

UPORABA METOD RAČUNALNIŠKEGA VIDA ZA POMOČ PRI TELESNI VADBI

BLAŽ ČERNI, BORUT BATAGELJ

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
blaz.cerni@hotmail.com, borut.batagelj@fri.uni-lj.si

Pred nekaj leti smo bili priča hudi krizi, povezani s koronavirusom, ki je močno vplivala tudi na človeške navade pri telovadbi in izvajanju aktivnosti na prostem ali v fitnes centrih. Zaradi nezmožnosti obiskovanja teh prostorov se je vse več ljudi odločalo za različne oblike vadbe doma. Pri tem se je pojavilo vprašanje pravilne in varne izvedbe vadbe, še posebej pri začetnikih. V tem delu analiziramo, kako nam lahko metode računalniškega vida pomagajo pri prepoznavanju človeške drža in telesnih delov ter kako lahko te informacije uporabimo za usmerjanje in pomoč pri pravilni izvedbi telesnih vaj. Predstavimo tudi našo implementacijo aplikacije, imenovane Sweatpose, ki omogoča zaznavanje in sledenje telesnih delov ter hkrati uporabniku nudi pomoč pri pravilni izvedbi vaj.

DOI
<https://doi.org/10.18690/um.feri.2.2025.6>

ISBN
978-961-286-960-1

Ključne besede:

računalniški vid,
človeška drža,
fitnes,
ključne točke človeškega
telesa,
tensorflow

Prispevek temelji na:

Črni, B. (2024). *Uporaba metod računalniškega vida za pomoč pri telesni vadbi*. magistrsko delo, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko.



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.2.2025.6](https://doi.org/10.18690/um.feri.2.2025.6)

ISBN
978-961-286-960-1

Keywords:

computer vision,
pose estimation,
fitness,
human body key points,
tensorflow

The proceedings is based on: Černi, B.(2024). *Uporaba metod računalniškega vida za pomoč pri telesni vadbi*. master's thesis, University of Ljubljana, Faculty of Computer Science and Informatics.

USE OF COMPUTER VISION METHODS FOR GUIDANCE IN PHYSICAL EXERCISES

BLAŽ ČERNI, BORUT BATAGELJ

University of Ljubljana, Faculty of Computer and Information Science, Ljubljana,
Slovenia
blaz.cerni@hotmail.com, borut.batagelj@fri.uni-lj.si

A couple of years ago, we witnessed a severe crisis related to the coronavirus. This crisis significantly affected people's exercise habits and activities, whether outdoors or in fitness studios. Due to the inability to visit such places for exercise, more and more people opted for various forms of home workouts. This posed a challenge regarding the proper and safe execution of these exercises, especially for beginners. In this work, we analyze how computer vision methods can assist in recognizing human posture and key body points and how this information can guide and support individuals in performing exercises correctly. Finally, we present our implementation of such an application, which, with the help of pose estimation, provides users with assistance in executing their workouts properly. We named the application Sweatpose.



1 Uvod

Metode računalniškega vida, ki lahko na podlagi modelov ocenjujejo ključne točke človeškega telesa in človeško držo, se imenujejo metode za ocenjevanje človeške drže (angl. *pose estimation*). Gre za kompleksno področje v računalniškem vidu, ki omogoča napravam natančno določanje položaja anatomskih točk, kot je na primer lokacija kolena neke osebe na sliki. Rezultati ocenjevanje človeške drže, so napovedane ključne tičke, ki so označene z enoličnim identifikatorjem in največkrat vsebujejo tudi oceno zaupanja med 0 in 1. V zadnjih letih se je razvilo več različnih pristopov za ocenjevanje človeške drže. Nekateri pristopi se osredotočajo samo na določene dele telesa (npr. roke ali stopala), drugi pa na ocenjevanje človeške drže za samo eno osebo ali pa več oseb hkrati. Pri slednjih se pojavijo večji izzivi zaradi naraščajoče kompleksnosti, kar je posledica večjega števila oseb. Med osebami pa lahko prihaja tudi do prekrivanja (Černi, 2024).

Za implementacijo metod, ki ocenjujejo človeško držo večih oseb hkrati, sta znana predvsem dva načina implementacije:

- Pristop od zgoraj navzdol (angl. *top-down approach*). Ta metoda vključuje detektor oseb, ki najprej prepozna osebe in lokacijo njihovega telesa na sliki, potem pa se ocenjevanje ključnih točk telesa izvede za vsako osebo posebej.
- Pristop od spodaj navzgor (angl. *bottom-up approach*). V nasprotju s pristopom od zgoraj navzdol, pa ta metoda prepozna vse dele telesa na sliki naenkrat in potem poveže dele telesa z ustreznimi osebami.

