

Zeitkontinuierliche Schätzung der Absprung- und Flugparameter von Skispringern in Videoaufnahmen anhand automatischer zeitkontinuierlicher Erkennung von Körper- und Skiposen

(AZ 072009/18)

Rainer Lienhart¹ (Projektleitung), Ina Fichtner³, Dan Zecha¹, Dr. Sören Müller², Dr. Sascha Kreibich²

¹ Universität Augsburg, Institut für angewandte Informatik,
Lehrstuhl für Multimedia und maschinelles Sehen

² Institut für Angewandte Trainingswissenschaft, Leipzig
Fachgruppe Skispringen/Nordische Kombination

³ Institut für Angewandte Trainingswissenschaft, Leipzig, Fachbereich MINT

1 Problemstellung

Händisch ausgewertete Videoaufzeichnungen von Skispringerinnen und Skispringern bilden aktuell eine wesentliche Grundlage für deren Leistungsdiagnose, Leistungsverbesserung und Trainingssteuerung. Leider limitiert der damit verbundene hohe personelle Aufwand, wie häufig diese so gewinnbringende Auswertung durchgeführt werden kann. Eine automatisierte Videobildauswertung bietet daher eine enorme Erleichterung. Durch sie beschleunigt sich der Prozess der Bestimmung kinematischer Flugparameter, sodass die Messergebnisse Trainern und Sportlern deutlich schneller zur Verfügung gestellt werden könnten. Die freigegebenen Kapazitäten könnten einerseits für eine intensivere Ergebnisinterpretation und umfangreichere Trainergespräche genutzt und andererseits für die Erhebung und Bearbeitung umfangreicherer Datensätze eingesetzt werden. Ziel dieses Service-Forschungsprojekts war es deshalb, vollautomatisch – d. h. mittels Verfahren des maschinellen Sehens ohne mühsame und arbeitsintensive manuelle Auswertung – von einem Skispringer auf der Schanze mittels softwarebasierter Auswertung der Videoaufnahmen zeitkontinuierlich Informationen über charakteristische Gelenkpunkte (Körperpunkte für Fuß-, Knie-, Hüft- und Schulter-, Ellenbogen-, Handgelenke) und damit über die Pose der Athleten zu ermitteln. Zusätzlich wird ein

Sportgerätemodell vorgeschlagen, welches die Detektion der Skier des Sportlers (Skispitze und Skiende) ermöglicht. Die automatische, zeitkontinuierliche Detektion der Körperpunkte ermöglicht unter Zuhilfenahme eines anthropometrischen Modells die Ermittlung des Körperschwerpunktes, dessen Flugbahn ein wichtiges Hilfsmittel zur qualitativen Bewertung der Absprung- und Flugtechnik bietet. Zusätzlich werden diverse Flug- und Anstellwinkel aus Körper- und Skipunkten berechnet. Abschließend wurde untersucht, ob aus den bestimmten Körperpunkten die normalerweise mit zusätzlichen Kraftsensoren im Sprungtisch abgeleiteten physikalischen Absprungkräfte mit Verfahren des maschinellen Lernens geschätzt werden können.

2 Methode

Im Rahmen des Serviceforschungsprojekts wurde das System DeepSkiJumpPose entworfen und implementiert, welches den Skispringer detektiert, kontinuierliche und robust dessen Pose schätzt sowie Winkel- und Flugparameter ableitet und ausgibt. Alle gefragten Parameter werden aus Videodaten bestimmt, die verschiedene Phasen eines Skisprungs von der Seite zeigen (Beispielansicht siehe Abb. 1). Bis dato werden diese Videoaufnahmen von Trainingswissenschaftlern bzw. -wissenschaftlerinnen manuell ausgewertet. Diese markieren händisch

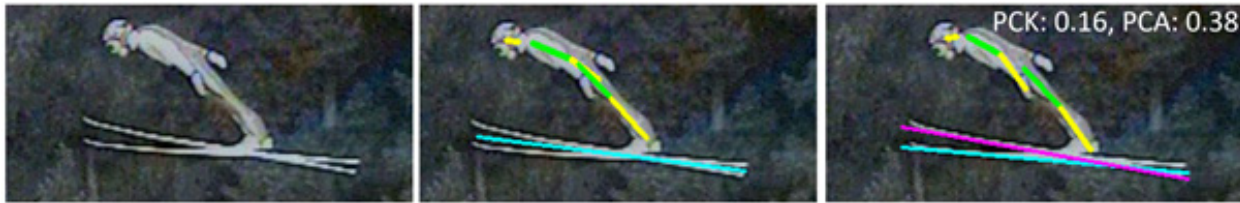


Abb. 1: Ein qualitativer Vergleich zwischen Expertenannotationen und automatischen Schätzungen
 Links: Ausschnitt aus einem Kamerabild; Mitte: Annotation eines Experten; Rechts: Schätzung des DeepSkiJumpPose Systems

