

SIAK-Journal – Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis



Labudde, Dirk

Das digitale Rig als intelligentes, bildbasiertes, forensisches Instrument

SIAK-Journal – Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis (4/2023), 28-39.

doi: 10.7396/2023_4_C

Um auf diesen Artikel als Quelle zu verweisen, verwenden Sie bitte folgende Angaben:

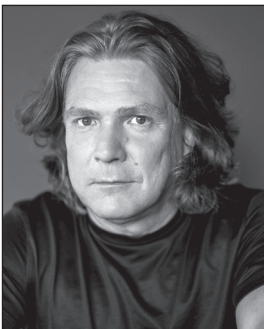
Labudde, Dirk (2023). Das digitale Rig als intelligentes, bildbasiertes, forensisches Instrument, SIAK-Journal – Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis (4), 28-39, Online: https://dx.doi.org/10.7396/2023_4_C.

© Bundesministerium für Inneres – Sicherheitsakademie / Verlag Österreich, 2023

Hinweis: Die gedruckte Ausgabe des Artikels ist in der Print-Version des SIAK-Journals im Verlag Österreich (<https://www.verlagoesterreich.at/>) erschienen.

Online publiziert: 2/2024

Das digitale Rig als intelligentes, bildbasiertes, forensisches Instrument



DIRK LABUDDE,

Hochschule Mittweida, Leiter des FoSIL (Forensic Science Investigation Lab) und des Fraunhofer Lernlabors Cybersicherheit.

Aus der Geschichte der Forensik, der Lehre von der Verbrechensaufklärung, geht deutlich hervor, dass es Kriminellen immer wieder gelingt, durch den Einsatz neuer Methoden unbekannte Begehensweisen für ihre kriminellen Handlungen zu nutzen und Strategien zu verwenden, die einer Aufklärung entgegenwirken. Ein aktuelles Beispiel ist die Vermummung bzw. Verschleierung von Täterinnen oder Tätern während der Tatbegehung in Bereichen, die von öffentlichen Kameras überwacht werden. Dieses gezielte Verhalten der Täterinnen oder Tätern muss eine kontinuierliche Weiter- und Neuentwicklung von Techniken und Konzepten für die Aufklärung bedingen. Am Beispiel des künstlichen, digitalen Rigs¹ und dessen individualisierenden Eigenschaften wird ein neues biometrisches Merkmal vorgeschlagen. In dieser Arbeit wird ebenfalls das dazu korrespondierende Verfahren erläutert und an einem konkreten Fallbeispiel detailliert beschrieben. Im Mittelpunkt steht der Root Mean Square Deviation (RMSD) als Ähnlichkeitsmaß für den Vergleich zweier Rigs. Die Rigs können mit Hilfe von intelligenten Systemen aus Tatvideos abgeleitet werden und mit den Rigs von tatverdächtigen Personen verglichen und bewertet werden. Die Diskussion des künstlichen Rigs als neues biometrisches Merkmal mit dem Ziel der Identifizierung von Personen setzt jedoch eine Auseinandersetzung mit dem Begriff der Spur, auch in ihrem historischen Kontext, voraus.

1. DER ZERFALL EINER SPUR IN EINEN STATISCHEN UND DYNAMISCHEN ANTEIL AM BEISPIEL DES KÜNSTLICHEN, DIGITALEN RIGS EINER PERSON

Eine der Hauptaufgaben der Forensik ist die Rekonstruktion des Tathergangs. Die Betrachtung der Spur in diesem Kontext eröffnet die Frage nach der zeitlichen und örtlichen Einordnung einer Spur in den gesamten Hergang. Dabei wird schnell deutlich, dass analoge Spuren – wie aus der digitalen Forensik bekannt – oft keinen Zeitstempel besitzen. Häufig ist nur eine Einordnung in die Chronologie des Tat-

hergangs möglich. Diese chronologische Einordnung kann über einen hypothesengetriebenen Analyseprozess erfolgen. In der Arbeit von Becker u.a. (vgl. Becker et al. 2021) wird der klassische hypothesengetriebene Ansatz mit Methoden der computergestützten Modellierung kombiniert, mit dem Ziel, die Chronologie der entstehenden Blutspuren darzustellen. Die Verbindung beider Ansätze wird hier durch eine Interpretationsfunktion generisch dargestellt und liefert eine zeitliche Einordnung der entstandenen Spuren (vgl. ebd.; Bevel 2002; Bergmann et al. 2022). Die Interpretation der an einem Tatort

gefundenen und zu einer späteren Zeit eingeordneten Spuren ist für die Rekonstruktion der Umstände und des Hergangs von entscheidender Bedeutung. Der Großteil der praktischen Arbeit beinhaltet auch heute noch menschliche Entscheidungen, die sich auf den Vergleich von Mustern beziehen, z.B. Werkzeugspuren und daktyloskopische Spuren. Die Entscheidung, auf der Basis einer Musterähnlichkeit, besteht darin, festzustellen, ob die beiden Muster ähnlich genug sind, um von demselben Spurenverursacher zu stammen (vgl. Albright 2022). Parallel zum reinen Mustervergleich ist es möglich, die Entstehung von Spuren zu interpretieren.

Somit liegt der Betrachterin oder dem Betrachter einer Spur eine Aufnahme von der Sicherung vor, die analysiert und im nächsten Schritt mit einer Vergleichsspur verglichen wird. In vielen Fällen ermöglicht eine Separation der Spur in einen statischen und dynamischen Anteil die Entwicklung neuer Hypothesen.

In diesem Zusammenhang bezieht sich der statische Teil auf das Aussehen, den Zustand und die Geometrie der Spur zum Zeitpunkt ihrer Erhebung. Der dynamische Teil einer Spur kann Informationen liefern, die sich auf eine mögliche Entstehung beziehen. Das bedeutet, dass man sich überlegen muss, warum eine Spur an einem bestimmten Ort gefunden wurde und wie diese entstanden ist. Die Dynamik der Entstehung einer Spur liefert oft einen Kontext über die Beziehung der verschiedenen Spuren zueinander. So ist es auch möglich, die Umstände eines Verbrechens zu rekonstruieren. Im Schritt der Interpretation der Spuren und Spurenkomplexe wird deutlich, dass eine Aufspaltung in einen statischen und dynamischen Spurenteil wesentlich mehr Informationen liefern kann (vgl. Becker et al. 2021).

