

Računalniški vid in inteligentna robotika za napredne oblike komuniciranja

Luka Čehovin Zajc¹, Matej Dobrevski¹, Peter Mlakar¹, Matej Račič¹,
Bine Rovansek¹, Jurij Slabanja¹, Rok Štemberger², Maks Vrščaj¹, Lojze Žust¹,
Urša Vodopivec³, Urška Saletinger⁴, Matej Kristan¹, Danijel Skočaj¹, Vesna Žabkar²

¹Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, Ljubljana

²Univerza v Ljubljani, Ekonomska fakulteta, Kardeljeva ploščad 17, Ljubljana

³Tehniški muzej Slovenije, Tržaška cesta 2, Ljubljana

⁴Infinum si, računalniške storitve, d.o.o., Litostrajska cesta 56, Ljubljana

E-pošta: luka.cehovin@fri.uni-lj.si

Computer vision and intelligent robotics for advanced forms of communication

The project, presented in this paper, investigates the possibilities of using computer vision and intelligent robotics solutions to interact with customers and visitors in public spaces, such as fairs, expos, and museums. The project focused on three diverse existing platforms, a multi-touch surface, a mobile robot, and Android mobile phones that were extended with functionalities that would enable such interaction. We have also conducted a study of commercial viability of using such technologies in interaction context.

1 Uvod

V moderni družbi, prenasičeni z informacijami, je potrebno razviti nove inovativne poti za napredne oblike komuniciranja med ponudniki informacije in porabniki. Na primer, podjetja želijo do potrošnikov pristopati na opazen, obenem pa čim bolj naraven in intuitiven način. Enako velja za obiskovalce sejmov in drugih prireditev ali razstav v muzejih in galerijah. Slednjim je treba omogočiti hiter in intuitiven dostop do bistvenih informacij o različnih eksponatih. Na podlagi svetovnih trendov sklepamo, da bodo inteligentni interaktivni vmesniki postali vedno pomembnejši v oglaševanju in obveščanju javnosti, zato smo v okviru projekta "InKom: Računalniški vid in inteligentna robotika za napredne oblike komuniciranja" preučili možnosti uporabe metod računalniškega vida ter inteligentne robotike za izboljšanje učinkovitosti ter uporabnikove izkušnje v komuniciranju s potrošniki in obiskovalci javnih prostorov.

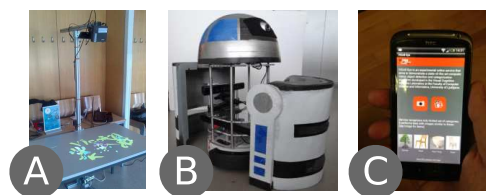
V okviru projekta je kot gospodarski partner nastopalo podjetje Infinum d.o.o. Gre za visoko tehnološko podjetje, ki se ukvarja z razvojem naprednih rešitev, tudi takih, ki imajo velik potencial v tržnem komuniciranju, tako da je projekt obogatilo z ekspertnim znanjem tako s tehnološkega kot s komunikološkega vidika. Kot družbena organizacija je bil v projekt vključen Tehniški muzej Slovenije, ki redno uporablja napredne metode komuniciranja in predstavljanja vsebin, katere pogosto gradi v sodelovanju s tehniškimi in naravoslovnimi fakultetami. Tako muzej predstavlja odličen poligon za definiranje kon-

kretnih zahtev in testnih aplikacij. Z razvitimi sistemi bi lahko obiskovalcem omogočili dostop do dodatnih informacij in poglobljeno razumevanje predstavljenih tematik tehniške dediščine.

2 Opis problema

V okviru projekta smo preučevali nove metode komuniciranja, podprte z modernimi visoko-tehnološkimi rešitvami. Slednje se bistveno razlikujejo od obstoječih uveljavljenih načinov komuniciranja s porabniki storitev in imajo zato potencial biti bolj opažene ter bolj usmerjene, s tem pa tudi učinkovitejše.

