größe ist eine genauere Auswertung in Tabelle 8.1 notiert.

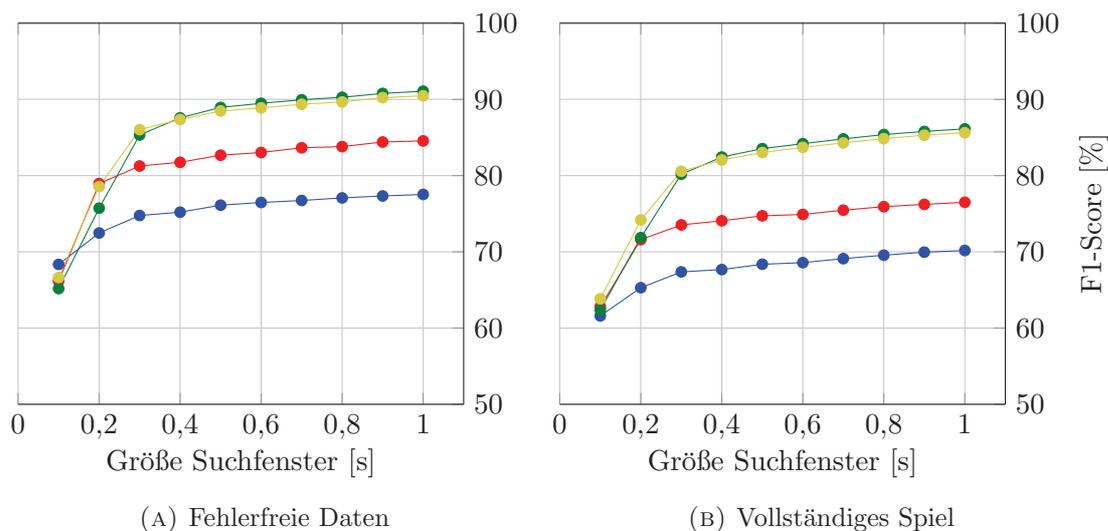


ABBILDUNG 8.7: F1-Scores für die Erkennung des Ballbesitzwechsels. IBP: Naïve Physical (blau), Kick Detection (rot). IBA: Ball Prediction (grün) und Player Prediction (gelb). Die Erkennung erfolgt in Abhängigkeit von verschiedenen Fenstergrößen, jeweils für fehlerfreie Daten (A) und das komplette Spiel (B).

	Recall	Precision	Timeline
IBP (Intervallbeginn: Naïve Physical, Ende: <i>entfällt</i>)			
Fehlerfreie Daten	88,5	67,3	82,5
Vollständiges Spiel	76,0	62,5	73,9
IBP (Intervallbeginn: Kick Detection, Ende: <i>entfällt</i>)			
Fehlerfreie Daten	80,1	86,1	87,5
Vollständiges Spiel	73,0	76,9	80,3
IBA (Intervallbeginn: Kick Detection, Ende: Ball Prediction)			
Fehlerfreie Daten	86,7	92,4	93,6
Vollständiges Spiel	81,9	86,6	87,7
IBA (Intervallbeginn: Kick Detection, Ende: Player Prediction)			
Fehlerfreie Daten	85,9	92,1	92,9
Vollständiges Spiel	81,3	86,2	87,3

TABELLE 8.1: Erkennungsraten der vorgestellten Verfahren innerhalb der IBP- und IBA-Modelle; Fenstergröße: 0,6 s

Obwohl sich für die IBA sechs Kombinationen aus den vorgestellten Verfahren ergeben würden, wurden alle IBA-Intervallanfänge in Tabelle 8.1 durch die Kick Detection bestimmt. Bis auf eine Ausnahme für ein kleines Suchfenster (Abbildung 8.7, A, 0,1s) erreicht dieses Verfahren sowohl bessere Recall- als auch Precision-Werte als Naïve Physical. Damit erscheinen auf Basis dieser Daten ebenso die Kick Detection wie auch die Ball Prediction als die erfolgversprechendsten Übersetzungen der formulierten Modelle.

Auf Basis des Timeline-Maßes kann eine Übereinstimmung von 93,6 % auf fehlerfreien Daten innerhalb des IBA-Modells und 87,5 % im IBP-Modell erreicht werden. Da die Ballposition höhere Ungenauigkeiten aufweist, wird für den Recall der Kick Detection (IBP-Bestimmung) ein Rückstand von 7,1 %-Punkten ausgewiesen: Durch Einbeziehung des vollständigen (laufenden) Spiels sinkt dieser von 80,1 % auf 73,0 %. Bezogen auf die IBA (Ball Prediction) haben perfekte Daten 4,8 %-Punkte Vorsprung bzgl. des Recall und 5,9 %-Punkte bzgl. der Precision. Mit Werten von 81,9 % und 86,6 % versprechen beide Modelle praxistaugliche Anwendungen auch auf fehlerbehafteten Daten.

8.3.6.3 Evaluation der Ball Control

Das Bayes'sche Netz zur Klassifizierung der IBA Intervalle nach Ballkontrolle wurde anhand der im Tracking-Verfahren als fehlerfrei klassifizierten Daten innerhalb der zweiten Halbzeit trainiert. In den Testdaten verblieben 337 IBA-Intervalle, wobei in 25 Intervallen keine IBC vorlag. Die erste Halbzeit steht damit als Testdatensatz zur Verfügung. Die Verhältnisse sind hier mit 305 zu 37 als ähnlich anzusehen. Der Aufbau des Netzes wurde mit einem Elter pro Knoten vorgegeben und anschließend durch den K2-Algorithmus [Cooper u. Herskovits, 1992] trainiert. Die Wahrscheinlichkeiten wurden direkt aus den Daten geschätzt.

		Ground Truth	
		IBC	Keine IBC
Bayes'sches Netz	IBC	300	15
	Keine IBC	12	10

TABELLE 8.2: Konfusionsmatrix zur IBC

Für Intervalle ohne Ballkontrolle wurde ein Anteil von 95 % korrekt klassifiziert (siehe Tabelle 8.2). Insgesamt sind somit 92 % der Intervalle korrekt zugeordnet. Liegt jedoch keine Ballkontrolle vor, konnten nur 40 % der Intervalle korrekt klassifiziert werden, entsprechend liegt das κ nach [Cohen, 1960] bei 0,38. Mit nur 25 Nicht-IBC-Intervallen ist die Trainingsmenge für die Differenzierung zwischen IBA und IBC allerdings relativ klein. Darüber hinaus zeigte auch ein Interrater-Reliabilitätstest zwischen zwei menschlichen Beobachtern auf einer Teilmenge ($\# = 98$) der Intervalle ebenso keine vollständige Übereinstimmung ($\kappa = 0,72$). Dies zeigt nicht nur, dass eine Bestimmung auf Grundlage der Positionsdaten als schwierig angesehen werden kann - es ist auch fraglich, ob eine vollständige Objektivierung dieses Konstrukts letztlich möglich ist. Weitere Untersuchungen müssen hier Aufschluss über Möglichkeiten zur Erkennung der Ballkontrolle geben.

8.3.7 Vorstellung exemplarischer Anwendungen

Auf Basis der vorgestellten Ballbesitzmodelle können einige exemplarische Anwendungen vorgestellt werden. Die Berechnung der Intervalle erfolgt auf Basis der Kick Detection und Ball Prediction, da diese Ansätze in der Evaluation die besten Quoten erreicht haben. Wieder wird die bekannte Begegnung der Erstligamannschaften Verwendung finden.

Ballbesitz \ Spielzeiten	Σ	1. Hälfte (Heim / Gast)	2. Hälfte (Heim / Gast)
PT brutto [m:s]	90:12	45:00	45:12
PT netto [m:s]	57:56	28:24	29:32
TBP [m:s]	54:03	19:45 / 6:31	18:20 / 9:27
TPM [m:s]	47:12	18:38 / 4:12	16:46 / 7:36
TBC [m:s]	29:48	11:16 / 2:57	10:42 / 4:53
TBC [#]	1 291	481 / 149	442 / 219

TABELLE 8.3: Ergebnisse der Ballbesitzererkennung für TPB, TPM, TBC aufgeteilt nach Halbzeiten

Wie aus Tabelle 8.3 ersichtlich, ist die Summe der TBP mit 54:03 etwas kürzer als die Nettospielzeit beider Mannschaften. Hierbei wird dem Fakt Rechnung getragen, dass in Bereichen, welche als fehlerhaft identifiziert wurden (siehe Kapitel 7.3), keine Zuordnung stattfindet. Zum anderen werden Ballbesitzphasen direkt nach Spielunterbrechungen nicht betrachtet. Dies ist ein Kompromiss, um Randerscheinungen innerhalb der sportwissenschaftlichen Disziplin zu vermeiden. Die betrachtete Spielzeit wird dadurch geringfügig kürzer, bleibt aber repräsentativ. Der Ballbesitzanteil der Heimmannschaft (70%) wird 5%-Punkte geringer berechnet als durch Opta angegeben (Tabelle 7.2). Da nicht in jeder TBP Ballkontrolle vorliegt, ist die TPM mit 47:12 kürzer. In der Summe ist das Heimteam sowohl bezüglich Ballbesitz als auch Ballkontrolle überlegen.

