

Vijeću Prirodno-matematičkog fakulteta

IZVJEŠTAJ KOMISIJE O PODOBNOSTI TEME MAGISTARSKOG RADA KANDIDATA STRAHINJE RADMANA

Vijeće Prirodno-matematičkog fakulteta na sjednici održanoj 10. 9. 2019. imenovalo je mentora i Komisiju za ocjenu podobnosti teme za izradu magistarskog rada pod nazivom "Mašinsko učenje u optimizaciji videa", kandidata Strahinje Radmana, u sastavu:

dr Milenko Mosurović redovni profesor - član

dr Savo Tomović, vanredni profesor - član

dr Vladimir Božović, vanredni profesor - mentor.

Nakon uvida u podneseni materijal, a u vezi sa članom 24 Pravila studiranja na postdiplomskim studijama, podnosimo sljedeći

IZVJEŠTAJ

Strahinja Radman, specijalista računarskih nauka, prijavio je temu magistarskog rada pod nazivom "Mašinsko učenje u optimizaciji videa". Tema spada u oblast računarskih nauka za koju je matičan Prirodno-matematički fakultet. Dokumenta podnesena za prijavu teme sadrže: biografiju kandidata, naziv i kratku razradu teme, kao i kratko obrazloženje predmeta istraživanja i strukture rada.

Podaci o kandidatu

Strahinja Radman je rođen 15. 10. 1988. godine u Nikšiću. Završio je specijalističke studije Odsjeka za računarske nauke Prirodno-matematičkog fakulteta Univerziteta Crne Gore 2012. godine. Nakon završenih specijalističkih studija, Strahinja Radman je nastavio sa praktičnim usavršavanjem u domaćim i inostranim firmama u kojima je i danas radno angažovan. U septembru 2016. godine upisuje magistarske studije odsjeka za Računarske nauke Prirodno-matematičkog fakulteta. U periodu od septembra 2016. do oktobra 2018 položio je sve ispite predviđene programom studija.

Aktuelnost teme

Čulo vida je čulo preko kojeg dobijamo veliku količinu informacija o svijetu koji nas okružuje. Nekada je fokus bio na plijenu i predatorima, dok je danas to u sferi zabave i filmske umjetnosti. Shodno tome, napravljeni su veliki intelektualni i naučni naponi u pravcu istraživanja oka, odnosno shvatanja kako zapravo funkcioniše ljudski vizuelni sistem. Ljudsko oko, znamo, ima svoja ograničenja. Na primjer, oko ne može da registruje talasne dužine svjetlosti ispod 400nm i one iznad 700nm. Ovo znači da ljudsko oko ne može da vidi pun spektar boja koji je prisutan u prirodi. Takođe, ljudski vizuelni sistem je u stanju da prati stvari koje se brzo kreću, kao što je na primjer antilopa koja trči, ali pri tom zanemarujući detalje u okruženju van centra fokusa. Moderni digitalni aparati su u stanju da registruju cijeli skup talasnih dužina svjetlosti, ali ih proizvođači namjerno filtriraju kako bi slika bila što sličnija onoj koju mi vidimo. Dugogodišnja istraživanja su pokazala da je ljudsko oko vrlo loše u registovanju različitih nijansi boja, ali da je veoma dobro u registovanju promjena osvjetljenja. Upravo je to važna činjenica koja ključno opredjeljuje načine kompresije video signala. Kako je nekompresovani video u memoriji predstavljen sa jednom komponentom osvjetljena i dvije komponente boje odnosno 4:4:4 sistem, prirodno je rešenje da se kompresuje tako što se uvodi sistem 4:2:0, čime se zauzima jedna četvrtina od početnog prostora potrebnog za boju. Ipak, kvalitet slike, sa stanovišta ljudskog oka, će ostati isti, odnosno oko neće moći da registruje razliku. To bi bio prvi nivo optimizacije, drugi nivo bi bio u dijelu gdje ljudski sistem prati brze predmete, a zanemaruje okolinu. Ova optimizacija se odvija u enkoderima koji dalje kompresuju video signal. Naime, scene u kojima je prisutna veća količina kretanja, enkoder će da kompresuje više, a samim tim će opasti broj detalja, jer oko ne može da ih registruje. Za razliku od prethodnog, u scenama gdje je sadržaj statičan, kompresija će se držati na razumnom nivou kako bi se što više detalja očuvalo. Naravno, kompresije koje se obavlja u enkoderima se računaju po već utvrđenim generalizovanim parametrima. Da bi se poboljšao nivo kompresije, rješenje se traži u mašinskom učenju. Jedno od predloženih rješenja koje je dao Netflix. Rješenje koje koristi u svojim data centrima za obradu videa sastoji se od uvođenja nove mjere kvaliteta - VMAF, koja ima bolju korelaciju sa ljudskom percepcijom kvaliteta u odnosu na PSNR i SSIM koje se koriste u enkoderima. (PSNR je odnos između novonastalog šuma usled kompresije i referentne slike, a SSIM je srednja vrijednost razlike piksela kompresovane i referentne slike). VMAF je treniran uz pomoć mašinskog učenja gdje su ljudi gledajući različite video klipove ocjenjivali da li je scena kvalitetna ili ne. Tako dobijenu metriku Netflix koristi na sledeći način. Scene (djelovi videa) se enkodiraju sa različitim parametrima i svaka scena se ocjenjuje sa VMAF metrikom. Na kraju se bira scena enkodirana sa parametrima koji daju najbolji VMAF rezultat. Ova sistem zahtijeva ogromno procesorsko vrijeme jer se ista scena mora enkodirati veći broj puta kako bi se našli optimalni parametri za enkoder, što nije najoptimalniji pristup. Takođe, ovo se može koristiti samo u slučaju video sekvenci koju je potrebno arhivirati, dok je prenos uživo gotovo i nemoguć. Google je takođe pokušao da se bavi sličnom tematikom. Naime,

