



Zadání bakalářské práce

Název:	Analýza a zpětná vazba správnosti provádění cviků
Student:	Daniel Pilař
Vedoucí:	Ing. Jakub Novák
Studijní program:	Informatika
Obor / specializace:	Znalostní inženýrství
Katedra:	Katedra aplikované matematiky
Platnost zadání:	do konce letního semestru 2023/2024

Pokyny pro vypracování

Cílem práce je automatické zhodnocení správnosti provádění cviků a poskytnutí zpětné vazby cvičícímu.

Ideální představa je nahrazení trenéra inteligentním kamerovým systémem, který by byl schopen cvičícímu po provedeném cviku odhalit konkrétní chyby provádění daného cviku a jednoduše by mu je vizualizoval. Práce vyžaduje využití reálného kamerového vybavení laboratoře ImproLab a využití sofistikovaných technik hodnocení průběhů dat získaných z obrazu.

Úkoly:

- 1) Proveďte rešerši v oblasti zpracování obrazových dat s cílem získání skeletonu člověka, seznámte se s možnostmi hodnocení pohybové aktivity pomocí takových dat.
- 2) Sestavte měřicí kamerový systém pro snímání člověka během cvičení.
- 3) Vytvořte metodiku měření a nasnímejte reálná data vybraných cviků.
- 4) Definujte parametry ideálního průběhu cviku pro další porovnávání.
- 5) Navrhněte algoritmy pro získání hledaných parametrů z obrazu (skeletonu) a jejich porovnávání s ideálním průběhem.
- 6) Otestujte výsledky navržených algoritmů a zhodnoťte úspěšnost.
- 7) Vizualizujte výsledky algoritmu na obrazových datech.

Bakalářská práce

ANALÝZA A ZPĚTNÁ VAZBA SPRÁVNOSTI PROVÁDĚNÍ CVIKŮ

Daniel Pilarš

Fakulta informačních technologií
Katedra aplikované matematiky
Vedoucí: Ing. Jakub Novák
10. května 2023

České vysoké učení technické v Praze
Fakulta informačních technologií

© 2023 Daniel Pilař. Všechna práva vyhrazena.

Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení, je nezbytný souhlas autora.

Odkaz na tuto práci: Pilař Daniel. *Analýza a zpětná vazba správnosti provádění cviků*. Bakalářská práce. České vysoké učení technické v Praze, Fakulta informačních technologií, 2023.

Obsah

Poděkování	vii
Prohlášení	viii
Abstrakt	ix
Úvod	1
1 Rešerše	3
1.1 Klasifikace kvality cviků na posilovacích strojích	3
1.2 Analýza postavení těla při cvičení s využitím modelu lidského skeletonu	4
1.3 Vyhodnocování kvality rehabilitačních cviků pomocí zařízení Microsoft Kinect	5
1.4 Evaluace rehabilitačních cviků pomocí fuzzy pravidel	6
2 Teoretické zázemí	9
2.1 Reprezentace lidského těla	9
2.2 Metody umělé inteligence pro získání lidského skeletonu	10
2.2.1 BlazePose	10
2.2.2 MotionBERT	11
2.3 Dynamic Time Warping	11
2.4 Algoritmy pro vyhlazení časových řad	12
2.4.1 Klouzavý průměr	12
2.4.2 Lokální regrese	13
2.4.3 Butterworthův filtr	13
2.5 Rozhodovací stromy	14
2.5.1 Inference	14
2.5.2 Trénování	15
3 Analýza	17
3.1 Návrh postupu	17
3.2 Snímací soustava pro získání videozáznamů	18
3.2.1 Nasnímání vybraných cviků	18
3.3 Ideální průběh cviků	19
3.3.1 Ideální technika dřepu	19
3.3.2 Ideální technika kliku	20
3.3.3 Ostatní cviky a jejich důležité parametry	21
4 Realizace	23
4.1 Data	23
4.2 Implementace	23
4.3 Získání lidského skeletonu z videí	24
4.4 Předzpracování časových řad souřadnic	27

4.5	Získ parametrů ze souřadnic pro popis cviků	29
4.5.1	Volba parametrů pro reprezentaci	29
4.5.2	Vyhlazení a porovnání získaných parametrů z modelů pro získání skeletonů	29
4.5.3	Segmentace trénovacích cviků	29
4.5.4	Klasifikace cviku v neznámé sekvenci	31
4.5.5	Automatická segmentace jednotlivých opakování v sekvencích k vyhodnocení	32
4.5.6	Vyhodnocení správnosti cviků	35
5	Výsledky	37
5.1	Vyhodnocení cviků	37
5.2	Vizualizace	38
6	Diskuze	41
7	Závěr	43
A	Seznam použitých zkratk	45
	Obsah přiloženého média	49

Seznam obrázků

1.1	Jednotlivé kroky od zisku dat až po modelování [1].	3
1.2	Špatně provedený klik s vizualizovaným detekovaným skeletonem [2].	5
2.1	Ukázka typické podoby modelu lidského těla při kinematické reprezentaci [5]. . .	10
2.2	Ukázka zarovnání časových řad. Výsledná vzdálenost je určena jako součet rozdílů bodů, které jsou spojeny šedými spojnicemi [10].	12
2.3	Ukázka vyhlazení pomocí lokální regrese [13].	13
2.4	Příklad funkce amplitudové odezvy s $w_0 = 4$ a různými hodnotami n	14
3.1	Diagram popisující hlavní kroky při vyhodnocování.	17
3.2	Snímací soustava k záznamu cviků.	18
3.3	Ukázka pořízených snímků cviků. Na prvním obrázku je prováděn rumunský mrtvý tah na jedné noze a na druhém klik.	19
3.4	Ukázka jednotlivých fází správně provedeného dřepu.	20
3.5	Průběh úhlu v kolenu (modře) a průběh úhlu mezi osou trupu a rovinou podlahy (červeně) při ideálním provedení dřepu.	20
3.6	Ukázka jednotlivých fází správně provedeného kliku.	21
3.7	Průběh úhlu v rameni (modře) a průběh úhlu v kyčli (červeně) při ideálním provedení kliku.	21
4.1	Ukázka různých orientací natáčení při provádění pánevního mostu.	24
4.2	Ukázka špatně provedených cviků pro testování. Vlevo je vidět chyba při provádění rumunského mrtvého tahu na jedné noze, kdy zadní noha není dostatečně zvednutá. Vpravo je špatně provedený klik, kdy není dodržen rozsah a ruce jsou příliš daleko od sebe.	24
4.3	Popis kroků při zpracování trénovacích dat.	25
4.4	Popis kroků při zpracování a analýze dat k vyhodnocení.	26
4.5	Ukázka získaných souřadnic pomocí modelu BlazePose při provádění bulharského mrtvého tahu na jedné noze. Vizualizováno pomocí knihovny <code>mediapipe</code> , ve které je implementován i samotný BlazePose. Oranžově jsou vyznačeny části těla náležící levé straně, modře části náležící levé části.	27
4.6	Ukázka doplnění chybějících hodnot časové řady souřadnice Y levého lokte při provádění sklapovačky.	28
4.7	Ukázka různých metod vyhlazení časových řad demonstrováných na souřadnici Y levého lokte při provádění sklapovaček.	28
4.8	Vizualizace vypočítaných a zkoumaných parametrů. Zvýrazněny jsou oranžově úhly a relativní vzdálenosti.	30
4.9	Porovnání získaných příznaků z různých modelů pro získání skeletonu. Demonstrováno na úhlu v levém kolenu při provádění výpadů natáčených ze šikma.	30
4.10	Ukázka manuální segmentace demonstrovaná na průběhu úhlu v levém boku při provádění sklapovačky. Svislé černé čáry ohraničují jednotlivá opakování.	31

4.11	Ukázka nalezených opakování při automatické segmentaci. Demonstrováno na cviku pánevní most při pohledu zepředu.	33
4.12	Ukázka nalezených opakování při automatické segmentaci. Demonstrováno na dřepu při pohledu zešikma.	34
4.13	Ukázka výsledků segmentace pro vybrané orientace všech druhů chyb. Znázorněny jsou skutečné počty opakování, počet nalezených opakování a průměrná časová odchylka hranic nalezených segmentů a skutečnosti. Zpracováno modelem BlazePose.	34
4.14	Ukázka výsledků segmentace pro vybrané orientace všech druhů chyb. Znázorněny jsou skutečné počty opakování, počet nalezených opakování a průměrná časová odchylka hranic nalezených segmentů a skutečnosti. Zpracováno modelem MotionBERT.	34
5.1	Ukázka vizualizace správně (vlevo) a špatně (vpravo) provedeného přitahu na hrazdě.	39
5.2	Ukázka vizualizace správně (vlevo) a špatně (vpravo) provedeného dřepu.	39
5.3	Ukázka vizualizace správně (vlevo) a špatně (vpravo) provedené sklapovačky. Na špatném příkladu je vidět, že úhel v levé kyčli byl vyhodnocen špatně (ve skutečnosti je jeho úhel v pořádku). Hranice nebyla ale přesáhnutá tolik, proto je sousední tělní segment obarven pouze oranžově.	40
5.4	Ukázka textového popisu prováděného cviku a aktuálního počtu provedených opakování při vizualizaci.	40

Seznam tabulek

1.1	Testovací přesnost vybraných modelů pro všechny reprezentace příznaků [3].	6
3.1	Nasnímané druhy cviků.	19
3.2	Důležité parametry cviků dle názoru trenéra.	22
4.1	Výčet nasnímaných chyb při cvičení a jejich popis.	25
4.2	Validační a testovací výsledky rozhodovacích stromů a náhodných lesů pro klasifikaci cviků na jednotlivých snímcích.	32
5.1	Přehled výsledků evaluace správných a chybných cviků.	37

Seznam výpisů kódu

Chtěl bych poděkovat především panu inženýrovi Jakubu Novákovi za spoustu užitečných rad, trpělivost a především velice přátelský přístup při vedení bakalářské práce. Dále bych rád poděkoval Janu Bernardovi, sportovnímu trenérovi, který umožnil natočení všech potřebných videomateriálů a poskytl odborné rady.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací. Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů, zejména skutečnost, že České vysoké učení technické v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 citovaného zákona.

V Praze dne 10. května 2023

Abstrakt

Práce se zaměřuje na analýzu a vyhodnocování techniky cviků pomocí metod počítačového vidění a strojového učení. Ze vstupních videí cvičenců byly získávány lidské skeletony a z nich spočítány parametry popisující ideální průběh cviku. Pomocí metody Dynamic Time Warping byly testované cviky porovnávány s ideálním průběhem a každý parametr vyhodnocován zvlášť. Porovnávány byly různé modely pro získání skeletonu, stejně tak různé pozice, ze kterých byl cvičenec natáčen. Pro vybrané orientace, ze kterých byly cviky natáčeny, dosahovalo hodnocení kvality cviků průměrnou přesností 90 % a průměrné f-skóre 83 %.

Klíčová slova počítačové vidění, Dynamic time warping, získání lidského skeletonu, strojové učení, cvičení, časová řada

Abstract

The thesis focuses on the analysis and evaluation of the technique of strengthening exercises using the methods of computer vision and machine learning. From the input videos of the exercisers, human skeletons were obtained and parameters describing the course of the exercise were calculated from them. Using the Dynamic Time Warping method, the tested exercises were compared with the ideal course and each parameter was evaluated separately. Different models for skeleton extraction were compared, as well as different positions from which the exerciser was recorded. For the selected orientations from which the exercises were filmed, the quality evaluation of the exercises reached average accuracy of 90 % and average f-score of 83 %.

Keywords computer vision, Dynamic Time Warping, human skeleton extraction, machine learning, exercise, time series

Úvod

Fyzická aktivita by měla jistě být součástí zdravého životního stylu. Jedním z nejzákladnějších způsobů cvičení je provádění jednoduchých posilovacích cviků, jak s vlastní vahou, tak se závažím. Důležitým předpokladem pro správnou efektivitu cvičení a pro prevenci zranění je správná technika při provádění samotných cviků. V dnešní době sice existují dostupné zdroje s návody, jak dané cviky provádět, avšak zde chybí jakákoliv zpětná vazba. Lepším řešením je vyhledat pomoc sportovního trenéra.

S trenérem lze cvičit osobně na individuálním či hromadném tréninku, v obou případech dává cvičenci okamžitou zpětnou vazbu, která je velice důležitá. Tento přístup ale může být pro někoho nedostupný, například z časových či finančních důvodů. Dalším stupněm je tedy oprostít se od nutnosti přítomnosti trenéra a zpětnou vazbu poskytnout nějak automatizovaněji.

Pro tento účel se využívají automatické systémy (snímají cvičence pomocí kamer či pomocí nositelných měřicích senzorů), které analyzují prováděné cviky a poskytují zpětnou vazbu. Přítomnost trenéra je nutná pouze ve fázi sestavování systému a následné použití už může probíhat bez něj. Přístup je výhodný především z hlediska téměř okamžité dostupnosti pro cvičence. Jedním ze způsobů, jak takový systém realizovat, je využití dat z videokamery a pomocí metod strojového vidění a sofistikovaných počítačových algoritmů dále kvalitu cviků vyhodnocovat.

Kapitola 1

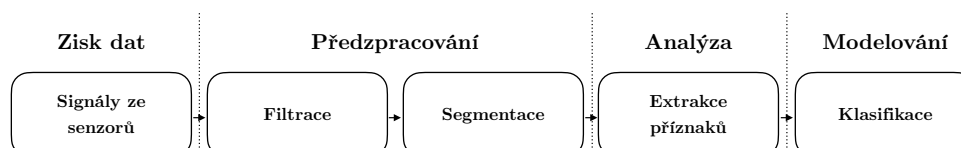
Rešerše

K problematice analýzy a ohodnocování kvality prováděných cviků existuje více přístupů, které se mohou lišit v získávání vstupních dat, interní reprezentace cviků či způsobem vyhodnocování jejich provedení.

1.1 Klasifikace kvality cviků na posilovacích strojích

Novatckhov a Batcha [1] se zabývali vyhodnocováním kvality cviků na posilovacích strojích. Neanalyzovali přímo pohyb lidského těla, ale hodnoty naměřené na samotných strojích, které byly vybaveny otáčkoměry a siloměry. Díky tomu byli schopni popsat průběhy cviků pomocí rychlosti pohybu, zrychlení a působení síly. V průběhu provádění daných silových cviků byly tyto atributy pravidelně zaznamenávány a ukládány jako časové řady.

Naměřená data byla následně předzpracována pro další fáze (Obrázek 1.1). Nejprve byl použit vysokofrekvenční filtr, který v signálu ponechal pouze nízké frekvence. Tímto způsobem byl potlačen šum z měřicích senzorů a zároveň byla zachována relevantní informace o průběhu daných veličin v čase. Jelikož byly zaznamenávány celé série cviků o více opakováních, data byla dále rozdělena podle nich. Toho bylo dosaženo algoritmem pro automatickou segmentaci, jehož princip spočíval v hledání lokálních extrémů.



■ **Obrázek 1.1** Jednotlivé kroky od zisku dat až po modelování [1].

Takto získaná opakování byla dále rozdělena do čtyř fází (extenze, flexe a dvě fáze držení) a z každé z nich extrahovány příznaky jako například délka trvání, průměrná rychlost, průměrné zrychlení atd. Původní signály probíhající v čase, které mohly být různě dlouhé, byly takto transformovány do nového prostoru příznaků jako vektory o stejné velikosti.

Všechna opakování byla v úvodní fázi také natáčena a následně profesionálními sportovními trenéry byla posouzena jejich kvalita, která byla rozdělena do tří tříd: správně provedené opakování, nesprávná flexe a nesprávná extenze. Toto ohodnocení umožnilo následné použití metod supervizovaného strojového učení. Cílem bylo tedy natrénovat model, který na základě extrahovaných příznaků pro všechny fáze opakování určí ohodnocující třídu. Data byla následně rozdělena na trénovací, validační a testovací. Použit byl model umělé neuronové sítě (ANN) a metoda podpůrných vektorů (SVM). ANN vykazovaly poměrně vysokou úspěšnost, korelace mezi skutečnou a predikovanou třídou se pohybovala nad 80 %. Kvalita výsledků SVM byla téměř shodná.

