

IZAZOVI PRI MODELOVANJU MAPE VIZUELNE PAŽNJE KOD ČOVEKA

Ana Gavrovska¹, Andreja Samčović²,

¹Univerzitet u Beogradu – Elektrotehnički fakultet, anaga777@etf.rs,

²Univerzitet u Beogradu – Saobraćajni fakultet, andrej@sf.bg.ac.rs

Sadržaj: *Standardne karakteristike mape vizuelne pažnje kod čoveka nije jednostavno definisati u ciju modelovanja. Različite promene vizuelne informacije usled poboljšanja kontrasta, tonskog mapiranja slika visokog dinamičkog opsega i osvetljenja generalno, utiču na usmeravanje pažnje. Objekti u pokretu, u centralnom delu scene, ljudi i drugi karakteristični objekti na slici značajno utiču na pažnju kod čoveka. U ovom radu su analizirani izazovi pri mašinskom modelovanju mapa vizuelne pažnje, kao i promene koje nastaju primenom tonskog mapiranja. Razmatranja vizuelnih mapiranja se smatraju značajnim pri razvoju novih vizuelnih tehnologija i industrijskih rešenja (video igre, autonomna vozila i slično).*

Ključne reči: *vizuelna informacija, ljudska pažnja, visoki dinamički opseg, detekcija ljudi, osvetljenje, modelovanje mape.*

1. Uvod

Vid je najznačajnije od pet ljudskih čula, budući da je odgovoran za skoro 80% svih informacija koje naš mozak dobija iz spoljnog okruženja. Jedan od važnijih ciljeva je interakcija sa okruženjem u kome živimo. U svakodnevnom životu suočavamo se sa hiljadama objekata, identifikujemo stotine lica, i prepoznamo brojne saobraćajne znake. Lakoća sa kojom obavljamo te zadatke ne ogleda se samo u jednostavnosti tih zadataka, već je i dokaz visoke razvijenosti ljudskog vizuelnog sistema HVS (*Human Visual System*). Ljudski vid pruža mnoštvo informacija mozgu iz spoljnog sveta i formira koherentno globalno vizuelno iskustvo.

Sa druge strane, računarska vizija (*Computer Vision*) predstavlja disciplinu koja za cilj ima da razume vizuelni svet koji ljudska vizija percipira kroz procesiranje, ekstrakciju i analizu informacija prisutnih na slikama. Računarska vizija se koristi u širokom spektru aplikacija. Mnoge aplikacije koje se koriste za grafiku, dizajn, kao i interakciju sa mašinom su bazirane na interpretaciji, odnosno modelovanju ljudskog vida. Tipične aplikacije računarske vizije uključuju: inteligentni video nadzor, navigaciju robota, medicinske slike, industrijsku kontrolu kvaliteta, video igre, a u poslednje vreme dobija svoje nezaobilazno mesto i u razvoju autonomnih vozila. Modelovanje uključuje

razmatranje metoda segmentacije slike i drugih metoda za obradu mirne slike i video zapisa, metoda za prepoznavanje objekata, kao i perceptivnog video kodovanja [1-5].

Sistemi za računarsku viziju inspirisani ljudskim vidom predstavljaju obećavajuću alternativu koja bi trebalo da dovede do robusnijih i još snažnijih rešenja. Da bi se razvila ova rešenja istraživači bi trebalo da odgovore na pitanje koje mehanizme uključuje ljudski vid koji su tako jednostavni za ljude, a još uvek kompleksni za računarske sisteme.

Jedno od važnijih pitanja koje se postavlja je na šta najpre prosečni posmatrač obraća pažnju na slici. Raspoložive referentne baze za testiranje uključuju mape dobijene vizuelnim posmatranjem. Njihovo modelovanje bi trebalo biti i konačni cilj mašinskog obučavanja. U ovom radu su analizirane mape pažnje (HAM - *Human Attention Map*) dobijene percepcijom na konkretnim primerima, kao i važniji postojeći modeli percepcije. Modeli percepcije služe za predikciju ljudske percepcije imajući u vidu i semantiku slike (*saliency map*).

