

**МУЛТИМЕДИЈАЛНА РЕПРЕЗЕНТАЦИЈА ЕЛЕМЕНАТА ЗНАКОВНОГ ЈЕЗИКА  
ЗАСНОВАНА НА МАШИНСКОМ УЧЕЊУ****MULTIMEDIA REPRESENTATION OF SIGN LANGUAGE ELEMENTS BASED ON  
MACHINE LEARNING**Марија Варга, Лидија Крстановић, *Факултет техничких наука, Нови Сад***Област – РАЧУНАРСКА ГРАФИКА**

**Кратак садржај** – У раду је реализован систем за мултимедијалну репрезентацију знаковног језика заснован на машинском учењу. Овакав систем представља добар основ за даљи развој система за превођење и посредовање у комуникацији са глувонемим особама. Препознавање покрета руке и разумевање шта направљен покрет значи у реалном времену представља велики изазов у компјутерској визији. Направљен је систем који детектује покрет шаке и препознаје значење начињеног покрета, а након тога покрет преноси на 3Д дигитални модел шаке у реалном времену.

**Кључне речи:** *Машинско учење, компјутерска визија, знаковни језик, рачунарска графика*

**Abstract** – *This paper describes implementation of a system for multimedia representation of sign language elements based on machine learning. This system represents a good basis for further development of a system for translation and mediation in communication with hearing impaired people. Recognizing hand movements and understanding what that movement means in real time is a major challenge in computer vision. A system was created that detects the movement of the hand and interprets the meaning of the movement, and then transfers the movement to a 3D digital model of a hand in real time.*

**Keywords:** *Machine learning, computer vision, sign language, computer graphics*

**1. УВОД**

Способност детекције и препознавања покрета руке може бити врло важна компонента у различитим технолошким апликацијама и платформама као и за потребе дигиталног асистента особама са оштећењем слуха. Направљен је систем који детектује покрет шаке, препознаје значење начињеног покрета, а након тога покрет преноси на 3Д дигитални модел шаке. Систем је обogaћен звуком па ће се након препознавања и разумевања начињеног покрета звучно представити шта покрет значи. Ово је корисно за превођење знаковног језика особама које се служе говорним језиком.

**НАПОМЕНА:**

**Овај рад проистекао је из мастер рада чији ментор је била др Лидија Крстановић, доцент.**

Препознавање знаковног језика помоћу алата који се користе у компјутерској визији је велики изазов због тродимензионалности покрета и положаја руку у односу на тело, као и због мимике лица која је врло важна приликом тумачења знаковног језика. Како би рачунар могао да интерпретира значење покрета потребно је применити алате машинског учења. Преношење знаковног језика на дигитални модел шаке може бити одличан почетак за развој едукативних апликација за децу са одређеним оштећењем слуха, за развој анимираних филмова и креирање базе анимација знакова знаковног језика.

**2. ПРОБЛЕМ ДЕТЕКЦИЈЕ И ПРЕПОЗНАВАЊЕ  
ЗНАКОВНОГ ЈЕЗИКА**

За знаковни језик српског језика не постоји велики број развијених апликација и интерактивних система који олакшавају споразумевање. Овакве апликације и интерактивни системи могу олакшати комуникацију између особа које се служе говорним и знаковним језиком. Значајне би биле и у области технологије јер би представљале додатан вид интеракција са уређајима и корисничким интерфејсом. Знаковни језик јесте језик који чине знакови који се формирају покретима руку и других делова тела, укључујући експресије лица, позицију тела и покрете усана којима се преноси одређена мисао. Важан сегмент у знаковном језику јесте и ручна азбука која може бити једноручна, дворучна или комбинована [1, 2].

Како би знаковни језик био препознат од стране рачунара у реалном времену потребно је извршити одређене операције над видео секвенцом, односно одрадити обраду слике. Неопходно је одрадити детекцију покрета руке и извући све потребне информације, а како би се знак препознао потребно је машинско учење.