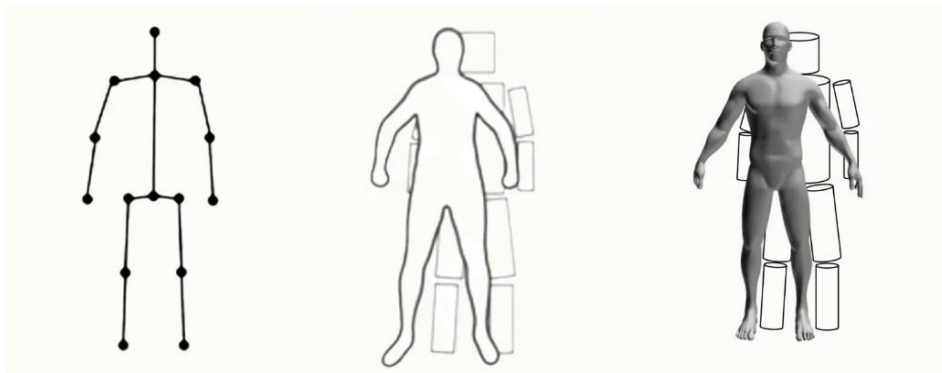
Pristop od zgoraj navzdol je lažji za implementacijo, ker je bolj sistematičen in direktno uporablja tehnike pristopa za ocenjevanje ključnih točk ene osebe. Slaba stran pa je, da je odvisen od uspešnosti detekcije oseb na sliki. Poleg tega pa se čas za izvedbo celotnega algoritma povečuje proporcionalno s številom oseb na sliki kar posledično onemogoča dobro pretočnost v realnem času. Iz tega vidika je pristop od spodaj navzgor v splošnem sprejet kot boljša opcija, ker omogoča visoko natančnost in pretočnost v realnem času, ne glede na število oseb na sliki.

Po uspešni identifikaciji oseb s pomočjo enega izmed zgoraj opisanih pristopov pa model za ocenjevanje človeške drže uporabi pridobljene koordinate, da sestavi predstavitev položajev ključnih točk. Ta predstavitev je lahko v 2D ali pa v 3D.

Prednost predstavitve v 2D je nižja računska zahtevnost, medtem ko nam ocenjevanje človeške drže v 3D prostoru, doda še koordinato Z (poleg X in Y), k napovedani lokaciji ključne točke, kar nam da informacijo o oddaljenosti osebe od kamere.

Sicer pa poznamo tri glavne tipe modelov za ocenjevanje človeške drže (Slika 1):

- Model na osnovi ogrodja (angl. *Skeleton-Based model*), v katerem ključne točke predstavlja množica sklepov in orientacij udov človeškega telesa. Ta model se lahko opiše tudi kot graf, kjer so vozlišča sklepi, poti med vozlišči pa so udi, ki povezujejo te sklepe.
- Model na osnovi obrisa (angl. *Contour-Based model*), v katerem množica ozbriso zajame povezavo delov telesa (značilno za 2D modele). Takšna predstavitev se je precej uporabljala v zgodnjih aktivnih modelih oblike (angl. *Active Shape Models*) ali krajše ASM (T.F. Cootes, 1995).
- Model na osnovi volumna (angl. *Volume-Based model*). Je naprednejši model, v katerem se človeško telo predstavi kot 3D volumen. Moderne predstavitve predstavlja mrežna oblika, ki se pridobi iz 3D skanov. Široko uporabljena modela na osnovi voluma sta SCAPE (Dragomir Anguelov, 2005) in SMPL (Matthew Loper, 2015).

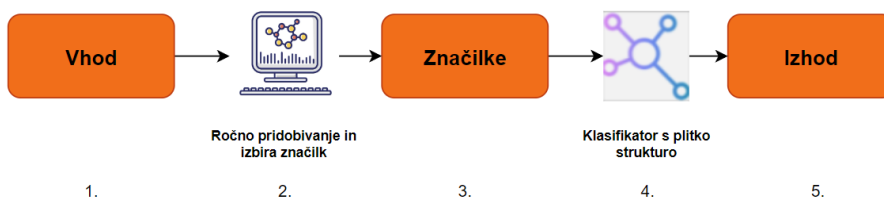


Slika 1: Tipi modelov za ocenjevanje človeške drže od leve proti desni: model na osnovi ogrodja, model na osnovi obrisa in model na osnovi volumna.

Vir: (Yucheng Chen, 2020).

2 Pregled področja

Začetni pristopi za ocenjevanje človeške drže temeljijo na tradicionalnih algoritmičnih strojnega učenja, ki tipično sledijo toku, ki je prikazan na sliki Slika 2. Diagram se začne z vhodnimi podatki (1.), v našem primeru gre za slike. V drugem koraku (2.) sledi ročno pridobivanje in izbiranje značilnk, ki so pomembne za obdelavo podatkov. Po obdelavi vhodni podatki postanejo značilke (3.), kar pomeni, da so pretvorjeni v številčne ali strukturirane oblike. Značilke nato preidejo skozi preprost klasifikator (4.), ki ima plitvo strukturo (na primer SVM ali odločitvena drevesa). Na koncu pa dobimo napoved oziroma izhod (5.), ki je lahko razvrstitev v različne kategorije. Postopek je potraten predvsem zaradi ročnega luščenja značilnk.

