Razvoj sistema zasnovanog na prepoznavanju gestova ruku za interakciju sa video igrama

Aleksandar Živković[1], dr Nemanja Ilić[2]

***Sadržaj* — U ovom radu razmatra se primena prepoznavanja gestova ruku za interakciju sa igrama, sa posebnim fokusom na pucačke igre iz perspektive prvog lica i igre koje koriste svetlosne pištolje. Predstavljen je sistem koji koristi softversku biblioteku MediaPipe za detekciju karakterističnih tačaka šake, dok se prepoznavanje gestova ostvaruje primenom tehnika dubokog učenja. U završnom delu rada sprovedeno je empirijsko testiranje sistema u različitim scenarijima primene, čiji rezultati ukazuju na potencijalnu primenu ovakvog pristupa u budućim sistemima za upravljanje video igrama.**

***Ključne reči* — Duboko učenje, Gestovi ruku, MediaPipe, Računarski vid, Svetlosni pištolj, Video igre iz perspektive prvog lica**

I. UVOD

SAVREMENI razvoj tehnologije sve više utiče na unapređenje interfejsa čovek–računar, što se ogleda u sve realističnijim i neposrednijim oblicima interakcije između korisnika i digitalnog prostora. Još početkom dvehiljaditih, pojavili su se komercijalni uređaji sposobni za praćenje pokreta korisnika u realnom vremenu. Jedan od takvih primera bio je *Wii Remote*, kontroler kompanije Nintendo, koji je zahvaljujući svojim senzorima stekao primenu u velikom broju različitih sistema, uključujući upravljanjem virtuelnim okruženjima, primenu u medicini u okviru rehabilitacije i različitim industrijskim procesima [1][2][3][4].

U međuvremenu, napredak u oblastima veštačke inteligencije i računarskog vida doveo je do razvoja softverskih rešenja koje prepoznaju gestove bez potrebe za dodatnim hardverom. MediaPipe je softverska biblioteka, koju je razvila kompanija Google, namenjena praćenju i analizi određenih tačaka na

[1] Aleksandar Živković, Računarski fakultet, Beograd, Srbija (email: azivkovic1723m@raf.rs)
[2] Dr Nemanja Ilić, Računarski fakultet, Beograd, Srbija (email: nilic@raf.rs).

telu u realnom vremenu. Zahvaljujući širokom potencijalu primene, MediaPipe je pronašla mesto u razvoju interaktivnih sistema u kojima se fizički gestovi korisnika mapiraju na odgovarajuće komande, uključujući i alternativne modalitete upravljanja u video igrama koji se oslanjaju na praćenje pokreta tela [5]. Na internetu se mogu naći projekti koji koriste detekciju pokreta ruku kao metod za upravljanje igrama, kao što je *Fruit Ninja*, dok je u *World of Warcraft* testiran alternativni oblik kontrolisanja pomoću kamere i naprednih algoritama za softversko prepoznavanje pokreta [6]. Iskustvo sa Nintendo Wii konzolom, koja je i dan danas vrlo aktuelna, pokazalo je da umanjivanjem distance interfejsa čovek–računar igrač stiče verodostojniji osećaj imerzije u igri. Kada pokreti tela postanu sastavni deo upravljanja, interakcija deluje neposrednije i prirodnije. Na taj način se briše oštra granica između korisnika i digitalnog okruženja, što doprinosi tome da igra deluje zabavnije, a doživljaj ubedljivijim.

Oslanjajući se na prethodno spomenute primere, ovaj rad ima za cilj razvoj prototipa sistema koji koristi gestove ruku za upravljanje video igrama, sa posebnim naglaskom na igre iz perspektive prvog lica. Pošto ovaj žanr već stavlja igrača u centar zbivanja, dodavanje fizičkih pokreta kao model upravljanja doprinosi tome da igrač ima osećaj da je stvarno deo onoga što se dešava na ekranu. Konkretno, rad se fokusira na igre pucačkog tipa, uključujući one koje koriste svetlosne pištolje. Za obradu gestova, pomoću biblioteke MediaPipe prikupljaju se sekvence karakterističnih tačaka ruku, koje se zatim analiziraju pomoću posebno konstruisanih modela predikcije za levu i desnu ruku, zasnovanih na tehnikama dubokog učenja. Prepoznati gestovi se dalje prevode u odgovarajuće komande unutar video igre simulacijom pritisaka tastera i pokreta miša.