**2.1. Детекција покрета**

Детекција покрета је техника у компјутерској визији и у обради слика где се неколико узастопних фрејмова видео секвенце међусобно пореде применом различитих метода како би се одредило да ли је дошло до покрета. Примена сегментације која издваја покретни објекат од стационарних омогућава лакше праћење и анализирање покрета. Постоји много техника за детекцију покрета као последица чињенице да различити методи нису подједнако погодни за различите примере. Не постоји стандардна метрика за процену резултата детекције пошто се користи квалитативна дефиниција добре сегментације по којој

региони добијени поступком сегментације треба да буду униформни и хомогени у односу на друге карактеристике као што су ниво сиве или текстура. Унутрашњост региона од интереса треба да буде једноставна и без малих рупа, а границе региона једноставне, глатке и просторно тачне. Проблем детекције покрета постаје већи када је потребно да се детекција извршава у реалном времену [3].

## 2.2. Обрада слике

Дигитална слика има коначан број елемената слике који се називају пиксели. За сваки пиксел, уређај за обраду слике бележи број или мали скуп бројева који описују неке особине овог пиксела. Вредности су распоређене у матрицу. Може се рећи да се под сликом подразумева било која континуална функција, односно сигнал, по две променљиве (1). Променљиве  $x$ ,  $y$  представљају просторне координате [4].

$$f(x, y) \quad (1)$$

На слици 1 приказана је репрезентација једног дела монохроматске (енг. *grayscale*) слике у виду матрице.



Слика 1. Дигитална слика

Обрада слика се састоји од низа корака у којима се над сликом обављају одређене операције у циљу добијања погодне слике за извлачење корисних информација. Дакле, обрада слика је анализа и манипулација дигиталном сликом у циљу њеног побољшања.

Кораци у детекцији покрета шаке који су имплементирани за потребе овог рада и који обухватају обраду слике су:

1. Одређивање позадине – проналажење шаке елиминисањем непотребних елемената на сцени. Једна од најефикаснијих метода за раздвајање позадине од објекта у покрету јесте коришћење методе рачунања просека. Код ове методе позадина се рачуна усредњавањем одређеног броја фрејмова.
2. Детекција покрета – на основу претходног корака, покрет се детектује одузимањем позадине од тренутног фрејма. Користи се позадина добијена усредњавањем у претходном кораку.
3. Праг слике (енг. *Thresholding*) – дефинисање прага на основу ког се врши сегментација, односно маскирање.
4. Екстракција контуре – претпоставка је да објекат од интереса има најдужу контуру, где је на

основу резултата претходног корака лакше пронаћи контуру.

## 2.3. Машинско учење

У општој перцепцији машинско учење представља дисциплину која се бави извођењем алгоритама из података, без експлицитног програмирања. Овај поглед наглашава његову практичну страну. Како дедуктивно и индуктивно закључивање имају важну улогу у природној интелигенцији, одговарајуће дисциплине које их аутоматизују – аутоматско расуђивање и машинско учење, имају важне улоге у вештачкој интелигенцији.

Методе засноване на логици, које се развијају у оквиру аутоматског расуђивања, погодне су у случајевима у којима је проблем могуће математички дефинисати. Обично су то проблеми које човек може лако да формулише, али их тешко решава. Са друге стране, машинско учење је посебно погодно и за проблеме које човек не може лако да дефинише, иако неке од њих лако решава. Један пример таквог проблема јесте препознавање лица. Као и људи, методе машинског учења могу врло успешно да се носе са овим проблемом [5].