oni su pokušali da odrede CRF (konstantni nivo kvaliteta, parametar za enkoder koji sam određuje bitrejt kako bi očuvao kvalitet) na osnovu veličine frejma, broja frejmova u sekundi, broja bita iskorištenih za vektore kretanja, procenat prekočenih makroblokova, srednji kvantizacioni parametar. Svi ovi podaci su dobijeni od enkodiranja istog videa na ogromnom bitrejtu. Posle enkodiranja parametri se stavljaju kao ulazi u neuronsku mrežu koja pokušava da predvidi potreban CRF za određenu scenu. Ovaj sistem je u 65% slučajeva davao dobre rezultate. Iako je 65% dobar učinak, nije i dovoljan.

Ostala istraživanja u ovom polju mogu se podijeliti u dvije grupe: korišćenje mašinskog učenja u optimizaciji određenih djelova algoritama već postojećih enkodera i striktno enkoder koji je dobijen nekom od tehnika mašinskog učenja. Prva grupa se bavi smanjivanjem kompleksnosti enkodera mijenjanjem ručno dizajniranih funkcija onim dobijenim pristupom mašinskog učenja ili povećavanjem performansi enkodera sa metodama mašinskog učenja. U ovu grupu takođe spadaju metodi koji smanjuju rezoluciju videa, enkodiraju već postojećim metodama kompresije, a onda povećavaju rezoluciju do potrebne granice. U drugu grupu spadaju autoenkoderi koji pokušavaju da nauče funkciju $f(x) = x$, odnosno da na izlazu od mreže dobiju istu reprezentaciju koja je data na ulazu.

Cilj, struktura i metodologija rada

Procjena količine internet saobraćaja tokom 2019. je 2 ZB, što je 1.000.000.000 TB, odnosno to je 4.5 EB na dan. Kako su u pitanju ogromne količine podataka, onda je i potrebna ogromna procesorska moć za njihovu obradu, bilo da se radi o mrežnim svičevima koji obrađuju i do 60 Gbps ili o kućnim računarima koji prikazuju video snimke ogromnih rezolucija na još većem bitrejtu. Od ukupnih 4.5 EB dnevno, 94% je video sadržaj! Na drugoj strani, približavamo se teorijskom limitu broja tranzistora u procesoru. Na primjer, kanali kroz koje se kreću elektroni su samo dva atoma široki, a kapije na tranzistorima toliko male da može da dođe do proboja elektrona kroz kapiju i bez napona na bazi. Stoga, neophodno je raditi na alternativnim pristupima rješavanja internet "zagušenja". Kako je, kao što smo naglasili, 94% informacija koje opterećuju današnje procesore video sadržaj, onda je upravo to polje za moguću primjenu nekih metoda optimizacije.