II. VIDEO IGRE IZ PERSPEKTIVE PRVOG LICA

Za video igre iz perspektive prvog lica karakteristično je da igrač posmatra virtuelno okruženje iz ugla lika kojim upravlja. Igre ovog tipa obuhvataju različite žanrove, među kojima se ističu akcione igre pucačkog tipa, avanture i igre u kojima se razvija narativ kroz uloge (RPG). Zbog njihove popularnosti, dostupne su na velikom broju platformi, uključujući konzole, računare pa čak i mobilne uređaje. Kontrola se najčešće ostvaruje putem tastature i miša, kontrolera ili specijalizovanih uređaja poput svetlosnih pištolja.

A. Upotreba svetlosnog pištolja u video igrama

Igre koje koriste svetlosni pištolj (engl. *light gun*) predstavljaju posebnu grupu pucačkih igara u kojima se ciljanje i „pucanje” ostvaruje direktno

pomoću fizičkog uređaja. Interakcija se postiže primenom ručnog uređaja čiji izgled podseća nalik pištolju pa je i po njemu dobio ime svetlosni pištolj. Ovakve igre bile su posebno popularne u arkadnim salama i kućnim konzolama tokom devedesetih, ali i danas privlače pažnju, naročito među entuzijastima i ljubiteljima klasičnih igara. Iako znatno ređe, moguće ih je igrati i na računarima, najčešće putem emulatora ili kao prilagođene računarske verzije poznatih naslova poput *Duck Hunt* ili *House of the Dead* [7].

Način funkcionisanja svetlosnih pištolja menjao se i usavršavao u skladu sa razvojem tehnologije. Generalno postoji 4 osnovna načina njihovog rada:

- **Sekvencijalno osvetljavanje meta** - Detekcija pogodaka zasniva se na zatamnjenju ekrana i pojedinačnom osvetljavanju meta dok senzor u pištolju registruje prelaz sa tamnog na svetli deo slike [8].

- **Infracrvene kamere i emiteri** - Pištolj sadrži infracrvenu kameru koja detektuje pozicije svetlosnih izvora (LED dioda) postavljenih oko ekrana i na osnovu toga izračunava usmerenje uređaja [9].

- **Nišanjenje putem obrade slike -** Kamera ugrađena u pištolj prepoznaje konture ekrana, dok softver obrađuje te informacije i izračunava tačnu poziciju pokazivača u odnosu na ekran [10].

- **Analogno nišanjenje -** Pozicija pištolja određuje se merenjem otpora potenciometara povezanih sa mehaničkim osama, čime se registruje pomeraj nišana na ekranu [11].

B. Razvoj pucačkih igara iz perspektive prvog lica

Pucačke igre iz perspektive prvog lica (engl. *First-Person Shooter*, FPS) predstavljaju podžanr akcionih video igara u kojima igrač posmatra svet kroz perspektivu lika kojim upravlja, koristeći vatreno oružje kao osnovni način interakcije [12]. Prema globalnom istraživanju iz 2024. godine, FPS igre bile su drugi najpopularniji žanr, sa 51.9% ispitanika koji su naveli da su igrali igre ovog tipa tokom prethodne godine [13]. U ranim fazama razvoja računarskih igara dominirale su spore strateške i RPG igre sa postepenim napretkom kroz otkrivanje sveta. Prvi poznati koncept FPS igranja pojavio se sa igrom *Maze War*, razvijenom za NASA računare, koja je koristila jednostavne vektorske prikaze prostora. Ipak, igra koja je postavila osnovne mehanizme žanra i utemeljila FPS kao zaseban oblik jeste *Wolfenstein 3D* [14]. Za razliku od prethodnih igara koje su se oslanjale na izometrijsku ili bočnu perspektivu, *Wolfenstein 3D* omogućio je igračima da dožive svet iz perspektive glavnog

lika koji se slobodno kretao kroz zatvorene prostore u realnom vremenu, koristeći tehniku *raycasting-a* za stvaranje iluzije trodimenzionalnog prostora.

Veliki uticaj na dalju popularizaciju žanra imala je igra *Doom,* uvodeći revolucionarne inovacije u oblasti grafike, zvuka, mehanike igranja i prostorne navigacije [15]. Upotreba pseudo trodimenzionalnog prikaza, mogućnost učešća više igrača (engl. *multiplayer*), kao i modifikacije prema recenzijama korisnika, značajno su uticale na dalji razvoj žanra. Nivoi su bili zasnovani na dvodimenzionalnim mapama podeljenim na sektore sa definisanim visinama poda i plafona, dok je iluzija prostorne dubine bila postignuta tehnikama teksturalne projekcije i *backface culling-a*. Za optimizaciju prikaza korišćen je *Binary Space Partitioning* algoritam pomoću kojeg bi se prilikom čuvanja mape nivoa prethodno ona podelila na konveksne regione s ciljem bržeg određivanja vidljivih površina tokom igranja.

