Plan upravljanja istraživačkim podacima

| Opće informacije |
| --- |
|  | Ime i prezime predlagatelja  | Prof.dr.sc. Robert Cupec |
|  | Matična organizacija | Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek |
|  | Naziv projekta | COSPER (COmplex Scene PERception)Robotska percepcija stvarnog svijeta: razumijevanje složenih scena (IP-2019-04-6819) |
|  | Upravitelj podacima | Robert Cupec, robert.cupec@ferit.hrMateja Hržica, mateja.hrzica@ferit.hrMatej Džijan, matej.dzijan@ferit.hrValentin Šimundić, valentin.simundic@ferit.hr |
| 1. | Prikupljanje podataka i dokumentacija |
|  | Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka) | Tijekom istraživanja, prikupit će se dva skupa podataka:1. COSPER – indoor
2. COSPER – storage space
3. HoDoor
4. HoDoor 2

Podatkovni skup COSPER – indoor:3D sintetički podatkovni skup generiran na temelju dva postojeća podatkovna skupa, SUN RGB-D [1] i ShapeNet [2]. Generirat će se neobojeni oblaci točaka s pripadajućim anotacijama za 10 klasa (kada, krevet, polica, stolica, radni stol, komoda, noćni ormarić, kauč, stol, zahodska školjka). Generirani podatkovni skup se sastoji od 20000 scena i zauzima približno 73 GB.Podatkovni skup COSPER – storage space:Ovaj podatkovni skup sastoji se od dvije vrste podataka:1. Računalno generiran skup podataka
2. Skup podataka snimljen RGB-D kamerom
	1. originalni zapisi
	2. računalno poravnate scene s dominantom horizontalnom ravninom.

