

Prepoznavanje Cifara na Video Snimku

Opis Problema

Dati su video snimci na kojima se nalazi plava linija. Cifre se po video snimku periodično kreću sa leva na desno, gore prema dole. Neke od tih cifara koje prelaze preko pomenute plave linije su od znaka. Potrebno je detektovati takve cifre, kao i izvršiti softversko prepoznavanje cifre o kojoj je reč. Cifre su rukom pisane, a budući da čovek neće nikada istu cifru napisati na isti način poput mašine, potrebno je obučiti 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 izvršava prepoznavanje cifara, odnosno u ovom smislu - slike dimenzija 28×28 , gde svaka od njih predstavlja detektovanu cifru sa video zapisa. Nakon detekcije slike cifre sa frejma video snimka, ona se transformise u jednodimenzioni niz grayscale vrednosti piksela koji je predstavlja.

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

Opis Resenja

Prvenstveno se svaki video snimak analizira frejm-po-frejm, gde se nad prvim frejmom koristi Houghova Transformacija kako bi se očitala plava linija. Budući da je implementirana OpenCv funkcija HoughLinesP, neophodno je ograničiti minimalnu dužinu linije kako bi se zaobišla nevalidna rešenja. Nakon što smo otkrili lokaciju linije, posmatramo je kao linearnu jednačinu i izračunavamo koeficijente k i n , kako bi se ostvarila jednačina prave na kojoj linija leži.

Svaki sledeći 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 veće od 150 određuju 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 приметni samo brojevi. Kako bi se uklonio eventualni šum, vrši se dilatacija novonastale grayscale 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 jednodimenzioni niz. Prepoznavanje transformisane slike radi se pomoću prethodno obučenog Support Vector Machine algoritma.

Za svaku prepoznatu cifru se traži centar lokacije u frejmu, kako bi se utvrdilo da li cifra leži na gore utvrđenoj linearnoj jednačini. Budući da preklapanje cifre i linije može eventualno da prouzrokuje vrstu suma, cifre se uzimaju u obzir tek nakon što predju liniju i odmaknu 5 piksela dole-desno. Ukoliko su svi uslovi zadovoljeni, cifra se sumira sa ostatkom takvih cifara. Rešenje se konacno snima u datoteku out.txt

Zaključak

Tačnost rešenja je validirana obezbeđenom skriptom koja poredi ostvarene sume po video snimcima, i tačne (stvarne) sume svih video snimaka. Pokretanjem validacione skripte saznajemo da rešenje nudi tačnost od 86%. Greške pri prepoznavanju nastaju usled mnoštva faktora, od kojeg izdvajamo ljudsku grešku prilikom prikupljanja test podataka, neoptimalan način otklanjanja suma, kao i zanemarivanje granicnih slučajeva kao što su prepoznavanje cifre sa poslednjeg frejma koja nije presla "dovoljno daleko" kako bi se ostvarila razlika od 5 piksela. Uz optimizaciju rešenja verujem da je moguće ostvariti veću tačnost rezultata.