

**INTELIGENTNA EKSTRAKCIJA KVANTITETA IZ KONTINUALNOG  
MULTIMEDIJALNOG STRIMA****INTELLIGENT QUANTITY EXTRACTION FROM A CONTINUOUS MULTIMEDIA  
STREAM**Nikola Slijepčević, *Fakultet tehničkih nauka, Novi Sad***Oblast – ELEKTROTEHNIKA I RAČUNARSTVO**

**Kratak sadržaj** – U ovom radu opisan je problem inteligentne ekstrakcije kvantiteta i prepoznavanja ručno pisanih cifara iz multimedijalnog strima, kao i jedno rješenje opisanog problema. Problem je rješavan uz pomoć vještačke inteligencije, mašinskog učenja, neuronskih mreža i raznih metoda za obradu slike. Programsko rješenje ovog problema pisano je u programskom jeziku Python uz korišćenje pripadajućih biblioteka.

**Ključne reči:** Ekstrakcija kvantiteta, Prepoznavanje cifara, Digitalna slika, Mašinsko učenje, Neuronske mreže

**Abstract** – This paper describes the problem of intelligent quantity extraction and recognition of handwritten digits from a multimedia stream, as well as a solution to the described problem. The problem was solved with the help of artificial intelligence, machine learning, neural networks and various image processing methods. The software solution to this problem is written in the Python programming language using the associated libraries.

**Keywords:** Quantity extraction, Digit recognition, Digital image, Machine learning, Neural networks

**1. UVOD**

Od najranijih početaka ljudske civilizacije čovjekovo djelovanje bilo je usmjereno tehnološkom napretku. Danas se taj napredak posebno ogleda u razvoju računarskih tehnologija, koje ujedinjuju naučnike različitih usmjerenja te inženjere raznih struka čiji je posao otkrivanje, razvoj i oblikovanje novih pristupa raznim problemima. Skupu takvih oblasti pripada i razvoj neuronskih mreža i mašinskog učenja, odnosno naučno-tehnološka disciplina koja se bavi teorijom izrade vještačke inteligencije i programa koji se bave dobavljanjem objekata iz videa ili slika te njihovom ekstrakcijom. Temelj razvoja čini mogućnost računarske percepcije ljudskog vida i digitalizacije slike, a djelovanje je usmjereno automatizaciji i integraciji širokog spektra procesa te prezentaciji vizuelne percepcije.

Klasični problemi ovog područja su analiza pokreta, restauracija slike, rekonstrukcija događaja, prepoznavanje i ekstrakcija objekata.

**NAPOMENA:**

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Dragan Ivetić, red. prof.

U ovom radu predstavljeni su problemi i rješenje inteligentne ekstrakcije objekata iz multimedijalnog strima tj. preuzimanjem slika, kao frejma, iz video zapisa, zatim izdvajanjem regiona od značaja, potom prepoznavanje objekata, gdje su objekti Arapski ručno ili kompjuterski (mašinski) pisani brojevi.

Danas se većina dostupnih informacija nalazi u elektronskom ili u obliku fotografija i video zapisa. Pa i iz tog razloga razvijaju se sve više algoritmi digitalne obrade podataka. Algoritmi digitalne obrade i analize slike koji danas postoje uglavnom su orijentisani na automatsku detekciju i prepoznavanje jednog ili ograničenog skupa objekata ili kategorija, dok algoritam koji bi bio u stanju detektovati i prepoznavati proizvoljne objekte nisu još u potpunosti razvijeni.

Zavidni rezultati neuronskih mreža u brojnim zadacima vještačke inteligencije, čiju ilustraciju vidimo na slici 1, pokazuju kako je to područje s izrazito velikim potencijalom. Nezavisno o tome radi li se o obradi i analizi slika, zvuka, video zapisa, teksta ili bilo kog drugog ulaznog podatka, neuronske mreže sve značajnije prednjače kao najbolji odabir algoritma za učenje.