Position	PT (netto)[s]	IBP $\left[\frac{\#}{10 \text{ min}}\right]$	IBA $\left[\frac{\#}{10 \text{ min}}\right]$	IBC $\left[\frac{\#}{10 \text{ min}}\right]$	IBC [s]	IBC $\left[\frac{s}{10 \text{ min}}\right]$	IBC $\left[\frac{s}{\#}\right]$
Heim							
Torwart	3 444	3,3	3,3	3,2	27,6	4,8	1,5
Außenverteidiger	3 444	15,1	15,4	15,4	106,2	18,6	1,2
Außenverteidiger	3 444	9,3	9,8	9,8	79,5	13,9	1,4
Mittelfeld	2 986	19,9	21,2	21,0	142,1	28,6	1,4
Innenverteidiger	3 443	15,6	16,3	15,3	133,2	23,4	1,5
Innenverteidiger	3 444	17,5	19,3	18,9	190,7	33,4	1,8
Mittelfeld	3 194	14,1	14,9	14,7	146,5	27,6	1,9
Mittelfeld	3 444	26,7	28,8	28,4	217,5	38,1	1,3
Mittelstürmer	3 444	5,8	6,0	6,0	30,1	5,3	0,9
Mittelfeld	3 444	16,8	17,7	17,5	129,0	22,6	1,3
Mittelfeld	2 735	11,9	11,9	11,9	68,4	15,0	1,3
Mittelfeld	708	7,0	7,0	7,0	13,8	12,1	1,7
Außenverteidiger	250	10,2	10,2	10,2	7,1	18,1	1,8
Mittelfeld	458	29,8	29,8	29,8	26,5	35,8	1,2
Gast							
Torwart	3 444	3,5	4,7	4,2	69,5	12,2	2,9
Außenverteidiger	3 444	4,7	5,1	5,1	21,5	3,8	0,7
Außenverteidiger	3 444	6,5	6,5	6,0	32,2	5,6	0,9
Mittelfeld	3 444	6,5	6,5	5,6	29,3	5,1	0,9
Innenverteidiger	3 444	5,6	5,6	5,1	29,2	5,1	1,0
Innenverteidiger	2 986	5,0	5,2	4,8	31,9	6,4	1,3
Außenstürmer	3 444	7,2	7,4	7,4	59,7	10,5	1,4
Mittelfeld	3 444	6,1	6,7	6,1	45,3	7,9	1,3
Mittelstürmer	3 444	4,9	5,1	4,7	26,6	4,7	1,0
Mittelfeld	2 987	7,7	8,3	8,1	56,8	11,4	1,4
Mittelfeld	1 091	9,4	9,4	9,4	24,6	13,6	1,4
Außenverteidiger	2 352	8,8	8,8	7,0	8,2	6,9	1,0
Mittelfeld	456	7,5	8,2	7,9	35,3	10,9	1,4

TABELLE 8.4: Ergebnisse der Ballbesitzerkennung auf Spielerebene unter Berücksichtigung individueller Spielzeiten (PT). Ballbesitz pro zehn Minuten Nettospielzeit $\left[\frac{\#}{10 \text{ min}}\right]$, Sekunden pro zehn Minuten Nettospielzeit $\left[\frac{s}{10 \text{ min}}\right]$ und Sekunden pro Intervall $\left[\frac{s}{\#}\right]$.

Anmerkung 10. Neben Auswechslungen, die zu mehr als elf Spielern pro Mannschaft in der Tabelle führen, ist die Gesamtnettospielzeit der Gastmannschaft verkürzt, da in der 76. Spielminute eine Rote Karte gezogen wurde.

Der individuelle Ballbesitz ermöglicht es, jedem Spieler Kenngrößen zuzuweisen (Tabelle 8.4). Exemplarisch wurden hierfür die Anzahl an IBP-, IBA- und IBC-Phasen sowie die Länge der IBC innerhalb von 10 min Nettospielzeit gewählt. Diese Normalisierung erlaubt einen Vergleich über verschiedene individuelle Nettospielzeiten (PT) hinweg. Gleiches gilt für die Angabe der durchschnittlichen Länge einer IBC-Phase. Da auch dieser Wert nicht ohne Aussagekraft ist, wurde die Gesamtlänge der IBC ebenso angegeben.

Die deutlichen Unterschiede im Mannschaftsballbesitz schlagen sich auch auf Spielerbene nieder. So beträgt die größte IBC-Länge der Gastmannschaft 69,5 s (Torwart), während das Heimteam 217,5 s (Mittelfeld) ausweisen kann. Auch wenn dies nicht mit dem Erzielen von spielentscheidenden Ereignissen korrelieren muss — der Mittelstürmer (Heim) hat die kürzeste durchschnittliche Ballkontaktzeit seiner Mannschaft — können Aussagen über relativ-zeitliche Anteile am spieltaktischen Geschehen formuliert werden. Die Ballbesitzphasen des Gast-Torwarts sind beispielsweise im Durchschnitt etwa doppelt so lang wie die des Heim-Torwarts (2,9 s zu 1,5 s). Dies kann etwa als Spielverlangsamung oder ein Liberospiel (Heim-Torwart) interpretiert werden. Weitere Analysen und Daten werden zukünftig nötige Kontextinformationen bereitstellen.

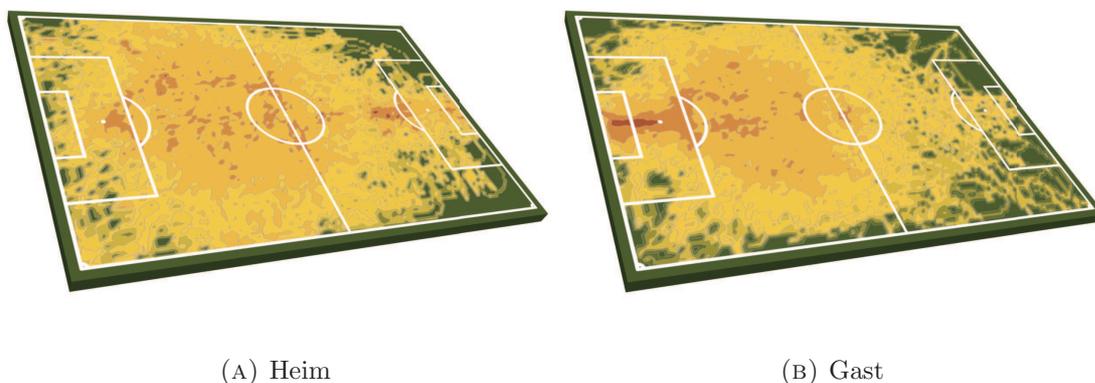
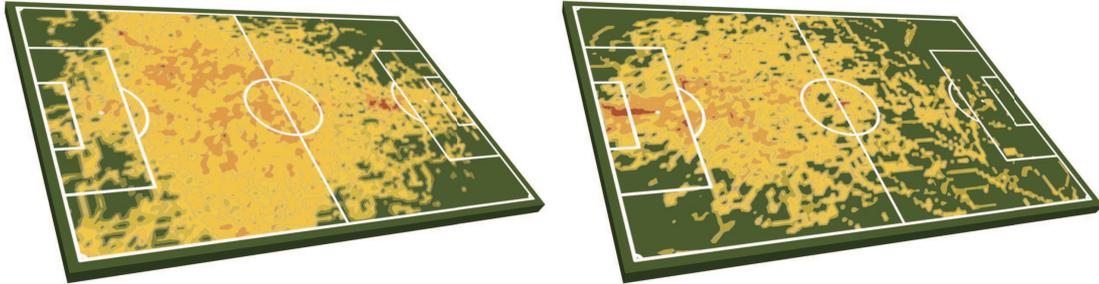


ABBILDUNG 8.8: Bewegungen der Feldspieler in der ersten Halbzeit. Die Heimmannschaft spielt von rechts nach links.

Neben einer statistischen Auswertung des Ballbesitzes erlaubt die Kombination aus Positionsinformationen und Ballbesitz auch die Erweiterung der bisher bekannten Heatmaps (Abbildung 8.8). Konnten in diesen zuvor nur die Trajektorien der Spieler aufbereitet werden, ergeben sich nun für visuelle Auswertungen dieser Art neue Einsatzmöglichkeiten. Durch die Verfügbarkeit zeitsynchroner Ballbesitz- und Positionsinformation können die Grafiken mithilfe von Aussagen über Spielerbewegung auf Ballinteraktion

eingeschränkt werden. Abbildung 8.9 zeigt die Heatmaps für die Ballkontrolle beider Mannschaften. Neben einer Ausdünnung der Bereiche fällt vor allem der geringe Anteil an Ballkontrolle der Gastmannschaft in der gegnerischen Spielfeldhälfte ins Auge.

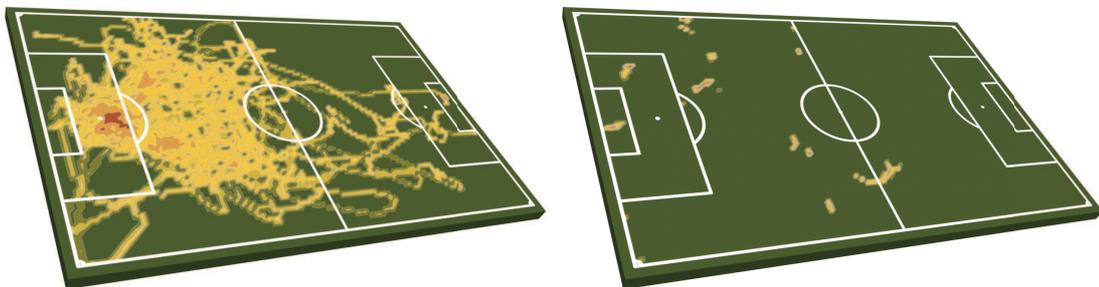


(A) Heim

(B) Gast

ABBILDUNG 8.9: Ballkontrolle der Feldspieler in der ersten Halbzeit. Die Heimmannschaft spielt von rechts nach links.

Die Betrachtung wird auf den Mittelstürmer des Heimteams während der ersten Hälfte der Begegnung eingeschränkt. Unter den Spielern seiner Mannschaft, die in dieser Halbzeit spielten, ist er derjenige mit der geringsten absoluten IBC-Länge.



(A) Bewegung

(B) Ballkontrolle

ABBILDUNG 8.10: Heatmaps des Mittelstürmers (Heim). Die Heimmannschaft spielt von rechts nach links.

Durch die Begrenzung der Heatmap auf Ballkontrollphasen erscheint die Grafik stark reduziert. Nur die Positionen zu rund 15s Ballkontrolle sind verblieben. Neben Möglichkeiten der sportwissenschaftlichen Interpretation können auf diese Weise auch bedeutende Ballinteraktionen des Spielers hervorgehoben werden. Aufgrund der geringen Datenmenge kann jede Zusammenhangskomponente mit den sie bedingenden Zeitpunkten des Spiels assoziiert werden. Die Region im Torraum (Abbildung 8.10, B) etwa korrespondiert mit einem erzielten Kopfballtor in der 24. Minute.

8.4 Erkennen von Spielereignissen

Spielereignisse werden wie der Ballbesitz während des Spiels manuell erfasst. Nach der Austragung werden diese Daten in einer weiteren Sichtung (durch CIPs) aufwändig manuell überprüft und korrigiert. Anhand der ausgewählten Begegnung sollen die gemäß DFL-Vorgaben [DFL, 2013] notierten Spielereignisse dargestellt werden.