Rad je organizovan na sledeći način. Nakon uvodnog dela, u drugom poglavlju je analizirana vizuelna pažnja i njen značaj. U trećem poglavlju ukratko su opisane neke od najvažnijih referentnih baza i tehnike koje se koriste za mapiranje vizuelne pažnje kod čoveka. Napravljen je i osvrt na promene nastale pri proširenju dinamičkog opsega. Četvrto poglavlje je posvećeno metodama koje se primenjuju u modelovanju mapa koje bi služile za predikciju ljudske percepcije. Rezultati eksperimentalne analize dobijene testiranjem predloženih metoda iz literature prikazani su u petom poglavlju. Najvažniji zaključci navedeni su u šestom poglavlju.

2. Značaj vizuelne pažnje i praćenje pogleda

Vizuelna pažnja (*human attention*) se odnosi na proces pomoću koga vizuelni sistem vrši odabir najvažnijih informacija iz neke scene. Merenje pažnje se dovodi u vezu sa pogledom (*gaze*) usmerenog ka pojedinim objektima ili delovima objekata u okviru scene. Tada se govori o mapama pažnje koje se dobijaju ljudskom percepcijom - HAM.

Koncept vizuelne pažnje može biti proširen na drugačiju vrsta mapiranja. Naime, generisani modeli za vizuelnu pažnju mogu biti matematički, računarski, odnosno mašinski modeli. Cilj razvoja ovakvih modela je predikcija vizuelne percepcije, a samim tim predviđanje pružanja konkretne informacije. Mape pažnje dobijene modelovanjem ljudske percepcije (*salient maps*) bi imale višestruke namene u različitim aplikacijama i sistemima, jer uzimaju u obzir poznavanje semantičkog značaja svakog piksela. Pikseli većeg intenziteta bi bili pridruženi onim delovima scene na koje se najviše obraća pažnja. Performanse savremenih modela vizuelne pažnje su ograničene u smislu preciznosti predviđanja i računarske kompleksnosti. Radi poređenja uspešnosti modela, trebalo bi ih uporediti sa merenjima koja odgovaraju realnoj percepciji.

Vizuelna pažnja se odnosi na sposobnost vizuelnih sistema da veoma efikasno selektuju najznačajnije podatke u sceni. U protekloj deceniji postoje značajna istraživanja u oblasti sistema za računarsku viziju zasnovanih na praćenju pokreta očiju (*eye-tracking*). Naime, praćenjem pokreta očiju i fiksacijom (*fixation*) pogleda na pojedine objekte može se opisati ljudska pažnja putem mapiranja. Performanse sistema zasnovanih na ljudskoj percepciji mogu biti poboljšane kroz procesiranje relevantnih informacija prisutnih u slikama i ignorisanjem irelevantnih (nevažnih) informacija. Relevantnost

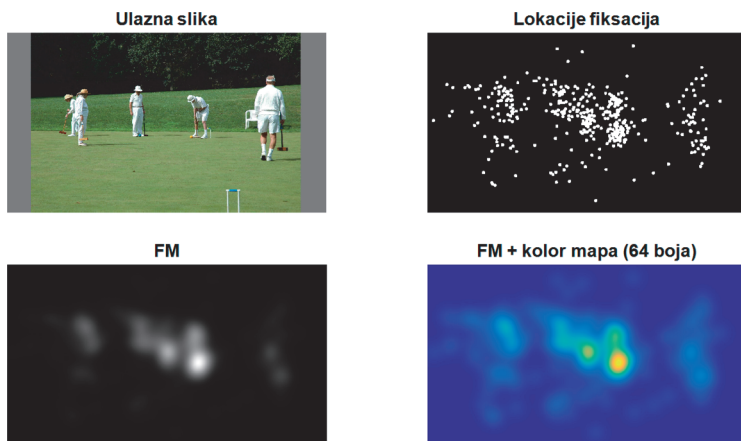
informacija se može direktno dovesti u vezu sa mapiranjem ljudske pažnje. Izborom relevantnih informacija na slici količina podataka koja bi trebalo da bude procesirana može biti znatno smanjena. To rezultira i smanjenom računarskom kompleksnošću, što je naročito važno za slike visoke rezolucije. Time je i olakšan rad u realnom vremenu [6-7].

