

Echtzeit-Videoanalyse im Fußball

Ein *Live*-System zum Spieler-Tracking

Marc Schlippling · Jan Salmen · Christian Igel

Online publiziert: 22. März 2013
© Springer-Verlag Berlin Heidelberg 2013

Zusammenfassung Die Automatisierung der Videoanalyse nimmt im Profisport eine immer wichtigere Rolle ein. Im Fußball kommt dabei der Auswertung der Laufwege der Spieler eine besondere Bedeutung zu. Der vorliegende Bericht dokumentiert unser Kooperationsprojekt zum computergestützten Spieler-Tracking auf Basis von Videobildern in Echtzeit. Wir beschreiben den Aufbau und diskutieren die Praxistauglichkeit des entwickelten Systems, das sich durch hohe Genauigkeit, Mobilität und Kostengünstigkeit auszeichnet.

1 Einleitung

Bildverarbeitung und künstliche Intelligenz erlangen in der Sportanalyse, d. h. der Wissenschaft der Analyse und Modellierung von Prozessen im Sport, eine immer größere Wichtigkeit. Sportarten mit hoher medialer Aufmerksamkeit verlangen nach neuen Wegen der systematischen Nachbetrachtung sowie objektiver Leistungsanalysen. Manager und Trainer machen Gebrauch von statistischen Größen und kategorisiertem Videomaterial der Sportler, um neue Erkennt-

nisse in die physische und taktische Ausbildung einfließen zu lassen.

Vor diesem Hintergrund wurde das hier beschriebene Kooperationsprojekt vom Bundesministerium für Wirtschaft und Technologie (*BMWi*) über einen Zeitraum von zwei Jahren gefördert. Die Arbeitsgruppe *Echtzeitfähige Bildverarbeitung* des *Institut für Neuroinformatik* (Ruhr-Universität Bochum) leistete im Projekt den zentralen wissenschaftlichen Anteil. Zwei Industriepartner waren verantwortlich für die Konzeption der Hardware und Videoaufnahme (*nissys GmbH*, Bochum) sowie für den nötigen sportwissenschaftlichen Hintergrund (*sports_analytics GmbH*, Dortmund).

Ziel des hier vorgestellten Projekts war die Entwicklung eines kamerabasierten Systems zur automatischen Spielanalyse. Das entwickelte System ist in der Lage, den Ort jeden Spielers zu jedem Zeitpunkt während des Spiels in Echtzeit zu ermitteln, wobei lediglich geringfügige manuelle Eingriffe nötig sind. Basierend auf diesen sogenannten *Tracking*-Daten lassen sich objektive Statistiken berechnen, das mannschaftliche Verhalten analysieren und Spielszenen kategorisieren. Ausgehend von der Anwendungsfallanalyse reichte der Entwicklungsprozess von der Definition der Hard- und Softwarearchitektur bis hin zu wissenschaftlichen Herausforderungen im Bereich Computersehen und Maschinellen. Das System wurde regelmäßig *live* und mit Videosequenzen aus sieben verschiedenen Stadien der 1. und 2. Fußball-Bundesliga getestet.

Das folgende Kapitel gibt einen Überblick über verwandte Arbeiten und kommerzielle Tracking-Systeme. Kapitel 3 erläutert den Systemaufbau. In Kap. 4 werden die zugrundeliegenden Lösungen der Objektklassifikation dargestellt. Kapitel 5 diskutiert zukünftige Entwicklungen der videobasierten Sportanalyse.

M. Schlippling (✉) · J. Salmen
Institut für Neuroinformatik, Ruhr-Universität Bochum,
44780 Bochum, Deutschland
e-mail: marc.schlippling@ini.rub.de

J. Salmen
e-mail: jan.salmen@ini.rub.de

C. Igel
Department of Computer Science, University of Copenhagen,
2100 Copenhagen Ø, Dänemark
e-mail: igel@di.ku.dk

2 Literatur

Automatische Analyse von TV-Bildern wurde in erster Linie zur Kategorisierung der gezeigten Sportart, des Kamerablickwinkels [16] und interessanter Ereignisse (z. B. Spielstand, Abseits [1]) untersucht. Aufgrund des beschränkten Blickfelds ist TV-Material für vollständiges Tracking jedoch nicht ausreichend gut geeignet.