In einer Vielzahl von Ermittlungen, die auf der Grundlage von Aufnahmen aus öf-

fentlichen Kameras beruhen, werden die oben diskutierten Annahmen deutlich. Unter der Annahme, dass es möglich ist, die Täterin oder den Täter trotz Maßnahmen der Verschleierung an individuellen Merkmalen zu erkennen und diese generisch zu beschreiben. Resultat einer solchen Merkmalsbeschreibung ist ein künstliches Rig.

Das künstliche Rig stellt ein Muster dar, welches das menschliche Skelett auf einer vereinfachten Abstraktionsebene repräsentiert. Durch die Ableitung des künstlichen Rigs für eine tatverdächtige Person und die Einordnung in ein 3D-Modell des Ereignisortes können Muster beobachtet und interpretiert werden. Diese Interpretation kann helfen, den exakten Tathergang zu simulieren bzw. zur Identifizierung einer tatverdächtigen Person beizutragen. Das künstliche Rig ist eine Abstraktion des menschlichen Skelettes und kann auf verschiedene Arten, manuell durch Markierung anatomischer Landmarken am Körper oder mit Hilfe eines intelligenten Systems, abgeleitet werden. Die Darstellung als Muster geht auf R. Martin zurück. Er liefert eine Beschreibung der Maße eines Menschen und deren Proportionen zueinander. Weiterführende Arbeiten auf diesem Gebiet und die Einordnung in die Biometrie sind unter anderem Dantcheva u.a. (vgl. Dantcheva et al. 2011) und Cole (vgl. Cole 2007). Die Rolle von Körpermaßen in der Forensik wurde von A. Bertillon (vgl. Bertillon/McClaghry 1896) diskutiert und stellte die Grundlage für die sogenannte Bertillonage dar. Weitere historische Arbeiten befassen sich mit dieser Thematik, so auch Schneickert (vgl. Schneickert 1923).

Durch die Einführung von Techniken zur Digitalisierung von Tatorten, der Aufzeichnung von Videos aus dem öffentlichen Raum und Softwareframes zur Erstellung von 3D-Szenen haben diese Arbeiten wieder an Bedeutung gewonnen.

Die Herausforderungen und die Grenzen des künstlichen Rigs sollen an einem realen Fallbeispiel erläutert werden.

Aufgrund des Tatorts kann davon ausgegangen werden, dass genügend Bild- und Videoüberwachungsmaterial zur Verfügung steht, das die Täterin oder den Täter am Tatort aufgenommen hat. Je nach verwendetem Überwachungssystem könnte sogar hochauflösendes Material verfügbar sein. Kurzum, es scheinen optimale Voraussetzungen für die weiteren Ermittlungen gegeben zu sein, insbesondere im Hinblick auf die Identifizierung der Täterin oder des Täters und die Beteiligung an einer Straftat. Allerdings agieren die Täterinnen oder Täter in der Regel mit einer Art von Verschleierung, um ihre Identität zu verbergen. In den meisten Fällen verschleiern sie ihr Gesicht und sogar ihren Gang. Konventionelle Ermittlungsansätze und Methoden wie Gesichtserkennung und Ganganalyse stoßen dann schnell an ihre Grenzen. Aus diesem Grund wurde ein Ansatz entwickelt, der auf anthropometrischen, personenspezifischen, digitalen Skeletten (digitales Rig) basiert. Mit Hilfe dieser Rigs ist es möglich, bekannte Identitäten mit unbekanntem Identitäten vergleichend zuzuordnen. Diese Methode soll computergestützte Möglichkeiten zur Identifizierung von Täterinnen und Tätern aufzeigen und bereits etablierte Verfahren unterstützen. Zum gegenwärtigen Zeitpunkt gibt es verschiedene Methoden, digitale Rigs von Personen abzuleiten. Diese Methoden werden ausführlich in Pistorius u.a. (vgl. Pistorius et al. 2023) dargestellt und verglichen.

In der vorliegenden Arbeit wird ein Rig-Vergleich zwischen einer tatverdächtigen Person und einer bzw. eines im Video agierenden Täterin oder Täters vorgestellt. Detailliertere Ausführungen zur Methode und den statistischen Auswertungen und

Beobachtungen sind in Becker u.a. (vgl. Becker et al. 2022) ausgeführt. Die im Beispiel gezeigten digitalen Rigs wurden mit dem intelligenten System OpenPose generiert (vgl. Andriluka et al. 2014; Cao et al. 2021; An et al. 2020).

Die Weiterentwicklungen auf dem Gebiet der künstlichen Intelligenz erlauben eine Neudefinition der Bewegungsanalyse im Umfeld der Forensik. Mit neuen Frameworks ist es möglich, kontextunabhängige Vergleiche von Personen in unterschiedlichem Videomaterial durchzuführen. Die Bewegung kann als Abfolge von Posen mit einem konstanten Rig angesehen werden. Als Pose wird in diesem Zusammenhang die eingenommene Körperhaltung einer Person zu einem bestimmten Zeitpunkt verstanden.

2. FALLBEISPIEL ZUR ANWENDUNG VON DIGITALEN RIGS FÜR EINE PERSONENFESTSTELLUNG

2.1 Auftrag

Im Jahr 2021 erhielt das Forensic Science Investigation Lab (FoSIL) den Auftrag zur Erstellung eines wissenschaftlich-technischen Sachverständigen-gutachtens im Zusammenhang mit dem Ermittlungsverfahren wegen versuchten schweren Raubes zum Nachteil eines Finanzhauses in Berlin; im Speziellen mit dem Auftrag zur Erstellung eines Bildidentifikationsgutachtens und einer Bewegungsanalyse durch Vergleich des übersandten Untersuchungsmaterials mit dem erkennungsdienstlichen Material (ED-Material) einer oder eines Tatverdächtigen (TV).