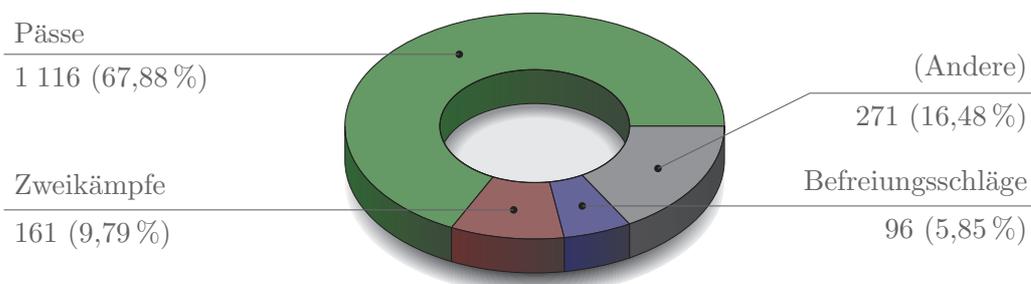


ABBILDUNG 8.11: Graphische Darstellung einer Teilmenge quantitativ herausragender Ereignisse

Das im Spiel am häufigsten eingetretene Ereignis ist der Pass — über zwei Drittel aller Ereignisse sind Pässe (siehe Abbildung 8.11). Dieser Anteil untergliedert sich weiter in erfolgreiche (ankommende) und nicht erfolgreiche Pässe. Auf Platz zwei folgt der Zweikampf, welcher explizit zwischen zwei Spielern definiert ist. Ein Kampf um den Ball zwischen drei Spielern ist nicht vorgesehen.

In Tabelle 8.5 werden alle 1 644 in der Begegnung vorkommenden Ereignisse kategorisiert. Während ballbesitz- und positionsabhängige Ereignisse hier die eindeutige relative Mehrheit darstellen, sind auch Ereignisse vertreten, welche nicht aus den Ballbesitzinformationen und den Positionen der Spieler ermittelt werden können. Hier ist etwa die Vergabe einer Gelben oder Roten Karten zu nennen.

Für die Bereitstellung eines Spieljournals, also einer Auflistung von spieler- und zeitbezogenen Ereignissen für eine Begegnung, soll eine Teilmenge von 1 362 Ereignissen (siehe Tabelle 8.5) durch einen Algorithmus automatisch bestimmt werden. Die Basis hierfür wird die IBP bilden.

Ereignis	Häufigkeit		Implementation
	abs.	rel. [%]	
KickoffWhistle	2	0,12	
Pass	1 116	67,88	X
ThrowIn	31	1,89	X
ShotAtGoal	27	1,64	
Foul	20	1,22	
FreeKick	20	1,22	X
CornerKick	18	1,09	X
BallContactWithoutControl	8	0,49	
Caution	2	0,12	
FinalWhistle	2	0,12	
GoalKick	10	0,61	X
OwnGoal	1	0,06	
KickoffAfterGoal	6	0,36	X
Substitution	5	0,30	
OtherPlayerAction	1	0,06	
Penalty	1	0,06	
TacklingGame	161	9,79	X
DefensiveClearance	96	5,84	
OtherBallContact	115	7,00	
Offside	1	0,06	
AbandonmentOfBall	1	0,06	

TABELLE 8.5: Detailübersicht der Spielereignisse

Das Verfahren zu Abbildung 8.12 ermöglicht das Schreiben eines Spieljournals auf Basis von Ballbesitzwechseln. Hierfür wird der aktuelle Ballbesitz in der Variablen P notiert, der letzte Spieler in Ballbesitz wird mit LP bezeichnet. Stimmen beide Variablen nicht überein, wird auf ein Ereignis geschlossen. Innerhalb einer Spielunterbrechung wird LP ein Platzhalter zugewiesen, um diesen Fall gesondert behandeln zu können. Auf eine Binnendifferenzierung der Ereignisse aus ruhendem Spiel wurde im Sinn der Durchführung einer Projektstudie verzichtet. Für den Fall korrekter Daten gestaltet sich dies bei den ausgewählten Ereignissen jedoch einfach.

Gemäß Abbildung 8.13 können aus der gewählten 83-prozentigen Teilmenge der Ergebnisse 69,2% mit den Ereignisdaten des CIP Opta identifiziert werden (Recall). Da diese Daten auf den fehlerhaften Positionen und Ballbesitzinformationen beruhen, ist diese Quote nicht überraschend, insbesondere da teilweise zwei Spieler für ein Ereignis korrekt bestimmt sein müssen. Mit $83\% * 69,2\% \approx 57\%$ ist für das gesamte Spiel jedoch bereits der größte Teil der Ereignisse validiert. Im Rahmen einer Qualitätssicherung können diese Ereignisse als korrekt angesehen und aus der manuellen Qualitätssicherung ausgeschlossen werden. Die aufwändige Fehlersuche kann somit verkürzt werden.

8.5 Diskussion und Ausblick

In diesem Kapitel wurden neue Ballbesitzmodelle eingeführt und algorithmisch umgesetzt. Die Korrektheit dieser Übersetzung wurde durch eine aufwändige Evaluation nachgewiesen. Hierzu wurden die Ballbesitzmodelle durch einen Sportwissenschaftler manuell verarbeitet und die fachlichen Modelle auf Basis von Positionsdaten algorithmisch nachempfunden. Die im vorigen Kapitel herausgestellten Fehler in den Daten schränken die resultierenden Ergebnisse nicht derart ein, als dass sie nicht für weitere Anwendungen sinnvoll nutzbar wären. Durch Kombination einer Ballbesitz- und Ereigniserkennung mit einer Korrektur der Positionsinformation kann die Qualität beider Systeme weiter verbessert werden. Denkbar ist etwa, dass bei Erkennung eines Eckballs gewisse Ballpositionen ausgeschlossen werden können.

Innerhalb der Ballbesitzerkennung wurde eine feste Prädiktionslänge von 1 s verwendet. Durch eine zweidimensionale Gridsuche bei einer größeren Eingangsdatenmenge kann die Genauigkeit evtl. vergrößert bzw. das Verfahren besser den sportwissenschaftlichen Annotationen angepasst werden. Ebenso kann es sinnvoll sein, dass ein Ballbesitz vor dem Zeitpunkt des lokalen Maximums der Kick Detection beginnt, da ein Spieler bereits zuvor die Möglichkeit des Ballbesitzes haben kann, etwa bei einem Pass in den Lauf eines Spielers. In weiteren Betrachtungen des individuellen Ballbesitzes wird sich herausstellen, welche zeitliche Bestimmung sich durchsetzt.

Aus den Ergebnissen des Ballbesitzes kann eine Vielzahl neuer Anwendungsmöglichkeiten für Vereine, Trainer und Dienstleister entstehen. Im Verlauf des Kapitels wurden Statistiken und Heatmaps gezeigt. Weiter denkbar ist beispielsweise ein auf Spielfeldregionen oder relativ zur gegnerischen Mannschaft ausgewerteter oder grafisch dargestellter Ballbesitz. Mit den gegebenen Informationen sind ebenso Auswertungen zu Anspielbarkeit, Deckungs- oder Pressingverhalten denkbar.

Als weitere Anwendung des Ballbesitzes wurde die Spielereigniserkennung vorgestellt. Hier konnten bereits 57 % der Ereignisse nach DFL-Anforderungen berechnet werden. Da die Qualitätssicherung für die manuelle Annotation von Fußballspielen einen großen zeitlichen Aufwand bedeutet, kann diesem Verfahren vor allem in diesem Gebiet Bedeutung beigemessen werden. Das Spieljournal wird mit einer weiterentwickelten Ballbesitzerkennung mehr Ereignisse finden. Derzeit sind jedoch noch Ereignisse, welche aus den Positionsdaten erkennbar sind, nicht implementiert, sodass auch Verbesserungen auf Basis der vorhandenen Ballbesitz-Informationen möglich sind. Insbesondere Befreiungsschläge, Abstöße, Torschüsse, Auswechslungen und andere Ballkontakte gilt es hier zu bearbeiten.

Kapitel 9

Abschließende Diskussion

In den vorangegangenen Kapiteln wurde ein möglicher Weg zur Bereitstellung von Weltkoordinaten zu Objekten in Sequenzen monokularer Bilder vorgestellt. Unter Kenntnis eines Umgebungsmodells wurde dieser durch eine Nachführung von Kameraparametern besprochen. Zu den in Kapitel 1.3 genannten Zielsetzungen kann damit festgehalten werden:

- Ausgehend von einer als rechteckig und homogen angenommenen Kalibrierfläche werden die Punkte eines Eingangsbilds bzgl. ihrer Zugehörigkeit klassifiziert (Kapitel 3). Zusätzlich erfolgt auch eine Betrachtung der Punkte hinsichtlich der Sichtbarkeit der Kalibrierfläche. Bezüglich eines manuell erfassten Ground-Truth-Datensatzes konnte die Robustheit der Klassifikation der Punkte nachgewiesen werden.
- Die Extraktion natürlicher, ortsfester Landmarken wird durch eine automatische Suche nach Geradensegmenten unterstützt (Kapitel 5). Der Schwerpunkt jenes Verfahrens liegt vor allem auf der automatischen Schwellwertbestimmung. Die in diesem Kapitel getroffene Behauptung der Anschaulichkeit der Parameterwahl der vorgestellten Erkennung von Geradensegmenten wird in Kapitel 6.7 bestätigt, indem sie für die Suche nach Geraden eingesetzt wird. Durch eine gewichtete Eintragung der gefundenen Geradensegmente in einen Hough-Raum liegen bereits die Parameter für die Geradensegmentsuche durch das Meta-Verfahren (die Geradensuche) vor.
- Die zum Ziel gesetzte Nachführung der Kameraparameter ist unter Berücksichtigung der Forderung ausgeführt, auch Bildfolgen in Bildsequenzen überbrücken zu können, in welchen eine für die Kalibrierung nicht ausreichende Zahl von Landmarken vorgefunden wird. Hierzu wird in einer Abwägung zwischen einer Kalibrierung

unter Ausnutzung eines gegebenen Modells und einer Nachführung von Modelllandmarken entschieden. Durch eine geeignete Auswahl von Homographien und der Bewertung bzgl. ihrer Quasi-Affinität konnten intrinsische und extrinsische Kameraparameter innerhalb einer synthetischen Kamerabewegung mit geringen Abweichungen bzgl. der Testebene nachgeführt werden.