Verschiedene Ansätze zur automatischen Positionsbestimmung im Fußball wurden bereits vorgeschlagen, für eine Übersicht siehe z. B. [5]. Alternativen mit einer Vielzahl im Stadion verteilter Kameras werden in [2, 11, 12] vorgestellt. Weitere für unser Projekt relevante Veröffentlichungen konzentrieren sich auf Detektion und Tracking mithilfe von Farb- und Tiefeninformationen [10] und die unüberwachte Merkmalsextraktion [9].

Die Wichtigkeit der Analyse verschiedener Farbräume zur Segmentierung wird in [15] herausgestellt. Als *Baseline* haben wir hier neben PCA-Merkmalen [14] auch Histogramme mit Ortsinformationen, sogenannte *SpatioGrams*, verglichen. Diese wurden bereits im Kontext von Objekt-Tracking erprobt [3]. Keiner der genannten Ansätze sind in der Lage, Spieleridentitäten aufzuschlüsseln, sondern erkennen lediglich die Mannschaftszugehörigkeit.

Erste kommerzielle Tracking-Systeme sind mittlerweile am Markt verfügbar: *Tracab*,¹ *AmiscoPro*² und *Vis.Track*.³ Diese verwenden bis zu 16 Kameras je Setup, basieren auf Ansätzen des Stereosehens bzw. benötigen fest installierte Kameras am Stadionsdach. Die entsprechenden Statistiken werden entweder mit hohem Personalaufwand *live* erfasst oder sind erst einige Tage später verfügbar. Von einem vollautomatischen Betrieb sind alle Systeme aktuell noch weit entfernt.

Zudem gibt es fürs *Live*-Klassifikations-Szenario (Erscheinung der Spieler ist erst kurz vor Beginn des Spiels bekannt) noch keine befriedigenden Lösungen. In der Literatur wird die Lernphase prinzipiell als *Offline*-Prozedur behandelt. Kommerzielle Systeme wirken dem mit hohem Personalaufwand entgegen und wählen entsprechende Farbrepräsentationen manuell. Dieses Vorgehen ist weder effizient noch robust.

In ihrem Überblick fassen D’Orazio and Leo zusammen, dass noch einige Arbeit bzgl. der Automatisierung der Systeme geleistet werden müsse, um manuelles Eingreifen und Fehler zu reduzieren [5] – und hier setzte unser Projekt an.

¹www.tracab.com.

²www.sport-universal.com.

³www.bundesliga-datenbank.de.

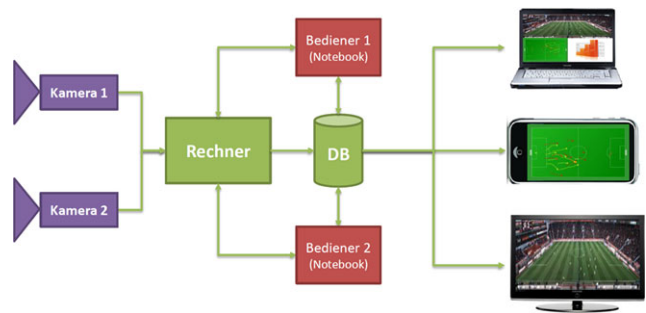


Abb. 1 Systemskizze

3 Systemarchitektur

Das vorgestellte System wurde unter Berücksichtigung folgender Design-Ziele entworfen:

Mobilität. Das System soll an beliebigen Spielorten, z. B. Stadion, Trainingsgelände oder Hallenplatz, aufgebaut und kalibriert werden können.

Genauigkeit. Hohe Zuverlässigkeit der Erfassung soll durch den Einsatz moderner Methoden des maschinellen Lernens und der Mustererkennung gewährleistet werden.

Automatisierung. Durch adaptive Bildverarbeitungsalgorithmen soll es möglich sein, das System durch zwei Bediener aufbauen und steuern zu lassen.

Kostengünstig. Die Anforderungen an die Systemhardware sind gering zu halten – z. B. zwei Industriekameras und ein handelsüblicher PC.