V okviru projekta smo preverjali uporabnost treh raznolikih senzorsko-robotskih platform v različnih scenarijih komuniciranja. Izbor je temeljil predvsem na cenovni ugodnosti strojne opreme, možnostih razširitve ter obstoju osnovnih gradnikov v okviru Laboratorija za umetne vizualne spoznavne sisteme na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Cilj projekta je bil obravnavane platforme nadgraditi oz. izboljšati z novimi funkcijami v smeri učinkovitega komuniciranja z ljudmi.



Slika 1: Preučevane platforme: A - interaktivna večdotična površina, B - mobilni robot, C - mobilni telefon.

Prva platforma je napreden interaktivni sistem, sestavljen iz projektorja in barvno-globinske kamere, ki lahko poljubno ravno površino ter prostor nad njo spremeni v intuitiven in zmogljiv **večdotični uporabniški vmesnik** [1]. Osnovni sistem je prikazan na Sliki 1A. Vizualna vsebina se s projektorjem projicira na projekcijsko površino, ki jo opazuje kamera. Uporabnik lahko nadzira računalniški sistem in spreminja projicirano vizualno vsebino preko kretenj, ki jih kamera zaznava in pretvarja v ukaze za spreminjanje slike. Na ta način dobimo interak-

tivno večdotično površino, ki poleg zaznavanja dotikov omogoča tudi nadzor v 3D prostoru nad površino. Sistem lahko z vertikalno postavitvijo projekcijske površine uporabimo tudi kot interaktivno tablo. V okviru projekta smo hoteli sistem še dodatno nadgraditi, poenostaviti njegovo namestitve in uporabo ter robustificirati njegovo delovanje in izboljšati uporabniško izkušnjo pri dolgoročni uporabi v javnem prostoru.

Druga platforma je robotski sesalec iRobot Roomba, nadgrajen v **avtonomnega mobilnega robota** z možnostjo opravljanja različnih opravil, vključno s premikanjem po prostoru, zaznavanjem okolice in preproste verbalne in neverbalne komunikacije (Slika 1B). Takšen robot je v scenarijih komuniciranja z obiskovalci muzejev ali sejmov lahko zelo uporaben za popestritev dogajanja, kot npr. vodnik po dogajanju, za nudenje dodatnih informacij ali zgolj kot atrakcija oz. vaba za povečanje zanimanja. Običajni mobilni roboti so za take namene občutno predragi. Ker pa smo v projektu uporabljenega robota zgradili iz naprav za široko potrošnjo, je njegova cena veliko nižja. Takega robota na Fakulteti za računalništvo in informatiko že več let uspešno uporabljamo v pedagoškem procesu [2], za uporabo v prostoru z večjim številom ljudi pa ga je bilo potrebno nadgraditi z možnostjo robustnega navigiranja v premikajoči se množici ter s sposobnostjo zaznave in ogovarjanja ljudi v bližini.

Tretja platforma je pametni telefon z operacijskim sistemom Android, ki v kombinaciji z visokozmogljivo kamero, ki jo imajo danes že vsi telefoni, predstavlja močan **mobilni zaznavni sistem** (Slika 1C). Tak sistem lahko v kombinaciji z naprednimi zaznavnimi algoritmi s področja računalniškega vida predstavlja enkraten medij za implementacijo inovativnih vizualno podprtih načinov komuniciranja, na primer prepoznavanja objektov na sliki ter izkušnjo obogatene resničnosti. Tako smo si za prvi cilj postavili učinkovito implementacijo razpoznavanja predmetov na slikah, ki jih uporabnik zajame s kamero mobilne naprave. Uspešno razpoznan predmet lahko nato sproži nadaljne akcije, kot so nudenje dodatnih informacij in podobno. Tak način povezovanja na dodatne vire informacij je veliko bolj prijazen, naraven in nevpadljiv kot npr. uveljavljena uporaba QR kod. Kot drugi cilj smo si zastavili izdelati splošno podporno tehnologijo za uporabo obogatene resničnosti na mobilnih napravah. Glavna ideja obogatene resničnosti je, da se na živi sliki, ki se zajema s kamero, detektira določene oznake. Te oznake definirajo ravnino na katero se lahko nato s pomočjo računalniške grafike izriše poljubni 3D model. Ob premiku kamere tako ta model vizualno ostane pritrjen na ravnino, kar uporabniku da zelo realistično izkušnjo obogatene resničnosti. Običajno so oznake, ki se uporabljajo za detekcijo oz. sledenje ravnini, v naprej določene in umetno namensko izdelane. Cilj tega projekta je bil omogočiti uporabo poljubnih oznak, ki se pojavljajo na ravnih površinah v okolju, kar je z vidika komuniciranja veliko prijaznejše in bolj naravno.