Slika 1. Vještačka inteligencija ilustracija

**2. NEURONSKE MREŽE**

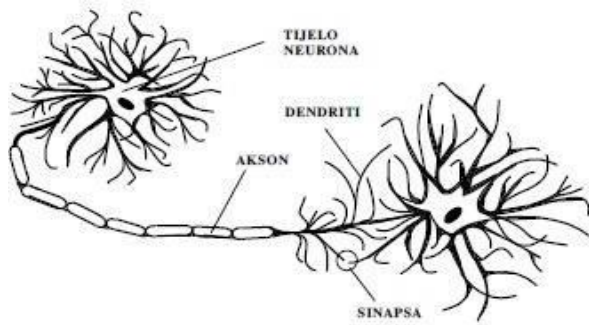
Neuronska mreža je skup međusobno povezanih jednostavnih elemenata obrade, jedinica ili čvorova. Sposobnost obrade mreže je posljedica jačine veza između tih elemenata, a postiže se kroz proces adaptacije ili učenjem iz već postojećeg skupa primjera [1].

Neuronska mreža je masovno paralelizovan distribuirani procesor sa prirodnom sposobnošću memorisanja iskustvenog znanja i obezbjeđivanja njegovog korišćenja.

## 2.1. Neuroni, biološki i vještački

Sastavni dijelovi vještačkih neuronskih mreža su vještački neuroni. Model vještačkog neurona nastao je pokušajem imitacije načina preuzimanja i obrade informacija biološkog neurona.

Biološki neuron, čiji je pojednostavljeni oblik prikazan na slici 2, sastoji se od tijela, aksona i velikog broja dendrita koji okružuju tijelo neurona.



Slika 2. Pojednostavljena slika biološkog neurona

Neuron je modul koji radi nelinearnu transformaciju ulaznog podatka. To je matematička funkcija koja ulazni vektor realnih brojeva transformiše parametrizovanom linearnom transformacijom, primjenjuje nelinearnu aktivacijsku funkciju i na izlazu daje jednu realnu vrijednost. To se može predstaviti jednačinom (1).

$$y = f(\vec{x} \cdot \vec{w} + b) = f\left(\sum_{i=1}^n x_i \cdot w_i + b\right) \quad (1)$$

Gdje je:

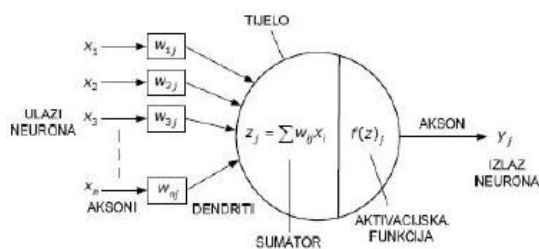
$X_i$  – ulazni vektor;

$W_i$  – vektor težina;

$b$  – prag neurona.

Analogija između biološkog i vještačkog neurona je sledeća: tijelo se zamjenjuje sa sumatorom, ulogu dendrita preuzimaju ulazi u sumator, izlaz iz sumatora je akson vještačkog neurona, a ulogu praga osjetljivosti biološkog neurona preuzimaju aktivacijske funkcije. Signali postaju numeričke vrijednosti, dok se vrijednosti sinapse opisuju težinskim faktorima.

Težinski faktori povezuju izlaze drugih neurona sa ulazima sumatora. Izlaz iz sumatora povezuje se na ulaz aktivacijske funkcije na čijem se izlazu reprodukuje izlaz vještačkog neurona. Na slici 3 vidimo prikaz vještačkog neurona.



Slika 3. Prikaz strukture vještačkog neurona

## 2.2. Aktivacijske funkcije

Na izlazima slojeva neuronskih mreža koriste se aktivacijske funkcije da bi se u model unijela nelinearnost. Bez njih bi mreža mogla naučiti samo linearnu granicu između podataka različitih klasa.