Podaci iz kategorije 1 i 2b se sastoje od datoteka spremljenih u *npy* i *npz* obliku. Scene namještaja za skladištenje u unutrašnjim prostorima pohranjene su u obliku oblaka točaka spremljenih u *npz* formatu. Svaka scena ima pridruženu *npy* numeričku datoteku koja sadrži referentne podatke graničnih okvira prostora za skladištenje za pripadni oblak točaka. Dodatno svaka scena sadržava i *npz* datoteku koja sadrži informacije o glasovima potrebnim za treniranje neuronske mreže *VoteNet*. Dodatno, u kategoriji 2 scene unutrašnjih prostora su spremljene i u obliku mreže trokuta u *ply* formatu. Podaci u kategoriji 2b sadrže i dubinske slike scena s označenom dominantom horizontalnom ravninom u *png* formatu te rotacijsku i translacijsku matricu u *npy* formatu potrebnu za transformaciju mreže trokuta iz kategorije 2a u 2b.Procjenjujemo da će za kategoriju 1 biti potrebno otprilike 11 GB, a za kategoriju 2 otprilike 4,5 GB prostoraPodatkovni skup HoDoor:Skup podataka u RGB-D formatu ormara i ladica u složenim scenama s označenim omeđujućim okvirima i stvarnim informacijama o načinu otvaranja jednih vrata ili jedne ladice (guranje, povlačenje i rukohvat). Slike su snimljene pomoću Intel RealSense L515 LiDAR dubinske kamere i pružene su iz iste perspektive s i bez ljudske ruke kako bi se demonstriralo otvaranje.Podatkovni skup HoDoor 2:Skup podataka u RGB formatu. Sastoji se dijela slika bez ljudske ruke iz HoDoor podatkovnog skupa. Na slikama su označeni granični okviri svih ladica i vrata prisutnih na slici. Granični okviri su poravnati s plohama na kojima se ladice i vrata nalaze.  |
|  | Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka) | Podatkovni skup COSPER – indoor:Korištenjem anotacija iz SUN RGB-D [1] podatkovnog skupa, generirat će se sintetički podatkovni skup stvaranjem 3D scena unutrašnjih prostora, renderiranjem tih scena virtualnom 3D kamerom te pretvaranjem dobivenih dubinskih slika u oblake 3D točaka. Scene će se generirati umetanjem 3D modela iz podatkovnog skupa ShapeNet [2] u granične okvire predmeta pripadajućih kategorija. Modeli će se birati s ciljem održavanja sličnih proporcija te će biti skalirani na veličinu graničnih okvira. Na scene će se dodati zidovi, čiji granični okviri nisu dostupni u originalnom podatkovnom skupu, već će se generirati na temelju dostupnih 2D anotacija, te ravan pod na dnu scene. Kako bi generirani oblaci točaka bili realističniji, renderiranim dubinskim slikama će se, prije pretvaranja u oblake točaka, dodati simulirane nesavršenosti kamere. Organizacija podataka je pobliže objašnjena u README datoteci priloženoj uz podatkovni skup.Podatkovni skup COSPER – storage space:Podaci iz kategorije 1 generiraju se automatizirano računalnim putem procedure objavljene u [1].Podaci iz kategorije 2a snimaju se 3D kamerom ASUS Xtion Pro Live, zatim se obrađuju računalnim putem procedurom objašnjenom u [3]. Nakon toga se dobivaju podaci u kategoriji 2b te se nad tim podacima ručno označuju granični okviri prostora za skladištenje pomoću programa izrađenog u [3]. Skup podataka pratit će i README dokumenti u kojima će se opisati hijerarhija direktorija.Podatkovni skup HoDoor:Podaci su prikupljeni snimanjem raznih prostora koji uključuju prostorije Fakulteta Elektrotehnike Računarstva i Informacijskih Tehnologija Osijek, osobne prostore suradnika na projektu te izložbene primjerke kuhinja, ormara i sličnih namještaja u prodavaonicama namještaja. Snimanje prostora suradnika na projektu je snimljeno uz njihovo dopuštenje, a snimanje podataka u prodavaonicama namještaja je učinjeno uz pismeno dopuštenje djelatnika. Za svaku scenu ručno su označeni granični okviri glavnih vrata ili glavne ladice, odnosno onih koji su u fokusu te snimke. Također je za svaki granični okvir označena klasa tih vrata ili ladice prema načinu otvaranja (guranje, povlačenje i rukohvat). Podatkovni skup je pobliže opisan u README datoteci priloženoj uz podatkovni skup.Podatkovni skup HoDoor 2:Podaci iz podatkovnog skupa HoDoor su nadograđeni označavanjem graničnih okvira svih ladica i vrata na svakoj od scena bez ljudske ruke. Označeni su i granični okviri poravnati s ravninom ladica ili vrata, odnosno označene su segmentacije na razini piksela za svaku ladicu i vrata. Anotacije su dostupne u coco formatu označavanja. Opis podatkovnog skupa i oznaka dostupan je u README datoteci koja je priložena uz podatkovni skup. |