Poleg same implementacije algoritmov in nadgradnje senzorskih in robotskih platform smo analizirali tudi uporabnost platform s trženjskega vidika. Zanimalo nas je,

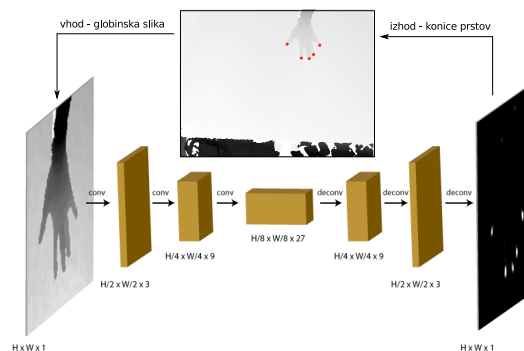
kako odločevalci v muzejih in galerijah ter organizatorji sejmov in drugih prireditelj sprejemajo takšne tehnologije in kakšne so omejitve delovanja, ki so za njih in uporabnike storitev še sprejemljive ter, kakšen je njihov tržni potencial in ekonomska upravičenost njihove uporabe za gospodarske in negospodarske organizacije.

3 Rezultati projekta

V tem poglavju bomo opisali rezultate projekta, razvite algoritme in prototipe ter glavne zaključke študije tržnega potenciala.

3.1 Večdotična površina

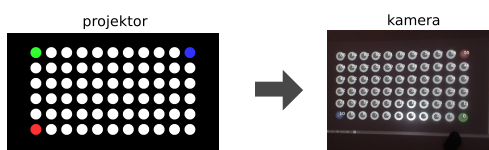
V okviru izboljšanja robustnosti interaktivne večdotične površine smo naslovili tri probleme. Prvi je bil nezanesljiva detekcija konic prstov v obstoječem sistemu, saj je algoritem za zaznavo temeljil na ročno nastavljenih pragovnih vrednostih. Cilj je bil zaznava konic, iz globinske slike, pridobljene z barvno-globinsko kamero Kinect, z metodami strojnega učenja, ter tako izboljšati obstoječi algoritem. Problema smo se konceptualno lotili kot segmentacije slike: za vsak slikovni element globinske slike skuša sistem avtonomno napovedati ali pripada konici prsta ali ne. Pripravili smo sintetično učno množico umetno generiranih slik rok, na kateri smo učili globoko konvolucijsko nevronske mreže za segmentacijo [3]. Iz segmentirane slike se nato izločijo koordinate konic prstov. Pristop smo preizkusili tudi na množici realnih globinskih slik in dobili zelo dobre rezultate. Pristop je prikazan na Sliki 2.



Slika 2: Zaznava konic prstov v globinski sliki s konvolucijsko nevronske mreže.