Može se reći da vizuelni sistemi imaju zadatak da drastično redukuju količinu vizuelnih informacija koje moraju biti kvalitetno obrađene radi izvršavanja složenih zadataka. Jedan od uobičajenih zadataka bi moglo biti prepoznavanje oblika u realnom vremenu. Vizuelna pažnja omogućava vozačima, kao i autonomnim vozilima, da brzo identifikuju i lociraju potencijalne rizike ili značajne vizuelne događaje u okviru vidnog polja. To mogu biti, recimo: prelasci pešaka preko ulice, upotreba traka, kretanje vozila u neposrednoj okolini, promena signalizacije na semaforu i slično. Jedan pristup u modelovanju bi bio da se razviju predikcioni modeli u skladu sa velikom količinom podataka, pogotovo karakterističnim za kritične situacije. Upotreba videa za vozače u laboratorijskom okruženju i prikupljanje podataka o pokretima očiju može biti značajno za identifikaciju rizika i relevantnih vizuelnih događaja prilikom vožnje. Trebalo bi imati u vidu da bi se na taj način otkrila područja od interesa za vozače, ali da se pažnja u konkretnoj situaciji u laboratorijskim uslovima može razlikovati od slučaja realne situacije [5].

3. Mape pažnje dobijene ljudskom percepcijom

Vizuelna pažnja kod ljudi je povezana sa njihovim pokretima očiju. Postoji potreba da se obavljaju obimna merenja pomoću odgovarajuće opreme, i da se ta merenja adekvatno interpretiraju. Cena hardvera koji omogućava praćenje pokreta očiju je višestruko manja u odnosu na prethodni period. Sa pojavom *Google* naočara i pristupačnijim nosivim (*wearable*) sistemima za praćenje pokreta očiju, monitoring ljudske pažnje može postati sveprisutan. Nosivi sistemi bi trebalo takođe da obezbede tehnologije koje će omogućiti precizniju estimaciju ljudske pažnje, kao i upotrebu mapa u okviru 3D okruženja [2]. Većina eksperimentalnog mapiranja trenutno se obavlja pomoću prikaza vizuelne informacije na displeju.

Mape pažnje (HAM - *Human Attention Map*) se često nazivaju i mape fiksacije (FM - *Fixation Maps*) ili mape gustine fiksacije (FDM - *Fixation Density Maps*), jer se baziraju na tačkama fiksacije izmerene u toku procesa praćenja pokreta očiju (*eye-tracking*). Drugim rečima, mapa ljudske pažnje se generiše u dva koraka: pokreti očiju velikog broja volontera se snimaju dok gledaju određenu sliku u prvom koraku, dok se u drugom koraku ti snimljeni zapisi (tačke fiksacije) pretvaraju u mapu pažnje. Namenski hardver se koristi za registrovanje tačaka fiksacije (*fixation locations*), gde su zapisi u binarnom formatu. Najčešće se registruje redni broj fiksacije, dve prostorne koordinate koje odgovaraju lokaciji i vremenske odrednice fiksacije (početak, kraj, trajanje) [6]. Referentna mapa za dalje analize, FM, može se generisati konvolucionim pristupom (konvolucija sa Gausovim filtrom u skladu sa fiksacionim lokacijama), kako bi predstavljala kontinualno prikazivanje informacija o pokretima očiju. Na Slici 1 prikazane su fiksacije lokacije nad kojima je primenjena operacija dilatacije (strukturni element je disk prečnika 10, radi boljeg prikaza lokacija). Prikazana je i generisana mapa fiksacije za ulaznu sliku iz [8, 9], kao i jedna njena kolor interpretacija.