Evidentní nevýhodou tohoto řešení je veliké omezení v počtu cviků, které se takto mohou zkoumat, neboť měření jsou prováděna na strojích. Popisuje ale důležité kroky, které jsou potřeba při zpracování jakkoliv naměřených údajů o pohybu, konkrétně odstranění šumu a segmentaci opakování.

1.2 Analýza postavení těla při cvičení s využitím modelu lidského skeletonu

Yang a kolektiv [2] se zaměřili na klasické silové cviky bez použití posilovacích strojů, zkoumali tedy držení těla cvičence v průběhu cvičení.

Vstupními daty byly videozáznamy cvičenců provádějící dřepy a kliky. Video byla pořízena z jednoho konkrétního úhlu. Následně z nich bylo potřeba extrahovat informace o poloze lidského těla. K tomu využili model OpenPose, což je vícestupňová konvoluční neuronová síť (CNN), která z obrazového snímku, na kterém se nachází postava člověka, získá 2D souřadnice jednotlivých bodů těla v obraze. Tělo bylo tedy vnitřně reprezentováno takzvaným kinematickým modelem.

Následně byla provedena regularizace, která měla odstranit odlišnosti ve velikosti získaného modelu, která může záviset jak na tělesných proporcích cvičence, tak na jeho vzdálenosti od kamery. Tato korekce spočívala v přepočítání velikostí končetin tak, že jsou proporční ku velikosti trupu v pixelech. Ta byla spočítána jako průměrná vzdáleností krku od pravé a levé kyčle.

Dané dva typy cviků byly vyhodnocované pomocí primitivních, manuálně definovaných pravidel. Ideální průběh dřepu byl definován pomocí úhlu v kolenu a vzdálenosti chodidel od kyčlí v dolní fázi. Pro dané opakování byl tedy spočítán úhel mezi vektorem určeným body reprezentující pozici kyčle a kolene a vektorem určeným pozicemi kolene a chodidla. Pokud byl úhel přibližně 90 stupňů, byl pohyb kolene považován za správný. Podobně byla měřena vzdálenost chodidel od kyčlí, kdy poměr této vzdálenosti a velikostí stehna musel být menší než jedna a větší než parametr τ . Při splnění této nerovnosti byla pozice kyčlí považována za správnou. Podobně jednoduše bylo i definováno správné provedení kliku, kdy vertikální souřadnice boků by měla být přibližně mezi vertikální souřadnicí chodidel a ramenou (Obrázek 1.2).

Toto řešení má výhodu v tom, že není závislé na žádných cvičebních strojích, neboť se analyzuje skeleton člověka získávaný z videí. Také poukazuje na způsoby, jak se vypořádat s rozdíly ve velikosti cvičence a jeho umístění v obraze. Nevýhodou je však to, že zkoumaný skeleton je



■ **Obrázek 1.2** Špatně provedený klik s vizualizovaným detekovaným skeletem [2].

reprezentovaný ve 2D a tedy hloubkovou informací není schopen jakkoliv podchytit. Zároveň pravidla, která jsou používána pro vyhodnocování jsou velice primitivní a je potřeba je manuálně sestavovat.

1.3 Vyhodnocování kvality rehabilitačních cviků pomocí zařízení Microsoft Kinect

Parmar a Morris [3] se zabývali vyhodnocováním kvality rehabilitačních cviků, které nevyžadovaly žádné cvičební pomůcky či stroje a analyzovali samotné postavení těla.

Cvičenci byli zaznamenáváni při provádění pěti typů cviků, které byly charakterizovány velkou amplitudou pohybu, tedy velice zřetelným pohybem končetin. Videozáznamy byly pořizované na zařízení Microsoft Kinect, které je vybaveno jak standardní RGB kamerou, tak hloubkovou kamerou fungující na principu infračervené projekce. Zařízení také poskytuje algoritmus pro trackování lidského těla, kdy výstupem je 20 souřadnic jednotlivých částí těla, tedy skeletem v 3D prostoru. Pro každý cvik byla zaznamenávána správná provedení, tak i špatná, v průměru 11 opakování pro každou třídu. Výstupem tohoto kroku byly časové řady lidských skeletů.

Následovalo předzpracování získaných dat. Prvním krokem bylo převzorkování časových řad, aby měly pevnou délku, konkrétně 160 snímků. Jelikož je každý člověk jinak veliký a může být různě daleko od kamery, souřadnice bylo potřeba přeškálovat, aby byly tyto rozdílnosti odstraněny. Každá ze souřadnic (x, y, z) byla namapována z původního naměřeného rozsahu do rozsahu 1 až 3. Jelikož se umístění člověka v obraze mohlo také lišit (jak vertikálně, tak horizontálně), souřadnice bylo nutné vycentrovat. Toho bylo dosaženo tak, že od souřadnic všech bodů byl odečten bod mezi kyčlemi a nové souřadnice byly tedy relativní vůči tomuto středovému bodu. Následně byly spočítány další příznaky, konkrétně úhly mezi různými končetinami. Vedle časové

■ **Tabulka 1.1** Testovací přesnost vybraných modelů pro všechny reprezentace příznaků [3].

	Časová oblast		Frekvenční oblast	
	Souřadnicová data	Úhlová data	Souřadnicová data	Úhlová data
SVM	90,13	87,41	90,89	87,63
Jednovrstvá ANN	90,65	79,09	56,60	55,96
Vícevrstvá ANN	80,21	74,44	78,35	68,89
Posílený rozhodovací strom typu AdaBoost	94,68	90,30	83,14	80,40
DTW	73,64	74,03	-	-

oblasti, byly úhly také přepočítány do frekvenční oblasti pomocí diskrétní kosinové transformace (DCT).

Problém hodnocení kvality cviků byl poté modelován jako problém binární klasifikace, neboť vstupní data byla popsána jako dobře provedené cviky a špatně provedené cviky. Testovány byly modely ANN, SVM, posílené rozhodovací stromy (boosted trees) a porovnávání příznaků v časové doméně pomocí metody dynamic time warping (DTW). Výše zmíněné modely (kromě DTW metody) byly natrénované jak na dobře provedených cvicích, tak špatně provedených cvicích a uměly mezi nimi rozlišovat. Jediná DTW metoda, což je způsob, jak určovat vzdálenost mezi časovými řadami, byla realizována tak, že ze správných opakování se sestrojilo průměrné opakování. Pokud byla vzdálenost cviku, který chceme klasifikovat, větší, než maximální vzdálenost mezi všemi správnými cviky od sestrojeného průměru, byl klasifikován jako špatný, v opačném případě jako dobrý.

Nejlepší přesnost v časové oblasti vykazoval model posílených rozhodovacích stromů, konkrétně typ AdaBoost. Ten byl nejpřesnější jak pro souřadnicová data, tak pro úhlová data. Ve frekvenční oblasti naopak nejpřesněji klasifikoval model SVM, ale obecně byla na tom frekvenční oblast hůře než časová (Tabulka 1.1).

Výhodou tohoto přístupu je především 3D reprezentace skeletonu, jehož přesnost je navíc podpořena informacemi z hloubkové kamery. Kromě vyhodnocování pomocí metody DTW jsou však ostatní modely nedostatečné v tom ohledu, že umí rozlišovat jen mezi správně provedenými cviky a špatně provedenými cviky a není jisté, jak by predikovaly, pokud by dostaly na vstup chybu, na které nebyly natrénovány. Také jsou cviky vyhodnocovány jako celek a neřeší nijak konkrétní chybné partie.

1.4 Evaluace rehabilitačních cviků pomocí fuzzy pravidel

V [4] se Huang a kolektiv zabývali vyhodnocováním rehabilitačních cviků zaměřených na starší populaci. Data získávali ze dvou zdrojů. Prvním byla klasická kamera, která zaznamenávala pacienta při cvičení a ze získaných videí byl potom extrahován lidský skeleton pomocí modelu OpenPose. Navíc, aby byl vyřešen problém překrývání některých částí těla (která pak nebyla viděna přímo) byli pacienti vybaveni také nositelnými inerciálními měřicími jednotkami, které zaznamenávaly pohyb jednotlivých částí těla.

Data získaná z uvedených zdrojů byla dále reprezentována jako časové řady. Na naměřené souřadnice byl poté použit Kalmanův filtr pro odstranění šumu a následně byly spočítány úhly

v končetinách, které se v daném cviku zkoumaly, nejčastěji úhly v ramenou či kolenou, stejně tak v čase měnící se úhlové rychlosti těchto končetin.

Hodnoty byly poté porovnány s ukázkovými daty a počítány rozdíly v každém časovém okamžiku. Celé sekvence byly poté ještě rozděleny do menších oken a pro každé okno spočítáno skóre, které vyjadřovalo kvalitu prováděného cviku.

Skóre bylo počítáno pomocí fuzzy inferenčního mechanismu. Postup vyžadoval nejprve určit fuzzy množiny a jejich funkce příslušnosti. Pro popis rozdílu úhlu a úhlové rychlosti to byly množiny NULOVÝ, MALÝ a VELKÝ. Pro finální skóre byly definovány množiny NÍZKÉ, STŘEDNÍ, VYSOKÉ A PLNÉ SKÓRE. Poté byla manuálně zkonstruována fuzzy pravidla, která naměřeným rozdílům přiřadila výsledné skóre. Jedním z pravidel bylo například: Pokud je rozdíl úhlu STŘEDNÍ a rozdíl úhlové rychlosti NULOVÝ, pak je výsledné skóre PLNÉ. Nakonec se výsledné množiny těchto pravidel defuzzifikovaly a byla získána číselná skóre.

Výhodou přístupu je zahrnutí určité vágnosti při vyhodnocování kvality daného cviku. Nevýhodou může být ale nutnost dané fuzzy množiny a fuzzy pravidla vytvářet manuálně a při větším množství parametrů a různých cviků může být proces velice náročný.

Teoretické zázemí

Pro realizaci je potřeba seznámit se s některými algoritmy, které byly při jednotlivých krocích analýzy a vyhodnocení používány. Následuje velmi stručný popis těch nejdůležitějších z nich.

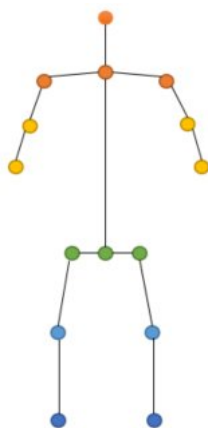
2.1 Reprezentace lidského těla

Dle [5] se pro reprezentaci lidského těla typicky používají následující tři typy modelů: kinematický, planární a volumetrický.

Kinematický model je z uvedených používán nejčastěji a je také nejjednodušší. Jedná se o množinu bodů, kde každý bod odpovídá určité části těla, například pravému zápěstí, levému koleni atd. Rozlišuje se 2D kinematický model, kde jsou body umístěné do roviny a popsány souřadnicemi x a y , a 3D kinematický model, kde se body nachází v trojrozměrném prostoru a popsány jsou pomocí souřadnic x , y a z . Model je schopen zachytit pouze vzájemné pozice částí těla z čehož lze určit například velikost končetin či úhly mezi nimi. Ačkoliv se mohou různé typy kinematických modelů lišit (zda obsahují například souřadnice očí, nosu, prstů atd.), typicky mají stejný základ pro reprezentaci trupu a horních a dolních končetin (Obrázek 2.1).

Planární model je schopen mimo vzájemné pozice částí těla zachytit i jejich plochu a aproximovat tak konturu člověka. Tělo je reprezentováno obdélníky (nejčastěji v rovině), kde každý z nich odpovídá určité končetině.

Nejsofistikovanější je volumetrický model, který se snaží o co nejpřesnější 3D aproximaci lidského těla. Umožňuje zachytit přibližnou tělesnou podobu člověka a především dynamiku lidské měkké tkáně. Nejčastěji je reprezentován množinou bodů v prostoru, které jsou vzájemně propojené do polygonů. Čím hustší je množina bodů, tím detailněji je lidské tělo zachyceno.



■ **Obrázek 2.1** Ukázka typické podoby modelu lidského těla při kinematické reprezentaci [5].

2.2 Metody umělé inteligence pro získání lidského skeletu

Pro získání lidského skeletu (v kinematické reprezentaci) z obrázků či videí se v dnešní době typicky používají metody hlubokého učení [5], neboť způsob dosahuje velice dobrých výsledků. Problém nalezení souřadnic kinematického modelu lze rozdělit na predikci 2D souřadnic nebo na predikci 3D souřadnic. Dále lze problém dělit na problém detekce a reprezentace jednoho lidského těla nebo detekci a reprezentaci více lidských těl v jednom snímku. V druhém zmíněném případě existují dva hlavní přístupy, které se liší v tom, zda se nejprve detekují osoby a poté se odhadují souřadnice nebo naopak. Takzvaný *top-down* přístup nejprve použije samostatný model pro detekci jednotlivých lidských těl v obraze a následně hledá souřadnice bodů těla v každé nalezené části obrazu samostatně. Druhý, takzvaný *bottom-up* přístup, nejprve odhadne souřadnice všech jednotlivých částí těla v obraze a následně je seskupuje do konkrétních skeletů.

Konkrétních architektur a modelů existuje velké množství, nejčastěji se ale využívají konvoluční neuronové sítě a jejich varianty, rekurentní neuronové sítě (především pro detekci skeletu ve videích) či takzvané transformery.

2.2.1 BlazePose

Konkrétním příkladem je model BlazePose od společnosti Google [6]. Jedná se o konvoluční neuronovou síť. Při inferenci se provádí dva kroky. Nejprve se využije model pro detekci lidského těla, čímž se vybere pouze určitá část vstupního snímku a následně se v něm hledají souřadnice jednotlivých částí těla.

Při trénování bylo využito přes 80 tisíc souřadnicemi anotovaných snímků lidských postav. Samotný proces trénování kombinoval dva přístupy. Jednak byl problém modelován jako regrese výstupních souřadnic, zároveň také jako takzvaný *heatmap* přístup. To znamená, že pro každou část těla je predikována pravděpodobnost jeho výskytu v jednotlivých pixelech v obrázku a při trénování je minimalizována střední kvadratická odchylka predikovaných distribucí a skutečných distribucí. Ty skutečné jsou vytvořeny tak, že v místě skutečné souřadnice je umístěno vícerozměrné Gaussovo rozdělení.

Základní podoba modelu počítá s 2D souřadnicemi, ale model byl snadno rozšířen na hledání 3D souřadnic [7]. Trénovací data byla dále anotována hloubkovou informací za pomocí modelu GHUM a výsledný model byl natrénován stejným způsobem, jako jeho základní verze.

2.2.2 MotionBERT

Model funguje na principu transformera a na rozdíl od prvního zmíněného modelu, není natrénován na získávání souřadnic přímo ze snímků lidských postav [8]. Hlavní procesem při tvorbě tohoto modelu bylo jeho předtrénování tak, aby z 2D lidských skeletů rekonstruoval sekvence 3D skeletů. Trénovací 3D sekvence byly na vstupu projektovány do 2D a po přidání šumu či odstranění některých částí skeletu měla síť za úkol zpět rekonstruovat odpovídající 3D sekvence. Model obsahoval dvě paralelní větve. První z nich se soustředila na prostorovou informaci o skeletu, neboť byla penalizována pouze odchylka predikovaných souřadnic. Druhá větev se zaměřovala na časově informaci a penalizovány byly odchylky rychlostí mezi po sobě jdoucími snímky. Výsledky obou větví byly poté spojeny na výstupu.

Pro použití modelu je tedy nejprve potřeba získat pomocí jiného způsobu 2D sekvence skeletů a ty ve specifickém formátu poskytnou jako vstup.