Im Rahmen des Anwendungsgebiets Fußball können weiterhin die folgenden Aussagen getroffen werden:

- In Kapitel 4 wird ein Verfahren zur Klassifikation von Bildern hinsichtlich des gezeigten Bildinhalts vorgestellt. Für aufbereitetes TV-Material ist dies in Verbindung mit einer Erkennung von Schnitten durch Kamerawechsel essentiell, um die Teilmenge an Bildern bestimmen zu können, aus welchen relevante Informationen (Objekt- bzw. Spielerpositionen) berechnet werden können. Mit einer Erkennungsrate der Einstellungsgrenzen der Totalen (Kamera-1-Einstellungen) von 98,2% und einer Klassifikationsrate von 96,3% werden alternative Verfahren übertroffen.
- Im Zusammenwirken mit der in Kapitel 7 besprochenen Objekt-Nachführung konnte für bodennahe Objekte gezeigt werden, dass die Kameraparameter- sowie die Objekt-Nachführung auch in langen Bildsequenzen anwendbar sind - ein herausragendes Merkmal der Kameraparameter-Nachführung, welches in einer aufwändigen Evaluation betrachtet wurde (Kapitel 6.8 und 7.4). Die durchschnittliche Abweichung über 540 700 betrachtete Bilder betrug zwischen 0,53 m und 0,82 m (MOTP in einem 3 m-Suchradius). Es bleibt eine Frage des Anwendungsszenarios, ob diese Abweichungen tolerierbar sind. Während für Bewegungen eines Torwarts im Torraum eine große Genauigkeit erforderlich ist, werden bei berechneten Laufleistungen über eine gesamte Spielzeit durchaus Differenzen im Kilometerbereich toleriert.
- Die in Kapitel 8 beschriebenen Methoden zur Ballbesitzererkennung wurden auf aufbereiteten CIP-Positionsdaten angewandt. Hierbei wurden die eingeführten Ballbesitzmodelle eingesetzt und evaluiert. Da es sich um die erste fachliche Unterscheidung von Graden des individuellen Ballbesitzes aus sportwissenschaftlicher Sicht handelt, muss jedoch auf einen Vergleich mit anderen Systemen verzichtet werden. Es kann allerdings festgehalten werden, dass die Genauigkeit der Bestimmung des Mannschaftsballbesitzes (TBP) die manuelle spielbegleitende Bestimmung dergleichen durch die CIPs übertrifft. Die auf dem individuellen Ballbesitz basierende Ereigniserkennung erkannte in einem Testszenario insgesamt 57% aller Spielereignisse übereinstimmend mit manuell annotierten Ereignissen, wobei für 83% der

auftretenden Ereignisse (relativ zur Anzahl der Ereignisse in der Begegnung) eine Implementation vorlag und die Richtigkeit der manuellen Annotierungen, wie im Live-Sport üblich, nicht als gesichert erachtet werden kann. Es wird das Ziel weiterer Betrachtungen sein, diesen Anteil zu steigern und damit die Qualität von automatisch erfassten Spieljournalen weiter zu erhöhen.

Wie [Lu et al., 2013] beschreiben, gestaltet sich insbesondere in Sportarten, welche im Freien ausgetragen werden, die Kamerakalibrierung als schwierig. Dafür verantwortlich machen sie unter anderem die Tatsache, dass in einem Einzelbild ohne Kenntnis genauer Parameter der Umgebung nur eine einzige Ebene — die Spielfeldebene — für die Zuordnung zwischen Bild und Modell zur Verfügung steht. Im Basketball beispielsweise gehört auch die Aufhängung des Korbs zu den spezifizierten Ebenen. Zudem sind weite Teile des Spielfelds oft verdeckt und aufgrund der Möglichkeit einer fortwährenden Änderung der internen Kameraparameter muss auch der klassische Ansatz zur Bestimmung von Kameraparametern verworfen werden.

Diese Probleme aufgreifend, wird in den Kapiteln 3 bis 6 ein Verfahren zur Kameraparameter-Nachführung für das Szenario gestörter planarer Eingangsdaten entwickelt. Als Störungen im Kontext Fußball werden zum einen die sich bewegenden Spielfeldobjekte (Spieler, Ball und Schiedsrichter) selbst betrachtet, welche für eine Kalibrierung notwendige natürliche Landmarken verdecken können. Als weitere Einflüsse zu nennen sind neben einer Vielfalt von zu berücksichtigenden Oberflächenbeschaffenheiten und variierenden Beleuchtungsverhältnissen auch unbeständige Wetterbedingungen. Weitestgehend ohne manuelle Eingriffe konnte dennoch für die überwiegende Mehrheit der Bilder eine zufriedenstellende Kalibrierung erreicht werden.

Offen bleibt neben weiteren Untersuchungen zur Ball-Nachführung in monokularen Bildern auch die Optimierung der Kameraparameter-Nachführung in Bildsequenzen, in welchen über größere Zeiträume keine ausreichende Anzahl an Feldlinien sichtbar ist und damit (nicht-lineare) Einflüsse wie Verzeichnungen zu einem Drift führen. Zukünftige Forschung muss sich diesen Problemen stellen. Mit der Verfügbarkeit von Spielerpositionen in Weltkoordinaten stellt sich ebenso neuer Forschungsbedarf in informationstechnischer und sportwissenschaftlicher Sicht hinsichtlich der Auswertung der gewonnenen Daten. Neben dem bereits in Kapitel 8 gegebenen Einblick sind Untersuchungen bezüglich der automatischen Erkennung von Spielsituationen (Pressing, Raumdeckung, ...) und Spielstilen (Aufbauspiel, Konter, ...) denkbar.

Anhang A

Kapitelresulte zur Kameraparameter-Nachführung für Bildbeispiele

Das in den Kapiteln 3 bis 6 vorgestellte Verfahren zur Kameraparameter-Nachführung wird nachfolgend anhand von Beispielen veranschaulicht. Das Ziel ist es dabei, einen Überblick über die Wirkung der eingesetzten Methoden bis hin zum Schritt der Spielfeld-Zuordnung zu ermöglichen (Tabelle A.1). Ebenso können Abhängigkeiten der Bearbeitungsschritte voneinander betrachtet werden, beispielsweise werden Geradensegmente nur in der Hüllenregion gesucht. Im Einzelnen abgebildet sind:

Abbildungen	Inhalt
A.1, A, B und A.2, A, B	Originalbild
A.1, C, D und A.2, C, D	Spielfeldhülle (Kapitel 3.3)
A.1, E, F und A.2, E, F	Grasmaske (Kapitel 3.4)
A.1, G, H und A.2, G, H	Geradensegmente (Kapitel 5)
A.1, I, J und A.2, I, J	Nachgeführtes Spielfeldmodell (Kapitel 6)

TABELLE A.1: Inhalte der Abbildungen dieses Anhangs

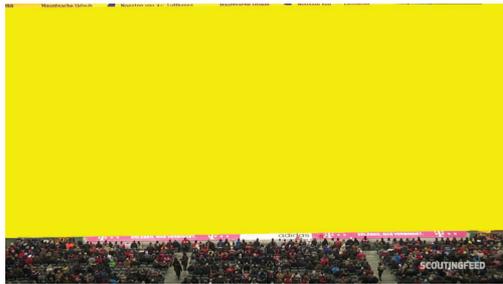
Es findet wiederum die Begegnung zwischen FC Bayern und Schalke 04 Verwendung, welche bereits in den Kapiteln 6 und 7 herangezogen wurden.



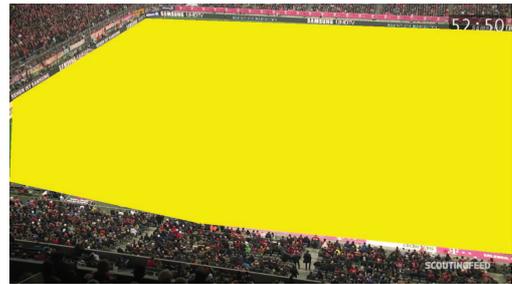
(A)



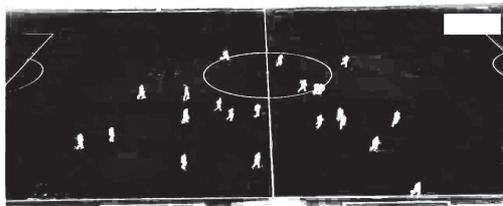
(B)



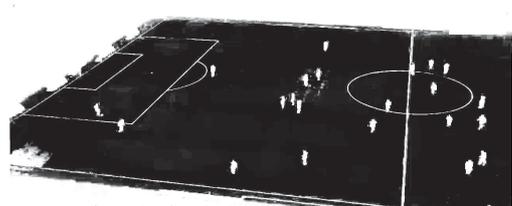
(C)



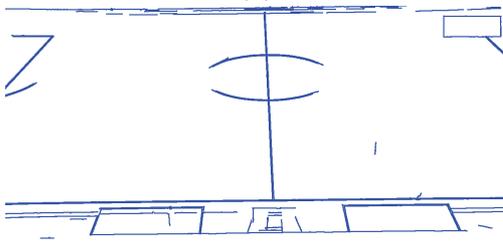
(D)



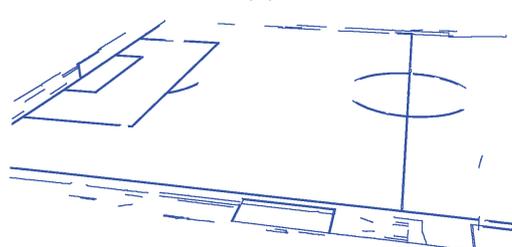
(E)



(F)



(G)



(H)



(I)



(J)