Funkcije aktivacije su matematičke jednačine koje određuju izlaz neuronske mreže. Funkcija je vezana za svaki neuron u mreži i određuje da li je treba aktivirati ili ne, na osnovu toga da li je ulaz svakog neurona relevantan za predviđanje modela

## 2.3. Vrste vještačkih neuronskih mreža

Kada bi jedan neuron radio samostalno imao bi velika ograničenja u pogledu realizacije nekog konkretnog zadatka. Međutim, povezivanjem više neurona na smislen i odgovarajući način, dobija se mreža koja može riješiti kompleksnije probleme uz zadovoljavajuće rezultate.

Vještačke neuronske mreže mogu biti jednoslojne i višeslojne [2]. Višeslojne mreže imaju ulazni i izlazni sloj, a između njih se nalazi jedan ili više skrivenih slojeva neurona. U odnosu na vrijeme razlikuju se vremenski-neprekidne i vremenski-diskretne neuronske mreže. Zavisno od područja primjene imamo perceptronske, asocijativne, dvostruko asocijativne, adaptivne neuronske mreže. Takođe, može se izvršiti podjela mreža prema smjeru u kojem signali putuju. Ukoliko signali putuju samo u jednom smjeru, od ulaznog do izlaznog sloja, riječ je o jednosmjernim neuronskim mrežama. U slučaju da postoji sloj ili dio mreže u kojem signal putuje u suprotnom smjeru tj. mreža ima povratnu petlju govorimo o povratnim neuronskim mrežama.

## 2.4. Učenje vještačkih neuronskih mreža

Za mrežu, učenje predstavlja proces mijenjanja težina, koji se odvija u odnosu na unesene ulazne i izlazne promjenljive.

Ukoliko su poznate željene vrijednosti izlaznih promjenljivih, govori se o nadgledanom učenju. Nadgledano učenje je iteracijski postupak koji podrazumijeva postojanje spoljašnjeg faktora koji koriguje rezultate mreže, sve dok se ne dobije željeno ponašanje mreže.

Kod nekih mreža nisu predstavljeni željeni izlazi mreže. Tada se govori o nenadgledanom učenju. Na ulaz mreže dovedu se ulazne promjenljive nakon čega mreža sama podešava težine po dobro definisanom algoritmu.

## 2.5. Primjena vještačkih neuronskih mreža

Možda najveća prednost vještačkih neuronskih mreža je u njihovoj mogućnosti da se koriste kao proizvoljan mehanizam aproksimacije funkcija koje uče na osnovu podataka

Danas se neuronske mreže primjenjuju u mnogim segmentima života poput računarskih nauka, medicine, bankarstva, fizike itd., najčešće za sledeće zadatke: prepoznavanje šablona i uzoraka, obrada slike, praćenje objekata, obrada govora, problemi optimizacije, nelinearno upravljanje, obrada nepreciznih i nekompletnih podataka, simulacije i sl. Iako je već dosta postignuto na području neuronskih mreža, daljim razvojem tehnologije i one će se dalje razvijati i širiti svoje područje primjene.

### 3. DIGITALNA SLIKA I NJENA OBRADA

Digitalna slika je numerička (najčešće cjelobrojna) reprezentacija dvodimenzionalne slike.

Vektorske slike, odnosno njihov sadržaj, opisan je prostim geometrijskim objektima, kao što su tačke, prave, krive i poligoni. Svaki od ovih objekata je predstavljen relativno malom količinom podataka, kao što su pozicija u koordinatnom sistemu, dimenzije, stil i boja linije i stil i boja objekta.

Kod rasterskih slika podaci su predstavljeni u vidu konačne matrice, u kojoj se svaki element naziva piksel. Svaki piksel, kao najmanja jedinica slike, sadrži podatke o boji i intenzitetu, dok je njegov položaj na slici određen položajem unutar matrice.

#### 3.1. Tipovi digitalnih slika

Digitalna slika je prikaz dvodimenzionalne slike sa konačnim skupom digitalnih vrijednosti koje se nazivaju pikseli. Pikseli se čuvaju u računarskoj memoriji ili hard disku kao rasterka slika ili rasterska mapa, odnosno dvodimenzionalni niz malih cjelina. Neke od najpoznatijih predstava slike su binarne slike, slike u nijansama sive, slike u boji, indeksirane slike.

