

Metode procjene gustoće mnoštva ljudi

Luka Banović, Matej Grcić, Mislav Jurić, Marin Sokol, Andrea Vodopija

27. siječnja 2019.

Poglavlje 1

Konceptualno rješenje zadatka

Osnovna ideja projekta je usporedba različitih načina ekstrakcije značajki iz slike u procjeni gustoće mnoštva ljudi. Metode ekstrakcije se uspoređuju na sljedeći način: iz skupa fotografija koje predstavljaju grupe ljudi različite gustoće generira se skup podataka, predstavivši svaku fotografiju kao skup značajki dobiven jednom od metoda. Nakon toga, odaberemo samo jedan model klasifikatora i uspoređujemo njegove ispitne rezultate dobivene nakon treniranja na svakom od dobivenih skupova značajki. Pretpostavka je da, ako klasifikator ostvaruje bolje rezultate nakon što je treniran nad skupom podataka dobivenim jednom metodom ekstrakcije značajki nego kad se treniran nad skupom dobivenim drugom metodom, tada je prva metoda pogodnija za ekstrakciju značajki za problem procjene gustoće gomile ljudi. Dakle, bolji rezultati klasifikatora – pogodnija metoda.

Dakako, uspješnost usporedbe ovisi i o odabranom klasifikatoru. U radu Rahmalan et al. (2006), gdje se se autori bavili navedenom problematikom, korištena je samoorganizirajuća mreža. Problem je što optimizacija ovog modela nije u zatvorenom obliku pa postoji mogućnost da će rezultati biti neupotreblijivi - u projektu se koristi relativno malen skup slika (60 primjeraka) što znatno umanjuje vjerojatnost da će postupak optimizacije konvergirati globalnom optimumu. Umjesto toga, odabran je SVM kao klasifikator jer je otporniji na navedene probleme. O bazi slika korištenoj u projektu bit će više rečeno u idućim poglavljima.

Metode procjene koje su uspoređuju su GLDM, MFD i TIOCM. Prve dvije metode predložene su u ranijim radovima (Marana et al. (1997), Marana et al. (1999)), a TIOCM je metoda koja je predstavljena u Rahmalan et al. (2006). U nastavku će biti objašnjen postupak pripremanja skupa slika za ekstrakciju značajki, detaljno predstavljene korištene metode i opisan način rada SVM-a.

1.1 Priprema skupa slika

Da bismo mogli koristiti pojedinu sliku u nekoj od metoda, potrebno ju je prilagoditi očekivanom ulazu za svaku od metoda. Zato, prvo svaku sliku iz skupa, koja je originalno u boji, transformiramo eliminirajući boje, ostavivši samo sliku s nijansama sive. Dobivši skup slika bez boje, generiramo i skup slika na kojima su samo obrisi likova sa slikom bez boje pomoću metode fazne kongruencije.

Fazna kongruencija (engl. *phase congruency*, Kovessy (2003)) postupak je pomoću kojeg se na slikama detektiraju obrisi oblika. Osnovna ideja je da, nakon što se signal razvije u Fourierov red, provjeri svaku komponentu tog razvoja, koja je sinusoida. Pokazuje se da su na mjestima gdje postoje obrisi, komponente razvoja u Fourierov red pokazuju visoku uređenost, odnosno u fazi su.

Ako je

$$F(x) = \sum_n A_n \cos(n\omega x + \phi_n)$$

razvoj signala u Fourierov red, tada se može definirati funkcija fazne kongruencije kao

$$PC(x) = \max_{\theta \in [0, 2\pi]} \frac{\sum_n A_n \cos(n\omega x + \phi_n - \theta)}{\sum_n A_n}.$$

Visoka vrijednost funkcije fazne kongruencije označava da su komponente razvoja u fazi, tj. ondje gdje je funkcija fazne kongruencije maksimalna, težinska standardna devijacija faznih kutova je minimalna.