ABBILDUNG A.1

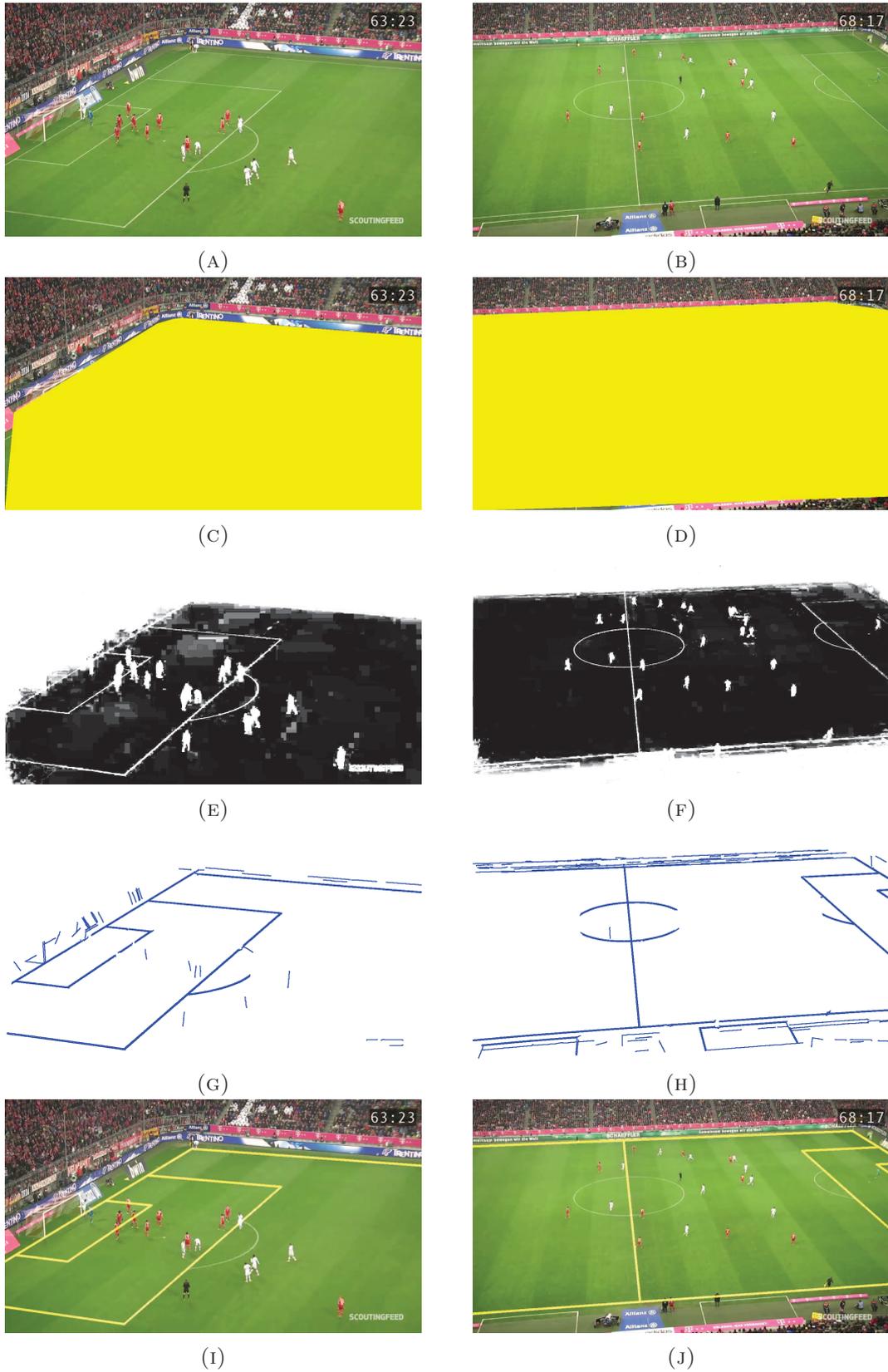


ABBILDUNG A.2

Anhang B

Einzelbilderergebnisse zur Kameraparameter-Nachführung

Zu allen in Kapitel 6.8.2 vorgestellten Testdatensätzen sind im Folgenden Einzelbilder und die zugehörigen Ergebnisse der Kameraparameter-Nachführung (Kapitel 6) in Form einer Einblendung im Bild (gelbes Spielfeldmodell) vorzufinden. Alle Bilder entstammen einer äquidistanten Abtastung. Die Abbildungen B.1, B.2, B.3 und B.4 sollen eine unverfälschte Einschätzung zum Verfahren geben. Die in den Abbildungen sichtbaren Timecodes sind bereits in den Originalbildern codiert.



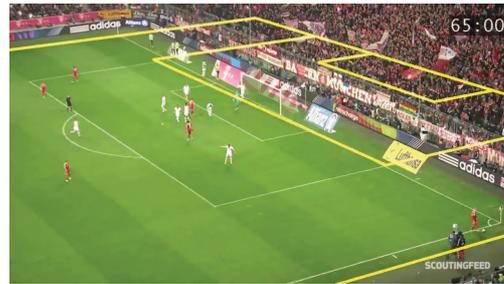
(A)



(B)



(C)



(D)



(E)



(F)



(G)



(H)

ABBILDUNG B.1: FC Bayern - Schalke 04



(A)



(B)



(C)



(D)



(E)



(F)



(G)



(H)

ABBILDUNG B.2: FC Bayern - Bayer Leverkusen



(A)



(B)



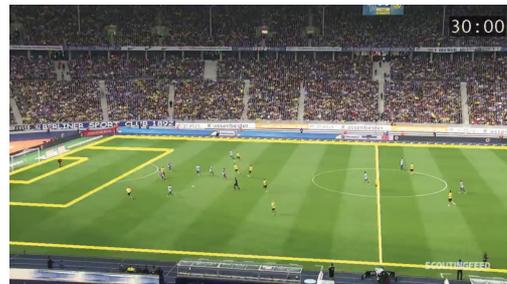
(C)



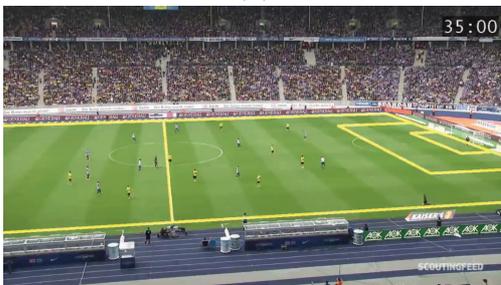
(D)



(E)



(F)



(G)



(H)

ABBILDUNG B.3: Hertha BSC - Borussia Dortmund



(A)



(B)



(C)



(D)



(E)



(F)



(G)



(H)

ABBILDUNG B.4: FC Augsburg - Paderborn

Quellenverzeichnis

- [de Agapito et al. 1998] AGAPITO, Lourdes de ; HAYMAN, Eric ; REID, Ian D.: Self-Calibration of a Rotating Camera with Varying Intrinsic Parameters. In: *BMVC*, 1998, S. 1–10
- [Akinlar u. Topal 2011] AKINLAR, Cuneyt ; TOPAL, Cihan: EDLines: A real-time line segment detector with a false detection control. In: *Pattern Recognition Letters* 32 (2011), Nr. 13, S. 1633–1642
- [Alemán-Flores et al. 2014] ALEMÁN-FLORES, Miguel ; ÁLVAREZ, Luis ; GÓMEZ, Luis ; HENRÍQUEZ, Pedro ; MAZORRA, Luis: Camera calibration in sport event scenarios. In: *Pattern Recognition* 47 (2014), Nr. 1, S. 89–95
- [Amiri u. Fathy 2010] AMIRI, Ali ; FATHY, Mahmood: Video shot boundary detection using QR-decomposition and gaussian transition detection. In: *EURASIP Journal on Advances in Signal Processing* 2009 (2010), Nr. 1
- [ARD/BR 2012] ARD/BR, Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland, Bayerischer Rundfunk: *Sportschau live: Fußball - DFB-Pokal-Achtelfinale vom 18.12.2012: FC Augsburg - FC Bayern München*. 2012
- [ARD/SWR 2010] ARD/SWR, Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland, Südwestrundfunk: *FIFA WM 2010 live vom 12.06.2010: Südkorea - Griechenland*. 2010
- [ARD/WDR 2013a] ARD/WDR, Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland, Westdeutscher Rundfunk: *Sportschau live: Fußball-Bundesliga vom 18.01.2013: FC Schalke 04 - Hannover 96*. 2013
- [ARD/WDR 2013b] ARD/WDR, Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland, Westdeutscher Rundfunk: *Sportschau live vom 18.01.2013: Frankreich - Deutschland*. 2013
- [Assfalg et al. 2003] ASSFALG, Jürgen ; BERTINI, Marco ; COLOMBO, Carlo ; DEL BIMBO, Alberto ; NUNZIATI, Walter: Semantic annotation of soccer videos: automatic highlights identification. In: *Computer Vision and Image Understanding* 92 (2003), Nr. 2, S. 285–305
- [Baber et al. 2011] BABER, Junaid ; AFZULPURKAR, Nitin ; DAILEY, Matthew N. ; BAKHTYAR, Maheen: Shot boundary detection from videos using entropy and local descriptor. In: *17th International Conference on Digital Signal Processing (DSP), 2011 IEEE*, 2011, S. 1–6
- [Barinova et al. 2012] BARINOVA, Olga ; LEMPITSKY, Victor ; KHOLI, Pushmeet: On detection of multiple object instances using hough transforms. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (2012), Nr. 9, S. 1773–1784
- [Bendale u. Talati 2012] BENDALE, Swati D. ; TALATI, Bijal J.: Analysis of Popular Video Shot Boundary Detection Techniques in Uncompressed Domain. In: *International Journal of Computer Applications* 60 (2012), Nr. 3
- [Besl u. McKay 1992] BESL, Paul J. ; MCKAY, Neil D.: Method for registration of 3-D shapes. In: *Robotics-DL tentative International Society for Optics and Photonics*, 1992, S. 586–606