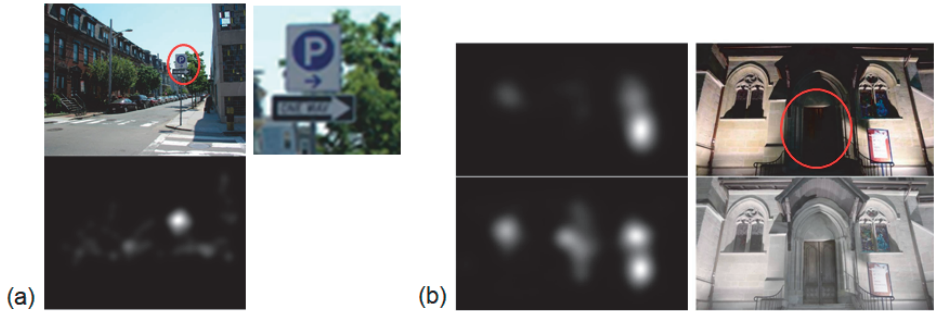
Slika 1. Lokacije fiksacije i mapa fiksacije (FM) za ulaznu sliku iz [9]

Vizuelna pažnja ne mora biti usmerena samo ka jednom objektu, već i ka drugim objektima u okviru scene. Naime, često postoji više objekata od interesa unutar jedne scene. Pažnja može biti usmerena samo ka pojedinim delovima objekta (npr. lica). Sa druge strane, sistem za praćenje pokreta očiju može da locira samo jednu lokaciju u sceni u datom trenutku. Na primeru sa Slike 1 se može videti da i pored blizine posmatraču određenih igrača, sama akcija je ta koja najviše privlači pažnju. Pažnja posmatrača je usmerena ka svim igračima, doduše nejednako. Prisutni su i sporedni pokreti pri analizi scene.

Dosadašnja testiranja su pokazala da posmatrači najčešće analiziraju lica, ljude i tekst, ali da i druge fiksacije postoje [10]. Tipični primeri su pojedini delovi tela na koje se posebno obraća pažnja (oči, ruke). Čak ne dolazi do fiksacije lica kao celine, već se posmatra samo određena količina podataka. Životinje, vozila, saobraćajni znaci, kulturni spomenici, karakteristični reljefi, takođe definišu područja od interesa. Na Slici 2 prikazan je primer usmeravanja pažnje ka saobraćajnom znaku. Jako osvetljeni delovi slike usmeravaju našu pažnju, pa i kontrast utiče na FM. Fiksacije se pri prelasku iz niskog u visoki dinamički opseg (*HDR - High Dynamic Range*), preraspodeljuju, generišući nove objekte od interesa, kao što je to slučaj sa vratima na Slici 2(b) [11-13]. Još jedna važna tendencija koja se ističe pri percepciji jeste usmeravanje pogleda ka centralnom delu slike i pored toga što se pojedinačni objekti od interesa ili konkretne akcije dešavaju van centralnog dela [9-10].

Ovakve mape služe kao polazna osnova za dalje analize. Evaluacija modela se sprovodi poređenjem sa FM dobijenim od strane većeg broja posmatrača. U bazi za testiranje HDR-a (*HDR-Eye* baza [11]) nalazi se 46 slika sa pratećim FM i pozicijama fiksacije snimljenim *Smart Eye Pro 5.8* opremom. Pri merenju je učestvovalo 20 posmatrača čiji je prosek godina oko 25. U bazi koja je korišćena u Judd et. al [10] nalazi se 1003 slučajno odabranih fotografija, koje su testirane od strane 15 posmatrača. Osim pozicija fiksacije i referentnih FM, na raspolaganju su i generisani modeli za predikciju (*salient* mape - SM). U bazi CAT2000 [9] nalaze se slike organizovane po 20 karakterističnih kategorija (*action, affective, art, ...*), koje se koriste kao stimulusi.

Ukupno 2000 trening FM slika dobijeno je od strane 18 posmatrača, a prate ih generisani podaci u vidu SM.



Slika 2. (a) Usmeravanje pažnje ka karakterističnim objektima (saobraćajni znak) i (b) promene mape pažnje usled visokog dinamičkog opsega