In einer Filiale eines Finanzhauses verschafften sich zwei Täterinnen oder Täter Zugang zum Tresorraum im Untergeschoß. Unter Einsatz einer Handfeuerwaffe sollte eine Angestellte den Tresor öffnen. Diese kriminelle Handlung wurde durch

die im Tresorbereich verbaute Videotechnik aufgezeichnet. Abbildung 1 zeigt ein Einzelbild, einen sogenannten Frame, aus dem Tatvideo. Deutlich erkennbar ist das Verschleierungsverhalten der Täterinnen oder Täter. Es sind keine biometrischen Merkmale bzw. Merkmalsträger, wie das Gesicht, erkennbar, die zu einer Identifizierung der Täterinnen oder Täter durch Vergleich mit Tatverdächtigen oder Einträge in Datenbanken führen könnten. Ermittlungen führten zu einem möglichen Tatverdächtigen, jedoch konnte auch hier kein belastbarer Vergleich durchgeführt werden. Klassische anthropologische Vergleiche zwischen dem ermittelten Tatverdächtigen und den beobachtbaren Täterinnen oder Tätern scheiterten an deren Kleidung.

Die Identifikation von Personen innerhalb von digitalem Bild- und Videomaterial kann nur mittels Identifikationsverfahren, wie z.B. Verfahren der Gesichtserkennung, erfolgen. Möglichkeiten der Gang- bzw. Bewegungsanalyse können eine Identifikation unterstützen. Aufgrund dessen, dass beide Täterinnen oder Täter im zugrunde liegenden Tatvideo ihre Identität gezielt verschleiern (durch Masken verummte Gesichter), sind keine verwertbaren Ergebnisse aus der Anwendung von Gesichtsidentifikationsverfahren zu erwarten. Für mögliche Analysen des Gangs bzw. spezifischer Bewegungen der Täterinnen oder Täter im Video fehlt es zur gutachterlichen Stellungnahme oft an aussagekräftigem Vergleichsmaterial. Durch das Tragen von Mänteln oder raumeinnehmender Kleidung wird eine Gang- oder Bewegungsanalyse deutlich erschwert.

2.2 Einsatz des digitalen Rigs

Am Beispiel dieses Tatvideos wird die Rig-Ableitung der beiden tatverdächtigen Personen demonstriert und eine statisti-

Quelle: Labudde (eig. Darstellung)



Abb. 1: Frame aus dem Tatvideo – Im Tresorraum der Filiale sind zwei widerrechtlich handelnde Personen (Täter 1 und Täter 2) beobachtbar, die eine Filialangestellte zwingen, den Tresor zu öffnen. Beide kriminell handelnde Personen sind durch Tragen von besonderer Kleidung verschleiert. Die Frames haben eine Auflösung von 576 px x 704 px.

sche Bewertung vorgeschlagen, um den ermittelten Tatverdächtigen mit den Täterinnen oder Tätern zu vergleichen und der kriminellen Handlung zuzuordnen. Der generelle Ablauf der Methode, einschließlich der Rekonstruktion des Tatorts, umfasst die folgenden Schritte:

- ▶ Bewertung und Analyse des vorliegenden Videomaterials und gegebenenfalls Aufbereitung und Verbesserung des Videomaterials;
- ▶ Erstellung des 3D-Modells des Ereignisortes mit virtueller Kamera;
- ▶ Erstellung der digitalen Rigs des(r) im Video wahrnehmbaren Täterin(nen) bzw. Täter(s) und der tatbeteiligten Personen;
- ▶ 3D-Modell der tatverdächtigen Personen und Ableitung der korrespondierenden Rigs der tatverdächtigen Personen mittels erweiterter erkenntnisdienstlicher Behandlung;
- ▶ Zusammenführung der unterschiedlichen Modelle und Superposition des Tatvideos;
- ▶ mathematischer Vergleich der Rigs im einheitlichen metrischen Raum;
- ▶ statistische Bewertung des Vergleichs.

In der Analyse der Ergebnisse und zum

Quelle: Labudde (eig. Darstellung)



Abb. 2: Gesamtmodell: Zusammenführung des 3D-Modells des Tatortes, des Modells der virtuellen Kamera und der Superposition des Tatvideos. A: Gesamtübersicht Modell des Tatortes, B: Modell des Tatortes mit Blick auf den Tresorraum, C: Tatvideo, D: Virtuelle Kamera und Sichtfenster der Kamera mit Video auf den Tatort, E: Zusammenführung von 2D- und 3D-Informationen.

Verständnis der abgeleiteten Resultate sind folgende Begriffe zu definieren: Pose, Körperhöhe und Körpergröße. Als Pose wird innerhalb eines Videoframes die eingenommene Haltung einer Person bezeichnet, die sich aus dem biomechanischen Bewegungsablauf des Gelenkapparates ergibt. Die Körperhöhe ist die messbare Strecke vom Boden bis zu einem wohldefinierten Kopfpunkt bei einer aktuell eingenommenen Pose. Dies steht im Gegensatz zur Körpergröße, die der maximal eingenommenen Körperhöhe einer Person bei geschlossenem Beinstand, geradem Kopf und ausgestreckter Körperhaltung entspricht. In Videoaufnahmen ist also nur in den seltensten Fällen die Körpergröße bestimmbar. Dadurch ist ein Vergleich mit Maßen aus der erkennungsdienstlichen Behandlung von tatverdächtigen Personen nur bedingt möglich und muss hinterfragt werden.

Der Ereignisort in der Bankfiliale im vorliegenden Fallbeispiel wurde mit einem herkömmlichen terrestrischen Laser Scanner (Typ FARO FOCUS 3D X130) aufgenommen und als 3D-Modell in die Software

Blender (vgl. Brito 2007; Flor 2011) überführt. Die Abbildungen 2A und 2B zeigen den Ereignisort als 3D-Modell. In das Tatortmodell werden virtuelle Kameras eingepasst und mit den dazu korrespondierenden Parametern versehen (siehe Abbildung 2D). Somit ist eine Betrachtung des Tatgeschehens im 3D-Modell möglich. Da das zweidimensionale Video (siehe Abbildung 2C) mit dem 3D-Modell des Ereignisortes über die virtuelle Kamera verknüpft ist, sind metrische Messungen auf dem Videomaterial möglich. Die Abbildung 2E zeigt das 3D-Modell mit dem eingepassten Überwachungsvideo.