2.3 Dynamic Time Warping

Jedná se o algoritmus, který pro dvě časové řady spočítá „vzdálenost“ $d \geq 0$. Hodnota d určuje, jak moc jsou si časové řady podobné [9]. Čím blíže je hodnota nule, tím podobnější si řady jsou a naopak. Výhodou algoritmu je, že řady nemusí být stejně dlouhé a lze tedy odhalit podobnost procesů, které se odehrávají při různých rychlostech. Uplatnění lze nalézt v mnoha oblastech. Od rozpoznávání řeči, po vytěžování dat či shlukování časových řad.

Nechť X a Y jsou vstupní časové řady s délkami, které obsahují N a M měření zaznamenaných v čase. Prvním krokem je spočítat vzdálenostní matici $C \in \mathbb{R}^{N,M}$, pro kterou platí:

$$c_{i,j} = \|x_i - y_j\|, \quad i \in \{1 \dots N\}, \quad j \in \{1 \dots M\} \quad (2.1)$$

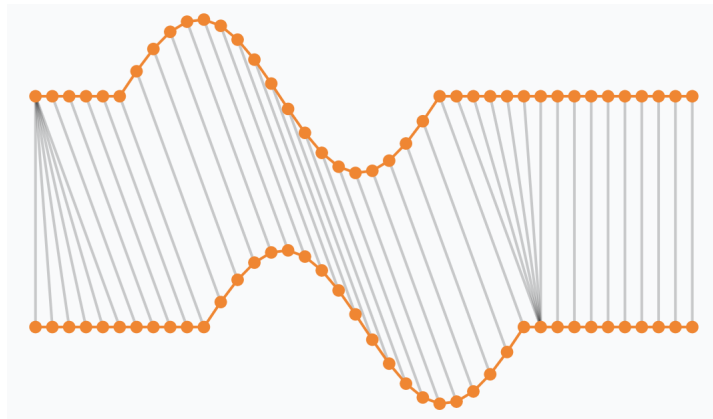
Jakmile je sestavena vzdálenostní matice, časové řady jsou poté „zarovnány“ tak, že je sestavena posloupnost $p = (p_1, p_2, \dots, p_k)$, kde $p_l = (p_i, p_j) \in \{1 \dots N\} \times \{1 \dots M\}$. Dále musí platit, že $p_1 = (1, 1)$ a $p_k = (N, M)$, indexy na první i na druhé pozici musí být v rámci posloupnosti neklesající a také, že $p_{l+1} - p_l \in \{(1, 1), (0, 1), (1, 0)\}$. Neformálně lze tedy říci, že každému bodu z jedné časové řady přiřadíme jeden nebo více bodů z druhé časové řady a naopak. Přiřazení si musí být počáteční a koncové body a musí být zachováno časové pořadí bodů. Posloupností, které splňují zmíněné vlastnosti, existuje více a cílem je nalézt takovou posloupnost, která minimalizuje součet vzdáleností bodů, které jsou si v posloupnosti přiřazeny. Minimalizován je tedy výraz:

$$\sum_{(p_i, p_j) \in P} c_{i,j} \quad (2.2)$$

Součet je také nazýván cenou zarovnání. Pro zarovnání, které ho minimalizuje, je jeho hodnota výslednou „vzdáleností“ časových řad. Cenu optimálního zarovnání lze spočítat z takzvané akumulované vzdálenostní matice D , pro kterou platí:

1. $d_{1,j} = \sum_{k=1}^j c_{1,k}$, $j \in \{1 \dots M\}$
2. $d_{i,1} = \sum_{k=1}^i c_{k,1}$, $i \in \{1 \dots N\}$
3. Pro ostatní prvky platí: $d_{i,j} = \min\{d_{i-1,j-1}, d_{i-1,j}, d_{i,j-1}\} + c_{i,j}$, $i \in \{2 \dots N\}$, $j \in \{2 \dots M\}$

Matrice je sestavena tak, že nejprve se přímo určí první řádek a první sloupec ze vzdálenostní matice C dle výše uvedených bodů 1 a 2. Poté se pro zvyšující se i a j vypočítají ostatní prvky matice podle výše uvedeného bodu 3. Prvek $d_{N,M}$ je roven ceně optimálního zarovnání. Algoritmus pracuje s časovou složitostí $\mathcal{O}(NM)$.



■ **Obrázek 2.2** Ukázka zarovnání časových řad. Výsledná vzdálenost je určena jako součet rozdílů bodů, které jsou spojeny šedými spojnicemi [10].

2.4 Algoritmy pro vyhlazení časových řad

Naměřené hodnoty v časových řadách bývají často zkreslené šumem a typicky je žádoucí šum odstranit pro další práci se získanými daty. Popsány jsou příklady algoritmů, které časové řady vyhladí a šum tak potlačí. Obecně jsou mohou být hodnoty v časových řadách měřeny v různých časových okamžicích, ne nutně stejně vzájemně vzdálených. Pro jednoduchost je dále uvažována situace, kdy jsou všechna měření ve stejně vzdálených časových okamžicích a je jich konečný počet. Lze časovou řadu vyjádřit jako vektor o N složkách.

2.4.1 Klouzávy průměr

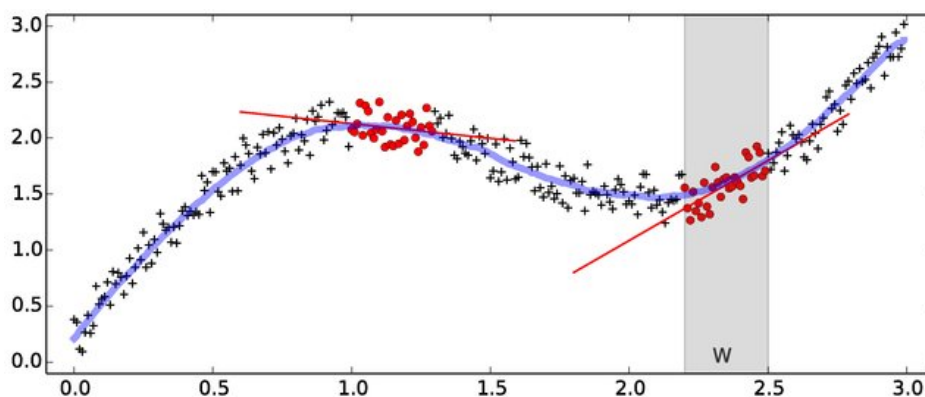
Jednou z nejjednodušších metod pro vyhlazení průběhu časové řady je klouzávy průměr [11]. Hodnota v daném bodě časové řady je vypočítána jako průměr okolních bodů. Počet sousedů, které se při výpočtu berou v úvahu je dán fixní velikostí takzvaného okna. Aby byl počet sousedů na obou stranách symetrický, velikost okna je typicky vypočítána jako $2k + 1$, kde k je zvolený „poloměr“ okna. Nechtě $X = (x_1, x_2, \dots, x_n)$ je časová řada s naměřenými hodnotami. Výpočet lze potom vyjádřit jako:

$$x_i = \frac{1}{2k+1} \sum_{j=-k}^k x_{i+k}, \quad i \in \{k+1, k+2, \dots, n-k\} \quad (2.3)$$

2.4.2 Lokální regrese

Lokální regrese, konkrétně metoda LOWESS je metoda pro vyhlazování časových řad, při kterém se pro každý bod opět určí nová hodnota bodu na základě nějakého jeho okolí [12]. Nechť $X = (x_1, x_2, \dots, x_n)$ je opět časová řada s naměřenými hodnotami. Prvním krokem je pro každý bod časové řady na indexu i určit nejbližších Q sousedů (nejbližších na ose x , tedy s ohledem na index měření), kde Q je dáno jako parametr. Poté je každému bodu na indexu j mezi těmito sousedy přiřazena váha, nejčastěji jako $w_j = (1 - |\frac{j-i}{d}|^3)^3$, kde d je rozdíl mezi i a od něj Q -tého nejbližšího indexu. Váhy jsou největší pro samotný bod na indexu i a s rostoucí vzdáleností od něj klesají.

Pro takto vážené body v daném okolí je poté provedena vážená regrese, nejčastěji lineární, metodou nejmenších čtverců. Nová hodnota bodu na indexu i je poté hodnota regresní křivky právě v tomto bodě. Vizualizace popsaného vyhlazení je vidět na obrázku 2.3.



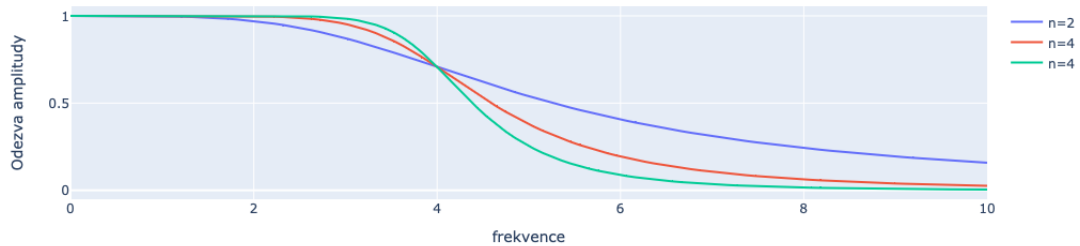
■ **Obrázek 2.3** Ukázka vyhlazení pomocí lokální regrese [13].

2.4.3 Butterworthův filtr

Algoritmus slouží k odfiltrování určitých frekvencí ze signálu, což lze použít i na časové řady. Filtr lze použít na potlačení nízkých frekvencí, vysokých frekvencí, či potlačení frekvencí mimo specifikovaný interval. Filtr lze charakterizovat takzvanou funkcí amplitudové odezvy. Ta definuje, jak moc bude amplituda při frekvenci potlačena. Pro vysokofrekvenční filtr (potlačuje vysoké frekvence) funkci lze vyjádřit ve tvaru:

$$f(\omega) = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_0}\right)^{2n}}}, \quad (2.4)$$

kde ω_0 je takzvaná mezní frekvence a určuje, od jaké hodnoty funkce začne značně klesat a tedy které frekvence budou více potlačeny. Hodnota n je takzvaný řád filtru a určuje, jak prudce funkce od mezní frekvence klesá. Příklad takové funkce je vidět na obrázku 2.4, čím nižší hodnota amplitudové odezvy, tím více bude daná frekvence potlačena. Hodnoty blízké 1 znamenají, že budou dané frekvence v signálu ponechány.



■ **Obrázek 2.4** Příklad funkce amplitudové odezvy s $w_0 = 4$ a různými hodnotami n .

Aplikace filtru na samotný signál je složitější a není zde detailně popsána. Nejčastěji je v praxi aplikován přímo v časové doméně signálu, také protože filtr lze implementovat hardwarově. Pro diskrétní data lze ale intuitivně aplikaci demonstrovat i ve frekvenční doméně [14]. Po provedení Diskrétní Fourierovy transformace jsou vstupní data vyjádřena ve frekvenční doméně, tedy závislost amplitudy na frekvenci. Vynásobí-li se amplitudy odpovídajícími hodnotami amplitudové odezvy filtru, relativně nezměněné zůstanou pouze amplitudy frekvencí menší, než mezní frekvence, ostatní budou potlačeny. Při použití inverzní Fourierovy transformace je obdržen signál, kde frekvence, pro které je hodnota amplitudové odezvy filtru blízká nule, nemají vliv na výslednou hodnotu.

2.5 Rozhodovací stromy

Jedním z nejjednodušších modelů supervizovaného strojového učení jsou rozhodovací stromy [15]. Používají se jak pro klasifikaci, tak pro regresi. Princip fungování je popsán na problému binární klasifikace.

2.5.1 Inference

Nechť $X = (x_1, x_2, \dots, x_n)$ je vektor příznaků a cílem je mu přiřadit jednu ze dvou cílových tříd 0 a 1. Při inferenci se začíná v kořeni stromu a dále se postupuje směrem k listům podle pravidel dělení. Pravidlo dělení v daném vrcholu je charakterizováno příznakem, ke kterému se vztahuje, a prahovou hodnotou, se kterou se porovnává. Nechť x_i je příznak, podle kterého se v aktuálním vrcholu dělí, a t je prahová hodnota daného pravidla. Poté, pokud je $x_i \leq t$, postupuje se dále do levého podstromu, pokud je $x_i > t$, postupuje se dále do pravého podstromu. Takto se postupuje dokud se nedojde do listu. Listy jsou charakterizovány tím, že obsahují informaci o cílové třídě, případně pravděpodobnostním rozdělení jednotlivých tříd pro daný vektor příznaků.

2.5.2 Trénování

Samotná inferenze je přímočará, avšak nejprve je potřeba na základě trénovacích dat strom zkonstruovat. Nechť X je matice trénovacích dat, kde v řádcích jsou zapsány jednotlivé vektory příznaků. Prvek $x_{i,j}$ tedy odpovídá j -tému příznaku i -tého trénovacího bodu. Strom je postupně konstruován od kořene směrem k listům a každé dělicí pravidlo je vytvořeno hladovým způsobem tak, že se vybere takový příznak a taková prahová hodnota, která maximalizuje takzvaný informační zisk:

$$IG(D) = H(D) - E(D_A, D_B), \quad (2.5)$$

kde D je množina všech bodů trénovací množiny, které se v daném vrcholu mají rozdělit, D_A a D_B jsou množiny bodů po rozdělení podle testovaného dělicího kritéria, H je takzvaná kriteriální funkce měřící neuspořádanost množiny, a E je funkce, která počítá vážený průměr kriteriálních funkcí H pro množiny D_A a D_B s ohledem na jejich velikost. Kriteriální funkce je funkce, která určuje, jak moc jsou množiny trénovacích bodů neuspořádané s ohledem na vysvětlovanou třídu. Pokud jsou cílové třídy všechny stejné, dosahuje minimální hodnoty, naopak, pokud jsou obě třídy zastoupeny stejně, dosahuje svého maxima. Příkladem takové funkce je binární entropie $H(D) = -p_0 \cdot \log(p_0) - p_1 \cdot \log(p_1)$, kde p_0 a p_1 jsou relativní frekvence bodů množiny D , jejichž cílová třída je třída 0 či 1. Kriteriální funkcí měřící neuspořádanost může být i takzvaný Gini index $H(D) = 2p_0 \cdot p_1$.

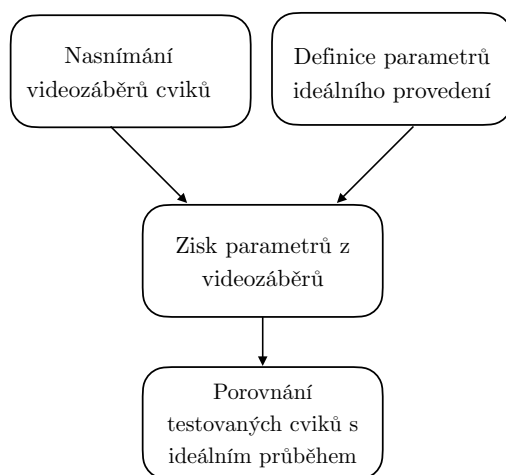
Takto se v každém kroku určí dělicí pravidlo, dokud se buď nedosáhne předem určené maximální hloubky, nebo v nějakém vrcholu zůstávají pouze prvky, které mají téměř všechny stejnou třídu (kriteriální funkce pro tuto množinu je dostatečně malá). Jakmile je tedy dosaženo kritérium pro zastavení dělení, aktuální vrchol se stane listem a o výsledné třídě při inferenci je rozhodnuto tak, že se zvolí majoritní třída bodů trénovací množiny, které pro vrchol po dělení zbyly. Případně může být vrácen odhad pravděpodobnosti jednotlivých tříd, určen relativním zastoupením dané třídy.

Přístup lze jednoduše rozšířit na problém vícetřídní klasifikace, neboť entropie i Gini index mohou být rozšířeny pro více tříd. Také pro regresi lze postup lehce modifikovat, a to tak, že jako kriteriální funkce je využít například rozptyl vysvětlované proměnné pro rozdělené množiny a cílová hodnota při inferenci je určena jako průměr vysvětlovaných hodnot trénovacích bodů v daném listu.