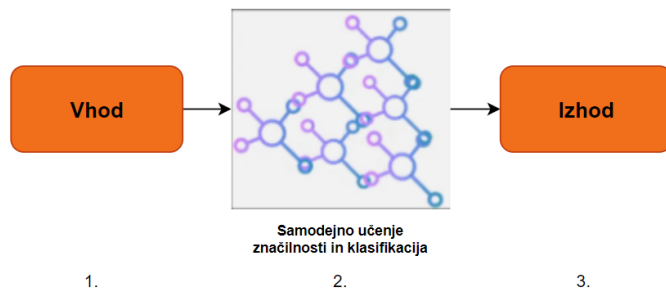


Slika 2: Postopek klasičnega pristopa za ocenjevanje človeške drže.

Vir: lasten.

Klasične pristope so uporabili avtorji v delu, kjer so poskušali problem reševati kot klasifikacijski problem s pomočjo naključnih gozdov (G. Rogez, 2008). Drugi (Felzenszwalb, 2005) pa so se ocenjevanja človeške drže lotili s pomočjo koncepta »Slik struktur« (angl. *Pictorial Structures*). Problem so zastavili kot optimizacijski problem, katerega cilj je bil najti najboljšo razporeditev in ujemanje različnih območij slik z videzom določenega dela telesa ter čim boljše geometrično razmerje med deli telesa (npr. bližina glave od preostalega dela telesa).

Iz tega razloga so se pristopi hitro osredotočili na umetne nevronske mreže in globoko učenje. Uteži na posameznih nevronskih povezavah, se tako preračunavajo na podlagi objektov in značilnk vhodne slike in znajo tudi razlikovati med njimi. Posebne in skrite vzorce lahko odkrijejo brez potrebnega predprocesiranja.



Slika 3: Postopek ocenjevanja človeške drže s pomočjo globokega učenja.

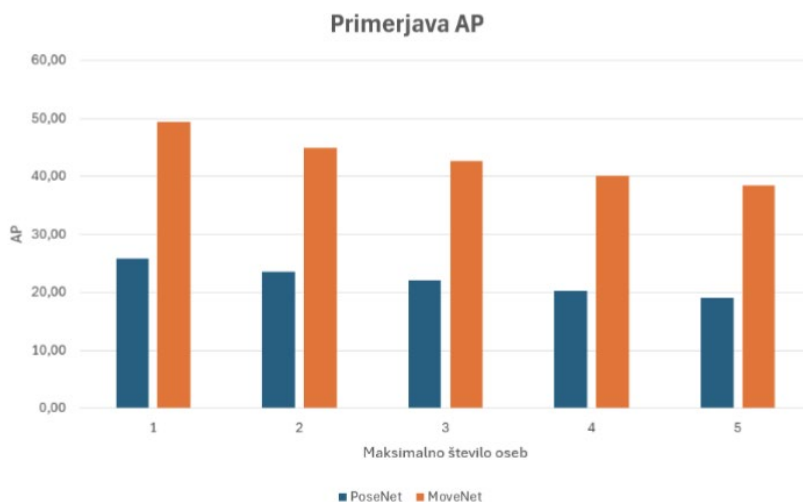
Vir: lasten.

Modeli, ki temeljijo na globokem učenju in nevronske mrežah, tipično prekosijo tradicionalne modele strojnega učenja v nalogah računalniškega vida kot je na primer ocenjevanje človeške drže. Med takšne modele štejemo Mask R-CNN (K. He, 2017), kjer gre v resnici za splošno ogrodje za detekcijo in segmentacijo objektov na sliki, katero pa omogoča tudi ocenjevanje človeške drže. Drugi znani modeli, ki temeljijo na globokem učenju so, AlphaPose (Hao-Shu Fang, 2023) in pa eden izmed najbolj znanih modelov s pristopom od spodaj navzgor OpenPose (Z. Cao, 2017). Postopek ocenjevanja človeške drže s pomočjo globokega učenja prikazuje Slika 3.

3 Izbor modela

Analiza modelov in pregled področja sta razkrila, da obstaja precej veliko različnih načinov in pristopov, s pomočjo katerih se lahko implementira aplikacija, ki s pomočjo ocenjevanja človeške drže, nudi ljudem pomoč in usmeritve pri izvajanju telesnih vadb. Za našo aplikacijo se osredotočimo na modele iz knjižnice Tensorflow.js (LeViet & Chen, 2021). Gre za odprtokodno programsko knjižnico za strojno učenje in umetni inteligenco, kjer za namen ocenjevanja človeške drže ponujajo modele MoveNet, PoseNet in BlazePose. Knjižnica omogoča izvajanje modela na strani odjemalca (na odjemalčevi strojni opremi), v sklopu spletne aplikacije. Vsi izmed naštetih modelov podpirajo izvajanje na grafični procesni enoti (GPU), kjer se v ozadju uporablja tehnologija WebGL ali pa WebGPU, prav tako pa tudi na centralni procesni enoti (CPU), z uporabo WebAssembly. Gre za kodo, ki je »bližje« strojnemu jeziku in zato omogoča precej boljše delovanje kot navadna verzija CPU in je v določenih primerih primerljiva z GPU verzijami (predvsem za primer uporabe pametnih telefonov). Za enega izmed ciljev pri implementaciji aplikacije si zastavimo podporo izvajanja vadbe več osebam hkrati, zato izmed izbora modelov