Igra *Quake* bila je prvi naslov koji je koristio pravu trodimenzionalnu geometriju za prikaz svih elemenata unutar igre, uključujući likove, oružja i celokupno okruženje [16]. Nivoi su bili prikazani u realnom vremenu pomoću unapred pripremljenih struktura koje su definisale granice vidljivih oblasti, dok je tehnika potencijalno vidljivih oblasti (engl. *potentially visible set*) korišćena za izračunavanje koje delove mape treba prikazati u određenom trenutku. Osvetljenje je rešeno kombinacijom unapred izračunatih svetlosnih mapa koje definišu statičke izvore svetla i dinamičkih efekata u pokretu. U domenu mrežnog igranja, *Quake* je uveo klijent-server arhitekturu, što je kasnije dodatno unapređeno verzijom *QuakeWorld*.

Nakon igara *Doom* i *Quake*, FPS žanr proširio se raznim inovacijama. Igra *Crysis* podigla je standarde realističnog prikaza okruženja, dok je *Call of Duty 4: Modern Warfare* uvela fokus na narativne misije i iscenirane događaje. *Half-Life 2* spojila je naraciju i igranje bez prekida, a *Mirror's Edge* je skrenla pažnju na kretanje kroz prostor umesto na borbu. Tokom daljeg razvoja, FPS igre su se podelile na više podpravaca. Neke su nastavile putem taktičke igre, kao što je igra *Counter-Strike* (CS), dok su se druge fokusirale na bržem tempu igranja i izraženijoj akciji poput naslova *Doom Eternal*. Takođe, sve veći broj naslova počeo je da uvodi dodatne elemente igranja. Igra *Titanfall* je kombinovala pucanje sa kretanjem po zidovima i upotrebom mehaničkih dodataka za telo, poput egzoskeleta. *Overwatch* i *Valorant* su, za razliku od uobičajenih taktičkih FPS igara u formatu 5 na 5 poput CS, uveli dodatnu stratešku složenost kroz likove sa posebnim ulogama. Danas FPS žanr predstavlja jedan od ustaljenih oblika igranja, a njegova sposobnost da integriše različite pristupe čini ga jednim od najraznovrsnijih i najprisutnijih savremenih formi u domenu video igara.

III. RAČUNARSKI VID

Računarski vid je oblast veštačke inteligencije koja se bavi razvojem tehnika kojima računari tumače i razumeju vizuelne informacije. Dok čovek lako prepoznaje boje, oblike i promene u svetlu, računarski vid se suočava sa brojnim preprekama zbog raznolikosti vizuelnih informacija. Vremenom se ovaj domen razvio iz potreba za automatskom analizom slike u različitim oblastima kao što su medicina, robotika i automatizovana vozila. Početna istraživanja bila su usmerena na jednostavnije zadatke poput detekcije ivica i osnovnih oblika, dok je razvoj računarske tehnologije i algoritamskih pristupa znatno proširio spektar primene ove oblasti. Mašinsko učenje, a naročito duboke neuronske mreže, danas predstavljaju osnovu savremenih sistema računarskog vida. Upotrebom ovih modela značajno su poboljšane mogućnosti sistema da prepoznaju složene strukture i obrasce u slikama i video zapisima [17].

A. Konvolucione neuronske mreže

Konvolucione neuronske mreže (engl. *Convolutional Neural Networks*, CNN) predstavljaju posebnu klasu neuronskih arhitektura osmišljenih za rad sa podacima koji sadrže lokalne prostorne ili vremenske korelacije, kao što su slike ili sekvence signala [17]. Osnovna operacija koja se koristi u ovim mrežama jeste konvolucija, matematički definisana kao suma proizvoda između vrednosti ulaznog signala i težinskih koeficijenata filtera.

Konvolucija se primenjuje tako što se filter pomera preko ulaznog podatka određenim korakom (engl. *stride*), pri čemu se na svakoj poziciji računa suma proizvoda između elemenata filtera i odgovarajućih vrednosti ulaza, kojoj se dodaje *bias* i prosleđuje kroz aktivacionu funkciju. Rezultat ove operacije je mapa karakteristika (engl. *feature map*) koja ističe specifične obrasce iz ulaza, poput ivica ili tekstura [17]. Primenom više različitih filtera nad istim ulazom, model istovremeno izdvaja različite osobine signala, jer svaki filter reaguje na drugačiji obrazac. Da bi se smanjila osetljivost na manje pomeraje i redukovala količina podataka, između konvolucionih slojeva često se primenjuje agregacija (engl. *pooling*), najčešće u vidu *max-pooling-a* ili *average-pooling-a*, pri čemu se iz svakog lokalnog regiona izdvaja najveća ili prosečna vrednost. Dobijene mape karakteristika niske rezolucije mogu se zatim reorganizovati u vektorski oblik (engl. *vectorization*) i proslediti potpuno povezanim slojevima koji na osnovu naučenih obrazaca vrše konačnu klasifikaciju [17].

