

Bildbasierte Messung und Analyse sportlicher Aktivitäten

Zusammenfassung

Im digitalen Zeitalter hat sich die Videoerfassung von sportlichen Aktivitäten als eine einfache, zuverlässige und kostengünstige Möglichkeit herausgestellt, um die Leistung von Athleten zu dokumentieren. Die Technologie zur Erfassung von Daten ist heutzutage so ausgereift, dass sie auch für extreme Szenarien wie Schwimmen und Skispringen geeignet ist und selbst die Synchronisierung mehrerer Kameras kein großes Problem mehr darstellt. Als Folge davon steht Trainern und Athleten eine große Menge von Daten sportlicher Aktivitäten zur Verfügung, die analysiert und ausgewertet werden können. Für diese Analyse ist immer noch die Kompetenz eines Trainers notwendig, um sportliche Fehler zu erkennen und Fortschritte des Sportlers zu erzielen. Diese Kompetenz basiert auf der empirischen Fähigkeit des Experten, zu „sehen“, was richtig ist und was nicht. Allerdings ist es im Hochleistungssport selbst für ein geschultes Auge oft sehr schwierig, die feinen Unterschiede zu erkennen, die zwischen Sieg und Niederlage entscheiden können. Deshalb werden hochpräzise Messungen benötigt, um Trainer und Hochleistungssportler zu unterstützen. Ein nahe liegender Gedanke ist, dass eine genaue Analyse von Videomaterial zu dem erforderlichen Maß an Präzision führen kann. In diesem Artikel werden wir sehen, dass dieser Gedanke richtig ist und Methoden vorstellen, wie sich diese Ziele erreichen lassen. In diesem Zusammenhang steht die automatisierte bildbasierte Analyse sportlicher Aktivitäten vor einer Reihe motivierender Herausforderungen: Definition der ausschlaggebenden Leistungsfaktoren (d. h., was gemessen werden soll), technische Rahmenbedingungen bei der Aufnahme (wie gemessen wird) und Anforderung an die Präzision dieser Messungen (wie genau die Messungen sein sollen). Während die erste Herausforderung am besten durch Sportexperten zu bewältigen ist, ist es dagegen möglich, die zweite und dritte durch Computer-Vision-Experten zu lösen. Anhand einer Reihe von Beispielen demonstrieren wir, welche Trackingverfahren zum aktuellen Stand der Forschung möglich sind und zeigen zukünftige Trends auf. In Bezug auf die Genauigkeit zeigen wir, wie vollautomatisches Tracking durch eine teilüberwachte Trackingmethode vorteilhaft ersetzt werden kann und führen die Notwendigkeit auf, einen Kompromiss zwischen Genauigkeit und der Menge an manuellen Benutzerinteraktionen zu finden. Abschließend stellen wir erste Ergebnisse aus unterschiedlichen Sportarten vor.

1 Einleitung

Die Analyse von Videosequenzen durch automatische Verfahren wird immer häufiger eingesetzt, um ein immer breiteres Spektrum an Aufgaben zu lösen. In der In-

dustrie können z. B. Kameras benutzt werden, um die Qualität der hergestellten Produkte vollautomatisch zu kontrollieren. Videokameras und entsprechende Trackingverfahren zur Erhöhung der Sicherheit der Bevölkerung werden in vielen Städten bereits angewendet. In vielen Situationen kann die Analyse der Bilder nicht nur den Menschen signifikant unterstützen, sondern unter Umständen auch Aufgaben übernehmen, die der Mensch allein nicht erledigen kann. In diesem Beitrag zeigen wir, dass sich die automatisierte Analyse von Videosequenzen auch im Bereich des Spitzensports anwenden lässt. Insbesondere stellen wir aktuelle Verfahren zum Objekttracking vor und zeigen, wie sie für die Messung wichtiger sportlicher Kriterien eingesetzt werden können.

2 Bildbasiertes Objekttracking

Unter Objekttracking wird die Analyse von Bewegungssequenzen über einen längeren Zeitraum hinweg verstanden. Informationen über diese Bewegungssequenzen können verschiedenen Sensoren eines Beobachters entnommen werden, z. B. einer Kamera, einem Laserscanner, einem Ultraschallsensor oder einem Odometer, wobei wir uns in diesem Artikel auf Kameras beschränken. Der Begriff des Trackings umfasst alle Verarbeitungsschritte, die der Verfolgung von (bewegten) Objekten oder Personen dienen. Ziel dieser Verfolgung ist die Extraktion von Informationen über den Verlauf der Bewegung und die Lage eines Objekts. Diese Informationen können die Geschwindigkeit der Bewegung, die Beschleunigung sowie Informationen bezüglich der Lage zu einem bestimmten, oft in der Zukunft liegenden Zeitpunkt sein. Je nach Art und Beschaffenheit des zu trackenden Objekts können Tracking-Algorithmen in unterschiedliche Kategorien wie folgt eingeordnet werden.