Die Vorhersage sogenannter Schlüssel- punkte am Bewegungs- bzw. Skelettapparat des TV im erweiterten ED-Material, der Täterinnen oder Täter im Tatvideo sowie innerhalb von Bildern eines Human-Solutions-Datensatzes im Rahmen eines Evaluationsschrittes erfolgte auf Grundlage der künstlichen Intelligenz (KI) OpenPose. OpenPose (vgl. Cao et al. 2021; Cao et al. 2017; Tomas et al. 2017; Shih-En 2016) war das erste Echtzeit-Mehrpersonen-System, das menschliche Körper-, Hand-, Gesichts- und Fußschlüssel- punkte auf einzelnen Bildern gemeinsam erkennen kann.

Für alle beobachtbaren Personen im Tatvideo werden die Gelenkpunkte mittels OpenPose in der eingenommenen Pose (Körperhaltung) ermittelt. Für die abgeleiteten Gelenkpunkte (siehe Abbildung 4, Seite 33) werden anschließend die Koordinaten berechnet. Das sich so ergebende Set an Gelenkkoordinaten bildet die Grundlage für den späteren Vergleich mit den Rigs möglicher tatverdächtiger Personen.

Die KI OpenPose ermöglicht nun das automatische Ableiten von digitalen Rigs. Die Abbildung 4 zeigt exemplarisch einen Videoframe, in dem das Rig einer Täterin oder eines Täters angezeigt wird. Aufgrund der Verknüpfung von Video und Kamera ist eine metrische Bestimmung der Maße

am Rig möglich. Parallel dazu wird von einer tatverdächtigen Person mit Hilfe von OpenPose in einem separaten Raum, der ebenfalls als 3D-Modell vorliegt, ein individuelles digitales Rig abgeleitet. Die Abbildungen 3B und 3C zeigen ein individuelles Rig einer tatverdächtigen Person, das mit OpenPose generiert wurde.

2.3 Bewertung

Neben der visuellen Betrachtung der Überlagerung beider Rigs (Tätverdächtiger und Täterin oder Täter), wobei dies lediglich der Kontrolle der Pose dient, geht es um einen mathematischen Vergleich. Aus dem mathematischen Vergleich lassen sich Ähnlichkeitsmaße ableiten. In der Software Blender ist es möglich, die Koordinaten einzelner Objekte im 3D-Raum zu erhalten. Dies gilt insbesondere auch für die integrierten Muster der Rigs. Durch die Verwendung von RMSD kann gemessen werden, wie weit die Koordinaten jedes verwendeten Gelenkpunktes durchschnittlich voneinander entfernt sind. Der mittlere Abstand der Gelenkpunkte zweier Rig-Muster berechnet sich wie folgt:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i^2)}$$

Dabei ist n die Anzahl der verwendeten Gelenkpunkte und d_i die individuellen Koordinaten. Der Wert d_i entspricht im Euklidischen Raum folgender Gleichung:

$$d(a, b) = \sqrt{(x_a - x_b)^2 + (y_a + y_b)^2}$$

Dabei sind x und y die Koordinaten der ausgewählten Gelenkpunkte.

Anhand von berechneten RMSD in verschiedenen Videoframes kann eine Statistik erstellt werden, die eine Beurteilung der

Quelle: Labudde (eig. Darstellung)

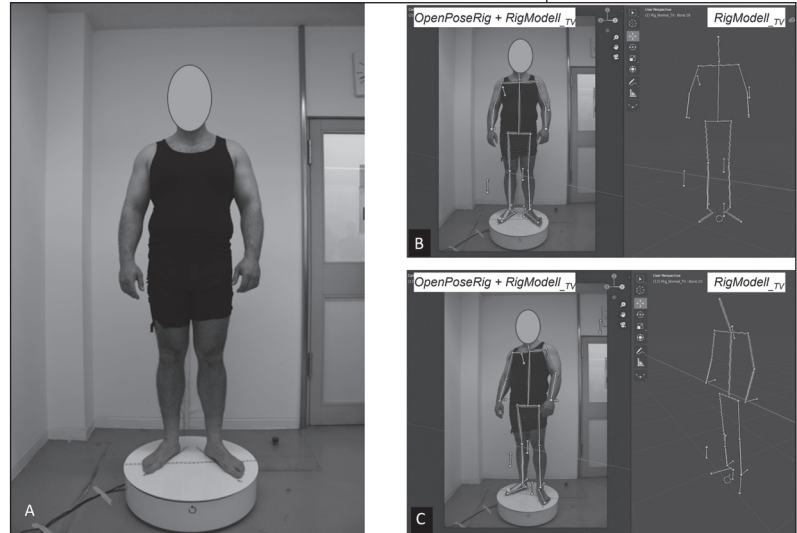


Abb. 3: Ableitung des Rig-Modells des Tatverdächtigen auf Grundlage von OpenPose. Die Basis für das Rig bilden mittels OpenPose vorhergesagte Schlüsselunkte. A: Der TV im Rahmen der erweiterten (photogrammetrischen) Behandlung, stehend auf einem Drehteller. B und C: Das Rig-Modell des Tatverdächtigen innerhalb von Blender aus zwei Perspektiven. Der virtuelle Raum, in dem das Rig erstellt wird, wird ebenfalls durch einen Laser-Scan in einen metrischen Raum überführt. Damit folgt dieses Rig der gleichen Metrik wie das 3D-Modell des Ereignisortes mit den virtuellen Kameras. Die mittels OpenPose vorhergesagten Schlüsselunkte an den Gelenken des TV sind bunt dargestellt. Diese dienen neben der Rig-Ableitung ebenfalls der Orientierung am Rig für den Abgleich mit anderen Rigs.