Nejprve bylo potřeba stanovit podstatu problému, navrhnout kamerovou soustavu pro získání dat a obecnou posloupnost kroků potřebných k vyhodnocování.

3.1 Návrh postupu

Prvním krokem bylo nasnímat obrazová data cviků. Dále bylo nutné definovat parametry správného průběhu cviku a najít způsob, jak je z obrazových dat získat. Posledním krokem bylo testované cviky pomocí těchto parametrů porovnat s ideálním průběhem a kvalitu provedení vyhodnotit (Obrázek 3.1).



■ **Obrázek 3.1** Diagram popisující hlavní kroky při vyhodnocování.

3.2 Snímací soustava pro zisk videozáznamů

Z podstaty snímaných cviků bylo důležité, aby cvičenec měl okolo sebe dostatek prostoru a aby při samotném pohybu bylo stále v obraze celé jeho tělo. S ohledem na co nejlepší kvalitu získaných záznamů také bylo vhodné, aby byly zřetelně viditelné kontury člověka a v záběrech nebylo žádné rušivé pozadí.

Kvůli uvedeným požadavkům probíhalo natáčení cviků v prostorné místnosti s bílým pozadím. Osvětlení bylo zajištěno z části denním světlem a z části interiérovými zářivkami. Záznamy byly pořizovány kamerou OAK-D-Lite¹ v rozlišení 3840 × 2160 pixelů a se snímkovací frekvencí 30 fps. Kamera byla umístěna na stativu přibližně 2,5 metru od cvičence ve výšce, která odpovídala úrovni pupíku cvičence. Prostor pro vykonávání cviků byl také ohraničen pomocnými kužely, aby cvičenec věděl, v jakém prostoru se může pohybovat, aniž by se pohnul mimo záběr. Výjimkou v popsaném uspořádání snímané scény byl jeden druh cviku, který byl prováděn na ribstolech a cvičenec se tedy nenacházel před čistě bílým pozadím. Kamerová soustava a natáčecí prostor je vidět na obrázku 3.2.



■ Obrázek 3.2 Snímací soustava k záznamu cviků.

3.2.1 Nasnímání vybraných cviků

Pro analýzu bylo vybráno více druhů cviků (Tabulka 3.1). Vybrány byly cviky, které se provádějí jak ve stoje, tak na zemi či na ribstolech. Mohlo tak být porovnáno, které cviky jsou pro analýzu

¹OAK-D-Lite - <https://docs.luxonis.com/projects/hardware/en/latest/pages/DM9095.html>



■ **Obrázek 3.3** Ukázka pořízených snímků cviků. Na prvním obrázku je prováděn rumunský mrtvý tah na jedné noze a na druhém klik.

■ **Tabulka 3.1** Nasnímané druhy cviků.

Český název	Anglický název (příp. jeho zkratka)
klik	pushup
dřep	squat
přítah na hrazdě	pullup
rumunský mrtvý tah na jedné noze	single leg romanian deadlift (slrdl)
sklapovačka	situp
výpad	lunge
pánevní most	glutebridge

a ohodnocení vhodnější a které méně. Pro zaznamenání cviků byl pozván profesionální sportovní trenér, aby byla zaručena ideální technika.

3.3 Ideální průběh cviků

Při cvičení by se vždy měla dodržovat správná technika, ať už kvůli efektivitě, nebo kvůli zranění, které by mohlo špatné provádění způsobit. Každý z vybraných cviků lze popsat klíčovými parametry, nejčastěji rozsahem pohybu nějakého kloubu (tedy jeho úhlem) nebo určitou vzdáleností konkrétních končetin, která má být dodržena. Pro příklad jsou představeny správné techniky vybraných dvou cviků a průběhy některých jejich parametrů.

3.3.1 Ideální technika dřepu

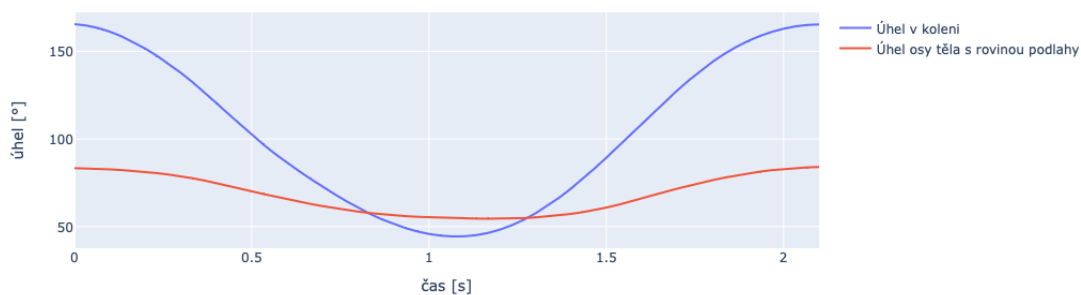
Jednotlivé fáze dřepu jsou znázorněny na obrázku 3.4. Dle experta jsou důležité tyto parametry: rozsah pohybu kolena (kolena pokrčena do minimálního možného úhlu), kolena v úrovni boků po celou dobu, záda rovná, s minimálním předklonem.

Z detailněji znázorněného průběhu úhlu v kolenu při jednom opakování dřepu (Obrázek 3.5) je vidět, že se začíná s propnutými koleny, postupně se jde dolů, minimální dosažený úhel je přibližně



■ **Obrázek 3.4** Ukázka jednotlivých fází správně provedeného dřepu.

44° a postupně se vrací do původní polohy. Úhel, který svírá osa trupu s rovinou podlahy, se mění pouze minimálně. Na začátku opakování se pohybuje kolem 85°, v dolní fázi dřepu dosáhne přibližně 55° a poté se opět vrací. Opakování trvá něco málo přes 2 sekundy.



■ **Obrázek 3.5** Průběh úhlu v koleni (modře) a průběh úhlu mezi osou trupu a rovinou podlahy (červeně) při ideálním provedení dřepu.

3.3.2 Ideální technika kliku

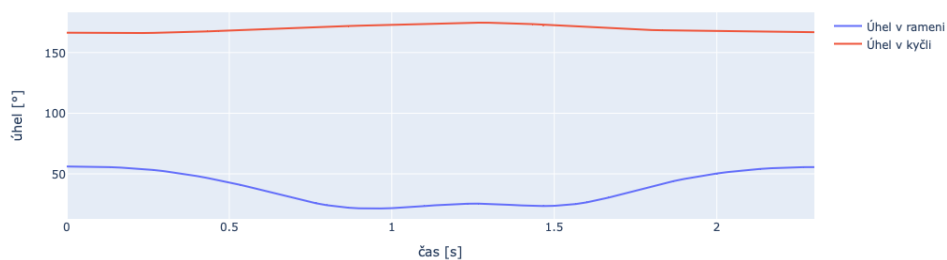
Podobným způsobem je popsána i správná technika kliku. Jednotlivé fáze jsou znázorněny na obrázku 3.6. Klik začíná v pozici s nataženými rukama a rovnými zády. Postupně se lokty

a ramena skládají podél těla a hrudník jde až těsně k zemi. Následně se opět lokty a ramena natahují a tělo jde do počáteční pozice. Důležité je mít ruce blíže k tělu, hrudníkem jít co nejnižše a záda mít s nohama v rovině.



■ **Obrázek 3.6** Ukázka jednotlivých fází správně provedeného kliku.

Při detailním znázornění průběhu úhlu v rameni (Obrázek 3.7) je vidět, že v úvodní pozici jsou ruce natažené, rameno svírá s bokem trupu úhel přibližně 55° , poté se v dolní fázi dostane až k úhlu mezi 20 a 25 stupni. Úhel v kyčli se naopak mění minimálně, pouze v rozsahu jednotek stupňů. V rámci celého opakování se drží okolo 170° . Opakování trvá přibližně 2,5 sekundy.



■ **Obrázek 3.7** Průběh úhlu v rameni (modře) a průběh úhlu v kyčli (červeně) při ideálním provedení kliku.

3.3.3 Ostatní cviky a jejich důležité parametry

Podobně jako výše detailněji popsané dva typy cviků byly definované i správné techniky všech ostatních. Mimo provedená opakování, která definovala ideální techniku, trenér detailněji popsal,

kteře parametry je důležité sledovat k vyhodnocení správné techniky daného cviku. Jak již bylo řečeno, u dřepu je důležité jít v kolenou co nejnižše, dodržovat rovná záda. U přítahu na hrazdě je důležitý pohyb rukou (lokty, ramena) a sklon trupu. Kompletní přehled důležitých parametrů pro jednotlivé cviky je v tabulce 3.2.

■ **Tabulka 3.2** Důležité parametry cviků dle názoru trenéra.

Cvik	Kritické parametry
klik	úhel v loktech, úhel v kyčlích, vzdálenost rukou, úhel v ramenou
dřep	úhel v kolenou, úhel v kyčlích, vzdálenost kolenou, úhel osy trupu se zemí
přítah na hrazdě	úhel v loktech, úhel v ramenou, úhel osy trupu se zemí
rumunský mrtvý tah na jedné noze	úhel spojnice kyčlí se zemí, úhel v kyčlích
sklapovačka	úhel v kyčlích, úhel v kolenou
výpad	úhel osy trupu se zemí, úhel v kyčlích vzdálenost chodidla od protějšího kolene
pánevní most	vzdálenost chodidel, vzdálenost pánve od země

Kapitola 4

Realizace

Pro samotné provedení analýzy a ohodnocení cviků bylo potřeba navrhnout, implementovat a porovnat různé algoritmy. Také bylo potřeba získat dodatečná data pro otestování ohodnocování.

4.1 Data

Nasnímáno bylo 7 různých druhů cviku. Ty byly nasnímané z více úhlů vůči kameře, aby mohlo být porovnáno, který úhel je pro analýzu a porovnání daného cviku vhodnější 4.1. Cviky byly natáčené zepředu, ze strany a ze šikma (natočení cvičence čelem přibližně 45 ° vůči ose snímání). Jedinou výjimkou byl přítah na hrazdě, kdy místo pozice ze předu byl cvik natáčen zezadu. V každé orientaci bylo natočeno 5 až 10 opakování, záleželo na obtížnosti daného cviku, neboť při náročných cvicích je těžké udržet správnou techniku se zvyšujícím se počtem opakování.

Mimo správné provádění cviků byly natočené i typické chyby, které se při cvičení často objevují 4.2. Pro všechny cviky to byly alespoň dva druhy chyb, pro přítah na hrazdě pouze jeden. Výčet chyb a jejich stručný popis se nachází v tabulce 4.1. Chyby byly taktéž natáčené ve více orientacích.

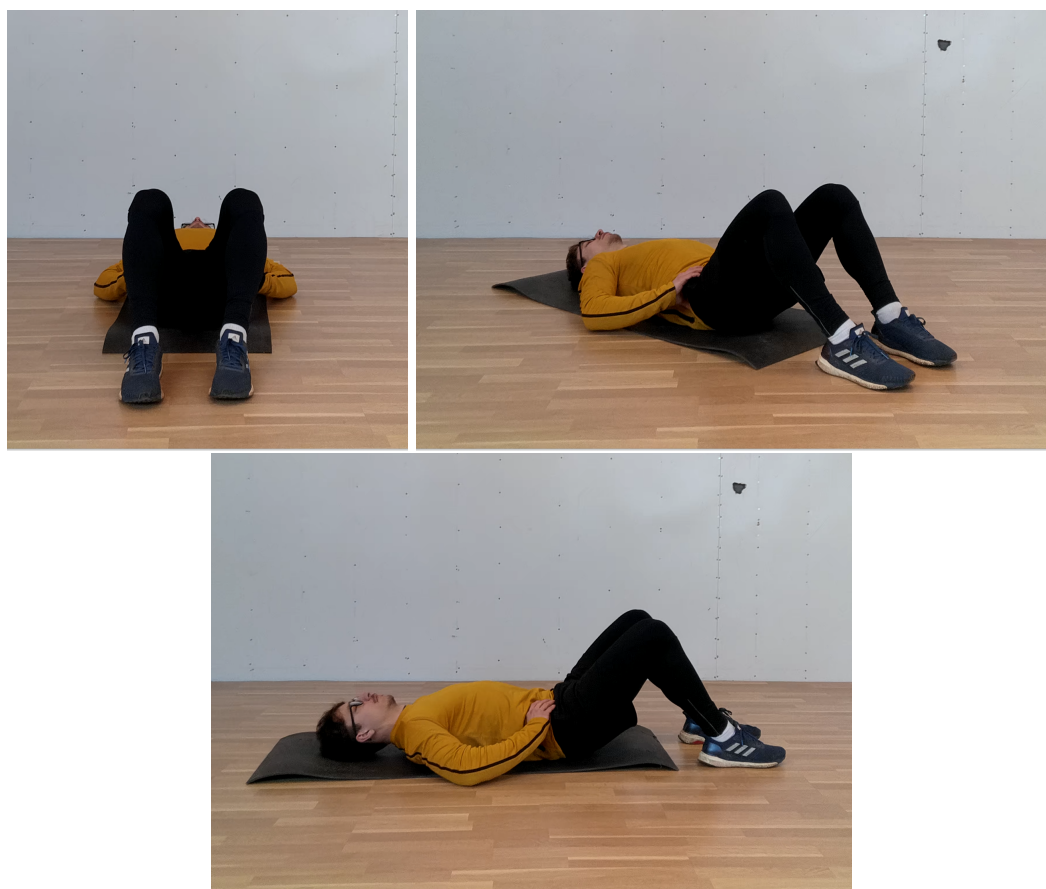
4.2 Implementace

Po získání vstupních dat bylo potřeba provést několik kroků, od zisku lidského skeletonu z videí, vypočítání parametrů k popisu průběhu cviků, segmentace, vyhodnocení až po výslednou vizualizaci. Trochu jiný postup byl vyžadován pro přípravu trénovacích dat (Obrázek 4.3) a pro ohodnocení testovaných dat (Obrázek 4.4). Všechny dílčí kroky byly vypracované v programovacím jazyce Python s pomocí jeho knihoven, především `numpy`¹, `pandas`² a `scipy`³.

¹NumPy - <https://numpy.org>

²Pandas - <https://pandas.pydata.org>

³SciPy - <https://scipy.org>



■ **Obrázek 4.1** Ukázka různých orientací natáčení při provádění pánevního mostu.



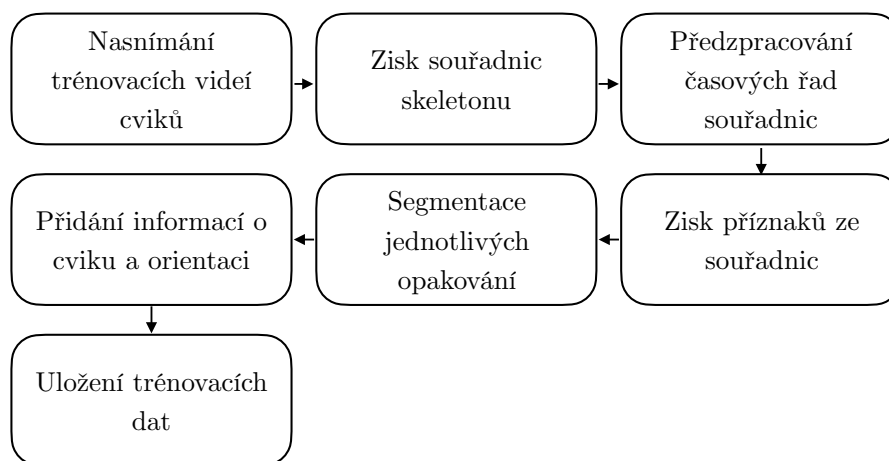
■ **Obrázek 4.2** Ukázka špatně provedených cviků pro testování. Vlevo je vidět chyba při provádění rumunského mrtvého tahu na jedné noze, kdy zadní noha není dostatečně zvednutá. Vpravo je špatně provedený klik, kdy není dodržen rozsah a ruce jsou příliš daleko od sebe.