2.1 Markerbasiertes Tracking

Erhält ein Computersystem ein vordefiniertes Bild, das in einer definierten Lage zu dem zu trackenden Objekt liegt, so kann dieses mithilfe einer Bildanalyse zur Positionsbestimmung der Kamera oder des Objekts verwendet werden. Dieses Bild wird als Marker, der Prozess als markerbasiertes Tracking bezeichnet. Abb. 1 zeigt unterschiedliche Formen eines planaren Markers. Der Vorteil von Markern ist, dass die genaue Position und Orientierung des Markers mit allen sechs Freiheitsgraden aus einem einzelnen Bild gewonnen werden kann und somit eine genaue Bestimmung der Position des an ihm befestigten Objekts möglich ist. Ebenso können mehrere verschiedene Marker gleichzeitig verwendet werden. Die Szene muss allerdings vor der Messung so präpariert werden, dass die Marker über den zeitlichen Verlauf der Aufnahme hinweg für die Kameras durchgehend sichtbar sind. Der Nachteil an diesem Ansatz ist, dass die Athleten durch die Marker möglicherweise behindert werden, was eine Messung in Wettkampfsituationen ausschließt.

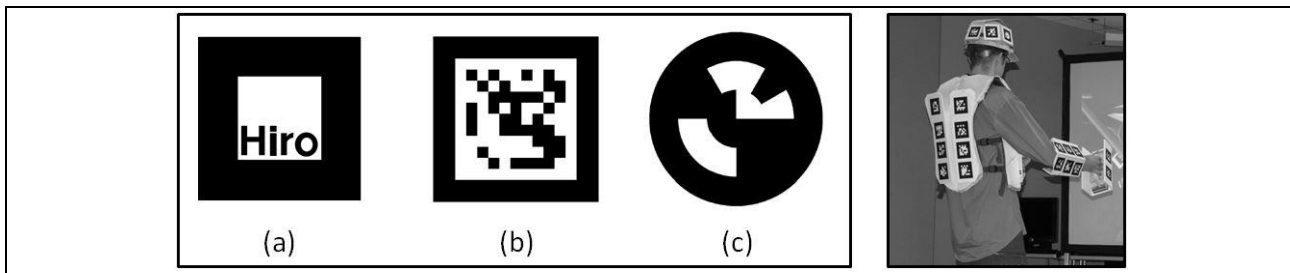


Abb. 1. Links: Unterschiedliche Designmöglichkeiten für Marker, a – ARToolkit Marker (Kato, Billinghamurst, Weghorst & Furness, 1999), b – ARTag Marker (Fiala, 2005), c – Zirkularer Marker (Pagani, Koehler & Stricker, 2011). Rechts: Eine mit Markern ausgestattete Person (Fiala, 2005).

2.2 Markerloses Objekttracking

Wenn Objekte oder Personen ohne spezielle Markierungen getrackt werden sollen, kommt die Technik des markerlosen Objekttrackings zum Einsatz. In klassischen Trackingverfahren wird das zu trackende Objekt oder die zu trackende Person am Anfang einer Videosequenz einmalig ausgewählt. In den meisten Fällen nimmt diese Auswahl die Form eines Rechtecks („Bounding Box“) an, welches das Zielobjekt umfasst. Ziel des Trackingverfahrens ist es, die Position, Größe und das Seitenverhältnis der Bounding Box über die Zeit hinweg zu aktualisieren, sodass das Objekt durchgängig und möglichst genau durch das Rechteck umfasst wird. Moderne Trackingverfahren sind mit einem dynamischen Model des Erscheinungsbilds des Objekts („Appearance Model“) ausgestattet, welches gewährleistet, dass das Objekt auch dann weiterverfolgt wird, wenn es sich durch Drehung, Bewegung und Änderung der Lichtverhältnisse leicht ändert. Die statistisch wahrscheinlichen zukünftigen Bewegungen des Objekts werden in einem Bewegungsmodel („Motion Model“) mitgeführt. Ein letzter wichtiger Aspekt eines Trackingverfahrens ist die Wahl der visuellen Merkmale, die für ein schnelles und robustes Tracking ausgewählt werden. Hier werden außer der Farbe noch Gradientenhistogramme sowie Texturmuster verwendet. Als Stand der Technik können MILTrack von Babenko, Yang und Belongie (2009), Tracking-Learning-Detection (TLD) von Kalal, Matas und Mikolajczyk (2009), Visual-Tracking-Decomposition (VTD) von Kwon und Lee (2010) sowie P-Channel Object Detection von Pagani, Stricker und Felsberg (2009) aufgeführt werden. Abb. 2 zeigt unterschiedliche Bilder einer Videosequenz, in denen eine Eisläuferin mit dem Verfahren VTD getrackt wird.