Um diesen Anforderungen gerecht zu werden, war eine systematische Integration von Hardware, Software, den Bedienern und letztendlich der Kundenschnittstelle nötig. Das verteilte System ist in Abb. 1 skizziert.

3.1 Hardware

Die Spiele werden von der Pressetribüne von zwei *Full-HD*-Kameras – eine für jede Platzhälfte – aufgenommen und per *Ethernet* an den Verarbeitungsrechner gesendet. Der Vierkern-PC enthält eine *CUDA*⁴-fähige Grafikkarte zur Parallelverarbeitung. Die erkannten Positionsdaten werden *live* an eine *SQL*-Datenbank übertragen. Zur Zuordnung der Spieler in Echtzeit sind die Bediener-Notebooks per LAN integriert. Die Daten können somit unmittelbar über verschiedene Nutzerschnittstellen verfügbar gemacht werden.

3.2 Bediener

Zu Beginn des Spiel ist es nötig, dass die Bediener den Klassifikator zur Unterscheidung der Trikots mit sortierten Trainingsdaten versorgen. Dieser Prozess wird von einem

⁴Compute Unified Device Architecture by *Nvidia*.

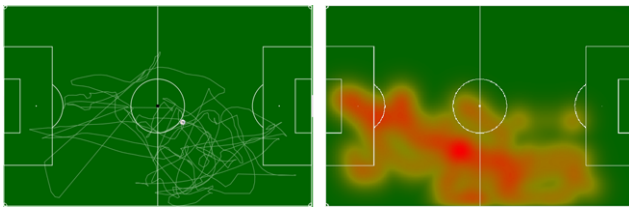


Abb. 2 Visualisierung der Tracking-Daten: Spielerpfad (*links*) und Verteilungsprofil, sog. Heatmap (*rechts*)

Cluster-Algorithmus unterstützt, der Trainingsbeispiele anhand der verwendeten Bildmerkmale (vgl. Abschn. 4.3) automatisch vorsortiert. Da während des Erkennungsprozesses nicht zwischen Akteuren gleicher Trikots unterschieden wird, müssen neue Tracks einem individuellen Spieler zugeordnet werden. Diese Aufgaben werden an einen Bediener übergeben, welcher daraufhin den entsprechenden Eintrag in der Datenbank unmittelbar ergänzen kann. Untersuchungen haben gezeigt, dass ein Bediener je Mannschaft ausreicht, um alle Spieler in Echtzeit zu tracken.

3.3 Nutzerschnittstelle

Die erhobenen Datensätze können unmittelbar für eine Reihe von Anwendungen eingesetzt werden: Fans rufen Statistiken *live* über ihr Smartphone ab, Medienberichte können durch interessante Fakten gestützt werden und letztendlich steht dem Trainerstab ein umfassendes Werkzeug zur Nachbetrachtung des Spiel zur Verfügung. Für diese Anwendungsbeispiele wurden die Daten entsprechend aufbereitet und visualisiert. Insbesondere gehen diese über konventionelle Tabellen und Diagramme hinaus (vgl. Abb. 2). Positionsdaten der Spieler können beispielsweise als Laufwege in einer virtuellen Spielfeldansicht (s. Abb. 6) oder integriert ins Kamerabild dargestellt werden (Abb. 7). In Kombination mit den Tracking-Daten wird auch das entsprechende Panoramavideo zur taktische Nachbetrachtung angeboten, welches einen echten Mehrwert gegenüber geschwenkten Aufnahmen bietet.

4 Bildverarbeitung

In diesem Abschnitt werden die entwickelten Lösungen der Spielererkennung vorgestellt. Abbildung 3 skizziert die Kette der entsprechenden Module.