4.3 Zisk lidského skeletonu z videí

Pro modelování pohybu byl zvolen kinematický model lidského těla. Byla zvolena 3D varianta, neboť ta poskytuje více informací než pouhá 2D reprezentace, především proto, že při použití

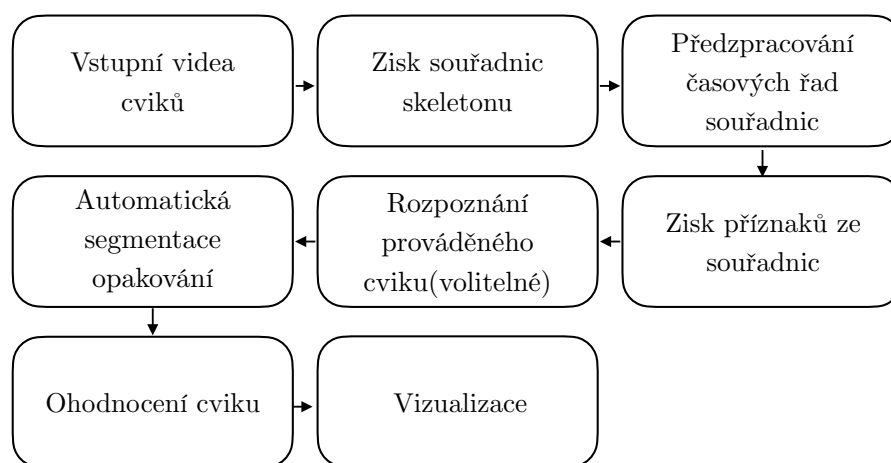
■ **Tabulka 4.1** Výčet nasnímaných chyb při cvičení a jejich popis.

související cvik	zkratka chyby použítá při implementaci	popis
dřep	squat_a	malý rozsah v kolenou, velký předklon
dřep	squat_b	jako v případě squat_a, kolena jdou příliš k sobě
výpad	lunge_a	velký předklon
výpad	lunge_b	koleno jde příliš blízko k chodidlu
rumunský mrtvý tah na jedné noze	srlrdl_a	velký předklon
rumunský mrtvý tah na jedné noze	srlrdl_b	vytáčení boků do strany
klik	pushup_a	prohýbání v bedrech
klik	pushup_b	malý rozsah pohybu ramen, dlaně příliš daleko od sebe
přítah na hrazdě	pullup_a	malý rozsah rukou, odklon od hrazdy
sklapovačka	situp_a	zvedání chodidel v důsledku pohybu kolen
sklapovačka	situp_b	ohnutá záda
pánevní most	glutebridge_a	příliš zvednutá pánev
pánevní most	glutebridge_b	málo zvednutá pánev, chodila příliš daleko od sebe



■ **Obrázek 4.3** Popis kroků při zpracování trénovacích dat.

2D modelu se ztrácí informace. I když by 2D souřadnice byly normalizovány a vycentrovány, rozdíl v rotaci cvičence vůči kameře nelze jednoduše odstranit. Jelikož bylo cílem analyzovat celý průběh cviku, ukládány byly souřadnice skeletonu v každém snímku a získané hodnoty byly dále reprezentovány jako časové řady. Potřeba byla tedy využít model, který ze snímku získá



■ **Obrázek 4.4** Popis kroků při zpracování a analýze dat k vyhodnocení.

souřadnice lidského skeletonu a aplikovat ho na vstupní videa. K tomuto kroku bylo využito dvou modelů: BlazePose a MotionBERT.

Model BlazePose je implementován v balíčku `mediapipe`⁴. Pomocí něj bylo možné video převést na časovou řadu souřadnic jak 2D (relativně vůči pixelům videa), tak 3D v souřadnicích, jejichž střed byl mezi kyčlemi nalezeného těla a měřítko odpovídalo odhadnutým rozměrům v metrech. Pro získání souřadnic z modelu MotionBERT byl potřeba jeden krok navíc. Model neočekává na vstupu přímo video či obrázek, ale sekvenci 2D souřadnic již extrahovaných ze snímků jiným modelem, navíc ve speciálním formátu. 2D souřadnice byly získané modelem AlphaPose [16], jehož výstup byl dle dokumentace⁵ s MotionBERT kompatibilní. Získané 2D sekvence byly poté dány na vstup MotionBERT modelu ke konstrukci sekvencí 3D souřadnic. Oba kroky vyžadovaly ke spuštění modelů použití knihovny `pytorch`⁶.

V této fázi bylo cílem dále uvedené modely srovnat ve dvou kritériích: v jednoduchosti použití a času inference.

Co se týká prvního kritéria, jednodušší instalaci a rozhraní pro spuštění měl model BlazePose. Naopak MotionBERT a potažmo i AlphaPose vyžadovaly knihovny, které zabírají více místa (jednotky GB) a instalace na lokálním stroji byla problematická, úspěšně byly spuštěny až v cloudovém prostředí Google Colab⁷.

V času inference byl jednoznačně lepší model BlazePose. Doba inference nepřekročila samotnou dobu trvání videa. Naopak inference MotionBERT v kombinaci s AlphaPose trvala velice dlouho, celková doba inference všech nasnímaných videí trvala jednotky hodin, což značně přesahovalo celkovou dobu trvání všech videí dohromady.

⁴MediaPipe - <https://mediapipe.dev>

⁵MotionBERT, dokumentace - <https://github.com/Walter0807/MotionBERT/blob/main/docs/inference.md>

⁶PyTorch - <https://pytorch.org>

⁷Google Colab - <https://colab.research.google.com>



■ **Obrázek 4.5** Ukázka získaných souřadnic pomocí modelu BlazePose při provádění bulharského mrtvého tahu na jedné noze. Vizualizováno pomocí knihovny `mediapipe`, ve které je implementován i samotný BlazePose. Oranžově jsou vyznačeny části těla náležící levé straně, modře části náležící levé části.

4.4 Předzpracování časových řad souřadnic

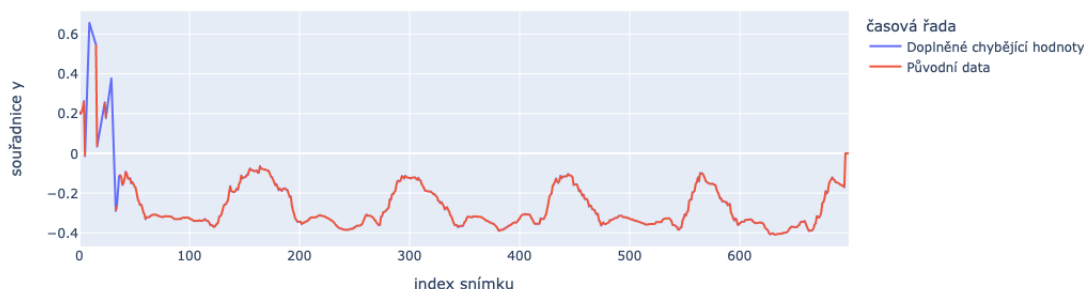
Časové řady bylo potřeba ještě před dalšími kroky předzpracovat. Jelikož byla všechna vstupní videa natáčena se stejnou snímkovací frekvencí, mohly být časové řady místo časové jednotky indexovány číslem snímku.

Nejprve byly doplněny chybějící hodnoty. V datech získaných ze zmíněných modelů se místy objevovaly chybějící hodnoty. To bylo zapříčiněno především tím, že v daném snímku si nebyl model dostatečně jistý, že detekoval danou část lidského těla. Tyto hodnoty byly nahrazeny metodou lineární interpolace, implementovanou v balíčku `pyts`⁸. Metoda byla použita v [17], kde se taktéž jednalo o pohybová data, konkrétně souřadnice částí těla při provádění znakové řeči, což jsou data stejného typu jako zde. Použitím této metody dosáhneme toho, že hodnoty budou na sebe plynule navazovat. Pokud se chybějící hodnoty nacházely na začátku či na konci časové řady, byly nahrazeny nejbližší nechybějící hodnotou.

Dále byly detekovány a nahrazeny odlehlé hodnoty. V některých případech se stalo, že se hodnota souřadnice některé části těla v jednu chvíli nacházela daleko od sousedních hodnot a to až o několik řádů. Tyto hodnoty se často také nacházely na konci časové řady v případě, kdy koncové snímky videa nebyly korektně zpracovány a všechny souřadnice byly nulové. Takové hodnoty jsou chybné a bylo potřeba je odstranit. Toho bylo dosaženo metodou Hampelova filtru, implementovaného v knihovně `hampel`⁹. Tento algoritmus označí za odlehlou hodnotu taková

⁸pyts - <https://pyts.readthedocs.io/en/stable/index.html>

⁹hampel - <https://pypi.org/project/hampel/>

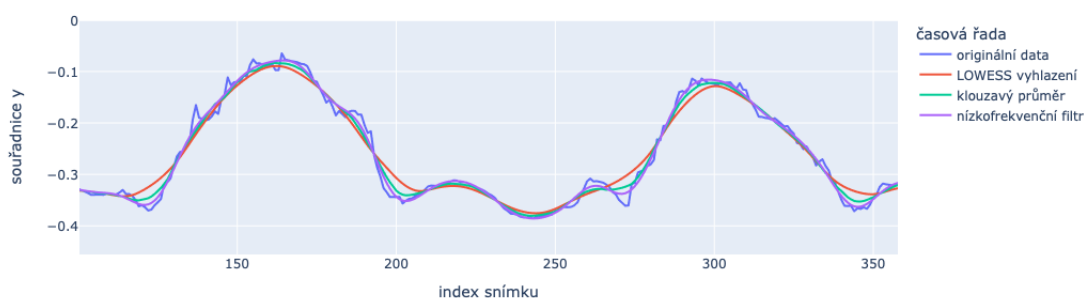


■ **Obrázek 4.6** Ukázka doplnění chybějících hodnot časové řady souřadnice Y levého lokte při provádění sklapovačky.

měření, jejichž vzdálenost od průměru v rámci pevně zvoleného okolí je větší než trojnásobek směrodatné odchylky. Okolí je určeno oknem, které je postupně posouváno přes časovou řadu. Velikost okna byla experimentálně zvolena jako 21 snímků. Takto detekované hodnoty jsou pak nahrazeny mediánem aktuálního okna.

Dalším důležitým krokem bylo získané časové řady vyhladit a odstranit tak vliv šumu, který by mohl dále překážet ve vyhodnocování. Otestovány byly tři různé přístupy: prostý klouzavý průměr, lokální regrese (LOWESS) a nízkofrekvenční filtr, konkrétně Butterworthův filtr. Při pohledu na ukázkou takových vyhlazení (Obrázek 4.7), je jako první vidět, že LOWESS algoritmus měl tendenci zplošťovat úzké lokální extrémy a pravděpodobně tak neodpovídal reálnému průběhu. Butterworthův filtr a metoda klouzavých průměrů dávaly výsledky se zanedbatelným rozdílem (pro jisté parametry de facto shodné) a pro jednoduchost byl tedy zvolen klouzavý průměr s šířkou okna 15 snímků jakožto metoda pro vyhlazení časových řad.

Výše zmíněnými kroky prošly jak záznamy cviků správných, tedy sloužících jako trénovací data, tak cviky k vyhodnocení.



■ **Obrázek 4.7** Ukázka různých metod vyhlazení časových řad demonstrováných na souřadnici Y levého lokte při provádění sklapovaček.

4.5 Získ parametrů ze souřadnic pro popis cviků

Získané souřadnice byly dále převedeny na potřebné parametry pro analýzu cviku, provedeno bylo také několik kroků jejich předzpracování.

4.5.1 Volba parametrů pro reprezentaci

Dále byly časové řady souřadnic převedeny na časové řady příznaků popisující úhly mezi končetinami, vzájemné relativní vzdálenosti určitých částí těla nebo úhly mezi rovinou podlahy a částí těla. Tato reprezentace odstranila problém neporovnatelnosti původních časových řad, který mohl být způsobem rozdílnou velikostí těl cvičenců, jejich vzdáleností od kamery, různá rotace v prostoru či umístění v obraze. Úhel mezi končetinami je nezávislý na všech těchto odlišnostech to samé platí pro vzdálenost, která je relativní vůči velikosti jiné (pevně zvolené) části těla. Každý model navíc poskytuje trochu jiný formát skeletonu (například BlazePose obsahoval i souřadnice očí, které byly pro další zkoumání zbytečné) a tato transformace tyto rozdíly odstranila. Dalším důvodem pro tuto reprezentaci je využití definovaných parametrů k vyhodnocování kvality daného cviku. Například ke správném provedení dřepu je potřeba malý úhel v kolenou (dostatečný rozsah pohybu). Tuto informaci snáze svázat s 1 parametrem reprezentující úhel v koleni, spíše než se souřadnicemi chodidla, kolene a kyčle (tedy dohromady s 9 příznaky). Pro reprezentaci byly zvoleny tyto parametry: úhel v kolenou, úhel v kyčlích, úhel v ramenou, úhel v loktech, relativní vzdálenost kolenou, chodidel, dlaní, pánve od podlahy, úhel mezi osou těla a rovinou podlahy a úhel mezi přímkou tvořenou kyčlemi a rovinou podlahy. Všechny úhly párových končetin byly ukládány pro každou stranu zvlášť, vzdálenosti byly vypočítávány relativně vůči průměrné vzdálenosti kyčlí a osa těla byla určena body, které ležely ve středech spojnic ramenou a kyčlí. Všechny vypočítané parametry jsou vidět na obrázku 4.8, kde jsou vizualizovány přímo na skeletonu.

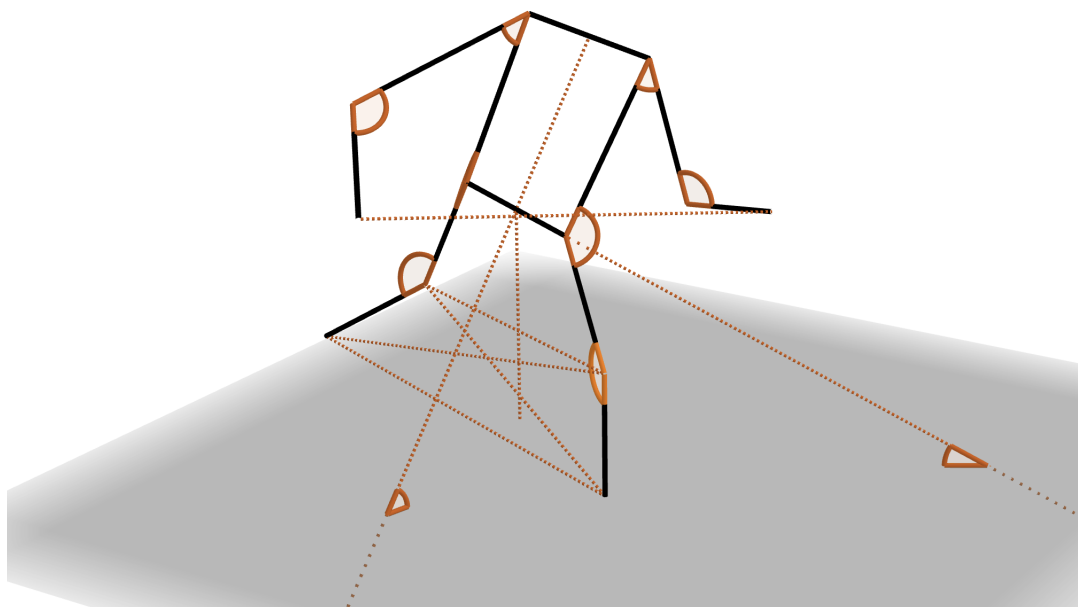
4.5.2 Vyhlazení a porovnání získaných parametrů z modelů pro získání skeletonů

Spočítané časové řady příznaků byly ještě jednou vyhlazeny pomocí metody klouzavého průměru s velikostí okna 7 snímků pro dodatečné potlačení šumu.