izločimo BlazePose, ki omogoča samo izvajanje za eno osebo. Med modeloma MoveNet in PoseNet pa se odločimo po izvedeni evalvaciji modelov na COCO podatkovni zbirki (Tsung-Yi Lin, 2014), kjer za naš primer uporabe iz zbirke 5000 testnih slik izberemo samo tiste, na katerih so ljudje. Natančneje, na katerih je ena (1045 slik), dve (1481 slik), tri (1749 slik), štiri (1897 slik) ali pa maksimalno pet oseb (2016) slik. Za ocenjevalno metriko izberemo uradno COCO metriko – povprečno točnost (AP – angl. *average precision*), ki temelji na metriki OKS (angl. *Object KeyPoint Similarity*).



Slika 4: Povprečna točnost (AP) za modela MoveNet in PoseNet, glede na naraščajoče maksimalno število oseb na slikah.

Vir: lasten.

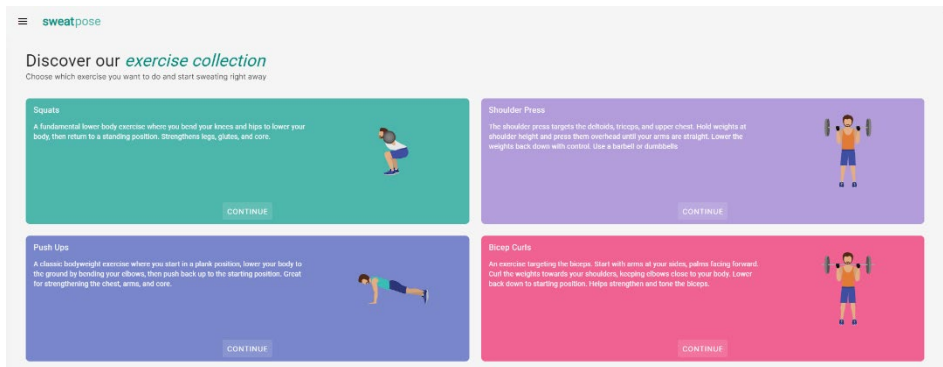
Rezultat evalvacije predstavlja Slika 4, na kateri vidimo, da model MoveNet dosega boljše rezultate zato, ga tudi izberemo za implementacijo aplikacije.

4 Aplikacija Sweatpose

V sklopu aplikacije Sweatpose, ki bo uporabnikom nudila pomoč in usmeritve pri izvajanju telesnih vadb podpremo štiri različne tipe telesnih vadb, ki jih lahko uporabniki izvajajo (Slika 5):

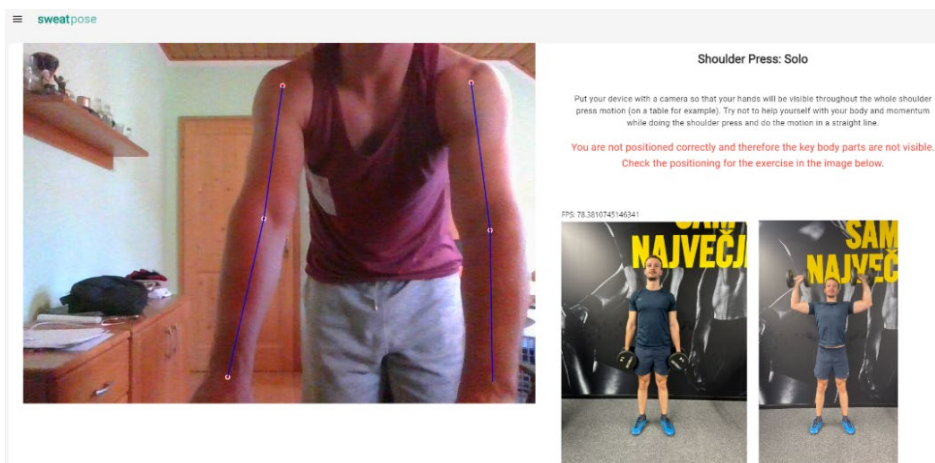
- Počepi – angl. *Squats*.

- Potisk nad glavo – angl. *Shoulder press* ali *Overhead press*.
- Skleca – angl. *Push ups*.
- Vzgib bicepsa – angl. *Bicep curls*.



Slika 5: Zaslonska slika zbirke telesnih vadb, ki jih podpremo v aplikaciji Sweatpose.

Vir: lasten.