Kod binarnih slika se koristi jedan bit za predstavljanje jednog piksela. Kako bit ima svega dva stanja, 0 i 1, to znači da je svaki piksel u jednoj od dvije predefinisane boje, a najčešće su to crna i bijela. Nemogućnost prikazivanja međunijansi sive je veliko ograničenje za binarne slike u prikazivanju fotografija.

Slike u nijansama sive, koje se često nazivaju i crno-bijelim slikama, sačinjene su od piksela čije boje su različiti intenziteti sive, u rasponu od bijele do crne boje. Najčešće se radi o 256 nivoa intenziteta sive, pa je za prikaz jednog piksela potreban jedan bajt. Imajući u vidu da je eksperimentima pokazano da ljudsko oko može da razlikuje najviše 200 nijansi sive, 256 nivoa je i više nego dovoljno da prikaže pun spektar između bijele i crne bez uočljivih prelaza, pa je ovaj tip digitalnih slika pogodan za prikaz fotografija.

Slike u boji čine pikseli koji pored informacija o intenzitetu sadrže i informacije o boji. Za vizuelno prihvatljive rezultate, neophodne su tri vrijednosti, odnosno kanala, koji opisuju boju svakog piksela u određenom prostoru boja. Najčešće korišćeni prostor boja je RGB, ali u upotrebi su i brojni drugi, kao što su HSV, CMYK, YCbCr itd. Kod RGB prostora se koristi 256 nivoa za svaku komponentu, tako da su potrebna tri bajta za prikaz jednog piksela. Ukupan broj boja koji se može prikazati u ovom prostoru je 16 777 216.

Indeksirane slike u boji imaju ograničenu paletu boja. Nazivaju se indeksirane zato što je boja piksela predstavljena indeksom, odnosno pozicijom te boje u predefinisanoj paleti. Kod prikazivanja fotografija pomoću indeksiranih slika dolazi do izvjesnog gubitka informacija, ali to u mnogome zavisi od broja i izbora boja u paleti, kao i od opsega boja fotografije. Zbog ograničenja starijih monitora koji su mogli da prikažu samo 256 boja, bilo je pokušaja da se formira paleta boja koja bi bila najbolje prilagođena tadašnjem hardveru. Izabrano je 216 boja, koje se dobijaju kombinacijom po 6 nijansi crvene, zelene i plave.

#### 3.2. Obrada digitalnih slika

Metode obrade slika se mogu podijeliti u dvije osnovne grupe, zavisno od vrste ulaznih podataka i rezultata obrade [3]. Prvu grupu čine metode čiji su ulazni i izlazni podaci slike, dok drugu grupu čine metode čiji ulazni podatak može biti slika, ali izlazni podaci su neki atributi izvedeni sa slike. U prvu grupu spadaju formiranje slika, poboljšanje, restauracija, obrada slika u boji, obrada pomoću talasića, kompresija, itd. U drugu grupu spadaju segmentacija, reprezentacija, prepoznavanje objekata. Morfološka obrada za izlaz može imati sliku ali i neke attribute, pa ona kao takva pripada i prvoj i drugoj grupi.

#### 3.3. Pretprocesiranje slike

Pretprocesiranje slike predstavlja pripremu slike za obradu. Sastoji se od različitih tehnika manipulacije nad ulaznom slikom kao što su skaliranje, poboljšanje kontrasta, detekcija ivica, segmentacija, otklanjanje šuma i sl. Ovaj korak predstavlja ključnu operaciju u mašinskom učenju, jer su ovdje izražene najvažnije osobine slike, dok su one manje bitne informacije i šum potisnute ili eliminisane.

Nova slika koja nastaje kao rezultat pretprocesiranja je slična originalnoj slici, a razlikuju se po tome što nova slika ima npr. manje šuma, bolji kontrast.