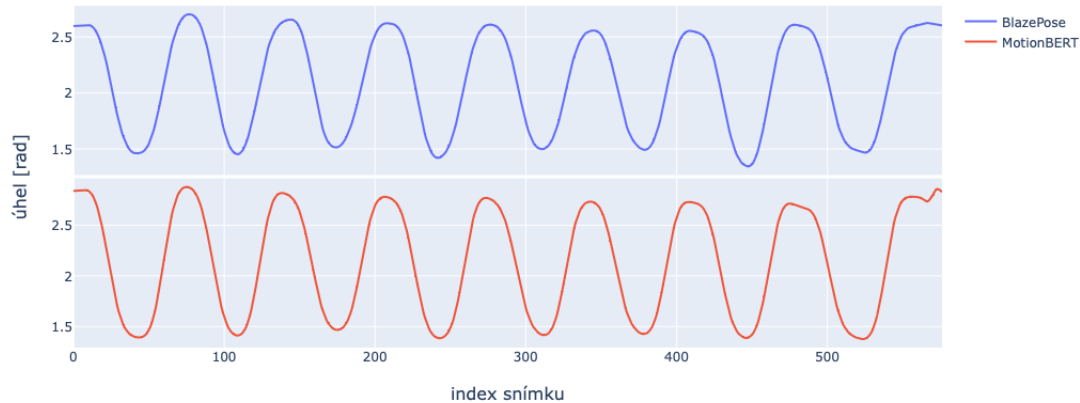
Poté byla provedena explorační analýza získaných příznaků v rámci jednotlivých modelů, které byly použity. Na obrázku 4.9 je vidět, že jak BlazePose, tak MotionBERT, podaly celkem obstojné výsledky. Časové řady úhlu v koleni při provádění výpadů vypadaly velice podobně a z průběhu byla jasně vidět jednotlivá opakování.

4.5.3 Segmentace trénovacích cviků

Jelikož bylo cílem ohodnocovat jednotlivá opakování samostatně, bylo potřeba získané časové řady segmentovat na dílčí opakování, aby mohla být samostatně porovnávána. Pro co nejpřesnější oddělení jednotlivých opakování byla zvolena metoda manuální segmentace.

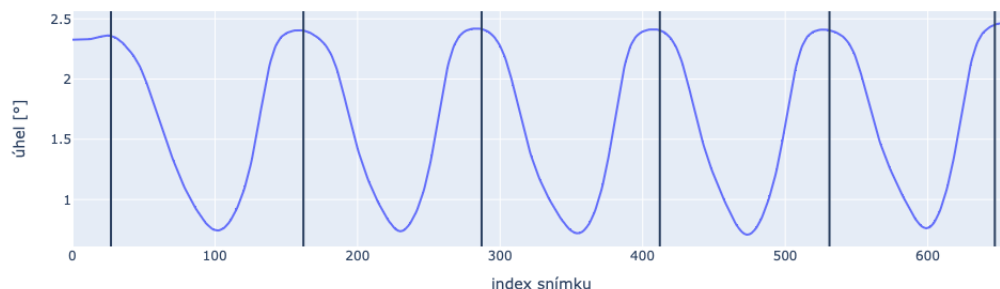


■ **Obrázek 4.8** Vizualizace vypočítaných a zkoumaných parametrů. Zvýrazněny jsou oranžově úhly a relativní vzdálenosti.



■ **Obrázek 4.9** Porovnání získaných příznaků z různých modelů pro získání skeletu. Demonstrováno na úhlu v levém kolenu při provádění výpadů natáčených ze šikma.

Pro každý záznam byl vykreslen graf získaných parametrů, konkrétně takových, pro které je rozsah hodnot v průběhu daného cviku velký a ze kterého je patrné, kde dané opakování začíná a kde končí. Například pro dřep byl pozorovaný příznak úhel v kolenu, pro přitah na hrazdě úhel v lokti atd. Z grafů byly odhadnuty indexy snímků, které odpovídaly hranicím opakování a podle nich byla sekvence segmentována (Obrázek 4.10).



■ **Obrázek 4.10** Ukázka manuální segmentace demonstrována na průběhu úhlu v levém boku při provádění sklapovačky. Svislé černé čáry ohraničují jednotlivá opakování.

Jednotlivá opakování byla poté uložena spolu s informací, o jaký cvik se jedná a z jakého směru byl záznam pořízen. U unilaterálních cviků (těch, které se provádí zvlášť na pravou a zvlášť na levou stranu) byla dále navíc uvedena informace, na kterou stranu bylo opakování provedeno. Příkladem takového cviku je třeba rumunský mrtvý tah na jedné noze, kdy jednou cvičenec zvedá za sebe levou a jednou pravou nohu.

4.5.4 Klasifikace cviku v neznámé sekvenci

Využitá metoda pro automatické segmentování již známých cviků, která je popsána v další části, vyžaduje informaci o tom, o jaký typ cviku se jedná. Ačkoliv může být tato informace uvedena explicitně, byl otestován i přístup, kdy se na základě získaných popisných příznaků zjistí typ cviku nějakým klasifikačním modelem.

Pro klasifikaci byly otestovány rozhodovací stromy a jejich varianty, což jsou jedny z nejjednodušších klasifikačních modelů, které navíc v [3] dosahovaly dobrých výsledků při klasifikaci dobře a špatně provedených cviků. Také v [18], kde se klasifikovalo 9 druhů cviků z pohybových dat, dosahovaly varianty rozhodovacích stromů jedny z nejlepších výsledků. Testován byl jak jednoduchý rozhodovací strom, tak náhodný les.

Vstupními daty nebyly celé časové řady, ale hodnoty spočtených příznaků v každém snímku zvlášť, tedy vektory délky 16. Jako trénovací data byla využita segmentovaná opakování správně provedených cviků. Ta byla rozdělena na jednotlivé snímky a spolu s informací, o jaký cvik se jedná, se použila pro učení rozhodovacích stromů.

Jak pro obyčejný rozhodovací strom, tak pro náhodný les bylo provedeno ladění hyperparametrů metodou křížové validace. Testovanými hyperparametry byla kritériální funkce (Gini index nebo entropie) pro dělení stromu a jejich maximální hloubka (2 až 20). Trénování bylo provedeno zvlášť pro příznaky spočítané ze souřadnic BlazePose a MotionBERT. Je vidět (Tabulka 4.2), že v obou případech získaných souřadnic byl lepší náhodný les, ačkoliv přesnost křížové validace se lišila jen minimálně. Dále byla testovací přesnost v obou případech stoprocentní.

Nutno podotknout, že tímto způsobem byla testována klasifikace pouze na jednotlivých snímcích a na datech, kde byly cviky prováděny správně, přitom požadavek byl, aby byly rozpoznány celé sekvence a také především cviky, které nejsou nutně provedeny správně. Celá sekvence byla klasifikována tak, že byly predikovány cviky v jednotlivých snímcích a výsledek byl vzat jako

nejčastěji se vyskytující typ cviku přes všechny snímky. Postup byl ještě vylepšen tak, že pro sekvenci klasifikací byl použit klouzavý modus. Tím byly potlačeny případy, kdy se v souvislých částech predikoval většinou stejný cvik. Ojediněle, na jednom či dvou snímcích, se predikce lišila. Jelikož je nepravděpodobné, že by člověk v pár snímcích dělal jiný druh cviku, než dělal doposud a potom se zase vrátil k původnímu cviku, byly tyto ojedinělé hodnoty přepsány na hodnoty modu v daném okně. Klasifikace sekvencí byla otestována na chybně provedených cvicích.

■ **Tabulka 4.2** Validační a testovací výsledky rozhodovacích stromů a náhodných lesů pro klasifikaci cviků na jednotlivých snímcích.

Původ souřadnic	BlazePose		MotionBERT	
Model	Rozhodovací strom	Náhodný les	Rozhodovací strom	Náhodný les
Nejlepší hyperparametry (kriteriální funkce, maximální hloubka)	Gini index, 12	Gini index, 9	entropie, 8	Gini index, 7
Přesnost křížové validace	99,8 %	100 %	99,9 %	100 %
Testovací přesnost lepšího z modelů	100 %		100 %	

Klasifikace chybných sekvencí, které pocházely ze souřadnic modelu BlazePose dosáhla přesnosti 85,7 % a pro souřadnice z modelu MotionBERT přesnosti 76,2 %. V obou případech byly nejčastěji chybné predikce způsobeny tím, že byly výpady predikovány jako bulharské mrtvé tahy na jedné noze, což není překvapivé, neboť při chybných prováděních těchto cviků jsou si polohy těla velice podobné.

4.5.5 Automatická segmentace jednotlivých opakování v sekvencích k vyhodnocení

Ve chvíli, kdy byl znám druh prováděného cviku (ať už zadán explicitně nebo predikován pomocí algoritmu, který byl uveden v předchozí sekci), bylo potřeba nalézt hranice jednotlivých opakování. To je potřeba z důvodu, že každý cvik může být provedený jinak kvalitně a musí být tedy posuzován samostatně. Opakování byla hledána tak, že bylo zkonstruováno průměrné opakování daného cviku a ve vstupní sekvenci byly nalezeny podsekvence, které se nejvíce shodovaly s průměrným opakováním. Míra shody byla definována jako DTW vzdálenost.

Průměrné opakování bylo zkonstruováno tak, že nejprve byla spočítána průměrná délka opakování daného cviku a na tuto délku byla převzorkována všechna opakování. Následně bylo možné zkonstruovat průměrný průběh všech těchto stejně dlouhých časových řad.

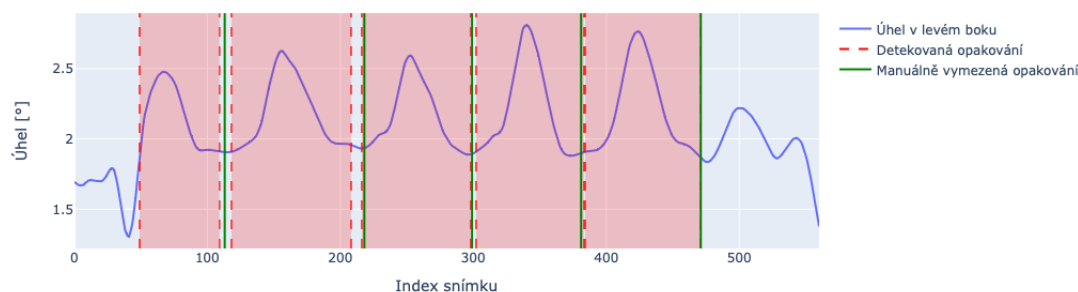
Při samotném hledání byly v sekvenci nalezeny podsekvence, které měly co nejmenší DTW vzdálenost od daného průměrného cviku. Algoritmus, který byl použit je implementován v knihovně `dtaidistance`¹⁰. Funkce nalezne k nejshodnějších podsekvencí podle rostoucí DTW vzdálenosti. Jelikož nebyl dopředu známý počet opakování, bylo potřeba nejprve nastavit k na dostatečně velké číslo (v tomto případě 20) a mezi nalezenými podsekvencemi nalézt pouze ty, které odpovídaly provedeným opakováním. Toho bylo docíleno tak, že pro nalezené podsekvence byl sestaven graf akumulovaného součtu jejich velikostí a v grafu byl hledán takzvaný loket, tedy místo zlomu, od kterého akumulovaný součet rostl už pomaleji.

¹⁰dtaidistance – <https://dtaidistance.readthedocs.io>

Bylo totiž pozorováno, že pokud byly nalezeny části, které neodpovídaly opravdu provedeným opakováním, byly výrazně menší svou velikostí (řádově jednotky až nízké desítky snímků). Také byl otestován způsob, kde byl loket hledán v grafu rostoucích DTW vzdáleností nalezených podsekvencí, neboť jakmile už se nejednalo o opravdová opakování, DTW vzdálenost prudce vzrostla. První ze zmíněných přístupů se ukázal na datech jako lepší. Pro hledání lokte byla využita implementace knihovny `kneed`¹¹. Pokud byla měřena DTW přes všechny příznaky, v některých případech nebyl algoritmus schopen najít téměř žádná opakování. Proto bylo otestováno i počítání DTW jen podle jednoho příznaku, který je význačný pro daný cvik (například úhel v koleni u dřepu). Přístup sice dobře fungoval v případech kde chyboval první popsáný přístup, ale výrazně chyboval v případech ostatních. Případy, kdy jeden přístup fungoval a druhý naopak ne, nebyly nijak závislé na zkoumaném typu cviku, ani na orientaci při provádění, a byly nejspíše způsobené nepřesnostmi v získaných parametrech. Výsledná segmentace byla provedena tak, že byly použity oba přístupy a vybrala se ta segmentace, která odhalila více opakování.

Testovány byly nejprve sekvence dobře provedených cviků. Zároveň byly porovnány detekce, kdy průměrné opakování bylo sestrojeno ze všech trénovacích opakování neohledě na orientaci, a také detekce, kde průměrné opakování bylo sestrojeno z opakování odpovídající orientace. Obecně byly znatelně lepší segmentace při detekci podle cviku odpovídající orientace.

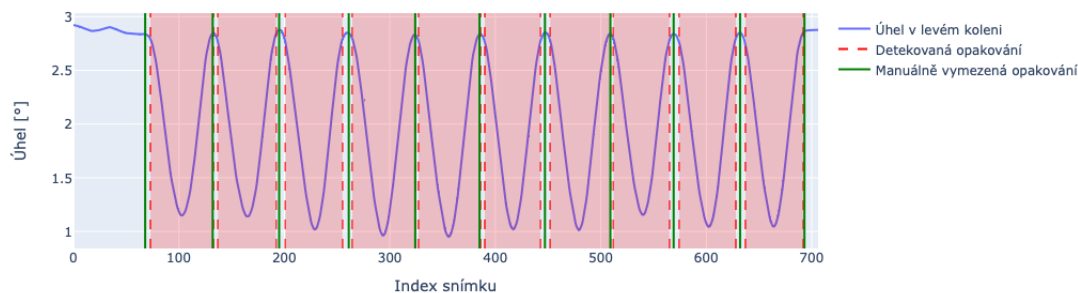
Co se týká orientace, která byla pro segmentaci nevhodnější, nedalo se říci, která by byla nejlepší pro všechny cviky, ale často to byl pohled ze strany nebo zešikma. Například na obrázku 4.11 jsou vidět výsledky u cviku pánevní most při pohledu zepředu. Nalezené segmenty přibližně odpovídají manuálně označeným hranicím opakování. Chybně byl nalezen menší segment na začátku, avšak opakování navíc není tak závažnou chybou, jako když není nějaké opakování detekováno vůbec. Naopak na obrázku 4.12 je ukázka segmentace při provádění dřepu a nalezené podsekvence odpovídají téměř přesně těm manuálně vyznačeným.



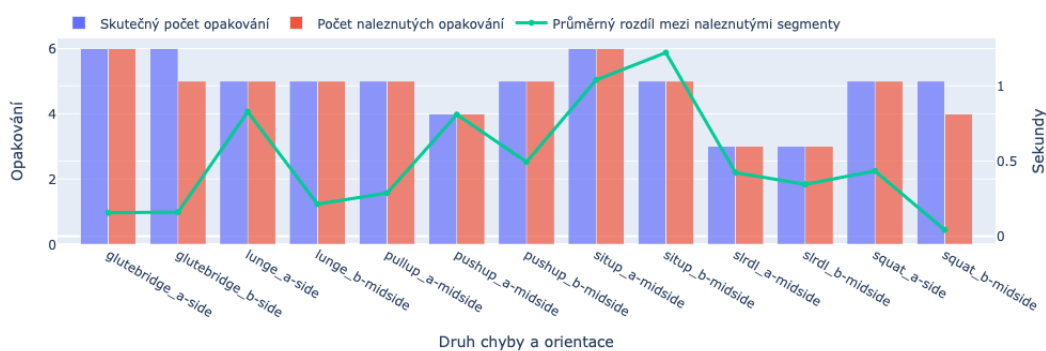
Obrázek 4.11 Ukázka nalezených opakování při automatické segmentaci. Demonstrováno na cviku pánevní most při pohledu zepředu.