2.3 Robustheit von klassischen markerlosen Trackingverfahren

Bei Wang, Chen, Xu und Yang (2011) werden 10 verschiedene Trackingalgorithmen miteinander verglichen. Es stellt sich heraus, dass die Leistung der Algorithmen sehr von der aktuellen Bildsequenz abhängig ist. Zwar konnte jeder Tracker für gewisse Sequenzen erfolgreich eingesetzt werden, aber kein Algorithmus war bei allen Sequenzen erfolgreich.



Abb. 2. Tracking einer Eisläuferin mit dem Verfahren VTD (aus Kwon & Lee, 2010)

Das zeigt, dass existierende Algorithmen stark spezialisiert sind und dass es bisher nicht gelungen ist, generalisierte Tracker zu entwickeln. Ähnliche Beobachtungen konnten wir auch bei einer internen Evaluierung machen.

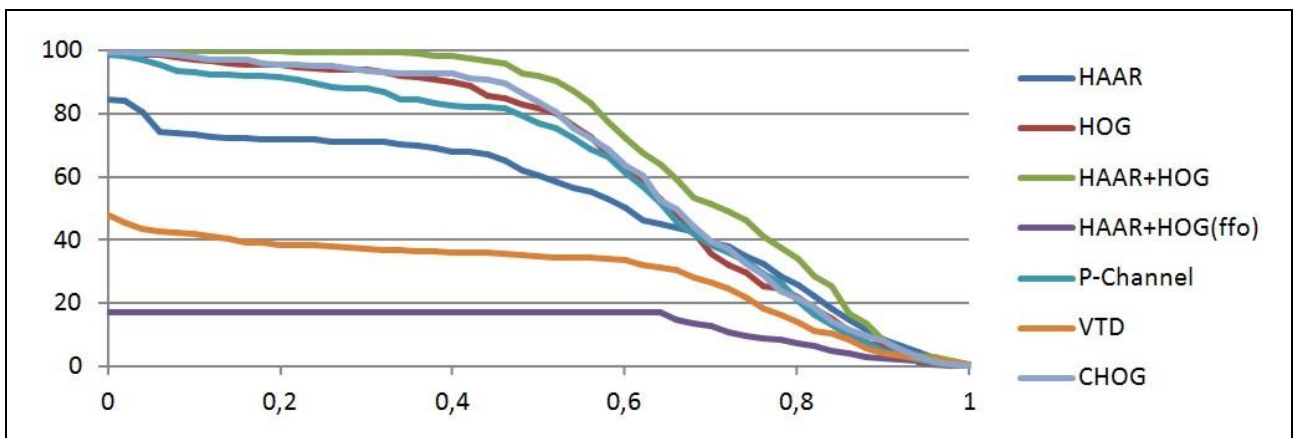


Abb. 3. Ergebnisse des Trackings mit unterschiedlichen Verfahren für die Sequenz „Lemming“