4.1 Bildtransformation

Zunächst wird das Panoramavideo, auf dem alle weiteren Schritte basieren, aus den synchronen Kamerabildern perspektivisch korrigiert und zusammengesetzt (s. Abb. 4). Dafür – und um für jedes Setup eine perfekte Abbildung von

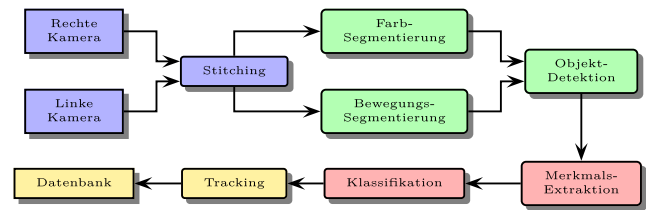


Abb. 3 Übersicht Bildverarbeitung: *Farben* kennzeichnen unabhängige Module des Abschn. 4

Bild- in Spielfeldkoordinaten zu garantieren – werden die Kameraperspektiven nach dem Aufbau (extrinsisch) kalibriert. Eine Kompensation der Linsenverzerrung ist in diesen Prozess integriert und erlaubt somit eine lineare Abbildung von Bild- in Weltkoordinaten [17]. Diese Homographie wird auf Basis von mindestens vier Spielfeldpositionen und deren korrespondierenden Punkten im Kamerabild per *Direct Linear Transformation* bestimmt [7, Kap. 4.1].

Durch die Annahme einer festen Kameraposition kann die Transformation in eine *Lookup*-Tabelle umgewandelt und das Panorama unter Einsatz paralleler Programmierung auf der GPU in Echtzeit berechnet werden.

4.2 Objektsegmentierung

Ziel der Segmentierung ist die Trennung in Vorder- und Hintergrund. Dazu wurden zwei pixelweise Merkmale untersucht (vgl. Abb. 4) und im Anschluss Vordergrund-Pixel zu Regionen (ROIs) zusammengefasst.

Zur Bewegungssegmentierung wird für jedes Pixel ein Hintergrundmodell angelegt. Dieses glättet die entsprechenden Farbwerte (RGB) exponentiell über einen kurzen Zeitraum und repräsentiert somit nicht-bewegten Hintergrund.

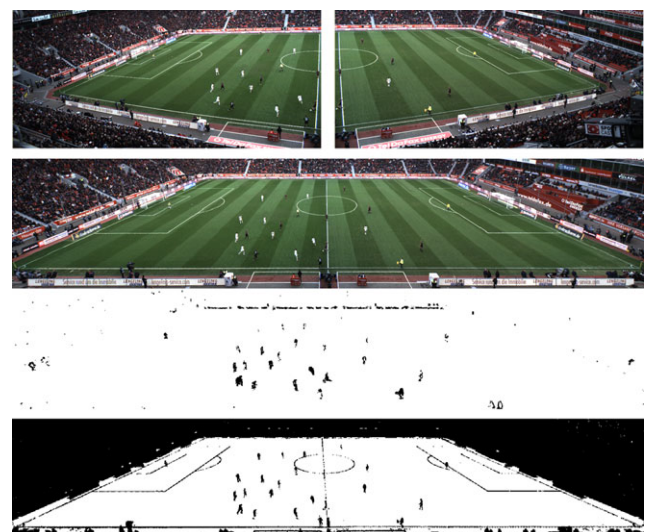


Abb. 4 Vorverarbeitung: Stitching der Eingabebilder zum Panorama, Bewegungs- und Farbsegmentierung (binär)

Die Vordergrundpunkte ergeben sich durch einen signifikanten Unterschied vom aufgenommenen Bild zum Modell. Nur die Hintergrundpixel werden fürs Modell-Update herangezogen. Für ein adaptiveres Modell zur Bewegungssegmentierung bei stark variierenden Beleuchtungsbedingungen wird ein zweiter, trägerer Glättungsparameter für die Vordergrundpixel berücksichtigt.

Als zweiter Hinweis zur Objektsegmentierung wird ein ortsunabhängiges Modell der Hintergrundfarbe gelernt. Dazu ermitteln wir die Farbverteilung aller Pixel im Bereich des Spielfelds, ignorieren Ausreißer (Farben außerhalb 3σ um μ) und wiederholen die Schätzung. Mit dieser Methode werden neben den Spielern auch Spielfeldlinien segmentiert. Da diese durch Regelwerk und Spielfeldkalibrierung bekannt sind, können entsprechende Bildbereiche von der Farbsegmentierung ausgeschlossen werden. Um robuster gegenüber Bildrauschen zu werden, folgt auf die Segmentierung eine morphologische Operation (*Opening*). Bis zu diesem Schritt werden alle Berechnungen in weniger als 10 ms für ein *HD-Panorama* mithilfe hoch-paralleler Ausführung auf der GPU durchgeführt.