Quelle: Labudde (eig. Darstellung)

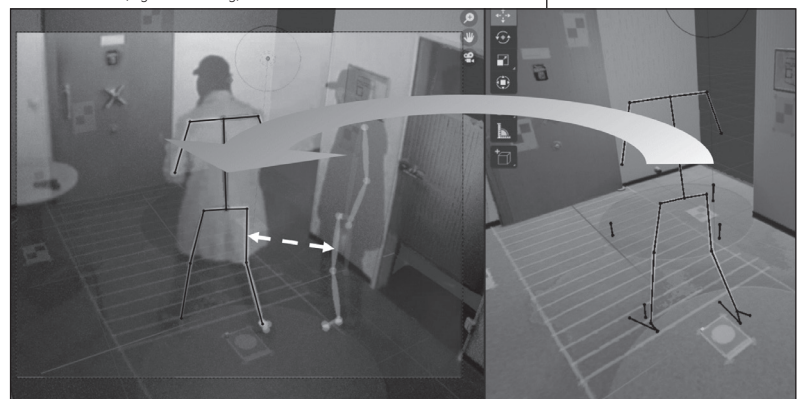


Abb. 4: Exemplarische Darstellung eines Rig-Abgleichs - Links: OpenPose-Rig bei den Täterinnen oder Tätern im Tatvideo. Das schwarze Rig innerhalb der Bildinformationen des OpenPose-Rigs bei Täter 1 im Tatvideo repräsentiert das Rig-Modell des Tatverdächtigen nach erfolgtem Rig-Abgleich (dicker Pfeil). Das Rig-Modell des Tatverdächtigen links im Bild wurde dabei stärker hervorgehoben, um die visuelle Beurteilung des hier Dargestellten zu unterstützen (schematische Darstellung). Rechts: Das Rig-Modell des Tatverdächtigen innerhalb des 3D-Modells des Tatorts mit der gleichen Pose wie Täter 1 im Tatvideo, dargestellt innerhalb von Blender.

Quelle: Labudde (eig. Darstellung)

Frame-Nummer	RMSD Täter/TV – cm ohne Hüfte und Knie	RMSD Täter/TV – cm ohne Knie
1	1,558613276	1,359808621
2	1,540231125	1,663639756
3	1,406421693	1,321645021
4	1,274278711	1,201112657
5	1,661191685	1,51226449
6	1,3223647	1,447416766
7	1,172628068	1,112452663
8	1,387239462	1,260022814
9	1,067913167	1,42632226
10	1,282004457	1,229655477
Mittelwert	1,36728863	1,35343405

Abb. 5: RMSD ohne die Koordinaten der Knie und der Hüfte für den Täter 1 aus dem Tatvideo mit einer tatverdächtigen Person

Übereinstimmung verschiedener Rigs ermöglicht.

Ein Schlüsselpunktpaar setzt sich dabei aus zwei definierten Punkten, z.B. linker Schulterpunkt am Rig-Modell der/des Tatverdächtigen und korrespondierender Punkt eines Rigs der Täterin oder des Täters im Video, zusammen. Folgende Schlüsselpunkte wurden dabei berücksichtigt: linke Schulter, Schultermitte, rechte Schulter, linker Ellbogen, rechter Ellbogen, linkes Handgelenk, rechtes Handgelenk, linkes Knie, rechtes Knie, linker Knöchel und rechter Knöchel. Der so berechnete RMSD wird auf das Pixel/cm-Verhältnis übertragen, um Personen, die durch unterschiedliche Pixelanzahl in einzelnen Frames dargestellt werden, zu normieren. Es ergibt sich ein RMSD in cm (siehe Abbildung 5, 2. Spalte), der auf allen durchgeführten Vergleichen anwendbar ist. Durch die Operationen erhält man eine Möglichkeit, digitale Rigs von Personen in unterschiedlichen Posen zu vergleichen und zu bewerten. Erst dadurch wird es möglich, kriminell handelnde Personen mit tatverdächtigen Personen auf eine Ähnlichkeit hin zu prüfen.

Die RMSD-Werte für den Täter 2 (siehe Abbildung 1, Seite 31) mit dem Rig des

Tatverdächtigen weisen einen Mittelwert von 10,196 cm auf. Somit kann ausgeschlossen werden, dass es sich bei Täterin oder Täter 1 um den ermittelten Tatverdächtigen handelt.

Studien haben gezeigt, dass das abgeleitete Rig bzw. die Positionen der abgeleiteten Schlüsselpunkte der Hüfte durch Hosen und Shirts beeinflusst werden können. In diesem Fall kommt hinzu, dass die Täterin oder der Täter einen Mantel getragen hat, der die Ableitung der Kniegelenkpunkte beeinflusst. Der Mantel verdeckt somit die Hüfte und zum Teil beide Knie. Aus diesem Grund wurden die Koordinaten der Hüfte und/oder der Knie aus der Analyse ausgeschlossen (vgl. Diederich 2023). Die Analyse zeigt, dass die Hüftkoordinaten konstant abgeschätzt wurden, jedoch eine große Variation der Kniegelenke vorhanden ist. Weitere Details finden sich in Pistorius u.a. (vgl. Pistorius et al. 2023) und Becker u.a. (vgl. Becker et al. 2022).

In der Arbeit von Heinke u.a. (vgl. Heinke et al. 2022) wurde ein anderes Ähnlichkeitsmaß verwendet, um die Individualität von Rigs zu beurteilen. Heinke u.a. (vgl. ebd.) befassten sich mit Ansätzen zur Schätzung von Identitätswahrscheinlichkeiten, die aus anthropometrischen Maßen abgeleitet werden. Hier wird eine neuartige Häufigkeitsschätzung anhand eines Datensatzes von 340 Personen und ihrer jeweiligen anthropometrischen Maße vorgestellt. Auf der Grundlage von Dichtekernschätzungen der Messungsungleichheit wird vorgeschlagen, dass die Identitätswahrscheinlichkeiten in der Größenordnung 10^{-15} bis 10^{-8} liegt. In der Arbeit Heinke u.a. (vgl. Heinke et al. 2023) wurden die Ansätze verbessert und stellen eine Grundlage der Berechnung von unterschiedlichen Kombinationen von Messstrecken im Vergleich von Rigs dar. Diese Ergebnisse sind in der Abbildung 6

zusammengefasst. Die Wahrscheinlichkeit, dass zwei Rigs von unterschiedlichen Personen identisch sind, variiert durch die Verwendung unterschiedlicher Kombinationen der Messstrecken, die händisch an den Probanden vermessen wurden. Dies ist bei der Verwendung von KI-Systemen nicht der Fall, zeigt jedoch deutlich das Potenzial der digitalen Rigs als biometrisches Merkmal in der Forensik.