CNN mreže se mogu konstruisati kao jednodimenzionalne, kada se koriste nad vremenskim signalima, ili dvodimenzionalne, što je slučaj kod slike. U

ovom radu koristi se jednodimenzionalna konvoluciona mreža, ali principi konstruisanja i treniranja ostaju slični kao kod dvodimenzionalnih arhitektura.

B. Rekurentne neuronske mreže sa dugoročnim pamćenjem

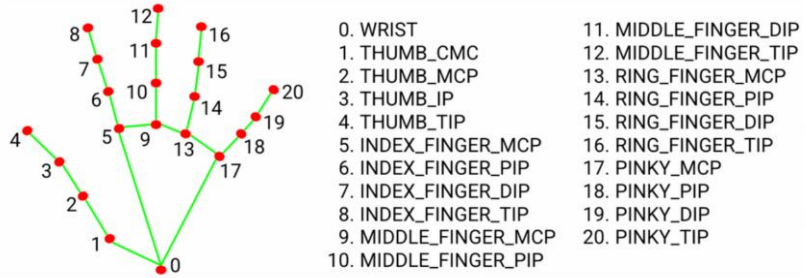
Rekurentne neuronske mreže (engl. *Recurrent Neural Networks*, RNN) koriste se za obradu podataka koji se javljaju kao nizovi, kao što su vremenske serije, govorni signali i tekst. Za razliku standardnih neuronskih mreža, RNN modeli sadrže povratne veze koje omogućavaju korišćenje informacija iz prethodnih koraka prilikom određivanja trenutnog izlaza, što ih čini pogodnim za modeliranje sekvencijalnih zavisnosti [18]. Ipak, kod jednostavnih RNN modela često dolazi do problema pri učenju dugoročnih zavisnosti, zbog pojave eksplodirajućih ili nestajućih gradijenata [18]. Kao rešenje tog ograničenja uveden je model sa dugoročnim pamćenjem (engl. *Long Short-Term Memory*, LSTM), koji se oslanja na specifičnu ćelijsku strukturu sa internim stanjem i više kontrolnih kapija. Ulazna, izlazna i kapija zaborava upravljaju protokom informacija kroz ćeliju odlučujući koje informacije se zadržavaju, ažuriraju ili odbacuju u svakom vremenskom koraku [18]. U okviru LSTM ćelije nalazi se i unutrašnje stanje (engl. *cell state*), koje se ažurira korišćenjem Hadamardovog množenja izlaza kapija sa odgovarajućim signalima. Ulazni podaci zajedno sa prethodnim informacijama iz sloja mreže koriste se za izračunavanje nove vrednosti stanja i izlaza, pri čemu se koriste nelinearne funkcije poput sigmoida i hiperboličkog tangensa, radi održavanja stabilanog prenosa gradijenata kroz više vremenskih koraka [18].

U poređenju sa standardnim RNN-ovima, LSTM pokazuje bolje rezultate pri učenju nad sekvencama koje zahtevaju memorisanje udaljenih konteksta. Zahvaljujući sposobnosti da obradi duže vremenske kontekste, ovaj model nalazi primenu u brojnim zadacima računarskog vida.

C. Detekcija karakterističnih tačaka ruku

Sa razvojem dubokog učenja, naročito konvolucionih neuronskih mreža, tradicionalni pristupi u računarskom vidu postepeno su zamenjeni novim modelima zasnovanim na dubokim mrežama, naročito u zadacima koji zahtevaju visoku tačnost i obradu u realnom vremenu. Jedan od takvih ranijih modela koji je napravio značajan iskorak ka detekciji objekata u realnom vremenu jeste YOLO (engl. *You Only Look Once*), koji je svojom arhitekturom i pristupom postavio osnove za kasnije sisteme [19]. U okviru ovog rada, za potrebe precizne detekcije i praćenja pokreta šake korišćena je softverska biblioteka MediaPipe, razvijena od strane kompanije Google [5]. Ova biblioteka otvorenog koda zasniva se na prethodno treniranim modelima

neuronskih mreža organizovanim u modularnu arhitekturu, namenjenu primeni u realnom vremenu. Detekcija položaja ruku u MediaPipe sistemu sastoji se iz dva osnovna koraka. U prvom koraku primenjuje se zaseban modul koji identifikuje prisustvo dlana i definiše region od interesa za detaljniju analizu. Nakon određivanja regiona, aktivira se sledeći modul koji detektuje ukupno 21 karakterističnu tačku na šaci, prikazano na Slici 1 [5]. Svaka od tih tačaka odgovara specifičnim anatomskim pozicijama kao što su zglobovi prstiju, centar dlana i vrhovi prstiju, pri čemu svaka tačka ima definisane koordinate i dubinu.