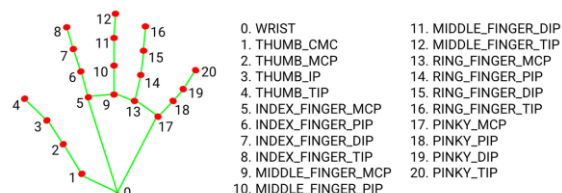
## 2.4. MediaPipe

Широк спектар потенцијалних апликација за машинско учење данас се ослања на неколико основних задатака машинског учења. На пример, детектори за навигацију и знаковни језик ослањају се на способност програма да идентификује и прати покрете руке. С обзиром на то да је изградња нечега попут модела праћења руку дуготрајна и рачунски интензивна, јавља се проблем у развоју и креирању свих апликација које се ослањају на праћење руку. Како би се овај проблем решио, компанија *Google* је осмислила *MediaPipe* моделе.

Ови модели су погодни за коришћење због тога што поједностављују и смањују време развоја многих апликација које се ослањају на знање из области компјутерске визије.

Омогућава приступ широком спектру моћних модела машинског учења који су креирани имајући у виду хардверска ограничења.

Оно што је за овај рад важно јесте детекција покрета шаке. *MediaPipe Hands* је решење за праћење покрета шаке и прстију високе тачности. Користи машинско учење да пронађе 21 3Д обележје шаке за сваки фрејм (Слика 2).



Слика 2. Обележја шаке *MediaPipe Hand* модела

За разлику од лица код ког се појављују региони који се разликују по контрасту од остатка лица (очи и уста), код шаке такви региони не постоје. Због тога је детекција одређених карактеристика и обележја шаке компликованија.

## 2.5. Алати

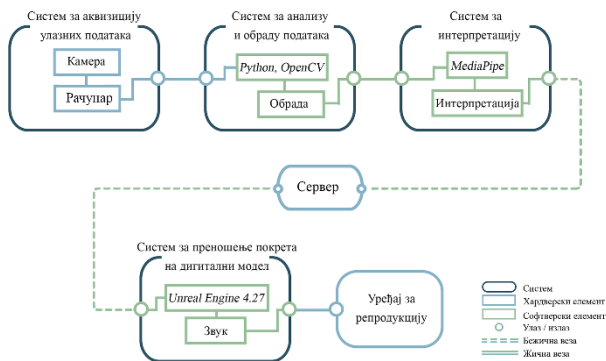
Приликом израде овог система коришћено је неколико метода и више различитих алата и софтвера. Коришћен је *Python* програмски језик, *OpenCV* библиотека, *MediaPipe* тренирани модел, затим „*Autodesk 3ds Max 2021*“ софтвер за потребе *rigg*-овања модела шаке, као и „*Unreal Engine 4.27*“ за преношење покрета на дигитални модел у реалном времену. За потребе локалног сервера коришћен је *JavaScript*, тачније *Node.js* окружење (енг. *environment*), док су подаци прослеђивани у *JSON (JavaScript Object Notation)* формату.

## 3. АРХИТЕКТУРА СИСТЕМА

Систем се састоји из софтверских и хардверских елемената који функционишу заједно ради извршавања задатака у реалном времену. Поред системских целина постоје и друге компоненте система које су намењене за комуникацију или репродукцију, а то су сервер и уређај за репродукцију. Све системске целине имају подједнако важну улогу у имплементацији система. Архитектура система приказана је шематски на слици 3. Креиран је систем који је подељен у четири системске целине:

1. систем за аквизицију улазних података
2. систем за анализу и обраду података
3. систем за интерпретацију
4. систем за преношење покрета на дигитални модел

Прва системска целина, систем за аквизицију улазних података, је хардверска целина која се састоји из камере и рачунара. Улазни подаци су секвенце слика чијом се обрадом у реалном времену добијају потребне информације за даље функционисање.



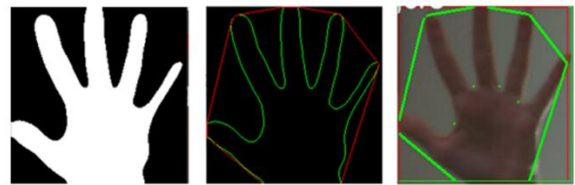
Слика 3. Шематски приказ архитектуре система

Систем за анализу и обраду података је софтверска целина чији је циљ анализирање улазних података и њихова обрада таква да се добију корисни подаци. Главни задатак овог система јесте детекција шаке и детекција покрета шаке. Како би се убрзао и оптимизовао рад система изабран је мањи регион у слици где ће се вршити детекција.