Mittels der Studie COMBI wurden 64 Individuen auf Basis von generierten OpenPose-Rigs miteinander kreuzweise verglichen (vgl. Bergman et al. 2022; Pistorius et al. 2023). Grundlage der Vergleiche bildet ebenfalls der RMSD. Aus den durchgeführten Versuchen konnte ein Schwellwert bestimmt werden, der die RMSD-Werte der identischen Probanden optimal von den RMSD-Werten der nicht-identischen Probanden teilt. Daraus ergibt sich, dass bei einem RMSD von unter 2 cm eine Ähnlichkeit angenommen wird. Andererseits sollte man von unterschiedlichen Individuen ausgehen. Die Verteilung der RMSD-Werte aus einem Kreuzvergleich zwischen 16 unterschiedlichen Probanden sind in Abbildung 7 (siehe Seite 36) visualisiert. Der minimal erreichte Wert, der bei einem Abgleich zwischen zwei divergenten Probanden ermittelt wurde, betrug 1,841. Als maximaler Wert zwischen zwei unterschiedlichen Probanden wurde ein RMSD von 26,523 berechnet. Der durchschnittliche RMSD zwischen zwei divergenten Probanden ist dabei 8,2436 und der Median 7,247. Bei den Abgleichen zwischen Rigs identischer Probanden ergab sich ein minimaler RMSD von 0,93, während ein maximaler Wert von 2,075 erreicht wurde. Der berechnete Mittelwert (durchschnittlicher RMSD) zwischen identischen Probanden war 1,5927 und der Median 1,6372.

Der RMSD (in cm) des Tatverdächtigen ohne die Einbeziehung der Koordinaten

Quelle: Labudde (eig. Darstellung)

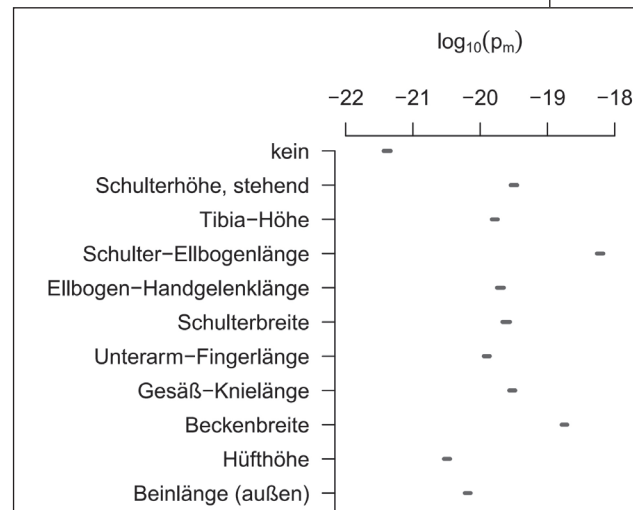


Abb. 6: Kombination von Messstrecken und resultierende Identitätswahrscheinlichkeiten. Dargestellt sind unterschiedliche Berechnungen der Identitätswahrscheinlichkeiten für den Fall, dass nicht alle Messstrecken berücksichtigt werden. So sinken die Identitätswahrscheinlichkeiten deutlich, wenn die Länge Schulter-Ellbogen nicht einbezogen wird.

der Knie betrug durchschnittlich 1,35 (siehe Abbildung 5, 3. Spalte, Seite 34). Zwischen den in den Videoaufnahmen durch die KI vorhergesagten Gelenkpunkten der Täterin oder des Täters und dem eingepassten KI-basierten Rig des Tatverdächtigen ergibt sich im Mittel ein RMSD von 1,35 cm. Bei einem RMSD-Wert unter dem erläuterten Schwellwert von 2 cm kann der Tatverdächtige nicht als Täter ausgeschlossen werden.

2.4 Einflussfaktoren

Einflussfaktor Laserscan: Mit einem systematischen Distanzfehler von ± 2 mm sowie einem Rauschen von max. 0,3 mm bis 0,4 mm (bei 90 % bis 10 % Reflektion) im Rahmen der vor Ort durchgeführten 10-m-Scans, ist dieser Einflussfaktor zu vernachlässigen.

Einflussfaktor OpenPose: OpenPose-Vorhersagen unterliegen einem zugrunde liegenden Vorhersagemodell, das auf Grundlage einer definierten Anzahl von Bildern trainiert und evaluiert wurde. Die Qualität vorhergesagter Rig-Schlüssel-

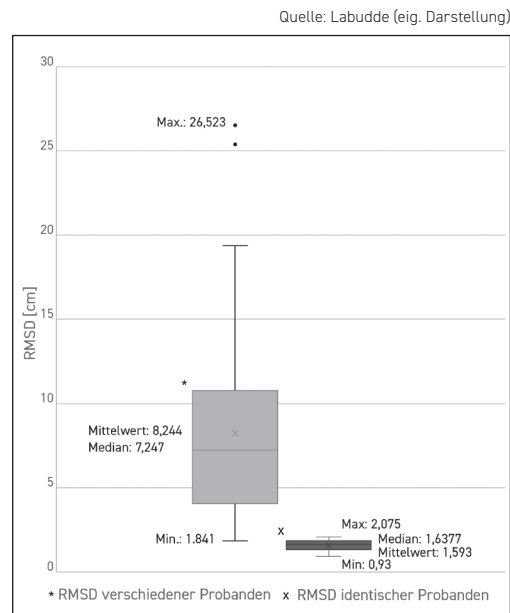


Abb. 7: Verteilung der RMSD-Werte von 16 Probanden aus dem Projekt COMBI. Dargestellt ist die Verteilung der RMSD-Werte aus den Kreuzvergleichen zwischen 16 unterschiedlichen Probanden im Projekt COMBI. Die y-Achse bildet die RMSD-Werte in cm ab. Hellgrau ist die Verteilung der RMSD-Werte zwischen unterschiedlichen Probanden und dunkelgrau die Verteilung der RMSD-Werte der identischen Probanden.