4. Mape pažnje dobijene modelovanjem ljudske percepcije

Svedoci smo sve većoj popularizaciji mašinskog učenja (*Machine Learning*), a u poslednje vreme i dubinskog učenja (*Deep Learning*). Uprkos impresivnom razvoju ove oblasti u poslednjoj deceniji trenutna rešenja računarske vizije su ispod mogućnosti HVS-a u pogledu robusnosti i performansi [14]. Istraživanja poslednjih godina koja se odnose na različite modele ljudske percepcije i definisanje metodologija za generisanje SM pokazala su da se i dalje zaostaje u modelovanju FM, kao i da se pažnja mašinskih pristupa značajno razlikuje od onoga što je rezultat direktne percepcije (FM) [14]. Na primer, dolazi do grešaka kada se pažnja usmerava ka onome što se drži u rukama, a ne toliko na lica koja drže predmet što se često dešava kod modela. Ako je veći intenzitet piksela usled dnevne svetlosti sa prozora, pažnja je prevashodno usmerena ka njima, iako se u centralnom delu nalazi krevet kao objekat, kome mašinski model daje veću prednost. I pored toga, smatra se da se došlo do maksimalno 88% tačnosti pri modelovanju vizuelne pažnje [10].

Različiti odabiri deskriptora se koriste u metodama [15-21] (metode: aim [15], aws [16], dva [17], gbvs [18], itti [19], sig [20], sun [21]). Na Slici 3 nalaze se primeri SM dobijenih različitim metodama [15-21]. Prikazane su SM za dve ulazne slike iz [22], gde se nalazi oko 3400 slika-stimulusa, od kojih su neke već preuzete iz postojećih izvora (doprinos slika-stimulusa (Pascal-S baza) je oko 850 slika). Test baze koje se koriste su zasnovane na fiksaciji i segmentaciji objekata (Bruce, Cerf, Ft, ImgSel, Judd, Pascal [10, 22-26]).



Slika 3. Primeri modelovanja SM pomoću sedam metoda [15-21].

Smatra se da je većina modela zasnovano na principu *bottom-up* analize [10]. To podrazumeva kombinovanje funkcija nižeg nivoa kao što su: kontrast, boja, intenziteti sive, kao i topološka struktura scene, orijentacija, simetrija i sl. *Bottom-up* karakteristike utiču na privlačenje pažnje posmatrača. Raspodela intenziteta sive, boje i orijentacija su sigurno jedna od osnovnih obeležja za modelovanje. Pored ovih *bottom-up* obeležja, koriste se danas i dodatna *top-down* obeležja koja utiču na privlačenje pažnje. Ova obeležja se dovode direktno u vezu sa kognitivnim faktorima, odnosno sa ciljevima posmatrača, motivacijom, željama i očekivanjima, kao i sa znanjem samog posmatrača. Zato su kognitivni aspekti i njihova analiza jedan od najvažnijih ciljeva pri modelovanju.

Ako se postavi konkretno pitanje/zadatak, može se ostaviti posmatraču da jedan po jedan deo izoštri (npr. povrti iz blura u prvobitno stanje izoštrjenog dela slike ili koristi *in-focus* i *out-of-focus* pristupe u analizi), kako bi došao do odgovora (npr. broja subjekata na sceni) [14]. Ako se postavi pitanje u vezi letećeg objekta, za očekivati je da će posmatrač prvo obratiti pažnju na gornji deo slike, odnosno deo slike koji pridružuje nebu. To znači da posmatranje može biti postavljeno u određeni kontekst. Veći broj obeležja i složeniji pristupi rezultiraju povećanjem računске složenosti. Postojeći modeli vizuelne pažnje sa visokom složenošću postižu bolju preciznost predikcije, ili pak slabiju predikciju uz bržu ekstrakciju relevantnih podataka iz slike.

Jedna od poznatijih metoda pri modelovanju SM je Judd et al. metoda [10], koja se zasniva na obeležjima niskog, srednjeg i visokog nivoa. Za obeležja niskog nivoa korišćena je piramidalna dekompozicija i podopsezi koji odgovaraju 3 skale i 4 orijentacije, kao i druga standardna ili već korišćena obeležja (histogramske reprezentacije, rastojanje do centra, *saliency* kanali [19], R/G/B kanali, kolor obeležja). Obeležja srednjeg nivoa, kao što su obeležja od značaja za detekciju horizonta, i obeležja visokog nivoa (za detekciju lica (*Viola-Jones*) i ljudi (*Felzenszwalb*)) dopunjuju obeležja niskog nivoa, za više informacija pogledati [10]. Još jedan prioritet je dodat u ovoj metodologiji, a to je prioritet centra, gde se uzima u obzir i daljina piksela od centra slike.