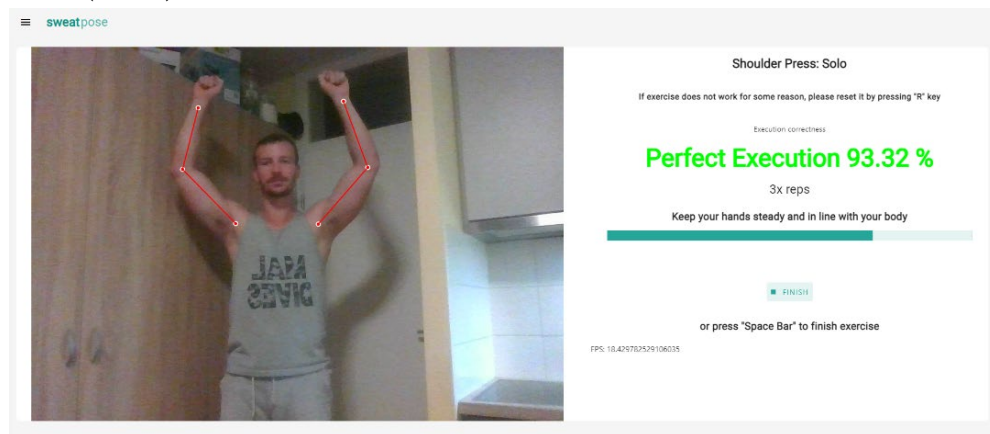
Slika 6: Prikaz za uporabnika pred začetkom vadbe, kjer je na levi strani slika iz kamere, na desni pa navodila: informacija o opravljeni oziroma nepravilni postavitvi, slika pravilne postavitve in posnetek pravilne izvedbe izbrane vadbe.

Vir: lasten.

Vsako vadbo je možno izvajati v tekmovalnem načinu dveh uporabnikov ali pa v načinu posameznika. Prav tako se izbere strojna oprema, na kateri se izvaja model za ocenjevanje človeške drže – grafična procesna enota ali centralna procesna enota

v kombinaciji z WebAssembly. Pred začetkom vadbe se mora uporabnik tudi pravilno postaviti pred kamero. Pomembno je, da so vidne vse ključne točke telesa, ki so za to vadbo glavne. Aplikacija uporabnika o tem tudi obvesti vizualno in tekstovno (Slika 6).

Ko uporabnik v pravilnem položaju stoji nekaj sekund, se vadba začne. Uporabniku se nato začnejo prikazovati informacije o natančnosti izvedbe vadbe v odstotkih in obarvanosti obvestila, število izvedenih ponovitev vadbe, grafični prikaz izvedbe posamezne ponovitve in usmeritve oziroma opozorila za bolj pravilno izvedbo vadbe (Slika 7).



Slika 7: Izvedba vadbe s prikazom slike iz kamere in prepoznanih delih telesa na levi strani in povratni informaciji o natančnosti izvedbe vadbe na desni strani.

Vir: lasten

Ko ima uporabnik dovolj vadbe, jo lahko konča s klikom na gumb »Finish« oziroma s tipko preslednice na tipkovnici. Po koncu vadbe se uporabniku prikaže pojavno okno, kamor lahko vpiše podatek o uporabljenih dodatnih utežeh oziroma svoji teži (na primer pri sklecah), za pomoč pri izračunu kalorij. Po potrditvi dodatnih informacij se uporabniku prikaže stran s povzetkom vadbe.

4.1 Implementacija izvajanja telesnih vadb

Implementacijo ločimo na tri dele, ki se izvajajo na čelnem delu (angl. *frontend*), kar pomeni, da jih poganja strojna oprema uporabnika (odjemalca):

- Inicializacija kamere in modela pred začetkom vadbe.
- Preverjanje ustreznosti postavitve uporabnika pred začetkom vadbe z uporabo modela MoveNet.
- Uporaba modela MoveNet med izvajanjem vadbe za pomoč in usmerjanje uporabnika k pravilni izvedbi vadbe in izračunavanje povratnih informacij, kot je točnost izvedbe v odstotkih.

Ob začetku posamezne vadbe se najprej pridobi zajem slike iz kamere. Kamera prenaša sliko v realnem času in jo hrani v HTML elementu *video*. Za dejanski prikaz uporabniku, pa uporabimo HTML element *canvas*, na katerega lahko potem tudi rišemo. Po inicializaciji kamere, se zgodi še inicializacija modela MoveNet. Glede na izbran tip strojne opreme, se pridobijo ustrezne definicije modelov iz oddaljenega Tensorflow repozitorija. V primeru izbora grafične procesne enote za način izvajanja, se pridobijo definicije modelov za izvajanje na grafični procesni enoti s podporo WebGPU (WebGPU, 2024). Gre za JavaScript API, ki ga podpirajo novejšje verzije brskalnikov in omogoča učinkovito izrabo grafične procesne enote. V primeru izbora centralne procesne enote, pa se pridobijo definicije modelov za izvajanje modelov s podporo WebAssembly, kjer pa gre za nizkonivojsko kodo, kar omogoča učinkovito izrabo strojne opreme. Inicializiran model vsebuje tudi sledilec (angl. *tracker*), ki omogoča sledenju več osebam, preko več zaporednih sličic. Vsaki zaznani osebi se na začetku priredi unikaten identifikator, s pomočjo katerega se skozi zaporedne sličice omogočata sledljivost in razumevanje, kdo je kdo.