Der Cluster-Algorithmus fasst verbundene „aktive“ Pixel unter Berücksichtigung minimaler und maximaler Spielergrößen zu abgeschlossenen ROIs zusammen. Dass ROIs im Bild überlappende Akteure als einzelne Detektion erfassen können, wird in der folgenden Klassifikation durch die Einführung einer zusätzlichen Klasse berücksichtigt.

4.3 Klassifikation

Neben den fünf offensichtlichen Klassen (vgl. Abb. 5) gibt es *Fehler* für nicht relevante Objekte und *Gruppe* für ROIs mit Feldspielern unterschiedlicher Mannschaften. Die Klassen sind bezüglich ihrer Auftretenshäufigkeit sehr ungleich verteilt, sodass ein rein unüberwachtes Lernverfahren keine ausreichend guten Ergebnisse liefert. Da Trikotkombinationen und Umgebungsbedingungen von Spiel zu Spiel stark variieren, muss der Klassifikator unmittelbar vor dem Spiel, z. B. während der Vorbereitungszeit oder beim Einlaufen der Teams, trainiert werden. Dabei muss die Methode mit wenigen Beispielen auskommen, einen schnellen Trainingsprozess ermöglichen und eine hohe Klassifikationsleistung für alle Klassen aufweisen.

Die Wahl geeigneter Bildmerkmale war durch einige Bedingungen eingeschränkt: Geringe Größe des Merkmalsvektors, effizient Berechenbarkeit und gute Klassendiskriminanz. Die Erkennung sollte unabhängig von der Spielerpose und seiner Orientierung zur Kamera sein. Da Trikots insbesondere für eine gute Unterscheidbarkeit ihrer Farbe ausgewählt werden, wurde entschieden, keine Form- oder Ortsinformationen einfließen zu lassen, sondern Histogramme diverser Farbräume zu betrachten. Ähnliche Merkmale wurden zur Objektklassifikation bereits erfolgreich eingesetzt [4, 18].



Abb. 5 Beispiele der sieben Klassen (*Spalte*) aus drei Datensätzen (*Zeile*): v. l. n. r.: Feldspieler 1/2, Torhüter 1/2, Schiedsrichter, Gruppe und Fehler

Zur Klassifikation wurde die frei verfügbare *Shark* Programmbibliothek genutzt [8]. In Kombination mit diversen Merkmalsvektoren wurden verschiedene Klassifikationsmethoden systematisch untersucht. Eine Klassifikationsrate von bis zu 98,2 % wurde erreicht, wobei die Trefferquote (*recall*) für die sechs relevanten Klassen bei nahezu 100 % lag. Die Klassifikationsleistung hing dabei in erster Linie von der Histogrammauflösung und weniger vom Farbraum ab. Die vorgeschlagenen Merkmale übertrafen *PCA-Features* und *SpatioGrams* [3] signifikant und bestätigten somit die Verwendung ortsunabhängiger Merkmale. Des Weiteren zeigten die Experimente mit einem *Nearest Neighbour* Klassifikator die beste Leistung in der initialen Phase, in der nur wenige Trainingsbeispiele zur Verfügung stehen. Nach einigen Minuten Sammelphase übertrifft eine *Support Vector Machine* (SVM) jedoch die alternativen Verfahren. Weitere Untersuchungen hierzu sind in [13] beschrieben.

4.4 Mehrziel-Tracking

Als Akteure klassifizierte ROIs werden anhand ihres Fußpunkts in Spielfeldkoordinaten transformiert. Um für je-

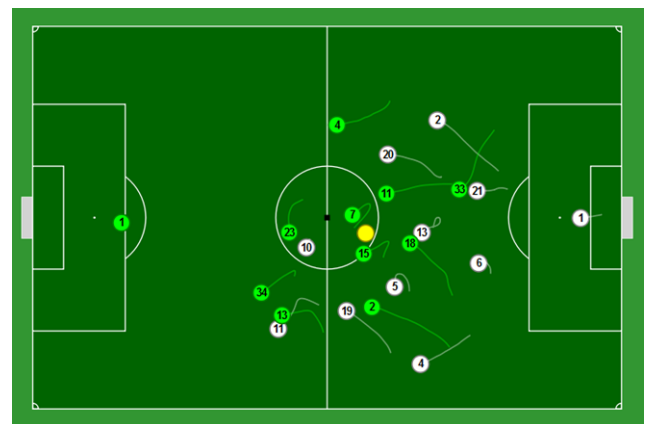


Abb. 6 Tracks aller Spieler in Feldkoordinaten (vgl. Abb. 7)