Razlog za to je što se najčešće kamera pozicionira u skladu sa objektom koji se snima, tako da je on relativno blizu centra u okviru scene. Korišćen je algoritam SVM (*Support Vector Machines*) za realizaciju modela na osnovu 903 trening slike i 100 test slika. Takođe, korišćena je i jedinstvena referentna SM, koja je dobijena zasebnim treniranjem za veliki broj slika (HGT SM - *Human Ground Truth Saliency Map*).

Osim modelovanja mapa vizuelne pažnje, poseban izazov predstavlja izbor metrika koje bi omogućile adekvatnu evaluaciju modela. Obično se za te svrhe koristi FM, ali i hijerarhijsko ispitivanje poklapanja sa referentnim modelom [9]. Kvantitativna analiza rezultata je važan zadatak, dok je kvalitativna, iako potrebna, zahtevna zbog velikih baza podataka. Jedan od problema kod metoda modelovanja na koji se nailazi je mogućnost snimanja lažno pozitivnih pogleda, recimo pogleda ka nevažnim područjima koje čine nebo, drveće i zgrade. Izazov je u tome kako da se odvoje lažno pozitivni pogledi od onih usmerenih ka konkretnom zadatku (npr. vožnji) [5]. Takođe, cena greške predikcije u određenim situacijama je praktično nepoznata. Aktuelni algoritmi za mašinsko učenje minimiziraju prosečnu grešku predviđanja, ne uzimajući u obzir kritične trenutke usled neadekvatnih interpretacija (npr. neprepoznavanje pešaka). Modeli vizuelne pažnje koje odlikuje visoka tačnost i preciznost, kao i niska složenost potrebni su i kod: procene kvaliteta slike i video zapisa, perceptualnog video kodovanja i u ostalim primenama gde postoji potreba za efikasnim smanjivanjem količine podataka zasnovanim na percepciji.

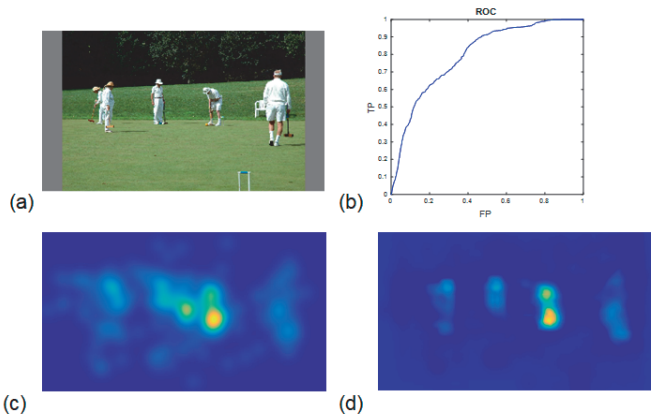
5. Eksperimentalna analiza mape pažnji

Izabrane su tri klase iz CAT2000 baze [9]: *Action*, *LowResolution* i *Social*. U prvoj klasi se nalaze primeri akcije, kao što je to slučaj sa golf igračima sa Slike 1 (0025.jpg). *LowResolution* klasa sadrži različite primere, dok su slike stimulusi neoštri. *Social* klasa poseduje isključivo primere gde su vidljiva lica i ljudi. Za svaku od slika iz klasa poređene su fiksacije i SM. Ukupno je testirano oko 300 slika stimulusa. Prvo su za svaku sliku iz baze izračunate ROC (*Receiver Operating Characteristic*) krive. Dobijeni TP (*True Positives*) i FP (*False Positives*) odgovaraju tome koliko je model mapiranja uspešan u odnosu na fiksacije. Za primer golfera, ROC kriva je prikazana na Slici 4, kao i korelaciona matrica između modela mape i merenih fiksacija. Izračunat je AUC (*Area under the Curve*) koji iznosi u ovom slučaju 0.8, dok je korelaciona mapa prikazana istom kolor mapom kao i FM radi kvalitativne analize. Izračunate vrednosti odnosa entropija (*entropy ratio*):