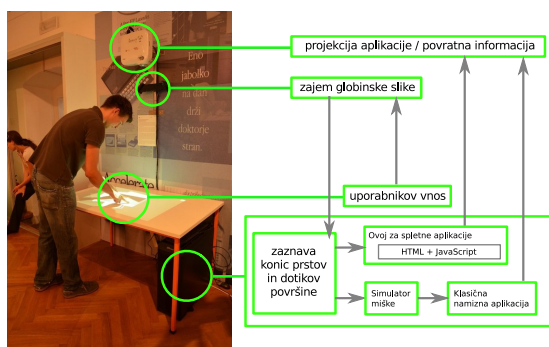
Drugi naslovljeni problem je geometrijska relacija med detektiranimi prsti in koordinatami slikovne vsebine. Sistem namreč projicira sliko na površino in na površini zaznava prste v 3D. Te 3D koordinate je potrebno preslikati nazaj v izvirne koordinate v sliki pred projekcijo, saj so v tem prostoru definirani elementi kot so gumbi in okna s katerimi preko projekcije uporabnik interaktira. Izračun geometrijskih relacij zahteva razvoj postopkov samodejne kalibracije ob zagonu ter po potrebi med delovanjem. Uspešnost je odvisna od mnogih dejavnikov v okolju, npr. svetlosti okolja ali odsevnost večdotične površine, zato smo se osredotočili predvsem na robustnost samodejnega postopka. Implementirani kalibracijski sistem omogoča prilagajanje parametrov generiranja kali-

bracijske predloge (Slika 3). Parametri so svetlost kalibracijskega vzorca, število elementov ter njihova velikost na vzorcu, vse skupaj pa je povezano v povratno zanko, ki samodejno išče optimalne parametre vzorca dokler ga ne zazna dovolj robustno.



Slika 3: Kalibracija sistema z referenčnim vzorcem, ki ga na površini ustvari projektor, detektira pa ga kamera.

Zadnji problem, ki smo ga obravnavali je razvoj aplikacij za večdotično površino. Z namenom, da bi bil ta čim bolj enostaven, smo razvili programski ovoj, ki omogoča razvoj aplikacij v tehnologijah spletnih strani (HTML in JavaScript). Poleg tega smo razvili tudi program, ki dotike prstov, zaznane s sistemom, pretvori v klike simulirane miške in s tem omogoča tudi delo z obstoječimi aplikacijami. Oba koncepta sta predstavljena na Sliki 4.



Slika 4: Aplikacija na večdotični površini.

3.2 Mobilni robot

Za mobilne robote je glavni problem pri navigiranju po prostoru polnemu ljudi nenehno spreminjanje njegove okolice, kar znižuje robustnost lokalizacije, to pa posledično otežuje njegovo premikanje. V okviru projekta smo navigacijski algoritem prilagodili tako, da se bolj zanaša primarno na lokalno oceno pozicije ter upošteva trenutno zaznane ovire. Ker v kontekstu interakcije z ljudmi v takem prostoru robotova globalna pozicija ni ključna, smo s takim pristopom dosegli, da se mobilni robot uspešno premika v množici, njegovo obnašanje pa je v splošnem bolj robustno.

Mobilni robot je bil v okviru projekta nadgrajen z dodatno kamero na premičnem stojalu (angl. pan-tilt), kar robotu omogoča, da bolj robustno zaznava in sledi ljudem v okolici. Zaradi hitrih premikov ljudi okoli robota in zaradi lastnega premikanja po prostoru je premična kamera v kombinaciji z modernim vizualnim sledilnikom [4] nujna, da robot dalj časa zadrži stik s posamezno osebo. V ta namen smo razvili algoritem za usmerjanje kamere ter krmiljenje robota. Sledenje je inicializirano na podlagi

predlogov, pridobljenih iz modula za zaznavanje obrazov, ki deluje na podlagi klasifikatorja SVM iz knjižnice Dlib [5].