Abb. 7 Spieler im Panoramabild mit zugehöriger Rückennummer und Abseitslinien

den Spieler entsprechende Tracks zu erhalten, werden diese Positionen zeitlich integriert. Dies geschieht mithilfe eines Kalman-Filters [6] (s. Abb. 6), der uns ermöglicht, die Position und Bewegung der Spieler explizit zu modellieren und präzisieren.

Daher bleibt in jedem Zeitschritt ein durch die Klassen-zugehörigkeit beschränktes Problem der Zuordnung neuer Messungen und bestehender Kalman-Prädiktionen zu lösen. Gruppen werden als „Joker“ behandelt und können zur Fortschreibung mehrerer Tracks verschiedener Klassen dienen. Zuletzt werden die gefilterten Positionsdaten in jedem Zeitschritt an die Datenbank übertragen.

5 Fazit

Der vorliegende Projektbericht skizziert unser videogestütztes Sportanalysesystem zur automatischen Erfassung individueller Spielerstatistiken in Echtzeit. Das System zeichnet sich durch hohe Erkennungsgenauigkeit, Mobilität und Kostengünstigkeit aus.

Alle entwickelten Module wurden regelmäßigen Tests von verschiedenen Bedienern mit sportwissenschaftlichem Hintergrund unterzogen. Dabei wurde sowohl aufgezeichnetes Videomaterial verwendet als auch Live-Tests in sieben Stadien durchgeführt. Folgende Aspekte wurden untersucht:

Kalibrierung. Aufbau des Systems zur Erzeugung des Panorama-Videos wurde für verschiedene Spielfelder (Fußball, Hockey, Tennis) erfolgreich getestet.

Segmentierung. Die adaptive Segmentierung in Echtzeit lief robust. Experimente wurden unter diversen äußeren Bedingungen (z. B. Wetter, Bodentextur) durchgeführt.

Genauigkeit. Ein Test mit vorgegebenen Laufstrecken mehrerer Zielpersonen bestätigte die Genauigkeit gegenüber Ground-Truth-Positionsdaten.

Feedback. Sportwissenschaftler begleiteten den Prozess der zielgerichteten Datenaufbereitung und gaben regelmäßiges Feedback zum Interface des Systems im Hinblick auf dessen Einsatz im Profisport.

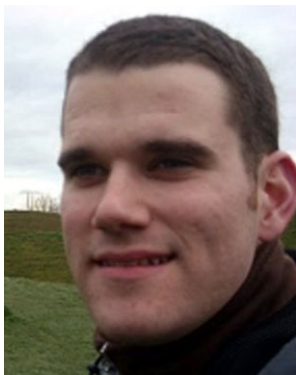
Kompletttest. Das Gesamtsystem, inkl. Bediener und Datenbank, wurde während eines Bundesligaspiels erfolgreich zum Einsatz gebracht.

Aktuelle und zukünftige Forschungsaktivitäten zielen auf die Erkennung von Spielerverhalten, taktischer Muster sowie auf die automatische Erfassung höherwertiger statistischer Daten, sog. *Ereignisdaten* (Standardsituationen, Zweikämpfe, spezielle Spielzüge). Dies wäre eine hilfreiche Unterstützung zur automatischen Szenenkategorisierung, systematischen Nachbetrachtung und differenzierten Fehleranalyse. Eine Rückkopplung dieser Daten, z. B. Aufenthaltsprofile, in den Erkennungsprozess und das Lesen von Rückennummern wären weitere Schritte zur vollkommenen Automatisierung. Für eine hohe Erkennungsleistung in dynamischen Einsatzumgebungen ist geplant, die Themen *Online-Lernen* und *Parameter-Adaption* weiter zu verfolgen.