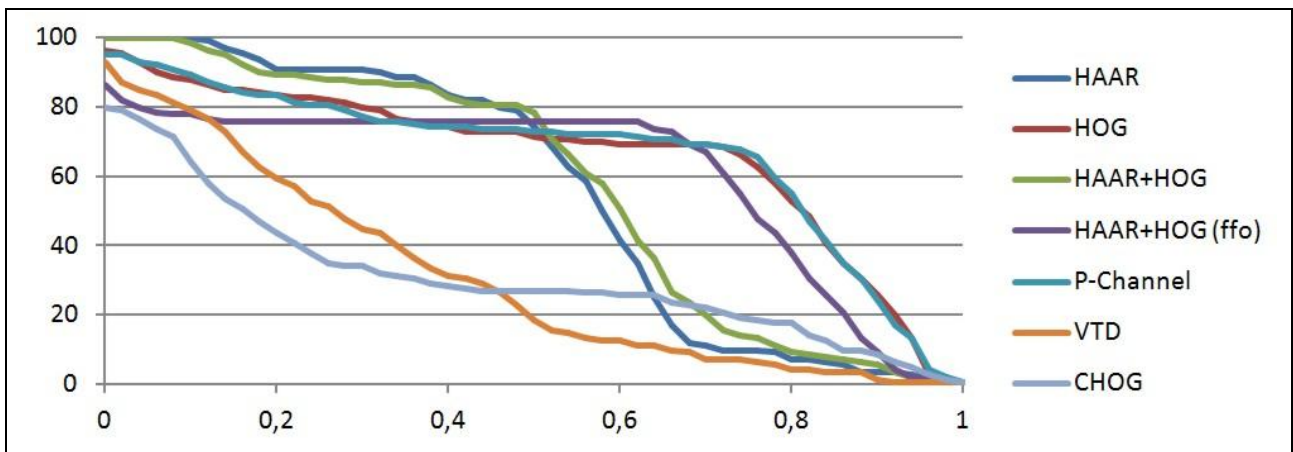


Abb. 4. Ergebnisse des Trackings mit unterschiedlichen Verfahren für die Sequenz „Board“

Abb. 3 und 4 zeigen die Ergebnisse unterschiedlicher Trackingverfahren für zwei Standardevaluierungssequenzen. Getestet wurden folgende Algorithmen: MILTrack mit HAAR Features (HAAR), MILTrack mit HOG Features (HOG), MILTrack mit HAAR+HOG Features (HAAR+HOG), MILTrack mit HAAR+HOG Features nur mit dem ersten Frame gelernt (HAAR+HOG ffo), P-Channel, VTD und MILTrack mit HOG Features auf drei Farbkanälen statt auf einem Schwarz-Weiß-Bild (CHOG). In diesen Diagrammen repräsentiert die Abszisse die Übereinstimmung der Ground-

Truth-Daten mit der ermittelten Region als Quotient von Schnitt- und Vereinigungsmenge. Die Ordinate repräsentiert den prozentualen Anteil der Frames eines Videos, die mindestens eine Überlappung in der Horizontalen haben (für $x = 0$ gilt: mindestens eine Berührung der Regionen). Diese zwei Überlappungsdiagramme zeigen einerseits, dass die Algorithmen je nach Sequenz unterschiedlich reagieren und andererseits, dass kein einziges Verfahren in der Lage ist, alle Frames korrekt mit einer Übereinstimmung von mindestens 50 % (d. h. größer als 0,5) zu tracken.

3 SmartTrack: ein semi-automatisches Tracking Framework

Aufgrund der Genauigkeitsanforderungen im Spitzensport sind hochpräzise und zuverlässige Trackingergebnisse unabdingbar. Um die Position von Objekten oder Körperteilen jederzeit mit hoher Genauigkeit messen zu können, dürfen die Trackingergebnisse bis auf minimale Toleranzen niemals von der tatsächlichen Objektposition abweichen. Diese strikte Bedingung erfordert es, das Konzept des voll-automatischen Videotrackings zu überdenken. In diesem Artikel stellen wir dazu ein Framework mit deutlich stärkerem Fokus auf halbautomatisches Tracking mit Benutzerinteraktion vor. Unser System „SmartTrack“ ist ein benutzerorientiertes, benutzerfreundliches, grafisches Trackingframework, das die Stärken moderner Trackingalgorithmen mit intelligenter, dafür aber minimaler Benutzerinteraktion kombiniert, um in kürzester Zeit korrekte Ergebnisse zu erzielen. In diesem Kapitel stellen wir die Hauptkomponente des Frameworks vor.

3.1 Trackingunterstützungsverfahren

Im Großteil der Literatur zur Objektverfolgung wird davon ausgegangen, dass das zu verfolgende Objekt nur in einem Frame markiert wurde. Dies ist ausreichend, um Trackingalgorithmen zu validieren, d. h., zu überprüfen, welcher davon am weitesten fehlerfrei arbeitet. Zur Erzeugung vollständig korrekter Trackingergebnisse für interaktive Videos ist es dagegen sinnvoll, wenn ein Objekt, insbesondere an schwierigen Stellen im Video, mehrfach markiert wird, um ein Abdriften des Trackers möglichst von vornherein zu limitieren. Wenn mehrere Instanzen des Objekts in einer Sequenz ausgewählt sind, können Regionen zwischen zwei markierten Frames auf zwei Arten getrackt werden:

1. *HVR*: Die 1. Hälfte wird in der Zeit vorwärts und die 2. Hälfte in der Zeit rückwärts getrackt. Dadurch muss das Objekt zwischen zwei Ground-Truth-Markierungen nur noch halb so weit verfolgt werden, wie bei üblichen Trackingverfahren, was die Fehlerwahrscheinlichkeit bereits stark verringert.
2. *FVFR*: Die Region wird vollständig vorwärts und vollständig rückwärts getrackt. Anschließend werden beide Trackingspuren miteinander verglichen und fusioniert. Sind beide Spuren innerhalb eines Frames ungefähr deckungsgleich, wird dem Benutzer ein Mittelwert als sicheres Trackingergebnis angezeigt. Ist dies nicht der Fall, wird das Trackingergebnis als unsicherheitsbehaftet markiert.