Ovako definirana fazna kongruencija vrlo je osjetljiva na šum u signalu. Međutim, postoje unaprijedene gotove implementacije detekcije obrisa pomoću fazne kongruencije koje sadrže parametar kojim se odabire osjetljivost na šum. Zato je korištena gotova implementacija u MATLAB-u uz parametar $k = 10$ (preporučena vrijednost za slike s većom količinom šuma, pretpostavljena vrijednost je $k = 2$). Parametar k označava broj standardnih devijacija od srednje vrijednosti energije šuma gdje postavljamo prag za šum.

//DATI SLIKU S DETEKTIRANIM RUBOVIMA

1.2 GLDM

Prva metoda ekstrakcije je GLDM (engl. *Grey Level Dependency Matrix*). Osnovna metoda je konstrukcija matrice uvjetnih vjerojatnosti pojavljivanja para određenih nijansi sive, ovisno o udaljenosti d i kutu θ kojim su dva piksela razdvojena. Prema Rahmalan et al. (2006), korišteni parametri u ovom projektu su $d = 1$ i $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. Neka je G broj nijansi sive koje se nalaze na slici bez boje. Tada se definiraju 4 veličine koje se računaju iz dobivene matrice, a opisuju promatranu sliku - kontrast, homogenost, energija i entropija, uz redom oznake C , H , Eg i Et .

$$\begin{aligned}
C(d, \theta) &= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j)^2 f(i, j|d, \theta) \\
H(d, \theta) &= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{f(i, j|d, \theta)}{1 + (i-j)^2} \\
Eg(d, \theta) &= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} f(i, j|d, \theta)^2 \\
Et(d, \theta) &= - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} f(i, j|d, \theta) \log f(i, j|d, \theta)
\end{aligned}$$

Kontrast je veličina koja predstavlja količinu lokalno kolebanje nijansi sive, iz čega slijedi da njegova visoka vrijednost može značiti prisutnost bridova ili šuma na slici. Homogenost, s druge strane, mjeri glatkoću distribucije nijansi sive na slici, što znači da je otprilike obrnuto proporcionalna s kontrastom.

Energija mjeri uniformnost distribucije nijansi sive na slici pa će slike s manjim brojem nijansi sive imati veću energiju. I za kraj, entropija je veličina koja mjeri neuređenost na slici - slično kao što su kontrast i homogenost povezani, povezani su i energija i entropija.

Ekstrakcija značajki odvija se na sljedeći način: slike dimenzija 400×400 prelazi se pomičnim prozorom veličine 20×20 uz korak $r = 10$. Za svaku sliku iz pomičnog prozora, generiraju se 4 matrice GLDM, prošavši kroze sve kombinacije d i θ . Iz svake se matrice izračunaju opisane veličine. To znači da za svaku podsliku dobivamo 16 značajki, a na svakoj slici prolazimo kroz 1521 podsliku. Slijedi da sa svake slike ekstrahiramo $16 \cdot 1521 = 24336$ značajki koje je opisuju. Iz tog razloga, potrebno je reducirati dimenzionalnost dobivenih vektora značajki, o čemu će biti više riječi kod analize rezultata.

1.3 MFD

Druga metoda koja je korištena u projektu je MFD (Minkowski Fractal Dimension, ili još znano kao Minkowski-Bouligand Fractal Dimension). Prvi je put ova veličina korištena za procjenu gustoće mnoštva ljudi u radu Marana et al. (1999).

Korištena je gotova implementacija u MATLAB-u gdje se ova veličina računa tako da se, prvo, slika na kojoj su detektirani obrisi binarizira koristeći određeni prag. Nakon toga, za $R = 1, 2, 4, \dots, 2^P$ konstruira se kvadrat čija je duljina stranice upravo R i broji koliko je takvih kvadrata potrebno da se pokriju svi bijeli pikseli na binariziranoj slici. Ovdje je P najmanja broj za koji vrijedi $\max\{x, y\} \leq 2^P$, gdje su x, y dimenzije slike.