Како би се шака добро детектовала потребно је да позадина буде што униформнија. Потребно је да се раздвоји шака од позадине, као објекат од интереса, и за то је неопходно користити бинаризацију на основу интензитета. Коришћена је Отсуова метода бинаризације што значи да је праг одређен аутоматски. Отсуова метода је техника заснована на варијанси за проналажење граничне вредности где је

пондерисана варијанса између пиксела у предњем плану и у позадини најмања. Метода обрађује хистограм слике, сегментирајући објекте минимизирањем варијансе на свакој од класа. Обично ова метода даје одговарајуће резултате за бимодалне слике. Резултат добијен након Отсуове бинаризације приказан је на слици 4. Ради отклањања шума на слици примењен је Гаусов филтар. Резултат оваквог нископропусног (енг. *Low-pass*) филтра јесте замућена слика са бољим ивицама.

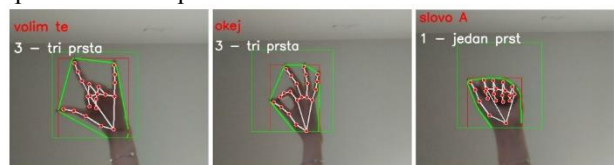
Наредни задаци овог система су проналажење контуре шаке и детекција прстију. Коришћење бинарне слике добијене у претходном кораку знатно олакшава проналажење контуре и даје прецизније резултате. Анализом контуре, конвексности криве и коришћењем математичких прорачуна дефинисан је праг угла којим ће се детектовати размак између прстију. Управо број размака омогућава систему да броји колико је прстију активно у систему (Слика 5).



Слика 4. а) Отсуова бинаризација, б) детектована контура, в) испитивање конвексности

Наредна системска целина је систем за интерпретацију која за циљ има иницијализацију *MediaPipe Hands* модела машинског учења и интерпретацију покрета шаке у реалном времену. Улазни подаци овог система се састоје из секвенце слика које су у претходној системској целини обрађене. Прво је потребно проверити да ли је шака детектована у фрејму, након тога се пролази кроз сваки фрејм где шака јесте детектована и чувају се координате детектованих обележја.

На основу тога се исцртавају обележја и следећи корак је препознавање покрета на основу претходно детектованих обележја. Иницијализован је модел *MediaPipe Hands* који може да препозна десет елемената знаковног језика. Коришћена је функција која узима листу обележја и враћа низ који садржи вредности вероватноћа за сваки до десет елемената. У том низу пронађена је максимална вредност на листи, односно издвојен је покрет за који је утврђено да има највећу вероватноћу. На слици 5 приказани су препознати покрети.



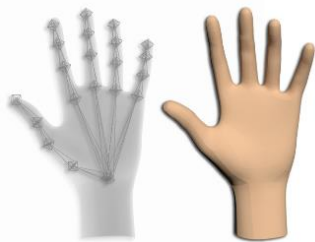
Слика 5. Исцртана обележја шаке и препознати покрети шаке

Излаз из овог система су подаци који се шаљу наредном систему тако да они буду читљиви. Због тога сва обележја као и назив препознатог покрета су смештени у *JSON* објекат. За слање ових података о

координатама обележја и називу препознатог покрета користи се локални сервер.

У контексту овог система, систем за интерпретацију шаље податке серверу, док је наредни систем, систем за преношење покрета на дигитални модел слуша податке од сервера. Дакле, оба система су клијенти, док је сервер посредник у комуникацији ова два система.