Selbst wenn kein sicheres Ergebnis angegeben werden kann, ist es möglich, das Resultat weiter zu verbessern, indem beispielsweise geschaut wird, ob die Tracker an der richtigen Stelle auf der anderen Seite der Region ankommen. Gesetzt den Fall, dass keine Entscheidung getroffen werden kann, wird für dieses Frame das gleiche Ergebnis wie in 1. verwendet. Dieser Ansatz stellt sich als besonders zuverlässig heraus, wenn beide Richtungen deckungsgleich sind. Ist dies nicht der Fall, stellt sich das Ergebnis im Schnitt besser als bei 1. dar.

Der Fusionsalgorithmus für Vorwärts-/Rückwärtstracking kann auch in verallgemeinerter Version verwendet werden, um beliebig viele verschiedene Trackingalgorithmen zu fusionieren. Für ein „sicheres“ Ergebnis ist hierbei mindestens eine ungefähre Deckungsgleichheit von 50 % der Algorithmen erforderlich.

3.2 Effiziente Validierung von Trackingergebnissen

Trotz guter Trackingalgorithmen und unterstützender Verfahren können Trackingfehler nicht vollständig vermieden werden. Das Finden solcher Fehler, insbesondere Kurzzeitfehler, bei denen ein Tracker ein Objekt nur kurzzeitig verliert und danach wiederfindet, kann sehr aufwendig sein, da der Benutzer sehr viele Frames überprüfen muss, um sicherzustellen, dass sich keine Fehler eingeschlichen haben. Dieser Aufwand konnte durch ein neuartiges interpolationsbasiertes Validierungsverfahren deutlich verringert werden. Der Algorithmus interpoliert hierbei die vom Benutzer markierten (und dadurch bereits bestätigten) Frames und vergleicht diese Interpolation mit dem Trackingergebnis. Die Abweichung wird als Farbcode angezeigt (Abb. 5). Dadurch ist es dem Benutzer möglich, das Trackingergebnis auf einen Blick auf seine Korrektheit zu überprüfen und Fehlerstellen schnell zu finden, einzugrenzen und zu behandeln. Ohne diese Interpolation müsste das Video mühevoll manuell nach Fehlerstellen abgesucht werden. Interpolationsabweichungen innerhalb korrekt verfolgter Sequenzen sind zwar möglich, können aber aufgrund der übersichtlichen Darstellung sehr schnell lokalisiert und beseitigt werden. Wurden diese beseitigt und entspricht die Interpolation grob dem Trackingergebnis, kann der komplette restliche Verarbeitungsschritt ohne weitere zeitintensive Überprüfung in einem Schritt bestätigt werden. Die Interpolation selbst wird nicht in das Ergebnis einbezogen, da der Tracker das Objekt in der Regel immer noch etwas genauer findet.



Abb. 5. Trackingzeitleiste mit eingeblendetem Fehlercode. Oben: direkt nach dem Tracking.
Unten: nach Bestätigung weniger zusätzlicher Frames (grüne Striche).
Da Interpolation und Trackingergebnis unten fast übereinstimmen, kann das Trackingergebnis nun bedenkenlos komplett übernommen werden.

Zusammenfassend lässt sich sagen, dass unsere Trackingunterstützungsverfahren in Kombination mit unserem Validierungswerkzeug sehr mächtige Instrumente sind,

mit denen sich auch mit suboptimalen Trackern sehr schnell brauchbare Ergebnisse erzielen lassen.