- [Beutelspacher u. Rosenbaum 1992] BEUTELSPACHER, Albrecht ; ROSENBAUM, Ute: Projektive Geometrie. In: *Von den Grundlagen bis zu den Anwendungen*. Verlag Vieweg, Wiesbaden (1992)
- [Bialkowski et al. 2014] BIALKOWSKI, Alina ; LUCEY, Patrick ; CARR, Peter ; YUE, Yisong ; SRIDHARAN, Sridha ; MATTHEWS, Iain: Large-Scale Analysis of Soccer Matches using Spatiotemporal Tracking Data. In: *IEEE International Conference on Data Mining (ICDM)* IEEE, 2014, S. 725–730
- [Bigontina et al. 2014] BIGONTINA, Andreas ; HERRMANN, Michael ; HOERNIG, Martin ; RADIG, Bernd: Human Body Part Classification in Monocular Soccer Images. In: *9-th Open German-Russian Workshop on Pattern Recognition and Image Understanding* (2014)
- [Boreczky u. Rowe 1996] BORECZKY, John S. ; ROWE, Lawrence A.: Comparison of video shot boundary detection techniques. In: *Journal of Electronic Imaging* 5 (1996), Nr. 2, S. 122–128
- [Bu et al. 2011] BU, Jiang ; LAO, Songyang ; BAI, Liang: Automatic line mark recognition and its application in camera calibration in soccer video. In: *IEEE International Conference on Multimedia and Expo (ICME)* IEEE, 2011, S. 1–6
- [Bundesministerium für Verkehr 1993] BUNDESMINISTERIUM FÜR VERKEHR: *Richtlinien für die Markierung von Straßen (RMS), Teil 1: Abmessung und geometrische Anordnung von Markierungszeichen (RMS-1)*. Forschungsgesellschaft für Straßen- und Verkehrswesen e.V., Arbeitsgruppe Verkehrsführung und Verkehrssicherheit, 1993
- [Burns et al. 1986] BURNS, J B. ; HANSON, Allen R. ; RISEMAN, Edward M.: Extracting straight lines. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 4 (1986), S. 425–455
- [Caballero et al. 2007] CABALLERO, Fernando ; MERINO, Luis ; FERRUZ, Joaquin ; OLLERO, Anibal: Homography based Kalman filter for mosaic building. applications to UAV position estimation. In: *IEEE International Conference on Robotics and Automation* IEEE, 2007, S. 2004–2009
- [Catapult 2015] CATAPULT, Catapult Sports Pty Ltd.: *OptimEye G5*. 2015
- [Chandraker et al. 2007] CHANDRAKER, Manmohan ; AGARWAL, Sameer ; KRIEGMAN, David ; BELONGIE, Serge: Globally optimal affine and metric upgrades in stratified autocalibration. In: *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007* IEEE, 2007, S. 1–8
- [Chen et al. 2003] CHEN, Shu-Ching ; SHYU, Mei-Ling ; ZHANG, Chengcui ; LUO, Lin ; CHEN, Min: Detection of soccer goal shots using joint multimedia features and classification rules. In: *MDM/KDD* 3 (2003)
- [ChyronHego 2005] CHYRONHEGO, ChyronHego Corporation: *Tracab*. <http://tracab.hegogroup.com/>, 2005–2015. – Abgerufen am: 15.09.2015
- [Cohen 1960] COHEN, Jacob: A coefficient of agreement for nominal scales. In: *Educational and Psychological Measurement* 20 (1960), S. 37–46. <http://dx.doi.org/10.1177/001316446002000104>. – DOI 10.1177/001316446002000104
- [Cooper u. Herskovits 1992] COOPER, Gregory F. ; HERSKOVITS, Edward: A Bayesian method for the induction of probabilistic networks from data. In: *Machine learning* 9 (1992), Nr. 4, S. 309–347
- [Criminisi et al. 2012] CRIMINISI, Antonio ; SHOTTON, Jamie ; KONUKOGLU, Ender: *Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning*. Now, 2012
- [Dalal u. Triggs 2005] DALAL, Navneet ; TRIGGS, Bill: Histograms of oriented gradients for human detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR Bd. 1* IEEE, 2005, S. 886–893
- [De Marco et al. 2013] DE MARCO, Tommaso ; LEO, Marco ; DISTANTE, Cosimo: Soccer Ball Detection with Isophotes Curvature Analysis. In: *Image Analysis and Processing-ICIAP 2013*. Springer, 2013, S. 793–802

- [Deltatre AG 2015] DELTATRE AG: *Tracking*. <http://www.bundesliga-datenbank.de/>, 2015. – Abgerufen am: 15.09.2015
- [Desolneux et al. 2007] DESOLNEUX, Agnès ; MOISAN, Lionel ; MOREL, J-M: *From gestalt theory to image analysis: a probabilistic approach*. Bd. 34. Springer Science & Business Media, 2007
- [Desolneux et al. 2000] DESOLNEUX, Agnès ; MOISAN, Lionel ; MOREL, Jean-Michel: Meaningful alignments. In: *International Journal of Computer Vision* 40 (2000), Nr. 1, S. 7–23
- [DFB 2014a] DFB, Deutscher Fußball-Bund e.V.: *Fußball-Regeln 2014/2015*. 2014
- [DFB 2014b] DFB, Deutscher Fußball-Bund e.V.: *Mitglieder-Statistik 2014*. 2014
- [DFL 2013] DFL, Deutsche Fußball Liga GmbH: *Definitionskatalog Offizielle Spieldaten Version 2.1*. 2013
- [Die Liga 2015] DIE LIGA, Die Liga — Fußballverband e.V.: *Lizenzordnung Spieler - LOS*. http://s.bundesliga.de/assets/doc/690000/684971_original.pdf, 2015. – Abgerufen am: 01.11.2015
- [Dimopoulos u. Winkler 2014] DIMOPOULOS, Michail ; WINKLER, Thomas: Imagewarmness—A new perceptual feature for images and videos. In: *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), 2013* IEEE, 2014, S. 1662–1666
- [Dubrofsky u. Woodham 2008] DUBROFSKY, Elan ; WOODHAM, Robert J.: Combining line and point correspondences for homography estimation. In: *Advances in Visual Computing*. Springer, 2008, S. 202–213
- [Ekin et al. 2003] EKIN, Ahmet ; TEKALP, A M. ; MEHROTRA, Rajiv: Automatic soccer video analysis and summarization. In: *IEEE Transactions on Image Processing* 12 (2003), Nr. 7, S. 796–807
- [Epanechnikov 1969] EPANECHNIKOV, Vassiliy A.: Non-parametric estimation of a multivariate probability density. In: *Theory of Probability & Its Applications* 14 (1969), Nr. 1, S. 153–158
- [Eurosport 2013] EUROSPORT, Discovery Communications Incorporated: *Fußball: Africa Cup of Nations 2013 in Südafrika vom 06.02.2013: Halbfinale: Burkina Faso - Ghana*. 2013
- [Farin et al. 2005] FARIN, Dirk ; HAN, Jungong ; WITH, Peter H.: Fast camera calibration for the analysis of sport sequences. In: *IEEE International Conference on Multimedia and Expo, 2005. ICME 2005* IEEE, 2005
- [Farin et al. 2003] FARIN, Dirk ; KRABBE, Susanne ; EFFELSBERG, Wolfgang et al.: Robust camera calibration for sport videos using court models. In: *Electronic Imaging 2004* International Society for Optics and Photonics, 2003, S. 80–91
- [Faugeras u. Toscani 1987] FAUGERAS, OD ; TOSCANI, Giorgio: Camera calibration for 3D computer vision. In: *Proceedings of International Workshop on Machine Vision and Machine Intelligence, Tokyo, Japan, 1987*
- [Faugeras 1993] FAUGERAS, Olivier: *Three-dimensional computer vision: a geometric viewpoint*. MIT press, 1993
- [Fernandes u. Oliveira 2008] FERNANDES, Leandro A. ; OLIVEIRA, Manuel M.: Real-time line detection through an improved Hough transform voting scheme. In: *Pattern Recognition* 41 (2008), Nr. 1, S. 299–314
- [Fielding 1985] FIELDING, R.: *The Technique of Special Effects Cinematography*. 4. Focal Press, 1985 (Library of communication techniques). – ISBN 9780240512341
- [FIFA 2014] FIFA: *Laws of the game*. Fédération Internationale de Football Association, 2014

- [Figuroa et al. 2006] FIGUEROA, Pascual J. ; LEITE, Neucimar J. ; BARROS, Ricardo M.: Background recovering in outdoor image sequences: An example of soccer players segmentation. In: *Image and Vision Computing* 24 (2006), Nr. 4, S. 363–374
- [Fischer 2001] FISCHER, Gerd: *Analytische Geometrie: eine Einführung für Studienanfänger*. Vieweg, 2001
- [Fischler u. Bolles 1981] FISCHLER, Martin A. ; BOLLES, Robert C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: *Communications of the ACM* 24 (1981), Nr. 6, S. 381–395
- [Förstner 2010] FÖRSTNER, Wolfgang: Optimal vanishing point detection and rotation estimation of single images from a legoland scene. In: *ISPRS Technical Commission III Symposium on Photogrammetry Computer Vision and Image Analysis*, 2010, S. 157–162
- [Förstner et al. 2009] FÖRSTNER, Wolfgang ; DICKSCHEID, Timo ; SCHINDLER, Falko: Detecting interpretable and accurate scale-invariant keypoints. In: *IEEE 12th International Conference on Computer Vision IEEE*, 2009, S. 2256–2263
- [Friedman et al. 1977] FRIEDMAN, Jerome H. ; BENTLEY, Jon L. ; FINKEL, Raphael A.: An algorithm for finding best matches in logarithmic expected time. In: *ACM Transactions on Mathematical Software (TOMS)* 3 (1977), Nr. 3, S. 209–226
- [Galambos 2000] GALAMBOS, C: *RAVL: Recognition And Vision Library*. 2000
- [Gall u. Lempitsky 2013] GALL, Juergen ; LEMPITSKY, Victor: Class-specific hough forests for object detection. In: *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, 2013, S. 143–157
- [Gedikli 2009] GEDIKLI, Suat: *Continual and Robust Estimation of Camera Parameters in Broadcasted Sports Games*, München, Techn. Univ., Diss., 2009, Diss., 2009
- [Grompone von Gioi et al. 2012] GIOI, Rafael Grompone v. ; JAKUBOWICZ, Jérémie ; MOREL, Jean-Michel ; RANDALL, Gregory: LSD: a Line Segment Detector. In: *Image Processing On Line* 2 (2012), S. 35–55. <http://dx.doi.org/10.5201/ipo1.2012.gjmr-lsd>. – DOI 10.5201/ipo1.2012.gjmr-lsd
- [Gonzalez u. Woods 2002] GONZALEZ, Rafael C. ; WOODS, Richard E.: *Digital image processing*. 2002
- [Gruenberg u. Weir 1977] GRUENBERG, K. W. ; WEIR, A. J.: *Linear Geometry (Graduate Texts in Mathematics)*. 2nd. Springer, 1977. – ISBN 9780387902272
- [Grunz et al. 2012] GRUNZ, Andreas ; MEMMERT, Daniel ; PERL, Jürgen: Tactical pattern recognition in soccer games by means of special self-organizing maps. In: *Human movement science* 31 (2012), Nr. 2, S. 334–343
- [Gupta et al. 2011] GUPTA, Ankur ; LITTLE, James J. ; WOODHAM, Robert J.: Using line and ellipse features for rectification of broadcast hockey video. In: *Canadian Conference on Computer and Robot Vision (CRV) IEEE*, 2011, S. 32–39
- [Guru et al. 2004] GURU, DS ; SHEKAR, BH ; NAGABHUSHAN, P: A simple and robust line detection algorithm based on small eigenvalue analysis. In: *Pattern Recognition Letters* 25 (2004), Nr. 1, S. 1–13
- [Han u. Poston 2001] HAN, Joon H. ; POSTON, Timothy: Chord-to-point distance accumulation and planar curvature: a new approach to discrete curvature. In: *Pattern Recognition Letters* 22 (2001), Nr. 10, S. 1133–1144
- [Hartley u. Zisserman 2003] HARTLEY, Richard ; ZISSERMAN, Andrew: *Multiple view geometry in computer vision*. Cambridge university press, 2003