Ako se pronađene točke ucrtaju na log-log grafu ovisnosti broja potrebnih kvadrata za pokrivanje obrisa o veličini pojedinog kvadrata, dobije se graf približno sličan pravcu. Neka je nagib tog približnog pravca označen s S , tj.:

$$S = \frac{d \log N}{d \log R}.$$

Ako je S približno konstantan za neki raspon od R , tada je on fraktalna dimenzija slike. U ovom se projektu koristi pretpostavka da je upravo to slučaj. S se koristi kao jedina značajka kod ove metode.

Uspješnost ove metode ovisi o uspješnosti načina kojim se detektiraju obrisi; zato se vrlo često događa da ne daje dobre rezultate ako na slici, osim mnoštva ljudi, postoje i drugi obrisi koji se mogu detektirati, a ne postoji valjan način kako ih eliminirati, primjerice tekstura podloge scene. //TODO UBACI SLIKU S DETEKTIRANOM PODLOGOM

1.4 TIOCM

Posljednja metoda korištena u ovom projektu TIOCM, što je kratica za Translation Invariant Orthonormal Chebyshev Moments (Rahmalan et al., 2006). Općenito, momenti su globalni deskriptori, odnosno opisuju sliku u cjelini. S druge strane, lokalni deskriptori opisuju dijelove dijelove slike, primjerice ključne točke. Zato se globalni deskriptori koriste kada se slika obrađuje na nižoj razini, primjerice kod detekcije objekata. Za prepoznavanje objekata bi se koristili lokalni deskriptori.

Momenti koji se koriste u ovom projektu su diskretni Čebiševljevi momenti uz modifikaciju prema radu Rahmalan et al. (2006) koja ih čini invarijantnima na translaciju. Konkretno, Čebiševljev moment reda $m + n$ za binariziranu sliku s detektiranim obrisima veličine $N \times N$ se računa kao:

$$T_{mn} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} t_m(i - i_c) t_n(j - j_c) f(i, j)$$

gdje su (i_c, j_c) koordinate centroida. $t_m, m = 0, \dots, N-1$ je Čebiševljev polinom

definiran rekurzijom:

$$\begin{aligned} t_m(q) &= \alpha_1 q t_{m-1}(q) + \alpha_2 t_{m-1}(q) + \alpha_3 t_{m-2}(q) \\ t_0(q) &= N^{-\frac{1}{2}} \\ t_1(q) &= \frac{\sqrt{3}(2q+1-N)}{\sqrt{N(N^2-1)}} \end{aligned}$$

$$\begin{aligned} \alpha_1 &= \frac{2\sqrt{4m^2-1}}{m\sqrt{N^2-m^2}} \\ \alpha_2 &= \frac{(1-N)\sqrt{4m^2-1}}{m\sqrt{N^2-m^2}} \\ \alpha_3 &= \frac{(m-1)\sqrt{2m+1}\sqrt{N^2-(m-1)^2}}{m\sqrt{2m-3}\sqrt{N^2-m^2}} \end{aligned}$$

Pokazuje se da je ovaj moment invarijantan na translaciju zahvaljujući oduzimanju koordinata centroida od koordinata svakog piksela. U projektu su, kao značajke, korišteni Čebiševljevi momenti od nultog do drugog reda, odnosno $m+n \in [0, 5]$, prema spomenutom radu.