4 Videotracking für den Sportbereich

4.1 Bedarf an automatischer Videoanalyse im Sportbereich

Im Bereich des Spitzensports wird die Videoaufnahme zunehmend als technisches Mittel benutzt, um die Performanz von Athleten zu verbessern. Die zahlreichen Herausforderungen, die sich beim Einsatz von Kameras für den Sport ergeben können, wurden größtenteils gelöst. Trainingswissenschaftler können heutzutage z. B. Fragen nach der Wahl der Kameras, der optimalen Platzierung oder der Synchronisierung gut beantworten. Dadurch kann zunehmend Videomaterial für zahlreiche Sportarten gesammelt werden. Dieses Rohmaterial muss zurzeit manuell bearbeitet werden, um die sportliche Aktivität zu evaluieren. Eine manuelle Bearbeitung kann ein Präzisionsniveau erreichen, das für eine gezielte Analyse ausreichend ist, ist aber oft langsam und kostenintensiv. Wünschenswert wäre daher eine automatische Bearbeitung des Videomaterials. Die zentrale Frage hierbei bleibt allerdings, ob die durch automatische Verfahren erzielte Genauigkeit ausreichend ist. Zur Beurteilung dieser Frage soll hier zwischen erwünschten Messungen und technisch möglichen Messungen unterschieden werden. Trainer und Trainingswissenschaftler haben die notwendigen Fachkenntnisse, um zu entscheiden, was gemessen werden soll (z. B. die Winkel der Gelenke, die Bewegung von Spielern oder der Flug eines Balls). Computer-Vision-Experten können dagegen erläutern, was mit dem heutigen Stand der Technik gemessen werden kann. Zum Beispiel kann der ganze Körper eines Spielers, sein Kopf oder ein Ball als starres Objekt verfolgt werden. Hier sind wir der Meinung, dass eine enge Zusammenarbeit zwischen Sportexperten und Bildverarbeitungsspezialisten zu einem besseren gegenseitigen Verständnis und zur Entwicklung dedizierter Systeme verhelfen kann.

4.2 Erste Ergebnisse

In einer Kooperation zwischen dem Institut für Angewandte Trainingswissenschaft (IAT) und dem Deutschen Institut für Künstliche Intelligenz (DFKI) wurden erste Videosequenzen unterschiedlicher Sportaktivitäten analysiert, deren wichtigste Ergebnisse wir hier vorstellen möchten.

Volleyball

Im Volleyball war die Erfassung der Flugbahn des Balls erwünscht, um den Flattereffekt beim Aufschlag zu analysieren. Mit unserem Trackingverfahren konnten wir die Position des Balls verfolgen (Abb. 6) und erfolgreich einzelne Spieler tracken (Abb. 7).

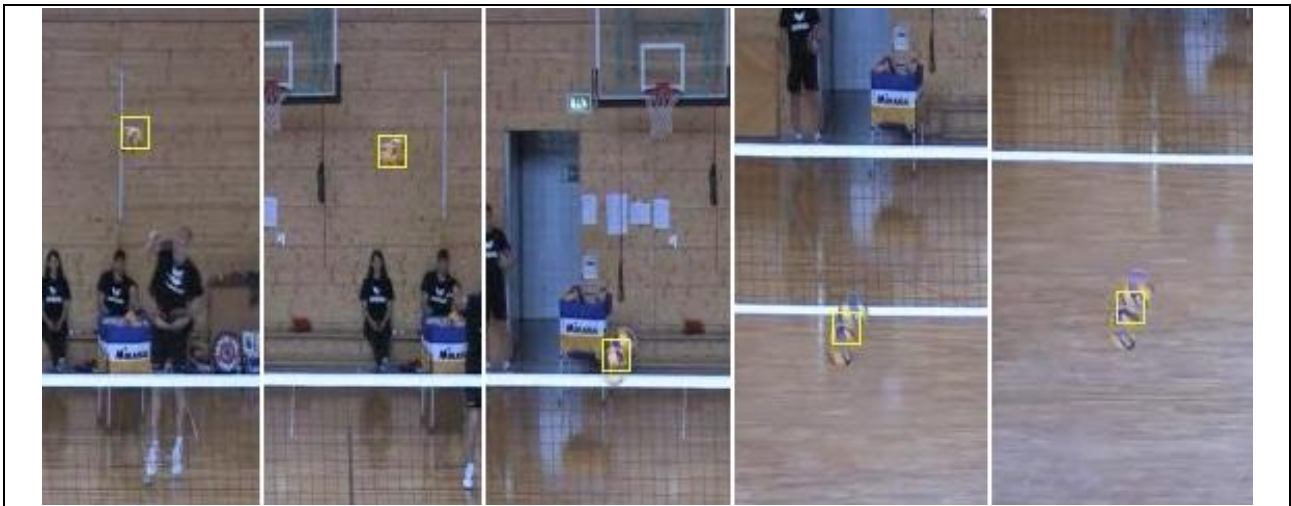


Abb. 6. Tracking des Ballflugs im Volleyball



Abb. 7. Tracking eines Spielers im Volleyball

Badminton

Die Technik des Spielertrackings wurde auch beim Badminton angewendet. Hier konnten wir zeigen, dass es möglich ist, einzelne Spieler sowie den Kopf der Spieler zu verfolgen (Abb. 8).



Abb. 8. Spielertracking im Badminton

Hochfrequenzkameras

In Trainingssituationen können Hochfrequenzkameras eingesetzt werden, deren Bildwiederholungsrate bei 500-1.000 Hz liegt. Damit ist es möglich, kleine, schnelle

Objekte zu tracken, wie z. B. den Fuß eines Spielers oder den Federball (Abb. 9 und 10).