V robotski sistem je bil integriran tudi sintetizator govora. Tako lahko robot, ko zazna osebo in ji začne slediti, to nakazuje tudi s poljubnim glasovnim pozdravom. Sistem je bil preizkušen v realnem scenariju v okviru dogodka Zotkini talenti v Cankarjevem domu, kjer je na ta način komuniciral z udeleženci v sprejemni dvorani, kot je prikazano na Sliki 5.



Slika 5: Nadgrajeni mobilni robot za interakcijo z ljudmi na dogodku Zotkini talenti.

3.3 Računalniški vid na mobilnih telefonih

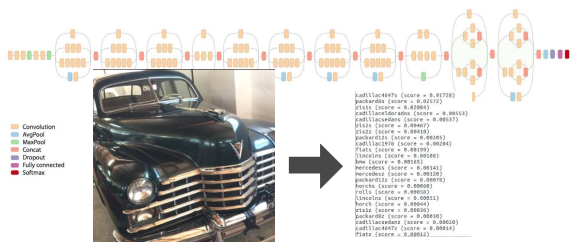
V okviru projekta smo razvili dve prototipni aplikaciji, ki naslavljata dve možnosti uporabe računalniškega vida. V okviru prve aplikacije smo implementirali podporni mehanizem za uporabo obogatene resničnosti, ki se izvaja na pametnih telefonih z operacijskim sistemom Android. Cilj je bil uporaba naravnih oznak, kot so razglednice in posterji, namesto klasičnih črno-belih kvadratnih oznak. Algoritem za razpoznavo temelji na programski knjižnici OpenCV [6], uporablja iskanje ujemanja s predlogo na podlagi detektorja značilnih točk ORB [7]. Aplikacija uporablja zbirko referenčnih slik (npr. razglednice, kot je prikazano na Sliki 6) in na njih detektirane značilnice. Ob uspešni detekciji, ki je prepočasna za uporabo v realnem času, sistem preklopi na ponovno detekcijo z optičnim tokom, kar omogoča sledenje v realnem času, ki pa je v fiksnih časovnih intervalih preverjeno s ponovno detekcijo objekta.



Slika 6: Aplikacija obogatene resničnosti na mobilnih telefonih.

Druga razvita aplikacija je namenjena detekciji in lokalizaciji ploščatih objektov v vnaprej naučeni zbirki pred-

metov. Uporabili smo knjižnico TensorFlow [8], v okviru katere je implementirano tudi razpoznavanje s konvolucijsko nevronske mreže [9]. Uporabljena je bila prednaučena mreža, ki pa jo je mogoče do-učiti in prilagoditi na posamezno domeno. V okviru projekta je bil izboljššan način učenja mreže na majhnem številu učnih primerov ter optimizacija nevronske mreže za izvajanje razpoznavanja na mobilnih napravah v skoraj realnem času (Slika 7). Del projekta je vključeval tudi zajem muzejske zbirke avtomobilov v Tehniškem muzeju Slovenije za namene učenja in testiranja.



Slika 7: Razpoznavanje znamke avtomobila iz zbirke TMS.

3.4 Analiza tržnega potenciala

V okviru projekta smo raziskali tudi možnosti trženja večdotične površine, mobilnih robotov in mobilnih aplikacij, podprtih z algoritmi računalniškega vida. Določili smo tržni potencial teh tehnologij na letni ravni. Osredotočili smo se na vpliv omenjenih tehnologij na sejmih in v muzejih ter na možnosti za obogatitev muzejske izkušnje in s tem povečanje letnega obiska. Na podlagi svetovnih trendov smo ugotovili, da bi z uporabo tehnologij na sejmih razstavljalci lahko pridobili prednost pred konkurenco, saj bi jih obiskovalci dojemali kot zanimivejše, inovativne, s tem pa bi pritegnili potencialne stranke. V ta namen je bil v času trajanja projekta opravljen pregled internetnih virov ter poglobljeni intervjuji s strokovnjaki, ki se ukvarjajo s komuniciranjem v muzejih (TMS) in sejemskimi predstavitvami. Muzeji v mednarodnem prostoru že uporabljajo tovrstne tehnologije; npr. 1% muzejev v ZDA uporablja elemente obogatene resničnosti znotraj predstavitev [10].