Sl. 1. Prikazuje sistem indeksiranja karakterističnih tačaka šake koji koristi MediaPipe biblioteka. Tačka sa oznakom 0 odgovara ručnom zglobu, dok se ostalih 20 tačaka raspoređuje duž svakog prsta, prateći njihovu anatomsku strukturu od korena do vrha. Izvor: Google Developers, MediaPipe Hands: Hand Landmarker.

Zahvaljujući svojoj modularnoj arhitekturi i optimizovanim modelima neuronskih mreža, MediaPipe dostiže visoke performanse u pogledu brzine i tačnosti detekcije. Te osobine čine ovaj pristup posebno pogodnim za primenu u zadacima analize pokreta i prepoznavanja gestova, što je upravo cilj ovog rada. Koordinate karakterističnih tačaka dobijenih MediaPipe-om koriste se kao ulazni podaci u sledećim fazama analize, u kojima rekurentne neuronske mreže vrše prepoznavanje vremenskih obrazaca i klasifikaciju gestova.

IV. Implementacija

Implementaciju sistema čine dva osnovna segmenta. Prvi obuhvata razvoj dva odvojena modela za prepoznavanje gestova, jednog za levu i jednog za desnu ruku. Iako koriste istu arhitekturu, modeli su obučeni nad različitim

skupovima podataka kako bi odgovarali specifičnim gestovima pokreta svake ruke. Drugi se odnosi na izgradnju arhitekture sistema koja povezuje izlaze modela sa odgovarajućim komandama u igri.

A. Model

Detekcija akcija u video zapisima često zahteva analizu više uzastopnih frejmova kako bi se ispravno prepoznala trenutna radnja. Na primer, razlikovanje između početka i završetka skoka može biti teško ako se posmatra samo jedan frejm u kome je osoba u vazduhu. Bez šireg konteksta, nije moguće sa sigurnošću utvrditi o kojoj se fazi radnje radi, što je prikazano na Slici 2. Iz tog razloga, gestovi u ovom sistemu se analiziraju u okviru sekvence frejmova, čija je dužina fiksirana na 30. Jedan od mogućih pristupa obradi takvih sekvenci jeste primena konvolucione neuronske mreže nad slikama frejma. Međutim, ovakvi modeli zahtevaju veliki skup podataka za treniranje, budući da je potrebno generalizovati prepoznavanje gestova u različitim uslovima snimanja. Osim toga, obuka ovih modela može biti vremenski zahtevna zbog složenosti ulaznih podataka.

Sl. 2. Ilustrativni primer problema prepoznavanja akcije na osnovu pojedinačnog frejma. Na slici B prikazan je trenutak kada je osoba u vazduhu, što može biti deo kako početne tako i završne faze skoka (slike A i C). Bez dodatnog konteksta iz prethodnih i narednih frejmova, teško je zaključiti o kom delu akcije se radi. Izvor: Guo, K., Ishwar, P. and Konrad, J. *Action recognition using log-covariance matrices.* (2013).

Drugi pristup, primenjen u ovom radu, zasniva se na analizi karakterističnih tačaka kroz sekvencu frejmova. Ovim pristupom problem se sa analize sirovih slika transformiše u analizu niza vektora koji predstavljaju pozicije karakterističnih tačaka u prostoru. Na taj način dimenzionalnost ulaznih podataka se značajno smanjuje, a sama priroda problema bliža je klasičnim zadacima prepoznavanja šablona u okviru dubokog učenja. Modeli koji koriste ovakvu reprezentaciju često zahtevaju manji skup podataka za obuku, a pri

čemu ipak postižu zadovoljavajuće performanse u identifikaciji različitih akcija.

B. Gestovi modela

Gestovi su definisani kao pozicije karakterističnih tačaka u prostoru koje opisuju oblik šake pri izvođenju pokreta. Zbog specifičnosti teme, korišćenje postojećih skupova podataka nije bilo izvodljivo. Kako svaki od modela prepoznaje različite gestove, bilo je potrebno snimiti poseban video zapis za svaki od njih. Tokom eksperimentalne faze razmatrana je i mogućnost uvođenja kompleksnijih gestova koji bi uključivali koordinisanu upotrebu obe ruke. Kao jedan od primera analiziran je pokret pri kojem se desna ruka formira u položaj pištolja[3], dok se leva postavlja ispod ili preko nje, čime bi se simulirala radnja ponovnog punjenja oružja. Međutim, takvi gestovi nisu dali zadovoljavajuće rezultate u praksi. Komponenta za detekciju karakterističnih tačaka ruku nailazila je na poteškoće u pravilnom prepoznavanju položaja i orijentacije šake u trenucima kada dolazilo do njihovog preklapanja ili kada njihova silueta nije bila u potpunosti vidljiva. Kao jedan od mogućih uzroka identifikovan je slab kvalitet kamere, što je dodatno otežavalo stabilno praćenje. Zbog toga se od implementacije dvoručnih gestova u ovom prototipu odustalo, a sistem je prilagođen tako da koristi isključivo jednostavne pokrete koji zahtevaju upotrebu samo jedne ruke.