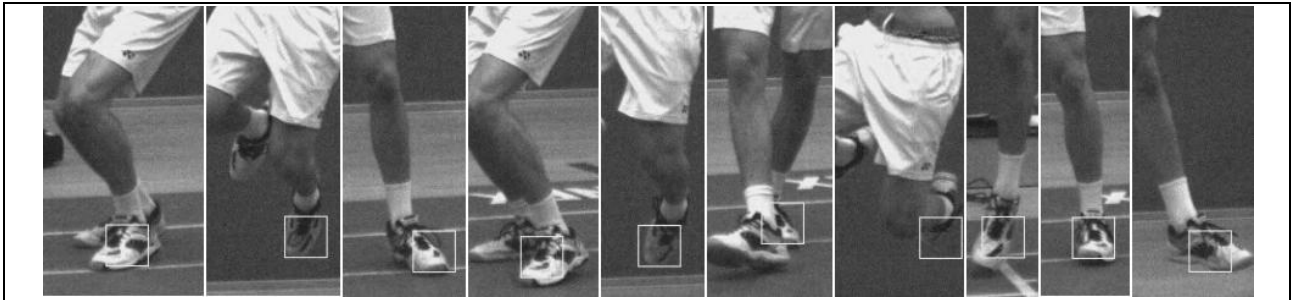


Abb. 9. Fuß des Spielers wird getrackt

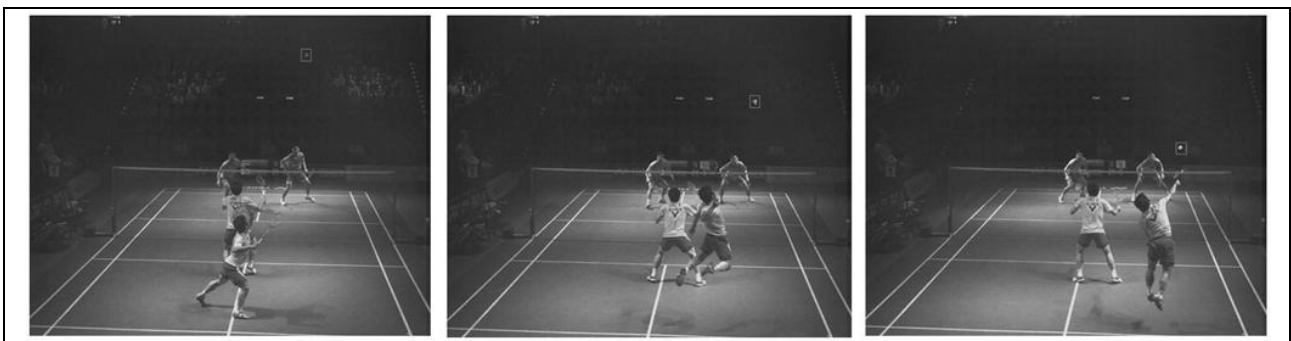


Abb. 10. Tracking des Balls mit einer Hochfrequenzkamera in einer Wettkampfsituation

4.3 Nächste Schritte

Trackingverfahren alleine können nicht genug Information liefern, um eine vollständige Analyse durchzuführen. Diese Verfahren liefern nämlich nur die Position des Objekts im Bildraum. Um genaue Messungen zu bekommen, muss diese zweidimensionale Information in den dreidimensionalen Raum konvertiert werden. In der Regel sind dafür mehrere Kameras notwendig, in einzelnen Fällen lassen sich diese 3-D-Messungen aber auch mit nur einer Kamera berechnen. In jeden Fall bedarf diese 3-D-Analyse eine genaue Kalibrierung der Kameraparameter. Auch in diesem Bereich verfügt das DFKI durch seine langjährige Forschung im Bereich der erweiterten Realität über weitreichende Erfahrung. Ziel ist es, hier aus einer 2-D-Bildanalyse 3-D-Messungen zu berechnen (Höhe eines Sprungs, absolute Geschwindigkeit des Balls), wie in Abb. 11 exemplarisch dargestellt.