Najpopularniji enkoderi današnjice su x264 i x265. Oba enkodera postižu slične performanse u kompresiji. Međutim, x265 daje bolje rezultate za rezolucije iznad 1080p, dok x264 daje bolje rezultate za one rezolucije ispod 1080p. Ono što karakteriše oba enkodera jesu modovi koji se koriste za kompresiju, a to su konstantni kvantizacioni mod (CQP), konstantni bitrejt (CBR), srednji bitrejt (ABR), srednji bitrejt u dva prolaza (2-pass ABR) i konstantni kvalitet (CRF). Svi ovi modovi zahtijevaju dobro poznavanje video tehnika i video sadržaja kako bi se mogli optimalno koristiti. Najčešće korišćeni mod jeste mod konstantnog kvaliteta ili CRF. U ovom modu enkoder jednostavno pokušava da očuva kvalitet videa na određenom nivou. Ovaj mod podrazumijeva automatsko podešavanje bitrejta što znači da bitrejt može

da bude, nepotrebno, ogroman. Nedostatak predefinisanih algoritama je u činjenici da je pojedine scene, na osnovu njihovog sadržaja moguće značajnije kompresovati uz neprimjetne gubitke kvaliteta. Dakle, pravom selekcijom parametara za enkoder može se postići najbolji odnos između kvaliteta i veličine konačnog videa. Time se smanjuje ukupna količina podataka koja je potrebna da se jedan video pošalje ili sačuva. Samim tim, smanjuje se i količina procesorske moći koja je neophodna da se isti enkodira.

Video je sačinjen od većeg broja scena, gdje je jedna scena skup sličnih frejmova. Pojam sličnosti frejmova je veoma težak za definisanje i ključan je kod raspoznavanja dvije različite scene. Promjena između scena može biti postepena i samim tim je razlika među frejmovima teško uočljiva za postojeće algoritme. Prethodno navedeni enkoderi rade na sledećem principu. Svaka scena se enkodira zasebno, odnosno postoji jedan glavni frejm (*Intra frame*) koji je najmanje kompresovan, a potom više uzastopnih njemu *sličnih* frejmova koji se na osnovu razlika mogu izračunati iz glavnog. Problem nastaje u trenutku kada se prekid scene ne detektuje i kada se umjesto novog intra frejma ubacaju frejmovi koji nemaju prevelike sličnosti sa glavnim. Ovo negativno utiče na proces kompresije video signala.

Skup atributa koji opisuje pojedinačne frejmove je u principu standardizovan. Neki od atributa koji se provjeravaju su odavno u upotrebi (prosječno osvijetljenje, apsolutna razlika osvijetljenja dva uzastopna frejma, prosječna razlika piksela dva frejma, varijabilnost, kontrast), dok su neki relativno neistraženi (horizontalni gradijent, vertikalni gradijent, blok gradijent, prostorno-vremenska složenost itd). Nažalost dosadašnja istraživanja su potvrdila da se jedan atribut ne može uzeti kao zlatni standard u detekciji promjene scene. Međutim, ispostavlja se da je moguće rangirati *značaj* postojećih atributa u odnosu na detekciju promjene scene.

Dakle, na osnovu već utvrđenih atributa, cilj istraživanja se sastoji i u određivanju metoda mašinskog učenja koji bi doveo do selekcije pravih parametara za enkoder kako bi dobili najmanju veličinu video fajla uz zadržavanja nivoa kvaliteta. Zapravo, potrebno je ispitati hipotezu - *mašinsko učenje je primjenljivo u oblasti optimizacije video signala*.

Verifikacija hipoteze se vrši tako što se uspostavlja sistem koji će koristiti navedene attribute primarno za detekciju promjene scene. Ustanovimo da je za to, dakle za detekciju promjene scene najpogodniji algoritam *nasumičnih šuma*. Nakon uspješno detektovane promjene scene, sistem će na osnovu prikupljenih srednjih vrijednosti odabranih atributa i jednog novog atributa iskoristiti *neuronske mreže* kako bi postigao unaprijed zadatu vrijednost kvaliteta videa. Samim tim će biti određena i CRF vrijednost na kojem je potrebno izvršiti enkodiranje scene. *Nasumične šume* će biti trenirane u statističkom paketu R, dok će se za treniranje *neuronskih mreža* iskoristiti MATLAB.

Rad će se sastojati od sledećih cjelina: Uvod, Osnovni principi videa, Mašinsko učenje, Mašinsko učenje u video enkodiranju, Zaključak.