

Prepoznavanje Cifara na Video Snimku

Opis Problema

Dati su video snimci na kojima se nalazi plava linija. Cifre se po video snimku periodично krecu sa leva na desno, gore prema dole. Neke od tih cifara koje prelaze preko pomenute plave linije su od znacaja. Potrebno je detektovati takve cifre, kao i izvrsiti softversko prepoznavanje cifre o kojoj je rec. Cifre su rukom pisane, a buduci da covek nece nikada istu cifru napisati na isti nacin poput masine, potrebno je obuciti program da klasifikuje cifre na osnovu korelacije podataka za treniranje.

Podaci

Razlikujemo podatke za testiranje i podatke za treniranje (obuku). Podaci za testiranje jesu sami video snimci nad kojima se izvrsava prepoznavanje cifara, odnosno u uzem smislu - slicice dimenzija 28×28 , gde svaka od njih predstavlja detektovanu cifru sa video zapisa. Nakon detekcije slicice cifre sa frejma video snimka, ona se transformise u jednodimenzionalni niz grayscale vrednosti piksela koji je predstavljaju.

Podaci nad kojima se vrsi obuka jesu analogno test podacima jednodimenzionalni nizovi, koji su dobijeni iz javno pristupnog "MNIST Original" skupa od 70.000 razlicitih cifara, gde za svaku cifru razlikujemo oko 7000 razlicitih zapisa.

Opis Resenja

Prvenstveno se svaki video snimak analizira frejm-po-frejm, gde se nad prvim frejmom koristi Houghova Transformacija kako bi se ocitala plava linija. Buduci da je implementirana OpenCv funkcija HoughLinesP, neophodno je ograniciti minimalnu duzinu linije kako bi se zaobisli nevalidna resenja. Nakon sto smo otkrili lokaciju linije, posmatramo je kao linearu jednacina i izracunavamo koeficijente k i n , kako bi se ostvarila jednacina prave na kojoj linija lezi.

Svaki sledeci frejm se pretvara u grayscale format, gde se prilikom inicijalizacije novog formata analizira trenutni frejm u boji. Prolazi se kroz sve piksele stvarnog frejma, gde svi oni kojima su R,G,B vrednosti vece od 150 odredjuju belu boju na novoj slici, dok u suprotnom uzimamo crnu boju. Vrednost od 150 je utvrđena eksperimentalno. Svrha ovoga jeste kako bi se ostvarila grayscale slika na kojoj su primetni samo brojevi. Kako bi se uklonio eventualni sum, vrsi se dilatacija novonastale grazscale slike. Nad sredjenom grayscale slikom se traže konture, i potom postavljaju granicni pravougaonici.

Konacno, prolazimo kroz sve prepoznate pravougaonike na sredjenom grayscale frejmu, i transformisemo ih u testni podatak, izdvajanjem piksela sa frejma i transformacijom u kvadrat 28×28 piksela od kojeg nastaje jednodimenzionalni niz. Prepoznavanje transformisane slike radi se pomocu prethodno obucenog Support Vector Machine algoritma.

Za svaku prepozнату cifru se trazi centar lokacije u frejmu, kako bi se utvrdilo da li cifra lezi na gore utvrđenoj linearnej jednacini. Buduci da preklapanje cifre i linije moze eventualno da prouzrokuje vrstu suma, cifre se uzimaju u obzir tek nakon sto predju liniju i odmaknu 5 piksela dole-desno. Ukoliko su svi uslovi zadovoljeni, cifra se sumira sa ostatkom takvih cifara. Resenje se konacno snima u datoteku out.txt

Zakljucak

Tacnost resenja je validirana obezbedjenom skriptom koja pored ostvarene sume po video snimcima, i tacne (stvarne) sume svih video snimaka. Pokretanjem validate skripte saznajemo da resenje nudi tacnost od 86%. Greske pri prepoznavanju nastaju usled mnostva faktora, od kojeg izdvajamo ljudsku gresku prilikom prikupljanja test podataka, neoptimalan nacin otklanjanja suma, kao i zanemarivanje granicnih slucajeva kao sto su prepoznavanje cifre sa poslednjeg frejma koja nije presla "dovoljno daleko" kako bi se ostvarila razlika od 5 piksela. Uz optimizaciju resenja verujem da je moguce ostvariti vecu tacnost rezultata.