#### 3.4. Segmentacija slike

Jedan od glavnih elemenata u prepoznavanju oblika na slici predstavlja pronalaženje ključnih elemenata slike koji reprezentativno predstavljaju njen sadržaj. Nakon što se ključni elementi pronađu stvaraju se uslovi da se ti elementi porede, klasifikuju i na drugi način koriste u daljoj obradi. Zadatak pronalaženja ovih elemenata ima segmentacija koja dijeli sliku na regione, ili objekte do određenog nivoa i predstavlja nezaobilazan korak pri automatskoj detekciji oblika i analizi slike.

Thresholding ili segmentacija pragom predstavlja izdvajanje objekata od pozadine pomoću jednog ili više pragova. Segmentacija sa jednim pragom predstavlja najjednostavniju metodu i zasniva se na tome da ukoliko je vrijednost intenziteta piksela veća od praga, onda se on svrstava u pozadinu, a ukoliko je vrijednost manja od praga, onda se svrstava u objekat.

#### 3.5. Histogram

Histogram koristimo kada je potrebno da dobijemo informaciju o distribuciji osvetljenosti piksela. Vrlo je koristan kada je potrebno odrediti prag za globalni threshold. Histogram je grafik koji prikazuje broj piksela na slici pri svakoj različitoj vrijednosti intenziteta pronađenoj na toj slici.

#### 3.6. Morfološke operacije

Kada se izoluje oblik kao binarna slika, u našem slučaju ručno ili mašinski pisan broj, oblici često imaju nesavršenosti i dolaze sa nepoželjnim šumovima i teksturama. Tada se koriste morfološke operacije i filteri kako bi se te nesavršenosti uklonile i dobio tačniji i precizniji oblik za dalju obradu. U morfološkim operacijama, vrijednost svakog piksela rezultujuće slike se zasniva na poređenju odgovarajućeg piksela na originalnoj slici sa svojom okolinom. Osnovne morfološke operacije su: erozija, dilacija, otvaranje, zatvaranje, istanjivanje i podebljavanje.

## 4. PROGRAMSKO RJEŠENJE

U ovom poglavlju će biti analizirani i izloženi rezultati prepoznavanja ručno pisanih brojeva koristeći vještačke neuronske mreže. Početni dio poglavlja posvećen je upoznavanju s programskim jezikom Python i pripadajućim programskim bibliotekama TensorFlow, NumPy, Keras i Matplotlib.

### 4.1. Python, TensorFlow, Keras, NumPy i Matplotlib

Python je interpretirani, interaktivni, objektno orijentisani programski jezik visokog nivoa i opšte namjene koji poseduje dinamičku sematiku. Python je „open source“, tako da je dopuštena distribucija i izmjena, što znači da ljudi širom svijeta mogu učestvovati u razvoju Python-a. Upravo takva fleksibilnost i otvorenost koda je doprinijela je rastućoj popularnosti Pythona u industrijskom i komercijalnom okruženju.

TensorFlow je biblioteka za numerička izračunavanja korišćenjem grafova toka podataka [4]. Među zapaženim primjenama su primjene u zadacima kao što su prepoznavanje govora, robotika, pronalaženje informacija, obrada jezika, pronalaženje geografskih informacija.

Osim TensorFlow-a, u izradi programskog koda korištena je i biblioteka NumPy. U njoj se nalaze esencijalne matematičke funkcije korištene u linearnoj algebri, razne transformacije, mogućnost stvaranja nasumičnih vrijednosti te naravno mogućnost rada s matricama.

Uz NumPy, važno je spomenuti i biblioteku Matplotlib, koja omogućava jednostavnu vizualizaciju podataka. Modul pyplot iz spomenute biblioteke je korišćen kod prikazivanja metrika učenja mreža korišćenih u programskoj podršci.

Baze podataka korištene u ovom radu za učenje i treniranje vještačkih neuronskih mreža su MNIST (*eng. Modified National Institute of Standards and Technology*) [5] i EMNIST (*eng. Extended Modified National Institute of Standards and Technology*).

### 4.2. Programska implementacija i rezultati

Prvo što je potrebno uraditi jeste video podijeliti na frejmove, da bi se potom izvršila transformacija u binarnu sliku, nakon čega će se izvršiti operacija erozije. Detekcija linije vrši se korišćenjem Hough transformacije. Iskorišćena je HoughLinesP transformacija iz OpenCV biblioteke.