V fazi preverjanja ustreznosti postavitve uporabnika pred kamero za izvajanje posamezne vadbe, model MoveNet že pridobiva ključne točke človeškega telesa. Tiste, ki so v uporabi se uporabniku tudi izrišejo in prikažejo. Faza preverjanja ustreznosti postavitve je pomembna iz dveh razlogov. Prvi je ta, da se takrat izvedejo določeni začetni izračuni oziroma začetne pozicije ključnih točk človeškega telesa, ki so potem v uporabi pri samem izvajanju vadbe. Drugi razlog pa je precej bolj praktične narave. Gre za to, da ko je uporabnik postavljen v pravilni, začetni držji za izvajanje določene vadbe, se dejansko začne izvedba vadbe, in ocenjevanje uporabnikovega izvajanja vadbe, na primer sklec. Aplikacija Sweatpose pravilnost postavitve pred začetkom vadbe, preverja na nivoju posamezne vadbe. Za vsako, posamezno vadbo sistem pričakuje določen položaj uporabnika oziroma bolje rečeno, določene ključne točke, ki morajo biti za začetek vadbe vidne in z dovolj

visoko oceno ocenjene (gre za oceno, ki jo ob zaznavanju ključnih točk vrača model MoveNet).

Primer takšnih pravil za izvajanje vadbe Potisk nad glavo:

- Vidnost in dovolj visoka ocena točnosti ključnih točk – zapestja na levi in desni roki, komolca na levi in desni roki in ramen.
- Pomembna je usmerjenost uporabnika neposredno v smer kamere.

Pravilna postavitve pred vadbo je izjemnega pomena za točnost ocenjevanja drže in ključnih točk s pomočjo modela MoveNet. Ker gre za model, ki ocenjuje v 2D, mora biti telo postavljeno tako, da omogoči pravilen izračun kotov med ključnimi točkami, ki se potem uporabljajo za ocenjevanje točnosti izvedbe vadbe.

Po pravilni postavitvi uporabnika pred kamero, se začne dejanska izvedba vadbe iz uporabnikovega pogleda. Model MoveNet na tej točki že izvaja ocenjevanje ključnih točk in drže človeškega telesa. Informacije o teh točkah uporabimo v sistemu, ki uporabniku nudi povratne informacije o tem kako izvaja posamezno vadbo. Povratne informacije so sestavljene iz procentualne točnosti izvedbe vadbe, števila ponovitev posamezne vadbe in predlogov ter opozoril, čemu se pri izvedbi vadbe izogibati in kako vadbo izvajati za boljši rezultat. Za vsako izmed štirih vadb implementiramo ločena pravila, ki se preverjajo za posameznega uporabnika. V ta namen najprej definiramo ključne točke, ki so pomembne za izvedbo vadbe posameznega tipa. Definirane ključne točke nato uporabimo za izračun kotov med njimi. Poleg kotov, ki jih izračunamo, za vsako vadbo definiramo tudi druge ključne spremenljivke, ki hranijo stanje izvedbe posamezne vadbe in parametre kot so na primer začetni in končni položaj vadbe, seznam vseh izračunanih procentualnih točnosti za vsako sličico zajetega posnetka ter število ponovitev izvedene vadbe. Za vsako izmed pravil izračunamo oceno točnosti v intervalu od 0 do 100. Ocene točnosti za pravila, se izračunajo na podlagi odstopanj kotov od idealnih kotov oziroma odstopanj koordinat od pričakovanih koordinat. Na podlagi ocen točnosti posameznih pravil, izračunamo končno oceno točnosti trenutne izvedbe vadbe v trenutni sličici, ki je enaka povprečju ocen točnosti posameznih pravil za nek tip vadbe.

Vadba Potisk nad glavo, je vadba, ki se jo primarno izvaja s pomočjo dodatnih uteži, ki se jih dviguje nad glavo. Aplikacija Sweatpose predpostavlja stoječo verzijo izvedbe vadbe, takšno kot prikazuje Slika 8.



Slika 8: Potisk nad glavo (angl. *shoulder press* ali *overhead press*) v stoje in z rdečo označen kot ter ključne točke, ki jih uporablja aplikacija za preverjanje pravilnosti izvedbe.

Vir: lasten.

Za to vadbo definiramo spodnje ključne točke človeškega telesa:

- Levo in desno zapestje.
- Levi in desni komolec.
- Leva in desna rama.

Definiramo torej šest ključnih točk človeškega telesa, pri ocenjevanju pa smo pozorni na to, da uporabnik hkratno premika obe roki. Zanimajo nas premiki vseh šest ključnih točk naenkrat. Definiramo tudi začetni in končni pričakovan kot izvedbe vadbe. Gre torej za kot, ki ga oklepajo zapestje, komolec in rama z vrhom v ramenu. Potem definiramo pravila na podlagi katerih se izračunavamo procentualno točnost izvedbe in nudimo nasvete in usmeritve uporabniku:

- Na podlagi x koordinat zapestij in komolcev preverjamo oziroma sledimo, da uporabnik z rokami ne zamahuje preveč levo ali desno in da ima roke

približno poravnane s telesom. Na takšen način lahko naredi potisk bolj mirno in naravnost navzgor. V primeru, da temu ni tako, uporabniku sporočimo, da ne drži rok dovolj mirno in v liniji s telesom.