- [Hartley 1997a] HARTLEY, Richard I.: In defense of the eight-point algorithm. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (1997), Nr. 6, S. 580–593
- [Hartley 1997b] HARTLEY, Richard I.: Self-calibration of stationary cameras. In: *International Journal of Computer Vision* 22 (1997), Nr. 1, S. 5–23
- [Hartley 1999] HARTLEY, Richard I.: Theory and practice of projective rectification. In: *International Journal of Computer Vision* 35 (1999), Nr. 2, S. 115–127
- [Hayet u. Piater 2007] HAYET, Jean-Bernard ; PIATER, Justus: On-line rectification of sport sequences with moving cameras. In: *MICAI 2007: Advances in Artificial Intelligence*. Springer, 2007, S. 736–746
- [Herrmann et al. 2014] HERRMANN, Michael ; HOERNIG, Martin ; RADIG, Bernd: Online Multi-player Tracking in Monocular Soccer Videos. In: *AASRI Procedia* 8 (2014), S. 30–37
- [Hess u. Fern 2007] HESS, Robin ; FERN, Alan: Improved video registration using non-distinctive local image features. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07* IEEE, 2007, S. 1–8
- [Hoernig et al. 2014] HOERNIG, Martin ; BIGONTINA, Andreas ; RADIG, Bernd: A Comparative Evaluation of Current HTML5 Web Video Implementations. In: *Open Journal of Web Technologies (OJWT)* 1 (2014), Nr. 2, S. 1–9
- [Hopper 2015] HOPPER, Martin: *Bildbasierte Erkennung des Balls in monokularen Aufnahmen von Fußballspielen*, Technische Universität München, Bachelor-Arbeit, 2015. – Themensteller: Prof. Dr. Bernd Radig, Betreuer: Michael Herrmann und Martin Hoernig
- [Hough 1962] HOUGH, Paul V.: *Method and means for recognizing complex patterns*. Dezember 18 1962. – US Patent 3,069,654
- [Hughes u. Franks 2005] HUGHES, Mike ; FRANKS, Ian: Analysis of passing sequences, shots and goals in soccer. In: *Journal of Sports Sciences* 23 (2005), Nr. 5, S. 509–514
- [Hung u. Hsieh 2008] HUNG, Mao-Hsiung ; HSIEH, Chaur-Heh: Event detection of broadcast baseball videos. In: *IEEE Transactions on Circuits and Systems for Video Technology* 18 (2008), Nr. 12, S. 1713–1726
- [Johansson u. Cipolla 2002] JOHANSSON, Björn ; CIPOLLA, Roberto: A system for automatic pose-estimation from a single image in a city scene. In: *IASTED International Conference on Signal Processing, Pattern Recognition and Applications* Bd. 15, 2002
- [Jones et al. 2004] JONES, PD ; JAMES, Nicolas ; MELLALIEU, Stephen D.: Possession as a performance indicator in soccer. In: *International Journal of Performance Analysis in Sport* 4 (2004), Nr. 1, S. 98–102
- [Kabel eins 2013] KABEL EINS, ProSiebenSat.1 Digital GmbH: *ran – Fußball: UEFA Europa League vom 14.02.2013: Borussia Mönchengladbach - Lazio Rom*. 2013
- [Kalman 1960] KALMAN, Rudolph E.: A new approach to linear filtering and prediction problems. In: *Journal of Fluids Engineering* 82 (1960), Nr. 1, S. 35–45
- [Kang et al. 2006] KANG, Chan-Hyun ; HWANG, Jung-Rae ; LI, Ki-Joune: Trajectory analysis for soccer players. In: *Sixth IEEE International Conference on Data Mining Workshops, 2006. ICDM Workshops 2006* IEEE, 2006, S. 377–381
- [Kim et al. 2011] KIM, Ho-Chul ; KWON, Oje ; LI, Ki-Joune: Spatial and spatiotemporal analysis of soccer. In: *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems* ACM, 2011, S. 385–388
- [Kim et al. 2010] KIM, Kihwan ; GRUNDMANN, Matthias ; SHAMIR, Ariel ; MATTHEWS, Iain ; HODGINS, Jessica ; ESSA, Irfan: Motion fields to predict play evolution in dynamic sport scenes. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* IEEE, 2010, S. 840–847

- [Kimball et al. 2013] KIMBALL, Spencer ; MATTIS, Peter ; NATTERER, Michael ; NEUMANN, Sven et al.: Gimp: Gnu image manipulation program. In: *Version 2* (2013), S. 12
- [Leal-Taixé et al. 2015] LEAL-TAIXÉ, Laura ; MILAN, Anton ; REID, Ian ; ROTH, Stefan ; SCHINDLER, Konrad: Motchallenge 2015: Towards a benchmark for multi-target tracking. In: *arXiv preprint arXiv:1504.01942* (2015)
- [Leavers 1993] LEAVERS, VF: Which hough transform? In: *CVGIP: Image understanding* 58 (1993), Nr. 2, S. 250–264
- [Lee et al. 2011] LEE, Dae-Ho ; LEE, Seung-Gwan ; CHOI, Jin-Hyuk: Region-based corner detection by radial projection. In: *Journal of the Optical Society of Korea* 15 (2011), Nr. 2, S. 152–154
- [Lee et al. 2006] LEE, Yun-Seok ; KOO, Han-Suh ; JEONG, Chang-Sung: A straight line detection using principal component analysis. In: *Pattern Recognition Letters* 27 (2006), Nr. 14, S. 1744–1754
- [Lienhart u. Maydt 2002] LIENHART, Rainer ; MAYDT, Jochen: An extended set of haar-like features for rapid object detection. In: *International Conference on Image Processing* Bd. 1 IEEE, 2002, S. 900–903
- [Lourakis 2005] LOURAKIS, Manolis I.: A brief description of the Levenberg-Marquardt algorithm implemented by levmar. In: *Foundation of Research and Technology* 4 (2005), S. 1–6
- [Lourakis 2006] LOURAKIS, M.I.A.: *homest: A C/C++ Library for Robust, Non-linear Homography Estimation*. <http://www.ics.forth.gr/~lourakis/homest/>, 2006. – Abgerufen am 17.12.2011
- [Lu et al. 2013] LU, Wei-Lwun ; TING, J-A ; LITTLE, James J. ; MURPHY, Kevin P.: Learning to track and identify players from broadcast sports videos. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2013), Nr. 7, S. 1704–1716
- [Luo et al. 2003] LUO, Ming ; MA, Yu-Fei ; ZHANG, Hong-Jiang: Pyramidwise structuring for soccer highlight extraction. In: *Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia* Bd. 2 IEEE, 2003, S. 945–949
- [Mackenzie u. Cushion 2013] MACKENZIE, Rob ; CUSHION, Chris: Performance analysis in football: A critical review and implications for future research. In: *Journal of sports sciences* 31 (2013), Nr. 6, S. 639–676
- [Maindonald 1984] MAINDONALD, John H.: *Statistical computation*. Bd. 1. John Wiley & Sons Inc, 1984
- [Marron u. Nolan 1988] MARRON, JS ; NOLAN, D: Canonical kernels for density estimation. In: *Statistics & Probability Letters* 7 (1988), Nr. 3, S. 195–199
- [MDR 2013] MDR, Mitteldeutscher Rundfunk: *Sport im Osten extra vom 29.05.2013: Relegation 3. Liga Hinspiel: RB Leipzig - SF Lotte*. 2013
- [Moisan et al. 2012] MOISAN, Lionel ; MOULON, Pierre ; MONASSE, Pascal: Automatic homographic registration of a pair of images, with a contrario elimination of outliers. In: *Image Processing On Line* 2 (2012), S. 56–73
- [Monnin et al. 2010] MONNIN, David ; BIEBER, Etienne ; SCHMITT, Gwenaél ; SCHNEIDER, Armin: An effective rigidity constraint for improving RANSAC in homography estimation. In: *Advanced Concepts for Intelligent Vision Systems* Springer, 2010, S. 203–214
- [Monroy et al. 2011] MONROY, Antonio ; EIGENSTETTER, Angela ; OMMER, Björn: Beyond straight lines—object detection using curvature. In: *18th IEEE International Conference on Image Processing (ICIP)* IEEE, 2011, S. 3561–3564
- [Moura et al. 2012] MOURA, Felipe A. ; MARTINS, Luiz Eduardo B. ; ANIDO, Ricardo De O. ; DE BARROS, Ricardo Machado L. ; CUNHA, Sergio A.: Quantitative analysis of Brazilian football players' organisation on the pitch. In: *Sports Biomechanics* 11 (2012), Nr. 1, S. 85–96