Literatur

1. Assfalg J, Bertini M, Colombo C, Bimbo AD, Nunziati W (2003) Semantic annotation of soccer videos: automatic highlights identification. *Comput Vis Image Underst* 92(2–3):285–305
2. Ben Shitrit H, Berclaz J, Fleuret F, Fua P (2011) Tracking multiple people under global appearance constraints. In: *Proceedings of the IEEE international conference on computer vision*, S 137–144
3. Birchfield ST, Rangarajan S (2005) Spatiograms versus histograms for region-based tracking. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, S 1158–1163
4. Chapelle O, Haffner P, Vapnik VN (1999) Support vector machines for histogram-based image classification. *IEEE Trans Neural Netw* 10(5):1055–1064
5. D’Orazio T, Leo M (2010) A review of vision-based systems for soccer video analysis. *Pattern Recognit* 43:2911–2926
6. Gelb A (1974) *Applied optimal estimation*, 1. Aufl. MIT Press, Cambridge
7. Hartley RI, Zisserman A (2004) *Multiple view geometry in computer vision*, 2. Aufl. Cambridge University Press, Cambridge
8. Igel C, Glasmachers T, Heidrich-Meisner V (2008) Shark. *J Mach Learn Res* 9:993–996
9. Liu J, Tong X, Li W, Wang T, Zhang Y, Wang H (2009) Automatic player detection, labeling and tracking in broadcast soccer video. *Pattern Recognit Lett* 30(2):103–113

10. Muñoz-Salinas R (2008) A Bayesian plan-view map based approach for multiple-person detection and tracking. *Pattern Recognit* 41(12):3665–3676
11. Poppe C, Bruyne SD, Verstocht S, Walle RV (2010) Multi-camera analysis of soccer sequences. In: *Proceedings of the IEEE conference on advanced video and signal based surveillance*, S 26–31
12. Ren J, Xu M, Orwell J, Jones GA (2010) Multi-camera video surveillance for real-time analysis and reconstruction of soccer games. *Mach Vis Appl* 21:855–863
13. Schlipsing M, Salmen J, Tenbült S, Tschentscher M, Igel C (2013) Adaptive pattern recognition in real-time video-based soccer analysis—eingereicht
14. Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3(1):71–86
15. Vandenbroucke N, Macaire L, Postaire J-G (2003) Color image segmentation by pixel classification in an adapted hybrid color space: application to soccer image analysis. *Comput Vis Image Underst* 90(2):190–216
16. Zhang N, Duan L-Y, Li L, Huang Q, Du J, Gao W, Guan L (2012) A generic approach for systematic analysis of sports videos. *ACM Trans Intell Syst Technol* 3(3):46:1–46:29
17. Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22(11):1330–1334
18. Zivkovic Z, Kröse B (2004) An EM-like algorithm for color-histogram-based object tracking. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, S 798–803



Marc Schlipsing erhielt 2006 den Master in Advanced Methods of Computer Science am Queen Mary College der University of London und 2007 das Diplom der Angewandten Informatik an der Technischen Universität Dortmund. Seit 2008 arbeitet er als wiss. Mitarbeiter am Institut für Neuroinformatik der Ruhr-Universität Bochum (RUB) und leitet seit 2011 zusammen mit Jan Salmen die Forschungsgruppe *Echtzeitfähige Bildverarbeitung*.



Jan Salmen erhielt 2006 sein Diplom in Informatik an der Technischen Universität Dortmund. Im Anschluss arbeitete er als wiss. Mitarbeiter am Institut für Neuroinformatik der RUB und leitete seit 2009 die Forschungsgruppe *Echtzeitfähige Bildverarbeitung*. 2013 wurde er von der Fakultät für Elektrotechnik und Informationstechnik (RUB) promoviert.



Christian Igel studierte Informatik an der Technischen Universität Dortmund, 2002 promovierte er an der Technischen Fakultät der Universität Bielefeld und wurde im Jahr 2010 an der Fakultät für Elektrotechnik und Informationstechnik der RUB habilitiert. Von 2002 bis 2010 war er Juniorprofessor für Optimierung Adaptiver Systeme am Institut für Neuroinformatik der RUB, seit 2010 ist er Professor an der Fakultät für Informatik der Universität Kopenhagen.