- Na podlagi kotov obeh rok in definiranih začetnih in končnih kotov, preverjamo, da uporabnik ob potisku navzgor ne zravna rok preveč oziroma da rok ob spustu ne spusti prenizko. V primeru, da potiska preveč navzgor ali spušča prenizko, se uporabniku sporoči ustrezna povratna informacija.
- Če uporabnik izvaja vadbo pravilno in stoji pri miru, se y-koordinata ramen med izvedbo vaje ne bi smela bistveno spreminjati. Zato preverjamo spremembe y-koordinate med zaporednimi sličicami. Če je sprememba večja od določene meje, sklepamo, da si uporabnik pri potisku pomaga tako, da se dvigne na prste in s tem pridobi dodatno moč za dvig. V tem primeru ga opozorimo.
- Med izvedbo vaje primerjamo kota obeh rok. Pri pravilni izvedbi se kota ne bi smela bistveno razlikovati, saj uporabnik uteži nad glavo potiska hkrati in usklajeno z obema rokama. Če odstopanje presega določeno mejo, uporabniku podamo povratno informacijo, da mora roke dvigovati bolj usklajeno.

Na podoben način določimo ključne točke in pravila tudi za preostale tipe vadb.

5 Rezultati in ugotovitve

Ena najpomembnejših lastnosti aplikacije za pomoč pri telesni vadbi je njena zadostna pretočnost oziroma delovanje v realnem času med izvajanjem vadbe. Zato izmerimo število sličic na sekundo (FPS) na različnih napravah in strojni opremi med vadbo. Rezultate prikazuje Tabela 1.

Tabela 1: Povprečno število sličic na sekundo (FPS), na različni strojni opremi in različnih napravah.

Naprava	Strojna oprema	FPS
PC	CPU	5
PC	CPU z WebAssembly	20
PC	Intel HD Graphics GPU	15
PC	NVIDIA GeForce RTX 2060 GPU	80
iOS pametni telefon	CPU z WebAssembly	20
Android pametni telefon	CPU z WebAssembly	15
Android pametni telefon	GPU	8

Vsi primeri, kjer se uporablja grafična procesna enota, temeljijo na WebGPU tehnologiji. Izvajanje s pomočjo WebGPU tehnologije je sicer možno na kakršnih koli grafičnih procesnih enotah. Ni nujno, da gre za ločeno grafično procesno enoto. Tudi integrirane grafične procesne enote, kakršne najdemo v prenosnih računalnikih so podprte. Prav tako tudi tiste, ki so integrirane v modernih pametnih telefonih. WebGPU je v resnici JavaScript API, ki omogoča brskalnikom, da na učinkovit način uporabljajo grafično procesno enoto odjemalca. WebGPU naj bi bil naraven naslednik precej bolj znane tehnologije WebGL saj zagotavlja boljšo združljivost z novjšimi grafičnimi karticami. Ker pa gre tukaj za novejšo tehnologijo, je podpora za WebGPU omejena tudi glede na (novejše) verzije brskalnikov, ki jih uporabljamo. Na spletni strani (Can I use WebGPU?, 2024) lahko preverimo natančno kakšna je trenutno podpora za WebGPU, v različnih brskalnikih. Chrome, Edge in tudi Opera podpirajo WebGPU z verzijami od začetka leta 2023 naprej, medtem ko je podpora omejena na brskalniku Firefox in Safari. Tudi uporaba WebAssembly je omejena z določenimi verzijami brskalnikov. Je pa ta podpora že precej bolj široka kot pa tista za WebGPU, kar lahko preverimo na spletni strani (Can I use WebAssembly?, 2024). Nižje vrednosti dosega le v primeru uporabe centralne procesne enote brez WebAssembly. Tukaj gre za primer, ko bi inicializacija modela padla tako za grafično procesno enoto kot tudi za centralno s pomočjo WebAssembly. Do tega bi lahko prišlo v primeru, da bi WebGPU neuspešno poskušal dostopati do grafične procesne enote uporabnika oziroma v primeru, ko brskalnik ne bi podpiral niti WebGPU niti WebAssembly. V takšni situaciji se model potem izvaja povsem brez dodatnih izboljšav na CPU. Gre za redek primer, do katerega v splošnem niti ne pride. Sicer pa pričakovano najvišje vrednosti FPS, aplikacija dosega na grafični procesni enoti višjega standarda, kot je omenjena NVIDIA GeForce RTX 2060, kjer je vrednost FPS v povprečju okrog 80 slik na sekundo.