punkte wird dabei durch eine Vorhersagewahrscheinlichkeit angegeben und wird im Bild durch unterschiedliche Radien abgebildet. Das bedeutet, dass der Durchmesser eines Rig-Schlüsselpunktes umso größer wird, je unsicherer auch die OpenPose-Vorhersage für diesen Punkt wird. Da jedes Pixel eines Rig-Schlüsselpunktes einen potenziellen Endpunkt für die Ableitung von dem Rig-Modell des Tatverdächtigen darstellt, wurden ausschließlich die Zentren der Rig-Schlüsselpunkte im Rahmen der Rig-Ableitung als Endpunkte berücksichtigt. Gleiches gilt für den Prozess des Rig-Abgleichs. Aufgrund der schlechten Videoqualität werden hier Rig-Schlüsselpunkte durch größere Radien abgebildet als bei den OpenPose-Vorhersagen des erweiterten (photogrammetrischen) erkennungsdienstlichen Bildmaterials. Im Rahmen des Rig-Abgleichs stellen demzufolge alle Pixel aller im Tatvideovorhergesagten Rig-Schlüsselpunkte mög-

liche Endpunkte dar, an dem sich das Rig-Modell des Tatverdächtigen dann orientieren kann. Damit wird der Toleranzbereich beim Rig-Abgleich größer. Das erhöht die Wahrscheinlichkeit, dass keine klare Personenzuordnung aufgrund von Mehrfachzuordnungen erfolgen kann, wenn es sich beim Abgleich um mehrere Personen mit ähnlicher Körperhöhe handelt. Im vorliegenden Fall ist dies jedoch vernachlässigbar, da beide Täterinnen oder Täter einen Körperhöhenunterschied von ≈ 10 cm zueinander aufweisen und beide damit nicht im selben Toleranzbereich liegen.

3. AUSBLICK UND WEITERE ARBEITEN

Die Bewegung als Abfolge von Posen rückt durch die Weiterentwicklung auf dem Gebiet der Schätzung der Pose von Personen in den wissenschaftlichen Fokus und erlaubt die Frage nach einer neuen Bewegungsanalyse des Menschen im forensischen Kontext. Die Beschreibung des Rigs als individuelles biometrisches Merkmal und die Herleitung der Duplikations-Wahrscheinlichkeit sowie die Möglichkeit der Detektion des individuellen Rigs erlauben diesen Weg. Arbeiten mit unterschiedlichen Koordinatensätzen unterstützen die Analyse der Posenauswertung. Der Gang als Sequenz von Posen stellt neue Ansätze und eine qualitative Weiterentwicklung dar.

Die Frameworks AlphaPose und Motion BERT stellen eine solche Weiterentwicklung dar. Präzise Ganzkörper-Positionsschätzung und -verfolgung von mehreren Personen ist ein wichtiges aber schwieriges Thema in der Computer Vision. Um subtile Handlungen von Menschen für eine komplexe Verhaltensanalyse zu erfassen, ist eine Ganzkörper-Positionsschätzung unerlässlich. AlphaPose ist ein neues System, das eine genaue Ganzkörper-Posenschätzung und -verfolgung in Echt-

zeit durchführen kann (Abbildung 8B). In AlphaPose werden neue Techniken verwendet, die eine präzisere Schätzung ermöglichen. Dazu gehören: Symmetric Integral Keypoint Regression (SIKR) für schnelle und feine Lokalisierung, Parametric Pose Non-Maximum-Suppression (P-NMS) zur Eliminierung redundanter menschlicher Erkennungen und Pose Aware Identity Embedding zur gemeinsamen Posenschätzung und -verfolgung. MotionBERT erzeugt eine einheitliche Perspektive aus verschiedenen menschenzentrierten Videoaufgaben durch das Lernen menschlicher Bewegungsdarstellungen aus großen und heterogenen Datenquellen. Konkret wird eine Prätrainingsphase durchgeführt (Abbildung 8C), in der ein Bewegungskodierer trainiert wird, um die zugrunde liegende 3D-Bewegung aus verrauschten 2D-Teilbeobachtungen zu ermitteln. Die Bewegungsdarstellungen, die auf diese Weise gewonnen werden, enthalten geometrisches, kinematisches und physikalisches Wissen über die menschliche Bewegung. Dieser Schritt wird durch die Implementierung eines

Dual-stream Spatio-temporal Transformer (DSTformer) abgebildet.

Das Konzept des künstlichen, digitalen Rigs zur Analyse und Bewertung von beobachtbaren Personen in Tatvideos verdeutlicht den Zerfall von Spuren. Ein Video, das eine kriminelle Handlung abbildet, kann als digitale Spur angesehen werden. Auch digitale Spuren lassen sich in einen statischen und dynamischen Anteil separieren.

Betrachtet man das Rig als Muster, so kann es bewertet und mit anderen Mustern verglichen werden. Betrachtet man Bewegung als Abfolge von Posen, die sich eindeutig mit einem Rig beschreiben lassen, so kann durch Interpretation der dynamische Anteil genutzt werden, um auch eine zeitliche Abfolge der Tathandlung zu beschreiben.

Die Abbildung 8C erlaubt eine Trennung der beobachtbaren Bewegung einer Person vom Videocontent. Somit wäre eine Beurteilung beider Inhalte separat möglich. Das unabhängige Rig kann dann für den Vergleich von Personen über Videos hinweg genutzt werden.

Quelle: Labudde (eig. Darstellung)



Abb. 8: Frameworks AlphaPose und MotionBERT – Anwendung der Frameworks auf Videomaterial aus der COMBI-Studie. A – Darstellung eines Beispielframes und B – abgeleitetes Rig mit AlphaPose. C stellt das Ergebnis von MotionBERT dar.

In dieser Arbeit wurden an einem konkreten Fallbeispiel die Möglichkeiten des künstlichen digitalen Rigs für die Forensik erläutert. Es wurden die einzelnen Schritte zur Erstellung eines künstlichen, digitalen Rigs an einem Beispielfall detailliert beschrieben. Die Arbeit beschreibt ein

Konzept zur Bestimmung der Ähnlichkeit zweier Rigs, hier das Rig einer Täterin oder eines Täters im sogenannten Tatvideo und das erstellte Rig eines Tatverdächtigen. Der Vergleich lieferte das Ergebnis, dass der Tatverdächtige nicht als Täter ausgeschlossen werden kann.