disch die Position einzelner Gelenke sowie der Skier in mehreren Einzelbildern unterschiedlicher Kameras. Anhand der Körperpunktannotationen können die gewünschten Winkel- und Flugparameter anschließend bestimmt werden. Diese Art der quantitativen Auswertung ist sehr zeitintensiv, da die Experten den Großteil der Einzelbilder eines Videos bewerten müssen.

Die im Folgenden vorgeschlagene Alternative zur vollautomatischen Bestimmung von Posen löst das Problem mittels Algorithmen aus dem Gebiet des maschinellen Sehens und Lernens. Dazu wird die Problemstellung auf folgende Betrachtungen reduziert: Aus einem konstanten Strom von Einzelbildern werden die Athleten mittels eines neuronalen Netzes zuerst in jedem Bild detektiert. Hierzu wird angelehnt an (Howard, 2017) ein tiefes neuronales Netz zur Detektion des Sportlers bzw. der Sportlerin im Absprungbereich und in der Flugphase trainiert. Die Anwendung eines solchen Modells¹ auf jedes Bild eines Skisprungvideos ergibt eine Schätzung für die Position des Sportlers. Damit wird die Aufmerksamkeit des Posenschätzers auf relevante Bereiche des Bildes fokussiert und Bereiche ausgeschlossen, die zu potentiellen Falschdetektionen führen können. Zur Detektion der Gelenkpunkte basierend auf (Wei, 2016) wurde ein tiefes neuronales Netz zur Schätzung der Körperpose trainiert. Dieser Ansatz funktioniert rein datengetrieben: Aus Trainingsbildern werden Ausschnitte extrahiert, die Gelenke in verschiedenen Teilen der Sprungphase zeigen. Zusammen mit der Information, welches Gelenk auf welchem Bildausschnitt zu sehen ist, dienen die Ausschnitte als Eingabe für ein neu-

ronales Netz. Mittels eines maschinellen Lernverfahrens werden die Parameter des Netzes angepasst, sodass der Vorhersagefehler für einen Bildausschnitt minimiert wird. Die Ausgabe des Netzes ist dann eine Wahrscheinlichkeit für die Anwesenheit jedes Gelenks in jedem Pixel des Bildes. In einem weiteren neuronalen Netz wird ein Modell zur Verifikation von Gelenkpositionen entwickelt, welches aus einer Sammlung von möglichen Gelenkdetektionen die wahrscheinlichste Konfiguration berechnet und, falls nötig, Gelenkpositionen korrigiert.

Beide Modelle zur Detektion der Skispringer und zur anschließenden Posenschätzung werden auf alle Einzelbilder eines Videos angewendet. Um eine möglichst konsistente Schätzung der Körperposen über die Zeit zu gewährleisten, wird eine robuste temporale Filterung auf den einzelnen Gelenken angewendet. Die entstehende Sequenz aus Posenschätzungen über ein Video hinweg wird verwendet, um die von Experten bestimmten Winkel und Flugparameter direkt zu berechnen.

Im Rahmen eines innovativen Arbeitspakets wurde ein temporales neuronales Netz zur Vorhersage der Absprungkräfte von Skispringern neu entwickelt und trainiert. Auch hier wurde datengetrieben gearbeitet: Die Trainingsdaten bestehen aus den geschätzten Körperposen und parallel durch Kraftmessplatten erfasste Absprungkräfte. Das neuronale Netz soll mittels der Abbildung von Posenkoordinatensequenzen auf Kraftmessesequenzen die Zusammenhänge zwischen beiden Modalitäten lernen und nach dem Training in der Lage sein, aus neuen Posensequenzen den entsprechenden Kraftverlauf vorherzusagen.

¹ Die Begriffe „Modell“ und „tiefes neuronales Netz“ werden im Folgenden synonym verwendet. Tatsächlich definiert die Struktur eines neuronalen Netzes das Modell mit dessen Modellparametern.