Kljub dobremu rezultatu delovanja aplikacije Sweatpose, pa ima le-ta še vedno nekaj pomanjkljivosti oziroma prostora za napredek in izboljšave. Glavna pomanjkljivost je v načinu kako model MoveNet izvaja ocenjevanje ključnih točk človeškega telesa. Ker smo želeli podpreti ocenjevanje človeške drže za več oseb hkrati, smo izbrali MoveNet model, ki pa v tem primeru izračunava in rezultate podaja v 2D koordinatah, kar so v resnici x in y koordinate pikslov na vsaki sličici zajema iz kamere. Tukaj se potem izgubi globina oziroma oddaljenost oseb od kamere. To se je ob definiranju pravil za izvajanje vadb, kjer smo upoštevali na primer razlike med koordinatami in podobno, izkazalo za pomembno pomanjkljivost. Pravilna postavitev pred kamero je zaradi pomanjkanja 3D koordinat v prostoru ključnega

pomena, prav tako lahko učinkovitost ocenjevanja točnosti izvedbe vadb malce varira glede na oddaljenost osebe, ki je postavljena pred kamero. Izboljšave bi se lahko lotili tako, da bi vzeli model, ki podpira 3D koordinate za posameznika potem pa bi z dodatnim modelom za detekcijo objektov in oseb na primer YOLO (Joseph Redmon, 2016) detektirali posamezne osebe in na vsaki posamezni posebej pognali model za 3D ocenjevanje ključnih točk človeškega telesa. To pa bi gotovo potegnilo za seboj drug problem, ker bi bilo težje takšno zadevo izpeljati v realnem času.

6 Zaključek

V delu predstavimo kaj je ocenjevanje človeške drže s pomočjo metod računalniškega vida in v kakšne namene ga uporabljamo. Izvedemo analizo različnih pristopov (klasični, globoko učenje, od spodaj navzgor, od zgoraj navzdol) in modelov za ocenjevanje človeške drže (model na osnovi ogrodja, model na osnovi obrisa in model na osnovi voluma). Za iskanje ključnih točk človeškega telesa, je najbolj uporaben model na osnovi ogrodja, ker je ravno ta sestavljen iz množice ključnih točk (po navadi sklepov) človeškega telesa, ki se lahko potem uporabljajo v 2D ali 3D predstavitevah.

Na podlagi analize in pridobljenega znanja izberemo primeren model za implementacijo aplikacije, ki bo uporabnikom nudila pomoč in usmeritve pri izvedbi fizičnih vadb. Izberemo model MoveNet iz zbirke Tensorflow.js in z uporabo tega razvijemo aplikacijo Sweatpose, ki omogoča izvedbo štirih različnih tipov vadb (počepi, skleca, potisk nad glavo in vzgib bicepsa), na različni strojni opreми (GPU ali CPU) in različnih napravah. Med izvedbo vadbe uporabnik v realnem času dobiva povratne informacije in nasvete o izvedbi vadbe. Izvajanje v realnem času pa tudi izmerimo s pomočjo vrednosti števila sličic na sekundo (FPS), hkrati pa izpostavimo tudi obstoječe pomanjkljivosti in možne izboljšave.

Viri in literatura

- Černi, B. (2024). Uporaba metod računalniškega vida za pomoč pri telesni vadbi, magistrsko delo, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko.
- Can I use WebAssembly?* (2024). Pridobljeno iz <https://caniuse.com/wasm>
- Can I use WebGPU?* (2024). Pridobljeno iz <https://caniuse.com/webgpu>
- Dragomir Anguelov, P. S. (2005). SCAPE: Shape Completion and Animation of People. *ACM Trans*, 408-416. doi:<https://dl.acm.org/doi/10.1145/1073204.1073207>
- Felzenszwalb, P. H. (2005). Pictorial Structures for Object Recognition. *International Journal of Computer Vision*, 55-79. doi:<https://doi.org/10.1023/B:VISI.0000042934.15159.49>

- G. Rogez, J. R. (2008). Randomized trees for human pose detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 1-8. doi:10.1109/CVPR.2008.4587617
- Hao-Shu Fang, J. L. (2023). AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7157-7173. doi:10.1109/TPAMI.2022.3222784
- Joseph Redmon, S. D. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2016.91
- K. He, G. G. (2017). Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961-2969.
- LeViet, K., & Chen, Y.-h. (2021). Pose estimation and classification on edge devices with MoveNet and TensorFlow Lite. *TensorFlow Blog*. Pridobljeno iz tensorflow.org: https://www.tensorflow.org/lite/tutorials/pose_classification
- Matthew Loper, N. M.-M. (2015). SMPL: a skinned multi-person linear model. *ACM Trans*, 16. doi:<https://doi.org/10.1145/2816795.281801>
- T.F. Cootes, C. T. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, 61(1), 38-59. doi:<https://doi.org/10.1006/cviu.1995.1004>
- Tsung-Yi Lin, M. M. (2014). Microsoft COCO: Common Objects in Context. (str. 740-755). Springer International Publishing.
- WebGPU. (2024). Pridobljeno iz <https://www.w3.org/TR/webgpu/>
- Yucheng Chen, Y. T. (2020). Monocular human pose estimation: A survey of deep learning-based methods. *Computer Vision and Image Understanding*. doi:<https://doi.org/10.1016/j.cviu.2019.102897>
- Z. Cao, T. S.-E. (2017). Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. *IEEE Conference on Computer Vision and Pattern Recognition*, 1302-1310. doi:10.1109/CVPR.2017.143