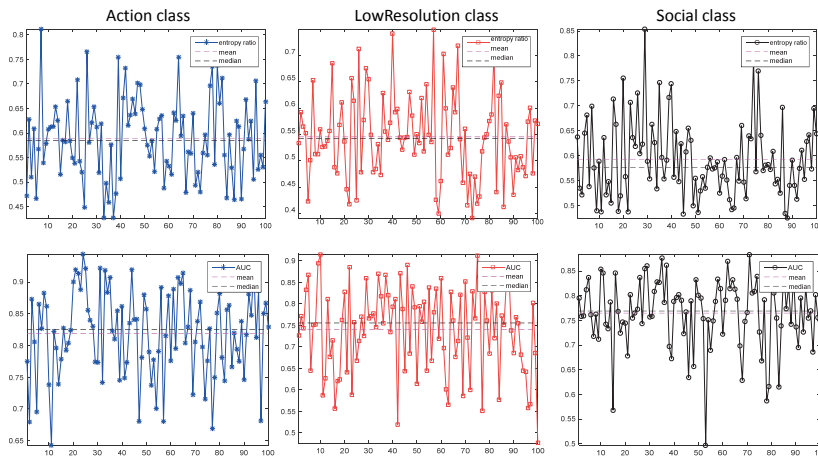
$$E = \text{entropy}(FM) / \text{entropy}(SM) \quad (1)$$

i AUC vrednosti prikazane su za sve tri klase na Slici 5.

Za testirane klase dobijene su sledeće vrednosti E : 0.5885 ± 0.0821 (medijana 0.5851), 0.5458 ± 0.0799 (medijana 0.5428), 0.5920 ± 0.0770 (medijana 0.5760), za prvu (*Action*), drugu (*Low Resolution*) i treću klasu (*Social*), respektivno. Vrednosti AUC su: 0.8186 ± 0.0680 (medijana 0.8253), 0.7399 ± 0.1002 (medijana 0.7559), 0.7647 ± 0.0699 (medijana 0.7690) za prvu, drugu i treću klasu, respektivno. Može se konstatovati da je mapiranje sa aspekta AUC najuspešnije za akciju, dok redosled uspešnosti za klase prati i odnos entropija.



Slika 4.(a) Slika iz klase Action (025.jpg), (b) ROC kriva, (c) FM, (d) korelaciona mapa.



Slika 5. Odnosi entropija (prvi red) i AUC vrednosti (drugi red) za tri klase

6. Zaključak

Performanse modela vizuelne pažnje utiču na rad sistema za računarsku viziju. Iako postoji zahtev za kompleksnim modelima, potrebni su računski manje kompleksni modeli koji će omogućiti da sistem računarske vizije radi u realnom vremenu. Zbog toga je važno razviti modele vizuelne pažnje koji imaju sposobnost otkrivanja *salient* regiona sa visokom tačnošću i preciznošću predikcije. Pri modelovanju potrebno je obaviti kognitivne eksperimente koji će omogućiti bolje razumevanje HVS-a i efikasnije generisanje SM. Osim ovakvih eksperimentalnih analiza, očekuje se u bliskoj budućnosti upotreba novih arhitektura mašinskog učenja koje će unaprediti efikasnost modelovanja.

Zahvalnica. Rezultati prikazani u ovom radu su dobijeni u okviru projekta TR32048.