У оквиру система за преношење покрета на дигитални модел коришћен је *Unreal Engine* који је врло погодан за репродукцију у реалном времену. У оквиру софтвера “*3ds Max 2021*” је одрађен *rigg* дигиталног модела шаке (Слика 6) [6]. *Rigging* је процес креирања хијерархијски повезаних костију. Кости стварају флексибилан систем који дозвољава покрете и доприноси квалитету и уверљивости покрета. Глатки и сложени покрети модела зависе од квалитета *rigg*-а.



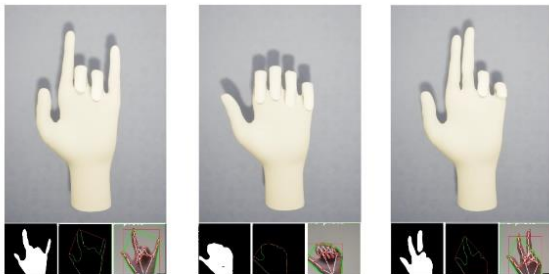
Слика 6. Хијерархија костију и модел шаке

Оно што је важно и на шта се покрети шаке свде јесу ротације зглобова, односно ротација прстију и зглобова прстију како би се формирали одређени знакови.

Оно што је систему прослеђено јесу координате. Углови које заклапају одређени зглобови се могу израчунати математички што се своди на израчунавање угла између три тачке, односно између два вектора. Дакле, проналажење угла се своди на следећи израз (2):

$$\theta = \arccos\left(\frac{\overline{BA} \cdot \overline{BC}}{|\overline{BA}| |\overline{BC}|}\right) \quad (2)$$

Систем је успешно имплементиран са задовољавајућим резултатима и отвореним могућностима за проширење система. На слици 7 могуће је видети неке од крајњих резултата.



Слика 7. Елементи знаковног језика пренети на дигитални модел

С обзиром на то да се у прослеђеном *JSON* објекту налази и назив препознатог покрета, унутар *Unreal Engine*-а направљена је функција која препознаје

назив покрета и у складу са тим пушта одговарајући звучни фајл.

#### 4. ЗАКЉУЧАК

Овај рад описује дизајн, архитектуру и имплементацију система за репрезентацију елемената знаковног језика заснованом на машинском учењу. Знаковни језик се ослања на покрете тела, руку и лица што креирање система за превођење чини врло компликованим за реализацију.

Креирање свеобухватног система, односно система који детектује покрете руку, тела и лица и препознаје их као одређени знак знаковног језика, био би велики корак у развоју асистивних технологија. Направити систем који је робусан и самосталан и који може да се користи на различитим платформама и уређајима, с тим да обухвата све елементе знаковног језика је велики изазов и велика мотивација за даље истраживање.

#### 5. ЛИТЕРАТУРА

- [1] Градска организација глувих Београда „Српски знаковни језик”  
<http://gogb.org.rs/srpski-znakovni-jezik/>
- [2] Савез глувих Србије и региона „Знаковни језик”  
<http://www.sgnsbg.com/index.php/znakovni-jezik>
- [3] Б. Бонцулић „Анализа резултата детекције покрета на слици”, Војна Академија, Београд
- [4] Ј. Крстановић „Видео у инжењерској анимацији”, Факултет техничких наука, Нови Сад, 2022
- [5] М. Николић, А. Зечевић, “Машинско учење”, Београд, 2019
- [6] Бесплатан модел шаке преузет је са адресе:  
<https://sketchfab.com>

#### Кратка биографија:



**Марија Варга** рођена је у Вршцу 1997. године. Основне студије Анимације у инжењерству на Факултету техничких наука у Новом Саду завршила је 2020. године. Мастер рад на Факултету техничких наука из области Рачунарске графике одбранила је 2022. године, Тренутно је запослена као сарадник у настави на катедри за Анимацију у инжењерству.



**Лидија Крстановић** завршила је основне и мастер студије на Природно-математичком факултету у Новом Саду, смер професор математике. Докторирала је 2017. године на Факултету техничких наука, смер Математика у техници. Године 2018. изабрана је у звање доцента на истом факултету на катедри за Анимацију у инжењерству.