Nakon toga potrebno je frejm po frejm izdvajati iz videa, pretvoriti svaki frejm u sliku u nijansama sive, a potom u binarnu sliku. Binarna slika se nakon toga kao parametar prosljeđuje u funkciju za selektovanje regiona tj. ekstrakciju cifara. Za svaki region je potrebno napraviti posebnu sliku dimenzija 28x28.

Za označavanje regiona korišćena je metoda `cv2.boundingRect(contour)`. Da bi se označili samo oni regioni od interesa oko brojeva koji pređu preko linije, potrebno je uz pomoć koordinata dobijenih pomoću Hough transformacije iskoristi jednostavan algoritam za njihovo izdvajanje. Taj algoritam se ogleda u tome da se izračunava udaljenost od gornje lijeve tačke konture broja do krajnjih tačaka detektovane linije.

Ukoliko je suma te dvije udaljenosti približna dužini linije ( $\epsilon=0.1$ ) uzima se da je broj prešao liniju.

Na osnovu konture broja se vrši isijecanje regiona slike u kome se broj nalazi. Posle označavanja regiona i njihovog smještanja u niz slika koje predstavljaju regione sortirane po rastućoj vrijednosti  $x$  ose. Dobijeni region se prosljeđuje neuronskoj mreži, obučenoj pomoću MNIST data seta, na prepoznavanje.

Za obučavanje neuronske mreže, potrebno je obezbijediti ulaze, a potom željene izlaze za date ulaze. Nakon toga slijedi definisanje parametara algoritma za obučavanje i obučavanje neuronske mreže. Postignuta tačnost ovog programskog rješenja je 92%.

## 5. ZAKLJUČAK

U okviru ovog rada opisan je problem ekstrakcije, obrade i prepoznavanja objekata iz kontinualnog multimedijalnog strima. Posebna pažnja posvećena je manipulaciji nad slikama tj. frejmovima koji su izdvojeni iz videa i prepoznavanju ručno ili mašinski pisanih cifara uz pomoć vještačkih neuronskih mreža.

Prije izrade samog rješenja proučeni su mnogi načini kojima bi mogla da se implementira prethodno opisana problematika. Nakon detaljno proučenih mogućnosti pristupilo se izradi projektnog rješenja koje je pisano u programskom jeziku Python uz oslonac na pripadajuće biblioteke kao što su TensorFlow, NumPy, Keras i druge. Korišćene su neke od najpoznatijih metoda za obradu slike kao što su invertovanje, konvertovanje, globalni i lokalni threshold, histogram, morfološke operacije.

Nastavak razvijanja i poboljšavanja rješenja ovog problema mogao bi teći u smjeru ekstrakcije i prepoznavanja preklapajućih cifara iz multimedijalnog strima, kao i smanjenju uticaja faktora okoline na tačnost prepoznavanja kvantiteta.

## 6. LITERATURA

- [1] Gurney K., „An introduction to neural networks“, University of Sheffield, London and New York, 1997.
- [2] Matthew D Zeiler, Rob Fergus, „Visualizing and understanding convolutional networks“, 2013, New York.
- [3] R. C. Gonzalez, R. E. Woods, „Digital Image Processing“, Prentice Hall, 2008.
- [4] Yuan Yu, Xiaoqiang Zheng. „Large-scale machine learning on heterogeneous systems“, 2015.
- [5] <http://yann.lecun.com/exdb/mnist/> (pristupljeno u oktobru 2020.)

### Kratka biografija:



**Nikola Slijepčević** rođen je u Trebinju 1996. god. Završio je gimnaziju u Gacku 2015. godine. Diplomski rad na Fakultetu Tehničkih Nauka u Novom Sadu odbranio je 2019. godine. Iste godine je upisao master studije u Novom Sadu na smjeru Računarstvo i automatika.

kontakt:  
nikolaslijepcevi081296@gmail.com