- [Mustafa 2001] MUSTAFA, Adnan A.: Boundary Signature Matching for Object Recognition. In: *VI 2001 Vision Interface Annual Conference*, 2001, S. 72–79
- [Nagabhushan et al. 2005] NAGABHUSHAN, P ; GURU, DS ; SHEKAR, BH: Eigen transformation based edge detector for gray images. In: *Pattern Recognition and Machine Intelligence*. Springer, 2005, S. 434–440
- [Nam u. Tewfik 2000] NAM, Jeho ; TEWFIK, Ahmed H.: Dissolve transition detection using B-splines interpolation. In: *IEEE International Conference on Multimedia and Expo, 2000. ICME 2000* Bd. 3 IEEE, 2000, S. 1349–1352
- [Ngo et al. 2002] NGO, Chong-Wah ; PONG, Ting-Chuen ; ZHANG, Hong-Jiang: Motion-based video representation for scene change detection. In: *International Journal of Computer Vision* 50 (2002), Nr. 2, S. 127–142
- [O’Callahan 2014] O’CALLAHAN, Robert: *HTML5 Video Correctness Across Browsers*. <http://robert.ocallahan.org/2014/11/html5-video-correctness-across-browsers.html>, 2014. – Abgerufen am: 15.09.2015
- [Ojala et al. 2002] OJALA, Timo ; PIETIKÄINEN, Matti ; MÄENPÄÄ, Topi: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002), Nr. 7, S. 971–987
- [Okuma et al. 2004] OKUMA, Kenji ; LITTLE, James J. ; LOWE, David G.: Automatic rectification of long image sequences. In: *Asian Conference on Computer Vision*, 2004
- [Pele u. Werman 2010] PELE, Ofir ; WERMAN, Michael: The quadratic-chi histogram distance family. In: *Computer Vision–ECCV 2010*. Springer, 2010, S. 749–762
- [Perform 2015] PERFORM, Perform Group Ltd.: *Opta*. 2015
- [Petersohn 2004] PETERSOHN, Christian: Fraunhofer HHI at TRECVID 2004: Shot boundary detection system. In: *TREC Video Retrieval Evaluation Online Proceedings, TRECVID*, 2004
- [Pollefeys u. Van Gool 1999] POLLEFEYS, Marc ; VAN GOOL, Luc: Stratified self-calibration with the modulus constraint. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (1999), Nr. 8, S. 707–724
- [Powell 1998] POWELL, MJD: Direct search algorithms for optimization calculations. In: *Acta Numerica* 7 (1998), S. 287–336
- [Prince et al. 2002] PRINCE, Simon J. ; XU, Ke ; CHEOK, Adrian D.: Augmented reality camera tracking with homographies. In: *Computer Graphics and Applications, IEEE* 22 (2002), Nr. 6, S. 39–45
- [Princen et al. 1990] PRINCEN, John ; ILLINGWORTH, John ; KITTLER, Josef: A hierarchical approach to line extraction based on the Hough transform. In: *Computer Vision, Graphics, and Image Processing* 52 (1990), Nr. 1, S. 57–77
- [Prozone 2015] PROZONE, Prozone Sports Ltd.: *Pro*. 2015
- [Reynolds 2009] REYNOLDS, Douglas: Gaussian mixture models. In: *Encyclopedia of Biometrics*. Springer, 2009, S. 659–663
- [Rousseeuw 1984] ROUSSEEUW, Peter J.: Least median of squares regression. In: *Journal of the American statistical association* 79 (1984), Nr. 388, S. 871–880
- [Russell u. Norvig 1995] RUSSELL, Stuart ; NORVIG, Peter: *Artificial intelligence: a modern approach*. 1995
- [Särkkä 2013] SÄRKKÄ, Simo: *Bayesian filtering and smoothing*. Bd. 3. Cambridge University Press, 2013

- [Schlipping 2014] SCHLIPSING, Marc: Videobasierte Leistungserfassung im Fußball. (2014)
- [Schlipping et al. 2013] SCHLIPSING, Marc ; SALMEN, Jan ; IGEL, Christian: Echtzeit-Videoanalyse im Fußball. In: *KI-Künstliche Intelligenz* 27 (2013), Nr. 3, S. 235–240
- [Schmid u. Zisserman 1997] SCHMID, Cordelia ; ZISSERMAN, Andrew: Automatic line matching across views. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* IEEE, 1997, S. 666–671
- [Seo et al. 1997] SEO, Yongduek ; CHOI, Sunghoon ; KIM, Hyunwoo ; HONG, Ki-Sang: Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick. In: *Image Analysis and Processing* Springer, 1997, S. 196–203
- [Shekar et al. 2006] SHEKAR, BH ; GURU, DS ; NAGABHUSHAN, P: Object recognition through the principal component analysis of spatial relationship amongst lines. In: *Computer Vision-ACCV 2006*. Springer, 2006, S. 170–179
- [Shi et al. 2010] SHI, Daming ; ZHENG, Liying ; LIU, Jigang: Advanced Hough transform using a multilayer fractional Fourier method. In: *IEEE Transactions on Image Processing* 19 (2010), Nr. 6, S. 1558–1566
- [Shinkai et al. 2009] SHINKAI, Hironari ; NUNOME, Hiroyuki ; ISOKAWA, Masanori ; IKEGAMI, Yasuo: Ball impact dynamics of instep soccer kicking. In: *Medicine and science in sports and exercise* 41 (2009), Nr. 4, S. 889–897
- [Siles Canales 2014] SILES CANALES, Francisco: *Automated Semantic Annotation of Football Games from TV Broadcast*, Technische Universität München, Diss., 2014
- [Smeaton et al. 2010] SMEATON, Alan F. ; OVER, Paul ; DOHERTY, Aiden R.: Video shot boundary detection: Seven years of TRECVID activity. In: *Computer Vision and Image Understanding* 114 (2010), Nr. 4, S. 411–418
- [Smeaton et al. 2006] SMEATON, Alan F. ; OVER, Paul ; KRAAIJ, Wessel: Evaluation campaigns and TRECVID. In: *Proceedings of the 8th ACM international workshop on Multimedia information retrieval* ACM, 2006, S. 321–330
- [Solem 2015] SOLEM, Ask: *Celery User Manual*. <http://www.celeryproject.org/>, 2015. – Abgerufen am: 15.09.2015
- [Sport1 2013a] SPORT1, Sport1 GmbH: *Fußball Live - Testspiel vom 08.07.2013: 1. FC Magdeburg - Borussia Dortmund*. 2013
- [Sport1 2013b] SPORT1, Sport1 GmbH: *Hallenfußball Live vom 02.02.2013: Dynamo Dresden - Werder Bremen*. 2013
- [Sturm u. Maybank 1999] STURM, Peter F. ; MAYBANK, Stephen J.: On plane-based camera calibration: A general algorithm, singularities, applications. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* Bd. 1 IEEE, 1999
- [Sutherland u. Hodgman 1974] SUTHERLAND, Ivan E. ; HODGMAN, Gary W.: Reentrant polygon clipping. In: *Communications of the ACM* 17 (1974), Nr. 1, S. 32–42
- [Swalwell u. Clarke 2012] SWALWELL, Derek ; CLARKE, Peter: *The Good House*. 2012. – Courtesy of Crone Partners
- [Tang et al. 2007] TANG, Cheng-Yuan ; WU, Yi-Leh ; HU, Pei-Ching ; LIN, Hsien-Chang ; CHEN, Wen-Chao: Self-Calibration for Metric 3D Reconstruction Using Homography. In: *MVA*, 2007, S. 86–89
- [Tomasi u. Kanade 1991] TOMASI, Carlo ; KANADE, Takeo: *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991

- [Tong et al. 2004] TONG, Xiao-Feng ; LU, Han-Qing ; LIU, Qing-Shan: An effective and fast soccer ball detection and tracking method. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004* Bd. 4 IEEE, 2004, S. 795–798
- [Tovinkere u. Qian 2001] TOVINKERE, Vasanth ; QIAN, Richard J.: Detecting semantic events in soccer games: Towards a complete solution. In: *IEEE International Conference on Multimedia and Expo, 2001. ICME 2001* IEEE, 2001, S. 833–836
- [Utsumi et al. 2002] UTSUMI, Okihisa ; MIURA, Koichi ; IDE, Ichiro ; SAKAI, Shuichi ; TANAKA, Hidehiko: An object detection method for describing soccer games from video. In: *IEEE International Conference on Multimedia and Expo, 2002. ICME'02* Bd. 1 IEEE, 2002, S. 45–48
- [Viola u. Jones 2001] VIOLA, Paul ; JONES, Michael: Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001* Bd. 1 IEEE, 2001
- [Von Gioi et al. 2012] VON GIOI, R. G. ; JAKUBOWICZ, Jeremie ; MOREL, Jean-Michel ; RANDALL, Gregory: LSD: a line segment detector. In: *Image Processing On Line* 2 (2012), Nr. 3, S. 722–732
- [Wan u. Xu 2004] WAN, Kongwah ; XU, Changsheng: Efficient multimodal features for automatic soccer highlight generation. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004* Bd. 3 IEEE, 2004, S. 973–976
- [Wang et al. 2014] WANG, Xinchao ; ABLAVSKY, Vitaly ; SHITRIT, Horesh B. ; FUA, Pascal: Take your eyes off the ball: Improving ball-tracking by focusing on team play. In: *Computer Vision and Image Understanding* 119 (2014), S. 102–115
- [Xie et al. 2004] XIE, Lexing ; XU, Peng ; CHANG, Shih-Fu ; DIVAKARAN, Ajay ; SUN, Huifang: Structure analysis of soccer video with domain knowledge and hidden Markov models. In: *Pattern Recognition Letters* 25 (2004), Nr. 7, S. 767–775
- [Xu u. Oja 1993] XU, Lei ; OJA, Erkki: Randomized Hough transform (RHT): basic mechanisms, algorithms, and computational complexities. In: *CVGIP: Image understanding* 57 (1993), Nr. 2, S. 131–154
- [Xu et al. 2004] XU, Ming ; ORWELL, James ; JONES, Graeme: Tracking football players with multiple cameras. In: *International Conference on Image Processing, 2004. ICIP'04* Bd. 5 IEEE, 2004, S. 2909–2912
- [Yu et al. 2005] YU, Xinguo ; HAY, Tze S. ; YAN, Xin ; CHNG, Engsiang: A player-possession acquisition system for broadcast soccer video. In: *IEEE International Conference on Multimedia and Expo, 2005. ICME 2005* IEEE, 2005, S. 522–525
- [Yu u. Tian 2003] YU, Xinguo ; TIAN, Qi: A novel ball detection framework for real soccer video. In: *Proceedings of the 2003 International Conference on Multimedia and Expo-Volume 1* IEEE Computer Society, 2003, S. 265–268
- [ZDF 2013a] ZDF, Zweites Deutsches Fernsehen: *ZDF SPORTextra vom 14.08.2013: Deutschland - Paraguay*. 2013
- [ZDF 2013b] ZDF, Zweites Deutsches Fernsehen: *ZDF SPORTextra vom 16.06.2013: Spanien - Uruguay*. 2013
- [Zeng et al. 2008] ZENG, Hui ; DENG, Xiaoming ; HU, Zhanyi: A new normalized method on line-based homography estimation. In: *Pattern Recognition Letters* 29 (2008), Nr. 9, S. 1236–1244
- [Zhang 2000] ZHANG, Zhengyou: A flexible new technique for camera calibration. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000), Nr. 11, S. 1330–1334
- [Zhang et al. 1995] ZHANG, Zhengyou ; DERICHE, Rachid ; FAUGERAS, Olivier ; LUONG, Quang-Tuan: A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. In: *Artificial intelligence* 78 (1995), Nr. 1, S. 87–119