C. Tipovi modela

Većina video igara zahteva istovremeno izvođenje više komandi, poput kombinacije hodanja i skakanja ili pomeranja i pucanja. Radi simulacije takvog ponašanja, u okviru ovog rada korišćena su dva odvojena modela, po jedan za svaku ruku, kako bi se nezavisno prepoznavali različiti gestovi i mogli koristiti paralelno.

Oba modela zasnivaju se na istoj arhitekturi, prikazanoj u Tabeli 1, koja koristi jednostavne LSTM mreže. Razlika među njima ogleda se u broju gestova koje svaki model klasifikuje.

TABELA 1: PRIKAZ STRUKTURE MODELA KOJI SE KORISTI ZA ANALIZU POKRETA DESNE RUKE.

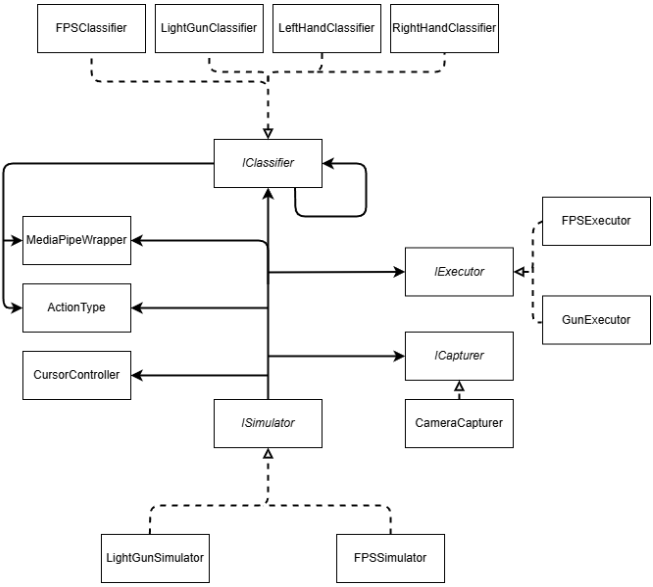
Type	Number of weights	Output shape
Convolution1D(kernel = 5)	64	26×64

[3] Položaj pištolja - kažiprst je ispružen, palac usmeren nagore, a ostali prsti savijeni ka dlanu.

LSTM	128	26 × 128
DropOut	-	26 × 128
LSTM	64	64
BatchNormalization	-	64
Dense	32	32
Dense	3	3

D.Arhitektura sistema

Arhitektura sistema osmišljena je tako da podrži dva režima rada: *FPSSimulator* i *LightGunSimulator*. Dijagram arhitekture sistema prikazan je na Slici 3. Režim *FPSSimulator* namenjen je standardnim igrama koje se kontrolišu tastaturom i mišem, pri čemu se analiziraju gestovi obe ruke u skladu sa opisom iz prethodne sekcije. U režimu *LightGunSimulator* desna ruka koristi se za nišanjenje pomoću ispruženog kažiprsta, dok leva služi za komande kao što su ponovno punjenje i ispaljivanje. Takođe, u okviru *FPSSimulator* režima može se dodatno konfigurisati da se nišanjenje desnom rukom isključi, u zavisnosti od potreba konkretne igre. Komponenta *MediaPipeWrapper* objedinjuje primenu Mediapipe biblioteke kao što je detekcija ruke i izdvajanje karakterističnih tačaka. U zavisnosti od režima, koristi se odgovarajuća implementacija *IClassifier-a*, *FPSClassifier* ili *LightGunClassifier*, koje interno pozivaju prilikom analize gesta klasifikator modela leve ruke *LeftHandClassifier* i klasifikator modela desne ruke *RightHandClassifier*. Modul *ActionType* mapira rezultate klasifikacije na odgovarajuće akcije pritisaka tastera i pokreta miša, koje se dalje prosleđuju komponenti *IExecutor*, radi njihovog izvršavanja. U zavisnosti od režima, koristi se *FPSExecutor* ili *GunExecutor*. Za dohvatanje video signala koristi se *ICapturer*, dok konkretna implementacija *CameraCapturer* preuzima snimak sa interne kamere. Ceo proces koordinisan je komponentom *ISimulator* koja u skladu sa režimom upravlja simulacijom korisničkog unosa u igri.