Dále byl algoritmus otestován na chybných cvicích. I zde byla většina opakování správně detekována a to jak pro příznaky získané z BlazePose, tak pro MotionBERT. Výsledky segmentace pro všechny druhy chyb v nejlepších orientacích jsou uvedeny na obrázcích 4.13 a 4.14. Uvedeny jsou počty skutečných a detekovaných opakování a dále je uvedena průměrná časová odchylka mezi hranicemi skutečnosti a nalezených úseků. Při výpočtu času byla vzata v úvahu snímkovací frekvence 30 fps, kterou mají všechna videa společnou. Výsledky jsou pro BlazePose a MotionBERT srovnatelné, liší se pouze v řádek milisekund. BlazePose správně detekoval o dvě opakování navíc a průměrná časová odchylka přes všechny druhy chyb byla 0,497 sekundy namísto 0,51 sekundy u MotionBERT.

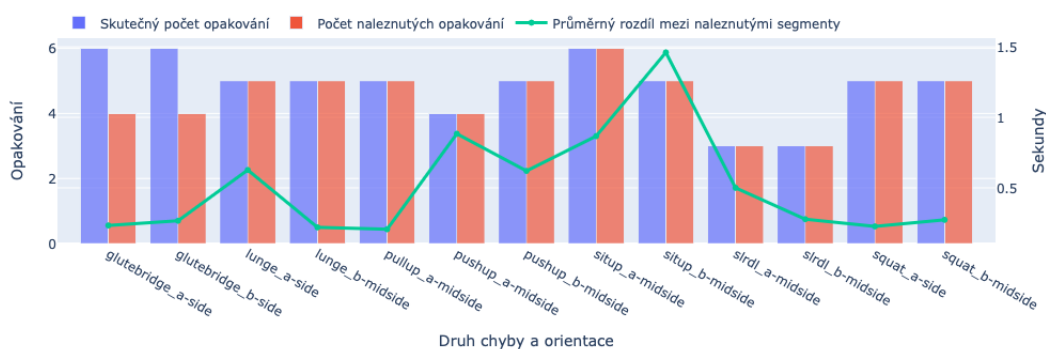
¹¹kneed – <https://kneed.readthedocs.io>



■ **Obrázek 4.12** Ukázka nalezených opakování při automatické segmentaci. Demonstrováno na dřepu při pohledu zešikma.



■ **Obrázek 4.13** Ukázka výsledků segmentace pro vybrané orientace všech druhů chyb. Znázorněny jsou skutečné počty opakování, počet nalezených opakování a průměrná časová odchylka hranic nalezených segmentů a skutečnosti. Zpracováno modelem BlazePose.



■ **Obrázek 4.14** Ukázka výsledků segmentace pro vybrané orientace všech druhů chyb. Znázorněny jsou skutečné počty opakování, počet nalezených opakování a průměrná časová odchylka hranic nalezených segmentů a skutečnosti. Zpracováno modelem MotionBERT.

4.5.6 Vyhodnocení správnosti cviků

Pro vyhodnocení, zda je daný cvik proveden správně bylo opět potřeba vypočítat průměrné opakování. V předchozí části bylo vidět, že porovnávání průměrného opakování sestaveného ze všech orientací s opakováním v nějaké konkrétní orientaci není vhodné. Při extrakci souřadnic částí těla byly pro každou orientaci získány trochu odlišné hodnoty, i když byla všechna opakování prováděna stejným způsobem. Pro každou orientaci bylo tedy vytvořeno průměrné opakování a hodnocení probíhalo pro každou orientaci zvlášť. Dále byla v rámci trénovacích dat změřena DTW vzdálenost každého opakování od vypočítaného průměrného opakování pro každý parametr zvlášť. Z těchto vzdáleností poté byl spočítán průměr:

$$\bar{d} = \frac{1}{n} \sum_{\mathbf{x} \in X} DTW(\bar{\mathbf{x}}, \mathbf{x}), \quad (4.1)$$

kde $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ je množina všech časových řad správných průběhů daného parametru a $\bar{\mathbf{x}}$ je průměrný správný průběh daného parametru.

Při samotném vyhodnocení správnosti byla spočítána DTW vzdálenost od průměrného opakování (odpovídající orientace) po příznacích. Vzdálenosti byly dále vyděleny průměrnými vzdálenostmi získanými z trénovacích dat. Takto získané číslo znamenalo, kolikrát byla vypočítaná vzdálenost větší (nebo menší) než průměr v trénovacích datech. Přístup je o něco odolnější vůči případným odlehlým hodnotám mezi trénovacími daty v některém příznaku než přístup, který byl uveden v [3]. Tento přístup porovnával vzdálenost testovaného cviku s maximální vzdáleností v rámci trénovacích dat. Pokud by ale pro některý příznak v trénovacích datech byla vzdálenost neobvykle vyšší než ostatní (například z důvodu nepřesných získaných souřadnic), i některé špatně provedené parametry by mohly být hodnoceny jako správně provedené.

Jakmile byl pro každý parametr spočítán podíl jeho vzdálenosti od průměrného opakování a průměrné vzdálenosti napříč trénovacími daty, bylo nutné ho porovnat s prahovou hodnotou, kterou když překročil, byl daný parametr považován za chybný. Parametr může být ponechán jako parametr určující toleranci při vyhodnocování. Čím větší je, tím větší musí být v daném příznaku odlišnost, aby byl označen za chybný. Pro detekci odlehlých hodnot na základně statistických měření se často používá hranice odlehlosti jako vzdálenost tří směrodatných odchylek od průměru [19]. V tomto případě se sice jedná o časové řady, ale přístup je intuitivně podobný. Pro účely testování a vizualizace byl tedy zvolen trojnásobek průměrné DTW vzdálenosti od průměrného opakování jako kritická hranice. Průběh testovaného parametru \mathbf{y} byl tedy označen jako špatný, pokud platilo:

$$\frac{DTW(\bar{\mathbf{x}}, \mathbf{y})}{\bar{d}} > \lambda, \quad (4.2)$$

kde λ je parametr tolerance a konkrétní hodnota byla v tomto případě zvolena jako $\lambda = 3$.

Následně byly takto ohodnocené příznaky omezeny jen na ty, které definovaly správné provedení daného cviku, ostatní byly ohodnoceny jako správné, neboť nemají na správnost daného cviku vliv.

Výsledky

Hlavními výsledky byly úspěšnosti při hodnocení kvality prováděných cviků a také vizualizace těchto hodnocení.

5.1 Vyhodnocení cviků

Úspěšnost ohodnocení byla testována na parametrech z modelu BlazePose i z modelu MotionBERT. Také byly srovnávány jednotlivé orientace. Výsledky byly popsány přesností klasifikace a také pomocí f1-skóre zaměřeného na minoritní třídu, což byly v tomto případě špatně provedené parametry.

Tabulka 5.1 popisuje pro každý druh cviku model a orientaci, která se ukázala jako nejlepší s ohledem na f1-skóre (kromě jediného případu byly vždy lepší i odpovídající přesnosti). Otestovány byly všechny typy chyb kromě chyby *situp_b*, neboť ohyb zad nebyl nijak zachytitelný pomocí kinematického modelu skeletonu. Přesnost pro BlazePose přes všechny cviky (každý ve své nejuspěšnějších orientaci) byla 90,9 % a f1-skóre 83,6 %. Pro model MotionBERT to bylo 89,4 % a 82,7 %.

■ **Tabulka 5.1** Přehled výsledků evaluace správných a chybných cviků.

cvik	nejlepší orientace	nejlepší model	přesnost [%]	f1-skóre [%]
sklapovačka	zešikma	BlazePose	100	100
dřep	zešikma	BlazePose	97,2	92,3
klik	zepředu	MotionBERT	96	90,9
pánevní most	zepředu	BlazePose	95	90,9
přítah na hrazdě	zešikma	MotionBERT	86	83,7
bulharský mrtvý tah na jedné noze	zešikma	MotionBERT	82	72,7
výpad	zešikma	BlazePose	79,1	72,7

Blíže jsou popsány výsledky hodnocení některých cviků, které oproti ostatním nedosahovaly tak vysoké testovací přesnosti. Při detailnějším rozboru hodnocení výpadů bylo vidět, že úspěšnost klesala nejvíce díky falešně negativnímu ohodnocení úhlu v kyčli a úhlu osy trupu pro chybu typu *lunge_b_l*. Dále také pro bulharský mrtvý tah na jedné noze platilo, že kvalitu hodnocení zhoršovalo falešně negativní hodnocení úhlu v kyčlích, které byly pro některé chyby prováděny správně. Naopak při hodnocení přitahu na hrazdě z pohledu zezadu (v tabulce neuvedeno), převažovaly falešně pozitivní parametry, které se týkaly úhlu v ramenou a loktech, a tedy nebyl odhalen rozsah pohybu oproti správné technice.

Srovnáním modelů pro získání souřadnic skeletonu s ohledem na kvalitu hodnocení lze říci, že na tom byly svou úspěšností obdobně. BlazePose byl lepší při hodnocení čtyř druhů cviků, kdežto MotionBERT při hodnocení tří druhů. Jde-li o srovnání úspěšnosti klasifikace a automatické segmentace, o trochu lépe vycházel model BlazePose, ale s pouze nepatrným rozdílem.

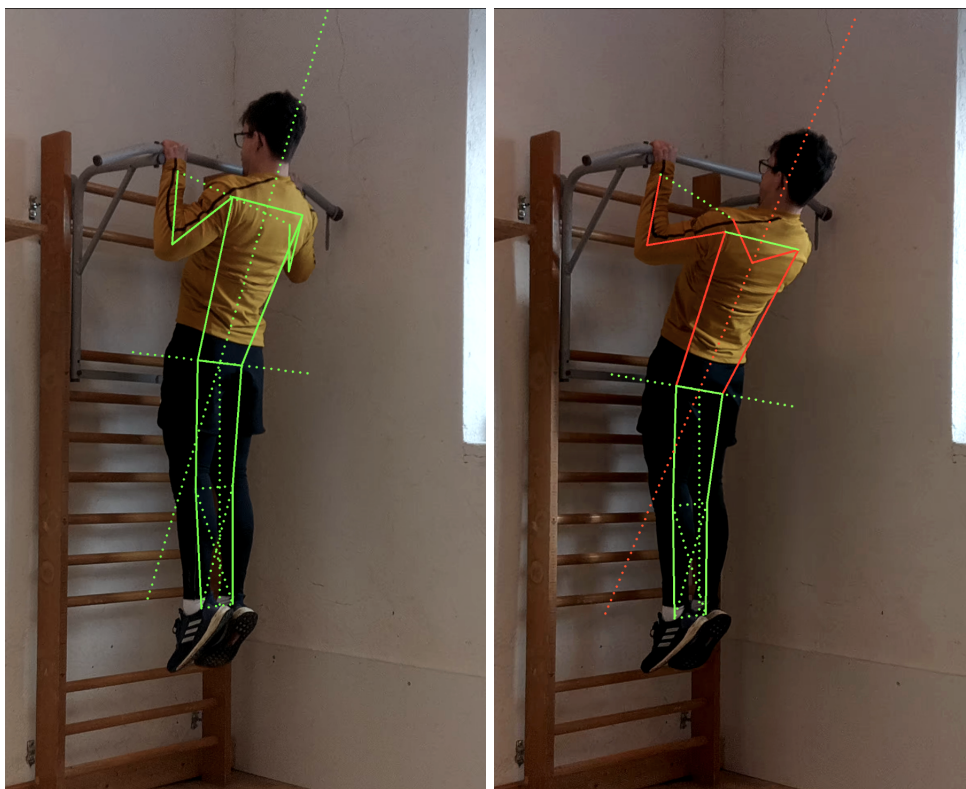
Co se týká srovnání jednotlivých orientací při hodnocení, nejúspěšněji vycházela pozice zešikma, nejlepší byla pro 5 druhů cviků. Pro dva druhy cviků byla nejlepší pozice zepředu. Pozice zezadu či ze strany nebyla ideální pro žádný ze cviků.

5.2 Vizualizace

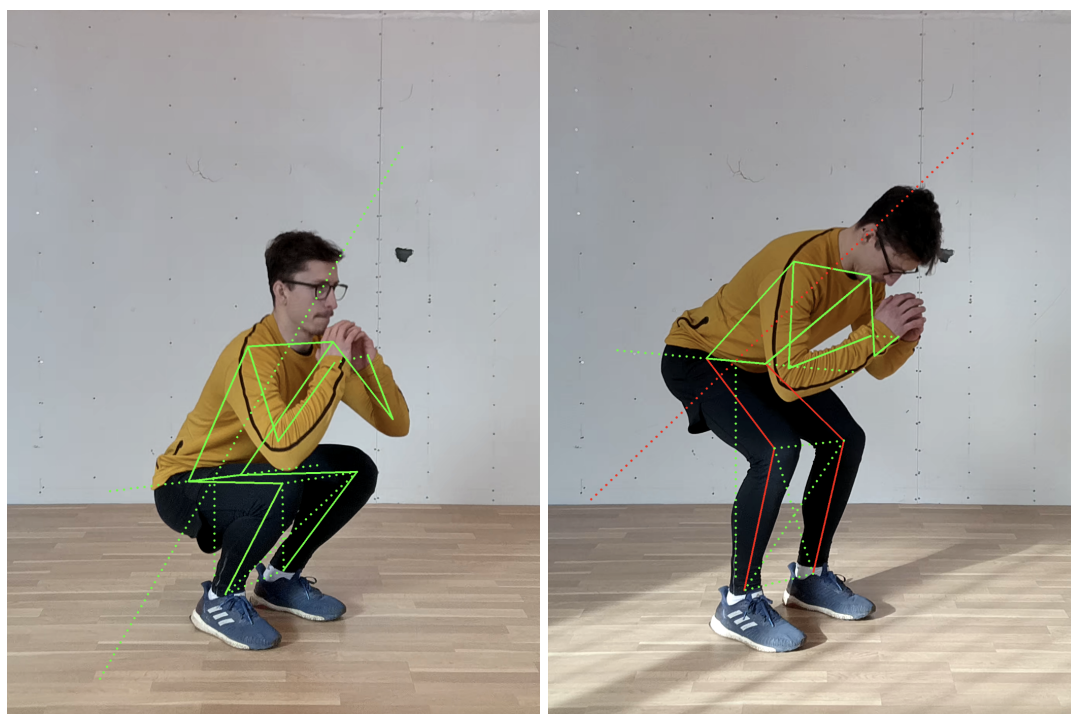
Po provedení všech předchozích kroků bylo na řadě výsledky vyhodnocení vizualizovat. Zvolenou metodou bylo zobrazení videa cvičence při provádění cviku a zobrazení skeletonu v místech, kde se nacházely souřadnice jednotlivých částí těla. Zakreslené byly části těla, které souvisely se získanými parametry – holeně, stehna, trup, paže a předloktí. Jelikož byly měřeny i další parametry, které nesouvisí přímo s částmi těla (například úhel osy trupu s rovinou země nebo vzdálenost kolenou), bylo potřeba i tyto parametry při vizualizaci zahrnout. Vzdálenosti částí těla byly zakresleny přerušovanými spojnicemi mezi odpovídajícími body na těle. Osy, u kterých byl zkoumán úhel, který svírají s rovinou podlahy, byly znázorněny taktéž jako přerušované čáry mezi částmi těla, kudy prochází. Krajiní body byly extrapolovány a přímka mezi tedy prodloužena, aby bylo jasné, že se nejedná o vzdálenost mezi částmi těla.