3 Ergebnisse und Diskussion

Im Folgenden werden die Ergebnisse des Serviceforschungsprojekts qualitativ und quantitativ bewertet. Die Springer werden mit einer Sensitivität und einer Genauigkeit von jeweils > 99 % in den Videos detektiert. Innerhalb einer Testmenge von 500 Bildern wurden lediglich zwei Bildbereiche fälschlicherweise als Athlet identifiziert. Die Erkennungsrate des verwendeten Sportlerdetektors ist damit mit der eines menschlichen Experten vergleichbar.

Ein qualitativer Vergleich zwischen Experte-annotation und Posenschätzung ist in Abb. 1 (Mitte und rechts) visualisiert. Da sich Fehler in der Detektion von einzelnen Gelenken bei der Berechnung von Winkelparametern fortpflanzen, ist in Abb. 2 der Fehler über die berechneten Winkel im Vergleich zu einem erfahrenen Experten visualisiert.

und orange) verglichen. Alle Experten erzeugen eine Abweichung von maximal einem Grad zwischen ihren Skiannotationen. Die automatische Schätzung schneidet hier etwas schlechter ab: Innerhalb von 2 Grad werden 90 % der Ski richtig detektiert. Dieser etwas schlechtere Wert lässt sich durch zwei Beobachtungen erklären. Zum einen muss zur Bestimmung der Abweichung aus beiden Skidetektionen der „mittlere“ Ski bestimmt werden (größerer Fehler durch Fortpflanzung), zum anderen kann es auch zu Fehldetektionen der Ski kommen.

Zur automatischen Bestimmung der Kraftverläufe aus wiederum automatisch geschätzten Posensequenzen des Absprungs lässt sich festhalten, dass die Vorhersageergebnisse von Kraftmessdaten aus Posensequenzen mittels neuronaler Netze äußerst vielversprechend sind. Allerdings konnte im Rahmen eines relativ kurzen Serviceforschungsprodukts keine konsis-

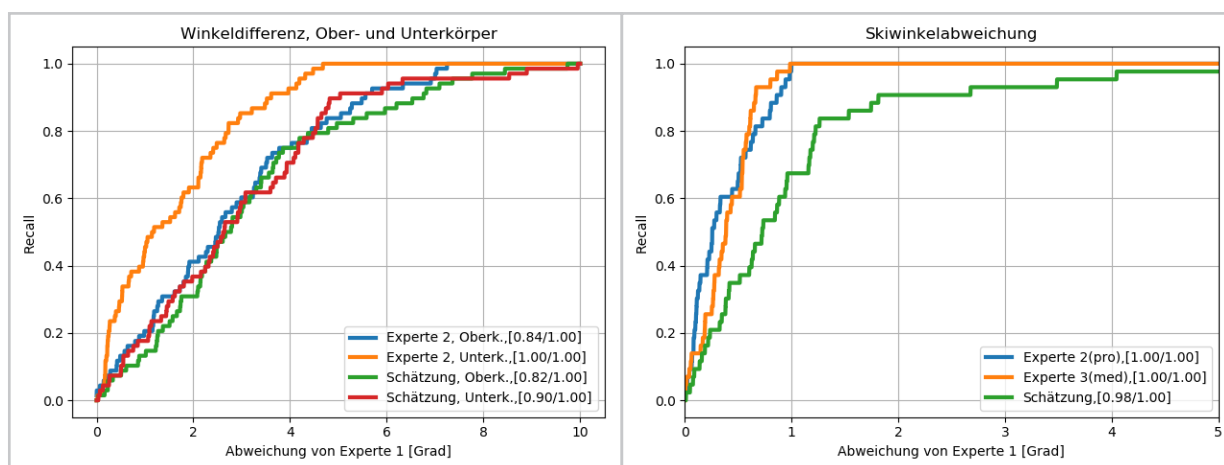


Abb. 2: Abweichung für Winkelparameter für Ober- und Unterkörper (links) und den Ski (rechts)

Auf der linken Seite der Abbildung werden die Schätzungen für Unterkörperwinkel (Fuß zu Hüfte) und Oberkörperwinkel (Hüfte zu Schulter) verglichen. Abgebildet ist jeweils die Abweichung zwischen Experten 1 und Experten 2 sowie Experten 1 und der Schätzung des Deep-SkiJumpPose Systems. Während die Lage des Oberkörpers sehr gut vom System reproduziert werden kann, wird die Lage des Unterkörpers vom Experten etwas besser geschätzt.