¹ Das digitale Rig stellte eine Abstraktion des menschlichen Skeletts dar.

Quellenangaben

Albright, Thomas D. (2022). How to make better forensic decisions, *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 119 (38), Online: <https://doi.org/10.1073/pnas.2206567119> (29.08.2023).

An, Weizhi et al. (2020). Performance Evaluation of model-based gait on multi-view very large population database with pose sequences, *IEEE Transactions on Biometrics Behavior, and Identity Science*, 2 (4), 421–430.

Andriluka, Mykhaylo et al. (2014). 2D human pose estimation: new benchmark and state of the art analysis, *IEEE conference on computer vision and pattern recognition, San Diego, CA, USA*, 3686–3693, Online: <https://doi.org/10.1109/CVPR.2014.471> (29.08.2023).

Becker, Sven et al. (2021). Hypothesis-driven Case Analysis in 3D-Space as a Support for Forensic Casework, *INFORMATIK 2021, Gesellschaft für Informatik, Bonn*, 835–846, Online: DOI: 10.18420/informatik2021-071 (29.08.2023).

Becker, Sven et al. (2022). COMBI: Artificial Intelligence for Computer-Based Forensic Analysis of Persons, *KI – Künstliche Intelligenz* (36), 171–180. <https://doi.org/10.1007/s13218-022-00761-x> (29.08.2023).

Bergman, Tommy et al. (2022). Automatic Classification of Bloodstains with Deep Learning Methods, *KI – Künstliche Intelligenz* (36),

135–141, Online: <https://doi.org/10.1007/s13218-022-00760-y> (29.08.2023).

Bertillon, Alphonse/McCloughry, Major R. W. (1896). *Signaletic instructions including the theory and practice of anthropometrical identification*, Chicago, New York, London.

Bevel, Tom (2002). *Bloodstain pattern analysis: with an introduction to crime scene reconstruction*, Boca Raton.

Brito, Allan (2007). *Blender 3D*, Novatec, New York.

Cao, Zhe et al. (2017). Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, *Proceedings of the IEEE conference on computer vision and pattern recognition (CPVR)*, 7291–7299, Online: <https://doi.org/10.48550/arXiv.1611.08050> (29.08.2023).

Cao, Zhe et al. (2021). OpenPose: Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43 (1), 172–186, Online: <https://doi.org/10.1109/TPAMI.2019.2929257> (29.08.2023).

Cole, Simon A. (2007). Twins, Twain, Galton, and Gilman: Fingerprinting, Individualization, Brotherhood, and Race in Pudd'nhead Wilson, *Configurations* 15 (3), 227–265, Online: <https://doi.org/10.1353/con.0.0036> (29.08.2023).

Dantcheva, Antitza et al. (2011). Bag of soft biometrics for person identification, *Multimed Tools Appl* (51), 739–777, Online: <https://doi.org/10.1007/s11042-010-0635-7> (29.08.2023).

Diederling, Jessica (2023). Analyse des Einflussfaktors Kleidung auf die Vorhersagege-

- nauigkeit Gelenk-spezifischer Punkte mittels Künstlicher Intelligenz, Bachelorarbeit, Hochschule Mittweida, University of Applied Sciences, Fakultät Angewandte Computer- und Biowissenschaften, Mittweida.
- Flor, Nick V. (2011). Technology corner: virtual crime scene reconstruction: the basics of 3d modeling, *Journal of Digital Forensics Security and Law*, 6 (4), 6, Online: <https://doi.org/10.15394/jdfsl.2011.1108> (29.08.2023).
- Heinke, Florian et al. (2022). A frequentist estimation of duplicate probability as a baseline for person identification from image and video material using anthropometric measurements, in: Demmler, Daniel et al. (Hg.), *INFORMATIK 2022, Gesellschaft für Informatik, Tagungsband, Bonn*, 91–98, Online: DOI: 10.18420/inf2022_07 (29.08.2023).
- Heinke, Florian et al. (2023). Analysing Distributions of Feature Similarities in the Context of Digital Anthropometric Pattern Matching Probability, accepted in: *INFORMATIK 2023. Gesellschaft für Informatik, Bonn*.
- Pistorius, El et al. (2023). The digital skeleton in modern video analysis – inter- and intraspecific comparison of individual rigs, accepted in: *INFORMATIK 2023. Gesellschaft für Informatik, Bonn*.
- Schneickert, Hans (1923). *Signalementslehre: Handbuch der Personenbeschreibung für Polizeibehörden, Gendarmerie- und Polizeischulen*, 2. Aufl., Reprint 2021, Berlin.
- Shih-En, Wie (2016). Convolutional pose machines, *CVPR*, Online: <https://doi.org/10.48550/arXiv.1602.00134> (29.08.2023).
- Tomas, Simon et al. (2017). Hand keypoint detection in single images using multiview bootstrapping, *CVPR*, Online: <https://doi.org/10.48550/arXiv.1812.00324> (29.08.2023).

Weiterführende Literatur und Links

- Bratton, Robin M. (2002). Bloodstain Pattern Analysis with an Introduction to Crime Scene Reconstruction, *Journal of Forensic Identification*, 52 (4), 477– 479.
- Fang, Hao-Shu et al. (2017). Regional Multi-person Pose Estimation, *International Conference on Computer Vision (ICCV)*, 2334–2343, Online: <https://doi.org/10.48550/arXiv.1612.00137> (29.08.2023).
- Fang, Hao-Shu et al. (2022). AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45 (6), 7157-7173, Online: <https://doi.org/10.1109/tpami.2022.3222784> (29.08.2023).
- Jiefeng, Li et al. (2019). Crowdpose: Efficient crowded scenes pose estimation and a new benchmark, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CPVR)*, 10863–10872, Online: <https://doi.org/10.48550/arXiv.1812.00324> (29.08.2023).
- Martin, Rudolf (1929). *Anthropometrie – Anleitung zu Selbständigen Anthropologischen Erhebungen*, Online: <https://doi.org/10.1007/978-3-662-31681-8> (29.08.2023).
- Zhu, Wentao et al. (2023). MotionBERT: A Unified Perspective on Learning Human Motion Representations, *Proceedings of the IEEE/CVF International Conference on Computer Vision*.