Sl. 3. Dijagram arhitekture sistema

E. Testiranje sistema

Testiranje sistema sprovedeno je sa pet ispitanika, sa raznolikim prethodnim iskustvom u domenu video igara. Ispitanicima je dodeljen zadatak da, isključivo primenom gestova ruku pomoću razvijenog sistema, dovrše prvi nivo u izabranim video igrama u optimalnom okruženju (jednobojna pozadina, blago sobno osvetljenje). U režimu za standardne FPS igre korišćena je igra *Quake*, dok je za režim svetlosnog pištolja korišćena je igra *Grand Shooter*.

U okviru *Quake-a*, testiran je opšti režim rada sistema. Prvi nivo igre, pod nazivom *The Slipgate Complex*, odvija se u linearno strukturisanom okruženju u kojem se igrač kreće, eliminiše neprijateljske agente i aktivira mehanizme kako bi stigao do portala koji označava kraj nivoa. U ovom režimu, ispitanici su koristili gestove obe ruke za interakciju sa igrom. Tokom testiranja, samo jedan ispitanik je uspešno završio nivo, dok je drugi bio blizu kraja. Iako ovaj režim upravljanja ocenjen kao intuitivan, neki ispitanici su kao potencijalni nedostatak naveli umor ruku zbog potrebe da šake stalno budu u vidokrugu kamere tokom dužih sesija igranja.

Scenario u igri *Grand Shooter* zahtevao je od ispitanika da reaguju na nasumične napade talasa neprijateljskih agenata koji se pojavljuju sa svih

strana ekrana. Ukupno se pojavljuje 15 neprijateljskih agenata raspoređenih u više faza, a igrač na raspolaganju ima tri života. U ovom režimu, desni kažiprst korišćen je za nišanjenje, dok su pokreti leve ruke služili za pucanje i ponovno punjenje oružja. Svi ispitanici uspešno su završili nivo. Opšti utisak bio je pozitivan, ali više njih je ukazalo na potrebu za intuitivnijim gestovima koji bi bolje dočarali fantaziju rukovanja oružjem, poput pokreta u kojima ruke imitiraju stvarnu upotrebu vatrenog oružja, kao što je dvoručni gest za ponovno punjenje koji je prethodno bio razmatran.

Tokom testiranja zabeleženi su i određeni propusti. Iako je sistem u celini imao dobar odziv, povremeno je dolazilo do slabijeg prepoznavanja gestova, što se može pripisati brzim pokretima, nepreciznim položajima šake ili ograničenjima kamere koja nije posedovala visoku rezoluciju. Takođe, pojedini ispitanici su u početku imali poteškoće pri upravljanju, ali su se, nakon kratkog perioda prilagođavanja, uspešno snašli. Kod igara opšteg FPS tipa, upravljanje gestovima nije moglo da se meri sa efikasnošću klasičnih vidova kontrole poput tastature i miša, posebno u pogledu preciznosti i brzine reakcije. U tom smislu, gestovna kontrola je uglavnom doživljena kao zanimljiv dodatak, a ne kao potpuna zamena postojećim rešenjima. Nasuprot tome, režim koji simulira rad svetlosnog pištolja pokazao se znatno uspešnijim. Zbog sličnosti sa originalnim načinom igranja u igrama koje su koristile svetlosne pištolje, ovaj režim je ocenjen kao prirodniji i realniji, što ga čini pogodnijim za dalje unapređenje i širu primenu. Generalni zaključak učesnika bio je da je interakcija putem gestova zanimljiva i savremena, sa potencijalom za primenu u budućim igrama.

V. Zaključak

U ovom radu je prikazan razvoj sistema zasnovanog na prepoznavanju gestova ruku za interakciju sa video igrama. Sa teorijskog stanovišta, obrađen je način rada svetlosnih pištolja i analiziran je istorijski razvoj žanra pucačkih igara iz perspektive prvog lica, sa akcentom na tehničke i konceptualne napretke koji su uticali na njihovu popularnost i širenje, naročito u kontekstu kompleksnih okruženja i narativne strukture. Razmotrene su i metode računarskog vida, sa fokusom modele dubokog učenja i posebnim osvrtom na prepoznavanju gestova. U praktičnom delu rada razvijen je sistem koji koristi gestove ruku za upravljanje u video igrama, sa fokusom na režime koji imitiraju svetlosne pištolje i standardne FPS igre.

Tokom testiranja ispitanici su pozitivno ocenili gestovni način upravljanja, naročito u kontekstu režima svetlosnog pištolja, dok je kod opštih FPS igara uočen manjak preciznosti u poređenju sa klasičnim metodama kontrole.