1.5 SVM

Kao što je već rečeno, u ovom je projektu kao klasifikator korišten SVM zbog svojih povoljnih karakteristika kod manjeg skupa podataka. Osnovni cilj kod SVM-a jest pronaći vektor težina \vec{w} koji na skupu za učenje maksimizira marginu - udaljenost između najbližih pripadnika zasebnih klasa. Formalno, decizijska funkcija SVM-a može se definirati kao:

$$g(\vec{x}) = \vec{w}^T \vec{x} + b$$

Da bi se našlo vektor \vec{w} koji maksimizira marginu, koristi se činjenica da je moguće skalirati vrijednost decizijske funkcije tako da je ona jednaka 1 za točke najbliže hiperravnini, iz čega slijedi da je potrebno maksimizirati:

$$r = \frac{1}{\|\vec{w}\|}, \text{ uz ograničenja: } d_i(\vec{w}^T \vec{x}_i + b) \geq 1, i = 1, \dots, N,$$

gdje su $d_i = 1$ ako je uzorak s pozitivne strane hiperravnine, a ako je s negativne, onda je $d_i = -1$. Maksimizator ovog izraza ujedno minimizira

$$J = \frac{1}{2} \vec{w}^T \vec{w}$$

uz već spomenuta ograničenja. Formulira se Lagrangeova funkcija

$$L(\vec{w}, b, \vec{\lambda}) = \frac{1}{2} \vec{w}^T \vec{w} - \sum_{i=1}^N \lambda_i [d_i(\vec{w}^T \vec{x}_i + b) - 1]$$

i zatim se ona minimizira. Optimalne vrijednosti Lagrangeovih multiplikatora dobivaju se prelaskom na dualni problem i maksimizacijom dobivenog izraza po multiplikatorima. Uz optimalne multiplikatore, optimalni vektor \vec{w} je jednak:

$$\vec{w} = \sum_{i=1}^N \lambda_i d_i \vec{x}_i$$

Odavde slijedi da je moguće decizijsku funkciju SVM-a formulirati i kao:

$$g(\vec{x}) = \sum_{i=1}^N \lambda_i d_i \vec{x}_i^T \vec{x} + b.$$

Ovakav model ima dvije bitne pretpostavke: linearan je i razredi su linearno odvojivi. Ako je potreban nelinearan SVM, to je vrlo lako postići zamjenom skalarnog umnoška iz dualne formulacije SVM-a s nekom jezgrenom funkcijom, ukoliko je ta funkcija Mercerova, odnosno predstavlja skalarni umnožak transformiranih ulaznih vektora.

SVM koji je robusniji na razrede koji nisu linearno odvojivi postiže se reformulacijom ograničenja koja moraju biti zadovoljena, dopuštajući određenu grešku:

$$\begin{aligned} d_i(\vec{w}^T \vec{x}_i + b) &\geq 1 - \epsilon_i \\ \epsilon_i &\geq 0 \end{aligned}$$

Ciljna se funkcija proširuje kaznom za uzorke u čijoj klasifikaciji postoji greška:

$$J = \frac{1}{2} \vec{w}^T \vec{w} + C \sum_{i=1}^N \epsilon_i$$

C je parametar kojim se određuje širina margine: njegovim povećavanjem više se kažnjavaju pogrešno klasificirani uzorci što marginu čini užom. Manji C više dopušta greške u klasifikaciji što čini marginu širom.

Bibliografija

Computing a fractal dimension with matlab: 1d, 2d and 3d box-counting.

<https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/13063/versions>

Pristup: 24. prosinca 2018.

Matlab and octave functions for computer vision and image processing.

<https://www.peterkovesi.com/matlabfns/PhaseCongruency/phasecong3.m>

Pristup: 24. prosinca 2018.

Peter Kovesi. Phase congruency detects corners and edges. In *DICTA*, January, 2003.

A. N. Marana, S. A. Velastin, L. F. Costa, and R. A. Lotufo. Estimation of crowd density using image processing. In *IEE Colloquium on Image Processing for Security Applications (Digest No: 1997/074)*, pages 11/1–11/8, March 1997.

A. N. Marana, L. Da Fontoura Costa, R. A. Lotufo, and S. A. Velastin. Estimating crowd density with minkowski fractal dimension. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, volume 6, pages 3521–3524, March 1999.

H. Rahmalan, M. S. Nixon, and J. N. Carter. On crowd density estimation for surveillance. In *2006 IET Conference on Crime and Security*, pages 540–545, June 2006.