Pro znázornění, zda se pohybovala nějaká konkrétní část těla správně či ne, byla použita barevná škála, kterou se daná část vykreslovala. Navíc zde nebyla barva rozdělena pouze binárně na „dobrou“ a „špatnou“, tak jak to bylo uvažováno v předešlé části, kdy se pouze vyhodnocovala úspěšnost rozpoznání chyb. Ve skutečnosti není hranice mezi dobře provedeným cvikem a špatně provedeným cvikem ostrá. Některé chyby mohou být menší, některé větší. Toho bylo využito i při volbě barvy částí těla. Při testování úspěšnosti se vycházelo ze zvoleného násobku (zde 3) průměrné vzdálenosti trénovacích opakování od průměrného správného opakování. Hranici bylo možné využít jako střed barevného přechodu. Pokud byla vzdálenost testovaného cviku pod touto hranicí dostatečně daleko, barva byla zelená. Pokud se vzdálenost blížila zvolené hranici, barva přecházela do žluté a od překročení hranice dál do červené. Pokud byl tedy daný parametr pohybu vyhodnocen jako chybný, ale hraniční vzdálenost od správného provedení byla překročena jen mírně, vizualizován byl žlutou či oranžovou barvou, které představují mírnější chybovost než barva červená.

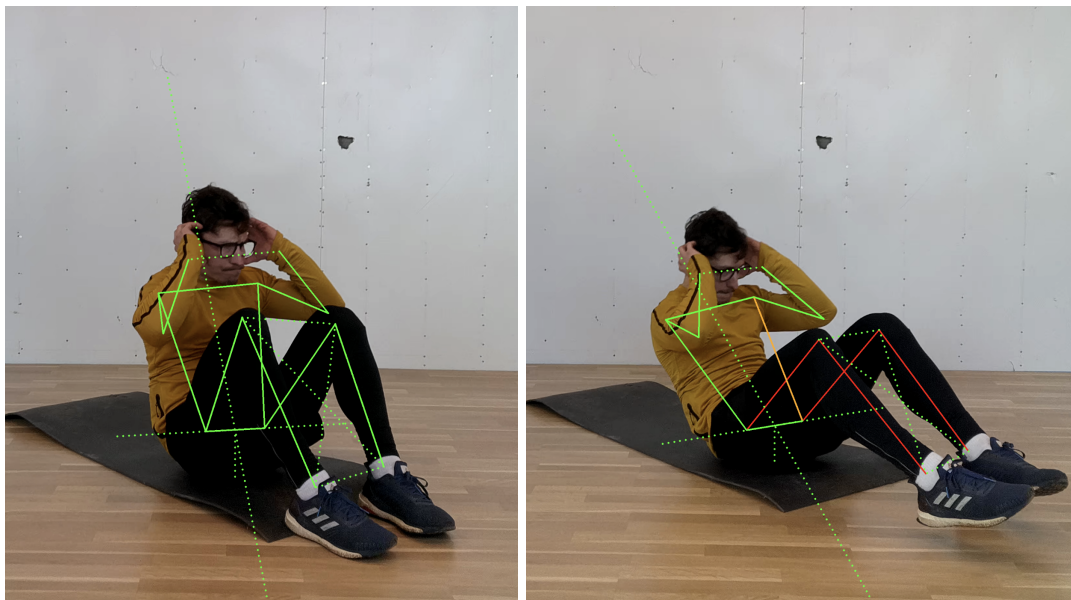
Mimo vizualizaci samotné kostry bylo navíc video doplněno o informaci o prováděném druhu cviku a aktuálním počtu opakování ve formě textu (Obrázek 5.4).



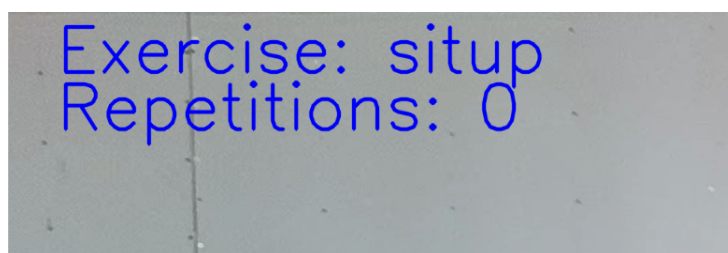
■ Obrázek 5.1 Ukázka vizualizace správně (vlevo) a špatně (vpravo) provedeného přítahu na hrazdě.



■ Obrázek 5.2 Ukázka vizualizace správně (vlevo) a špatně (vpravo) provedeného dřepu.



■ **Obrázek 5.3** Ukázka vizualizace správně (vlevo) a špatně (vpravo) provedené sklapovačky. Na špatném příkladu je vidět, že úhel v levé kyčli byl vyhodnocen špatně (ve skutečnosti je jeho úhel v pořádku). Hranice nebyla ale přesáhnutá tolik, proto je sousední tělní segment obarven pouze oranžově.



■ **Obrázek 5.4** Ukázka textového popisu prováděného cviku a aktuálního počtu provedených opakování při vizualizaci.

Kapitola 6

Diskuze

V práci bylo provedeno a otestováno několik dílčích kroků pro vyhodnocení správnosti cviku. Prvním krokem ke kvalitnějším výsledkům by bylo zaměřit se na každou dílčí část více do hloubky.

První možností by bylo otestovat jiné volby parametrů pro popis ideálního průběhu cviku. V úvahu by se mohly vzít názory různých trenérů, kteří by mohli doporučit sledovat pro daný cvik jiné parametry, například sklon hlavy, který zatím nebyl pro žádný cvik analyzován.

Dále by bylo možné porovnat jiné algoritmy pro analýzu časových řad a detekci špatně provedených cviků. Na problém rozpoznání špatně provedeného parametru cviku lze nahlížet jako na detekci anomálií mezi správně provedenými parametry. V problematice detekce anomálií existují další přístupy, jako například využití neuronových sítí nebo jednotřídních SVM modelů.

Hlavním faktorem ovlivňující výsledky je však samotný způsob získávání souřadnic částí těla a přesnost takového měření. Jednou z možností by bylo využití více kamer, které by zaznamenávaly cvik z více úhlů a dílčí data by poté byla syntetizována do jednoho modelu skeletonu. Tento způsob by jistě poskytl daleko přesnější měření, avšak by postrádal jednoduchost použití, kterou pro běžného uživatele představovala jedna RGB kamera, kterou může mít například i v mobilním telefonu. Také by bylo možné zkusit předřadit před model MotionBERT jiné modely pro získání 2D souřadnic a otestovat, zda by to jeho kvalitu nevylepšílo. Možností by byl zde testovaný model BlazePose, který sám o sobě dosahoval dobrých výsledků.

Dalším možným zkvalitněním analýzy by bylo natočení většího množství záběrů prováděných cviků, od různých cvičenců a s různou mírou nesprávnosti daného cviku, aby bylo možné otestovat, jak nepatrné rozdíly je možné odhalit, aniž by algoritmus penalizoval správnou techniku. Záběry by bylo dobré natáčet již jen z jednoho úhlu, a to z takového, který byl pro daný cvik nejvhodnější pro segmentaci a hodnocení.

V neposlední řadě by bylo zajímavé zkusit vyhodnocování záběrů v reálném čase. V takovém případě by ale bylo potřeba zaměřit se na rychlost dílčích kroků algoritmu a co nejvíce je optimalizovat. Použité metody se nezaměřovaly na rychlost, a proto by nemusely být pro použití v reálném čase vhodné. Hlavním faktorem, který celkovou rychlost ovlivňuje, je samotný model pro získání skeletonu. Ze zkoumaných modelů by jistě nešlo použít například model MotionBERT, jehož inference byla velice pomalá. Model by musel umět pracovat alespoň v 15 až 20 snímcích za sekundu, aby se vyhodnocování dalo považovat za plynulé. Co se týká klasifikace a segmen-

tace, nebylo by možné sekvenci zkoumat jako celek. Pravděpodobně by bylo potřeba zkoumat jen určité časové okno od posledního získaného snímku. Muselo by se otestovat, jak velké by takové okno muselo být, aby bylo možné s jistotou určit, o jaký cvik se jedná a zda bylo úspěšně detekováno jedno celé opakování.



Kapitola 7

Závěr

Práce se zabývá snímáním posilovacích cviků, jejich analýzou a následným vyhodnocením a vizualizací. Nasnímáno bylo několik druhů cviků za pomoci profesionálního sportovního trenéra. Zároveň byly nasnímány i typické chyby při cvičení, které dále sloužily k testování. Byly definovány parametry, podle kterých byla určena ideální technika cviku. Navrženy byly algoritmy pro získání těchto parametrů ze vstupních videí, jejich předzpracování, segmentaci opakování a následné vyhodnocení. V rámci všech kroků byly porovnávány různé modely pro získání souřadnic lidského skeletu a zkoumáno tak jejich vliv na dílčí výsledky.

Pro vybrané orientace, ve kterých byly cviky snímány, dosahoval navržený algoritmus pro hodnocení správnosti provádění průměrné přesnosti 90 % a průměrného f-skóre (s ohledem na minoritní třídu) 83 %. Výsledky byly vizualizovány a pomocí barevné škály byly odlišeny dobré a špatné parametry pohybu.

Všechny stanovené cíle byly splněny. Výsledky práce mohou sloužit jako zdroj pro další zkoumání problematiky.

Seznam použitých zkratek

ANN	umělá neuronová síť
CNN	konvoluční neuronová síť
SVM	metoda podpůrných vektorů
DCT	diskrétní kosinová transformace
DTW	dynamic time warping

Bibliografie

1. NOVATCHKOV, Hristo; BACA, Arnold. Machine learning methods for the automatic evaluation of exercises on sensor-equipped weight training machines. *Procedia Engineering*. 2012, roč. 34, s. 562–567. Dostupné z DOI: 10.1016/j.proeng.2012.04.096.
2. YANG, Lei; LI, Yingxiang; ZENG, Degui; WANG, Dong. Human Exercise Posture Analysis based on Pose Estimation. *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. 2021, roč. 5, s. 1715–1719.
3. PARMAR, Paritosh; MORRIS, Brendan. Measuring the Quality of Exercises. In: 2016, sv. 2016, s. 2241–2244. Dostupné z DOI: 10.1109/EMBC.2016.7591175.
4. HUANG, Yo-Ping; KUO, Wen-Lin; HAObIJAM, Basanta; LEE, Si-Huei. Evaluating Power Rehabilitation Actions Using a Fuzzy Inference Method. *International Journal of Fuzzy Systems*. 2021, roč. 23. Dostupné z DOI: 10.1007/s40815-021-01097-8.
5. ZHENG, Ce; WU, Wenhan; CHEN, Chen; YANG, Taojiannan; ZHU, Sijie; SHEN, Ju; KEHTARNAVAZ, Nasser; SHAH, Mubarak. *Deep Learning-Based Human Pose Estimation: A Survey*. 2022. Dostupné z arXiv: 2012.13392 [cs.CV].
6. BAZAREVSKY, Valentin; GRISHCHENKO, Ivan; RAVEENDRAN, Karthik; ZHU, Tyler; ZHANG, Fan; GRUNDMANN, Matthias. BlazePose: On-device Real-time Body Pose tracking. *CoRR*. 2020, roč. abs/2006.10204. Dostupné z arXiv: 2006.10204.
7. GRISHCHENKO, Ivan; BAZAREVSKY, Valentin; ZANFIR, Andrei; BAZAVAN, Eduard Gabriel; ZANFIR, Mihai; YEE, Richard; RAVEENDRAN, Karthik; ZHDANOVICH, Mat-svei; GRUNDMANN, Matthias; SMINCHISESCU, Cristian. *BlazePose GHUM Holistic: Real-time 3D Human Landmarks and Pose Estimation*. 2022. Dostupné z arXiv: 2206.11678 [cs.CV].
8. ZHU, Wentao; MA, Xiaoxuan; LIU, Zhaoyang; LIU, Libin; WU, Wayne; WANG, Yizhou. *Learning Human Motion Representations: A Unified Perspective*. 2023. Dostupné z arXiv: 2210.06551 [cs.CV].
9. SENIN, Pavel. Dynamic Time Warping Algorithm Review. 2009.
10. TAVENARD, Romain. *An introduction to Dynamic Time Warping* [online]. 2021. [cit. 2023-05-09]. Dostupné z: <https://rtavenar.github.io/blog/dtw.html>.
11. HYNDMAN, Rob J. Moving Averages. In: *International Encyclopedia of Statistical Science*. Ed. LOVRIC, Miodrag. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, s. 866–869. ISBN 978-3-642-04898-2. Dostupné z DOI: 10.1007/978-3-642-04898-2_380.
12. MORAN, Gary W. *Locally-Weighted-Regression Scatter-Plot Smoothing (LOWESS): a graphical exploratory data analysis technique*. 1984. Tech. zpr. NAVAL POSTGRADUATE SCHOOL MONTEREY CA.

13. RANGANATHAN, Pradeep; OLSON, Edwin. Locally-weighted homographies for calibration of imaging systems. *IEEE International Conference on Intelligent Robots and Systems*. 2014, s. 404–409. Dostupné z DOI: [10.1109/IR0S.2014.6942591](https://doi.org/10.1109/IR0S.2014.6942591).
14. ZHANG, Xunyu; JIANG, Shumin. Application of Fourier Transform and Butterworth Filter in Signal Denoising. In: *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*. 2021, s. 1277–1281. Dostupné z DOI: [10.1109/ICSP51882.2021.9408933](https://doi.org/10.1109/ICSP51882.2021.9408933).
15. ALPAYDIN, Ethem. *Introduction to Machine Learning*. 3. vyd. Cambridge, MA: MIT Press, 2014. Adaptive Computation and Machine Learning. ISBN 978-0-262-02818-9.
16. FANG, Hao-Shu; LI, Jiefeng; TANG, Hongyang; XU, Chao; ZHU, Haoyi; XIU, Yuliang; LI, Yong-Lu; LU, Cewu. *AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time*. 2022. Dostupné z arXiv: [2211.03375](https://arxiv.org/abs/2211.03375) [cs.CV].
17. CHEREPANOV, Alexander. *Motion Data Interpolation in Sign Language Synthesis* [online]. Praha, 2021 [cit. 2023-01-05]. Dostupné z: <https://dspace.cvut.cz/handle/10467/99105>. Diplomová práce. ČVUT v Praze, Fakulta elektrotechnická.
18. HUA, Andrew; CHAUDHARI, Pratik; JOHNSON, Nicole; QUINTON, Joshua; SCHATZ, Bruce; BUCHNER, David; HERNANDEZ, Manuel E. Evaluation of Machine Learning Models for Classifying Upper Extremity Exercises Using Inertial Measurement Unit-Based Kinematic Data. *IEEE Journal of Biomedical and Health Informatics*. 2020, roč. 24, č. 9, s. 2452–2460. Dostupné z DOI: [10.1109/JBHI.2020.2999902](https://doi.org/10.1109/JBHI.2020.2999902).
19. LEYS, Christophe; LEY, Christophe; KLEIN, Olivier; BERNARD, Philippe; LICATA, Laurent. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*. 2013, roč. 49, č. 4, s. 764–766. ISSN 0022-1031. Dostupné z DOI: <https://doi.org/10.1016/j.jesp.2013.03.013>.

Obsah přiloženého média

Soubory lze nalézt také v repozitáři na fakultním GitLabu: <https://gitlab.fit.cvut.cz/pilardan/human-pose-estimation-daniel-pilar>.

	<code>implementation</code> zdrojové kódy implementace, data
	<code>text</code> jednotlivé kapitoly práce ve formátu \LaTeX , obrázky použité v dokumentu
	<code>ctufit-thesis.cls</code> šablona práce ve formátu \LaTeX
	<code>README.md</code> informace k použité šabloně
	<code>ctufit-thesis.tex</code> hlavní zdrojová forma práce ve formátu \LaTeX
	<code>pilardan-thesis.pdf</code> text práce ve formátu PDF