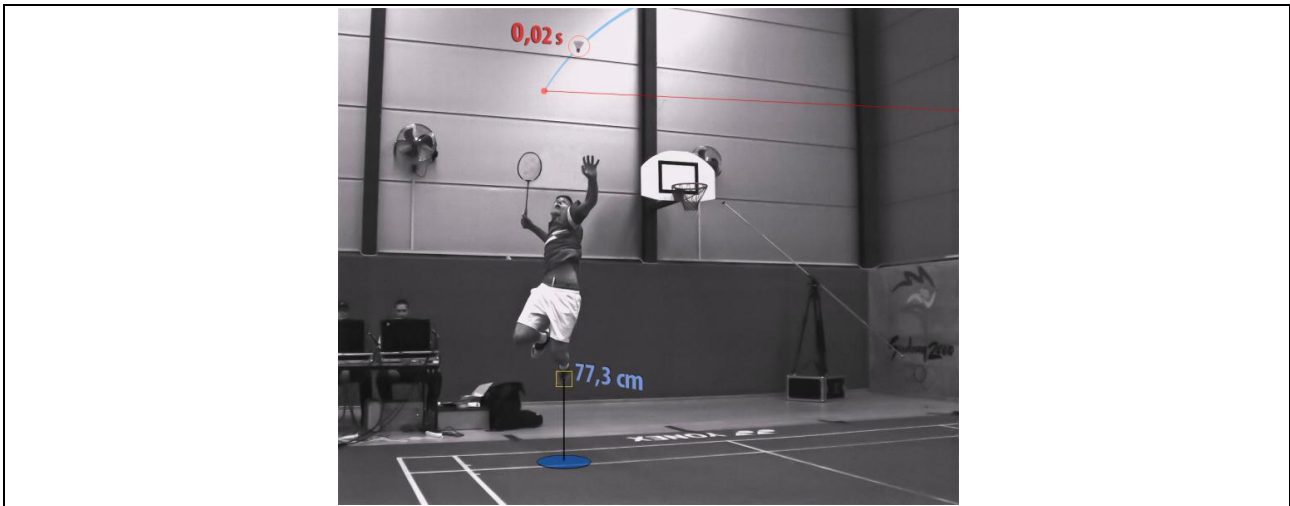


Abb. 11. Absolute Messungen nach Kamerakalibrierung

Fazit

In diesem Artikel haben wir unterschiedliche Möglichkeiten zur bildbasierten Analyse sportlicher Aktivitäten präsentiert. Nachdem verschiedene Trackingmethoden vorgestellt wurden, wurde auf markerlose Trackingverfahren näher eingegangen. Wir haben gezeigt, dass sich ein semi-automatisches Verfahren optimal einsetzen lässt, um die gewünschte Präzision mit minimalem Aufwand zu erreichen. Mit dem aufgeführten Satz an Technologien haben wir unterschiedliche Möglichkeiten vorgestellt, um in diversen Szenarien des Leistungssports Messungen in 2-D und 3-D realisieren zu können. Zusammenfassend können wir behaupten, dass die technischen Herausforderungen für eine bildbasierte Analyse im Sport nach und nach gemeistert werden und dass die Kooperation zwischen Sportexperten und Spezialisten der Bildverarbeitung zur Entwicklung nützlicher Systeme führen kann.

Literatur

- Babenko, B., Yang, M.-H., Belongie, S. (2009). Visual Tracking with Online Multiple Instance Learning. In *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June, 20-26, 2009). Miami: Institute of Electrical and Electronics Engineers.
- Fiala, M. (2005). Artag, a fiducial marker system using digital techniques. In *Computer Vision and Pattern Recognition 20-25 June 2005* (Vol. 2, pp. 590-596). Ottawa: Institute of Electrical and Electronics Engineers.
- Kalal, Z., Matas, J. & Mikolajczyk, K. (2009). Online learning of robust object detectors during unstable tracking. In *Computer Vision Workshops* (pp. 1417-1424). Kyoto: Institute of Electrical and Electronics Engineers.
- Kato, H., Billinghurst, M., Weghorst, S. & Furness, T. (1999). *A mixed reality 3d conferencing application*. Technical Report. Seattle. Human Interface Technology Laboratory, University of Washington.
- Kwon, J. & Lee, K.-M. (2010). Visual tracking decomposition. In *Computer Vision and Pattern Recognition*. San Francisco: Institute of Electrical and Electronics Engineers.

- Pagani, A., Koehler, J. & Stricker, D. (2011). Circular markers for camera pose estimation. In *Proceedings of the 12th International Workshop on Image Analysis for Multimedia Interactive Services (WMAMIS 2011)*. Netherlands: TU Delft.
- Pagani, A., Stricker, D. & Felsberg, M. (2009). Integral P-channels for fast and robust region matching. In *Image Processing (ICIP) 16th IEEE International Conference 7.-10. November 2009* (pp. 213-216). Cairo: Institute of Electrical and Electronics Engineers.
- Wang, Q., Chen, F., Xu, W., & Yang, M.-H. (2011). *An experimental comparison of online object tracking algorithms*. In M. Papadakis, D. van de Ville & V. K. Goyal (Eds.), *Proceedings of SPIE: Image and Signal Processing* (Vol. 8138). San Diego. SPIE Press.