Dodatno, više ispitanika izrazilo je želju za gestovima koji bi bolje dočarali doživljaj rukovanja oružjem, poput dvoručnih pokreta koji su prethodno razmatrani. Međutim, tehnička ograničenja pri detekciji takvih gestova, naročito u situacijama kada šake nisu jasno vidljive, dovela su do toga da se prototip bazira isključivo na jednostavnim jednoručnim komandama. Iako gestovna kontrola ne može u potpunosti zameniti klasične metode upravljanja u svim žanrovima, pokazala se kao prirodan i uverljiv oblik kontrole u simulaciji svetlosnog pištolja.

Dalji razvoj sistema trebalo bi usmeriti upravo ka tom režimu, kroz poboljšanje detekcije gestova, u vidu kombinovane analize karakterističnih tačaka i vizuelnog sadržaja slike prilikom interpretacije gesta, korišćenje bolje opreme i definisanje pokreta koji deluju prirodnije korisnicima.

LITERATURA

[1] Y. W. Chow, "3D spatial interaction with the Wii remote for head-mounted display virtual reality," Proc. World Acad. Sci. Eng. Technol., 2009.

[2] C. Eroğul, G. Tekkaya, and G. Vural, "Using head and finger tracking with Wiimote for Google Earth control," Middle East Technical University, 2007.

[3] M. Nielsen and M. Stenbacka, "Wii Remote Interaction for Industrial Use", M.S. thesis, Mälardalen University, Västerås, Sweden, 2009.

[4] E. Tsekleves, A. Warland, C. Kilbride, I. Paraskevopoulos, and D. Skordoulis, "The use of the Nintendo Wii in motor rehabilitation for virtual reality interventions: a literature review," in Virtual, Augmented Reality and Serious Games for Healthcare 1, Springer, 2014, pp. 321–344.

[5] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, et al., "MediaPipe: A framework for building perception pipelines," arXiv preprint arXiv:1906.08172, 2019.

[6] J. Bischoff, "Running hardcore Molten Core using Motion Capture," Wowhead, 2025. Available: https://www.wowhead.com/classic/news/running-hardcore-molten-core-using-motion-capture-software-the-rise-and-fall-of-372340

[7] S. Dor, "Emulation," in *The Routledge Companion to Video Game Studies*, Routledge, 2014, pp. 49–55.

[8] D. Teger, S. Rogowski, J. Dinerman, and K. Ramkishun, "DuckFeed: An embedded take," Columbia University Computer Science Department, Tech. Rep., 2011.

[9] S. De Amici, A. Sanna, F. Lamberti, and B. Pralio, "A Wii remote-based infrared-optical tracking system," *Entertainment Computing*, vol. 1, no. 3–4, pp. 119–124, 2010.

[10] Sinden Lightgun, "Information about the Sinden Lightgun," 2023. Available: https://www.sindenshop.com/pages/information-about-the-sinden-lightgun

[11] M. J. P. Wolf, *The Video Game Explosion: A History from PONG to PlayStation and Beyond*, 1st ed. Westport, CT, London: Greenwood Press, 2007, pp. 56–65.

[12] G. Voorhees, "Shooting," in *The Routledge Companion to Video Game Studies*, Routledge, 2014, pp. 272–278.

[13] Statista, "Global online games genre breakdown," 2024. Available: https://www.statista.com/statistics/240990/global-online-games-genre-breakdown/

[14] D. Kushner, *Masters of Doom: How Two Guys Created an Empire and Transformed Pop Culture*, 1st ed. New York, NY, USA: Random House, 2004, pp. 83–97
[15] F. Sanglard, *Game Engine Black Book: Doom,* 2019, pp. 155–285.
[16] F. Sanglard, "Quake Source Code Review," 2009. Available: https://fabiensanglard.net/quakeSource/quakeSourceRendition.php
[17] R. C. Gonzalez, *Digital Image Processing*, 4th ed. Boston, MA, USA: Pearson, 2019, pp. 900–993.
[18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MA, USA: MIT Press, 2016, pp. 367–415.
[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," arXiv preprint arXiv:1506.02640, 2016.

ABSTRACT

The topic of this paper is the application of hand gesture recognition for interaction with video games, with particular emphasis on first-person shooter games and the use of light guns. The proposed system employs the MediaPipe software library for detecting hand landmark points, while gesture recognition is performed using deep learning techniques. In the final part of the paper, empirical testing was conducted across various application scenarios, and the results indicate that this approach has potential for future use in video game control systems.

Development of a hand gesture recognition-based system for interaction with video games

Aleksandar Živković, dr Nemanja Ilić