Literatura

- [1] N. Ouerhani, et al., "Empirical validation of the saliency-based model of visual attention", *ELCVIA: electronic letters on computer vision and image analysis*, vol. 3, no. 1, pp. 13-24, 2004.
- [2] L. Paletta, et al., "3D attention: measurement of visual saliency using eye tracking glasses", in *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, ACM, pp. 199-204, 2013.
- [3] A. Kimura, R. Yonetani, and T. Hirayama, "Computational models of human visual attention and their implementations: A survey", *IEICE Transactions on Information and Systems*, vol. 96, no. 3, pp. 562-578, 2013.
- [4] N. Ouerhani, *Visual attention: from bio-inspired modeling to real-time implementation*. Université de Neuchâtel, Switzerland, 2004.
- [5] Y. Xia, et al., "Training a network to attend like human drivers saves it from common but misleading loss functions", *arXiv preprint arXiv:1711.06406*, pp. 1-15, 2017.
- [6] J. Chilukamari, *A computational model of visual attention*. Robert Gordon University, UK, 2017.
- [7] P. Zhang, *Towards Interpretable Vision Systems*. PhD diss., Virginia Tech, 2017.
- [8] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, "What do different evaluation metrics tell us about saliency models?", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1-24, 2018.
- [9] A. Borji, and L. Itti, "Cat2000: A large scale fixation dataset for boosting saliency research", *arXiv preprint arXiv:1505.03581*, pp. 1-4, 2015.
- [10] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look", in *12th IEEE international conference on Computer Vision*, pp. 2106-2113, 2009.
- [11] H. Nemoto, P. Korshunov, P. Hanhart and T. Ebrahimi, "Visual attention in LDR and HDR images", in *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Chandler, Arizona, USA, pp. 1-6, 2015. <http://mmspg.epfl.ch/hdr-eye> (last accessed 03.05.2018.)
- [12] U. Engelke, A. Maeder, and H. Zepernick, "Visual attention modelling for subjective image quality databases", in *IEEE International Workshop on Multimedia Signal Processing*, pp. 1-6, 2009.
- [13] A. Gavrovska, and A. Samčović, "Mogućnosti unapređenja kvaliteta slike i videa visokog dinamičkog opsega", *PosTel*, Beograd, str. 265-274, 2016.
- [14] A. Das, H. Agrawal, L. Zitnick, D. Parikh, and D. Batra, "Human attention in visual question answering: Do humans and deep networks look at the same regions?", *Computer Vision and Image Understanding*, 163, pp.90-100, 2017.
- [15] N. Bruce and J. Tsotsos, "Saliency based on information maximization", in *Advances in neural information processing systems*, pp. 155-162, 2005.
- [16] A. Garcia-Diaz, V. Leboran, X. R. Fdez-Vidal, and X. M. Pardo, "On the relationship between optical variability, visual saliency, and eye fixations: A computational approach", *Journal of Vision*, vol. 12, no. 6, pp.1-17, 2008.
- [17] X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments", in *Advances in neural information processing systems*, pp. 681-688, 2008.

- [18] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency”, in *Advances in neural information processing systems*, pp. 545-552, 2006.
- [19] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, 1998.
- [20] X. Hou, J. Harel, and C. Koch, “Image signature: Highlighting sparse salient regions”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no.1, pp. 194-201, 2012.
- [21] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, “Sun: A bayesian framework for saliency using natural statistics”, *Journal of Vision*, vol. 8, no. 7, pp.32-32, 2012.
- [22] *The Secrets of Salient Object segmentation*, <http://cbi.gatech.edu/salobj/> (last accessed 11.09.2018.)
- [23] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency-tuned salient region detection”, in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, 2009.
- [24] A. Borji, D. N. Sihite, and L. Itti, “What stands out in a scene? A study of human explicit saliency judgment”, *Vision research*, 91, pp. 62–77, 2013.
- [25] M. Cerf, J. Harel, W. Einhauser, and C. Koch, “Predicting human gaze using low-level saliency combined with face detection,” in *Advances in neural information processing systems*, pp. 241–248, 2008.
- [26] J. Li, M. D. Levine, X. An, X. Xu, and H. He, “Visual saliency based on scale-space analysis in the frequency domain,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 996–1010, 2013.

Abstract: *Standard characteristics of a human visual map for the purpose of modelling cannot be easily defined. Different changes in visual information can be found due to contrast enhancement, tone mapping of high dynamic range images and changes in lighting in general, and affect the direction of attention. Moving objects, objects in the central part of a scene, people and other specific objects significantly influence human attention. This paper analyzes the challenges with machine modelling of visual attention maps, as well as the application of tone mapping methods. Consideration of visual mapping, like with saliency maps, is important in the development of new visual technologies, as well as for the development of games and the industry of autonomous vehicles.*

Keywords: *visual information, human attention, high dynamic range, human detection, lightning, saliency map.*

CHALLENGES IN MODELING OF VISUAL HUMAN MAP ATTENTION

Ana Gavrovska, Andreja Samčović