|  | Koju ćete dokumentaciju i metapodatke ustupiti osim podataka? (navedite koje su informacije potrebne korisnicima kako bi mogli čitati i interpretirati podatke u budućnosti te koji će se standardi koristiti pri tumačenju podataka) | Skup ovi podataka pohranit će se u odabranom repozitoriju, popraćeni s README dokumentom sa sadržajem svih datoteka i opisom konvencije korištene za njihovo imenovanje. |
| 2. | Pravna i sigurnosna pitanja |
|  | Jeste li ograničeni sporazumom o povjerljivosti? Imate li potrebna dopuštenja za prikupljanje, obradu, čuvanje i dijeljenje podataka? Jesu li osobe čiji se podaci pohranjuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne kategorije osobnih podataka)? | Pri izvedbi ovog projekta neće se kršiti etička načela. |
|  | Kako će se regulirati pristup podacima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka?  | Podaci će se pohraniti u centraliziranom sustavu za pohranu Puh (<https://puh.srce.hr>). Pristup podacima upravlja se preko identiteta ustanove koji je siguran sustav i slijedi najbolje prakse u pogledu upravljanja identitetom.  |
|  | Kako ćete upravljati zaštitom autorskih prava i intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka? | Ne očekuje se da će rezultat istraživanja dovesti do patenta. Ostali problemi intelektualnog vlasništva će se u rješavati prema preporukama institucije Fakulteta elektrotehnike, računarstva i informacijskih tehnologija Osijek. Budući da podaci nisu podvrgnuti ugovoru, te se neće patentirati, objavit će se kao otvoreni podaci pod licencijom Creative Commons CC0. |
| 3. | Pohrana i čuvanje podataka |
|  | Kako će podaci biti pohranjeni i kako će biti napravljena sigurnosna kopija podataka (*backup*) tijekom istraživanja? Koji su kapaciteti čuvanja podataka kojim raspolažete? Kojim se procedurama koristite za sigurnosnu kopiju (*backup*)? | Podaci će se tijekom istraživanja s računala glavnog istraživača kopirati u sustav za pohranu Puh (<https://puh.srce.hr>) koji članovima projektnog tima omogućava pristup aktualnoj verziji podataka. Uz to, glavni istraživač posjeduje sigurnosnu kopiju podataka. |
|  | Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati? | Podatke ćemo čuvati trajno u centraliziranom sustavu za pohranu Puh (<https://puh.srce.hr>). Podatkovni skup COSPER – indoor:Podatke ćemo čuvati trajno u centraliziranom sustavu za pohranu Puh (<https://puh.srce.hr>). Anotacijski podaci se čuvaju u NPY i NPZ obliku, 3D scene u obliku oblaka točaka spremljene su u NPZ obliku, a u JSON obliku se čuvaju meta podaci o originalnim scenama i modelima.Podatkovni skup COSPER – storage space:Numerički podaci se čuvaju u NPY obliku, 3D scene u obliku oblaka točaka spremljene su u NPZ obliku, a u obliku mreže trokuta u PLY obliku.Podatkovni skup HoDoor:Podatke ćemo čuvati trajno u centraliziranom sustavu za pohranu Puh (<https://puh.srce.hr>). RGB i dubinske slike su spremljene u PNG formatu. Anotacijski podaci se čuvaju u yolo TXT formatu anotacije.Podatkovni skup HoDoor 2:Podatke ćemo čuvati trajno u centraliziranom sustavu za pohranu Puh (<https://puh.srce.hr>). RGB slike su spremljene u PNG formatu. Anotacijski podaci se čuvaju u coco JSON formatu anotacije. Također su dostupni i dodatni PLY (mesh) i YML podaci koji su korišteni za anotiranje podataka.  |
| 4. | Dijeljenje i ponovna uporaba podataka |
|  | Kako i gdje će se podaci dijeliti? Na kojem repozitoriju planirate dijeliti podatke? Kako će potencijalni korisnici doznati za podatke? | Konačnu verziju skupa podatka voditelj projekta podijelit će putem za pohranu Puh (<https://puh.srce.hr>). Podaci će biti objavljeni pod CC0 licencom. |
|  | Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Ne postoje takvi podaci. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima *FAIR-a*. | Potvrđujemo da se koristiti digitalnim repozitorijem koji je u skladu s načelima FAIR-a. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji održava neprofitna organizacija (ako ne, objasnite zašto ne možete dijeliti podatke na digitalnom repozitoriju koji nije komercijalan). | Potvrđujemo da se koristiti digitalnim repozitorijem koji održava neprofitna organizacija (CARNET) |

Ref:

[1] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva Learning Deep Features for Scene Recognition using Places Database Advances in Neural Information Processing Systems 27 (NIPS2014)
[2] Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., & Yu, F. (2015). ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR] Stanford University — Princeton University — Toyota Technological Institute at Chicago.

[3] Hržica M, Pejić P, Hartmann Tolić I, Cupec R. Detection of Household Furniture Storage Space in Depth Images. Sensors. 2022; 22(18):6774. <https://doi.org/10.3390/s22186774>