Die gleiche Auswertung kann auch für die Lage der Ski durchgeführt werden (Abb. 2, rechts). Hier wird die Lage der Ski von einem Referenzexperten vorgegeben und gegen die Leistung des Systems (grün) sowie zwei weitere Experten (blau

tent perfekte Kraftvorhersage erreicht werden: Während für einige Springer sehr gute Kraftverläufe reproduziert werden konnten, wurde in Experimenten auch deutlich, dass Vorhersagen für andere Skispringer nur unzureichend waren. Ein Grund hierfür ist die geringe Trainingsdatenmenge von 200 Videos. Des Weiteren müsste eine größere Datenmenge gewissenhaft von Experten annotiert werden, um eine noch bessere Posenschätzung im Absprungbereich zu ermöglichen und fehlerfortpflanzenden Effekte in der Vorhersage zu minimieren. Abb. 3 vergleicht die Ausgabe verschiedener neuronaler Netzstrukturen (spaltenweise) auf unterschiedlichen Kraftverläufen (zeilenweise).

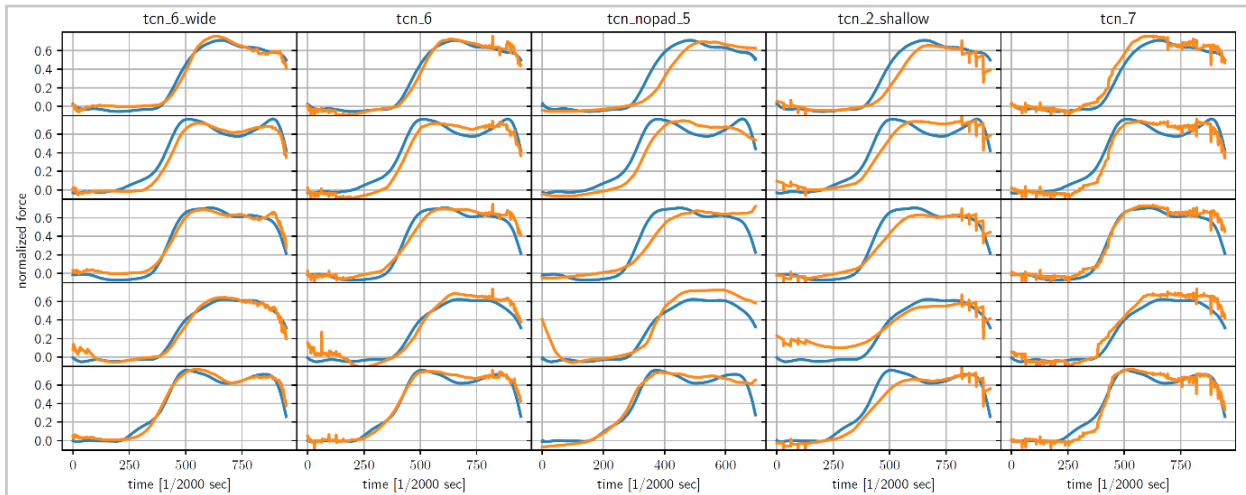


Abb. 3: Beispieldrohersagen für Kraftverläufe des in (Zecha, 2018a) präsentierten Vorhersagenetzes. Spalten zeigen jeweils die Ausgabe verschiedener Netzarchitekturen, Zeilen zeigen unterschiedliche Kraftsignale. Die Grundwahrheit ist in blau dargestellt, die Vorhersage des Kraftverlaufs in Orange.

In blau ist die mittels Sensoren gemessene Grundwahrheit angegeben, die geschätzten Kraftverläufe sind in orange dargestellt.

Ausführliche Ergebnisse zum Serviceforschungsprojekt wurden im Rahmen von zwei wissenschaftlichen Aufsätzen (Zecha 2018a, 2018b) bei renommierten internationalen Konferenzen publiziert.

4 Literatur

- Howard, A. et al. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. CoRR, abs/1704.04861. <http://arxiv.org/abs/1704.04861>, 2017
- Wei, S.-E., Ramakrishna, V., Kanade, T. & Sheikh, Y. (2016). *Convolutional Pose Machines*. CVPR 2016.
- Zecha, D. et al. (2018a). *A Convolutional Sequence to Sequence Model for Multimodal Dynamics Predictions in Skijumps*. ACM Multimedia: First International Workshop on Multimodal Content Analysis in Sports, 2018.
- Zecha, D. et al. (2018b). *Kinematic Pose Rectification for Performance Analysis and Retrieval in Sports*. Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018.