Pri naši analizi smo preverili tudi, kdo so ključni ljudje, ki odločajo o nakupu tovrstnih tehnologij v muzejih, in kdo bi lahko pozitivno vplival na nakup tovrstnih novih tehnologij. T.i. odločevalci imajo ključno vlogo pri odločitvi za nakup noviteta ali pa tudi za katero tematico in vsebino se bodo odločili ponuditi obiskovalcem v bližnji prihodnosti. Za izračun potenciala pri muzejih smo upoštevali število muzejev ter trajnost uporabe tehnologije. Potencial smo razdelili v kategorije večdotični uporabniški vmesnik, roboti in mobilni zaznavni sistem za potrebe muzejev in sejmov, na podlagi katerih smo opravili izračune. Izračuni so pokazali smiselnost oz. dovolj velik potencial za nadaljevanje aktivnosti, za začetek komercialne integracije omenjenih tehnologij pa priporočamo, da se delujoči modeli tehnologij preizkusijo na manjšem reprezentativnem vzorcu obiskovalcev muzeja,

na katerem se potem oceni odziv ter možnosti za nadaljnje korake v razvoju naprednih oblik komuniciranja.

4 Zaključek

V članku smo predstavili rezultate projekta, v okviru katerega smo preučevali možnosti uporabe računalniškega vida in inteligentne robotike v kontekstu naprednih oblik komuniciranja z ljudmi v javnih prostorih. Osredotočili smo se na tri platforme, večdotično površino, mobilnega robota in pametne telefone, vsako izmed platform pa smo nadgradili s funkcionalnostmi, ki so potrebne za njeno uporabo v omenjenem kontekstu.

Projekt je trajal pet mesecev in je izpostavil mnogo novih zanimivi problemov. Zato bomo sodelovanje med univerzo in organizacijami nadaljeval tudi po zaključku projekta. S podjetjem Infinum bomo tudi v prihodnje sodeloval pri adaptaciji tehnologij računalniškega vida in inteligentne robotike in iskanju možnosti njihove uporabe v komercialne namene na področju oglaševanja. Tehniški muzej Slovenije v jeseni pripravlja razstavo, ki jo načrtujejo podpreti z rešitvami, razvitimi v okviru predstavljenega projekta.

Zahvala

Projekt je bil financiran v okviru Javnega razpisa projektno delo z gospodarstvom in negospodarstvom v lokalnem in regionalnem okolju - po kreativni poti do znanja 2016/2017.



REPUBLIKA SLOVENIJA
MINISTRSTVO ZA IZOBRAŽEVANJE,
ZNANOST IN ŠPORT



JAVNI SKLAD REPUBLIKE SLOVENIJE
ZA RAZVOJ KADROV IN ŠTIPENDIJE



Nalozba v našo prihodnost
Ustanova za spodbujanje uporabe informacijske tehnologije

Literatura

- [1] Klemen Istenič, Luka Čehovin, and Danijel Skočaj. Multi-touch surface based on RGBD camera. In *HCI-IS 2014*, Oct 2014.
- [2] Luka Čehovin, Anže Rezelj, and Danijel Skočaj. Teaching intelligent robotics with a low-cost mobile robot platform. In *RiE 2015*, May 2015.
- [3] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, 2015.
- [4] A. Lukežič, L. Č. Zajc, and M. Kristan. Deformable parts correlation filters for robust visual tracking. *IEEE Transactions on Cybernetics*, PP(99):1–13, 2017.
- [5] Davis E. King. Dlib-ml: A machine learning toolkit. *JMLR*, 10:1755–1758, 2009.
- [6] G. Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [7] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: An efficient alternative to SIFT or SURF. In *ICCV2011*, pages 2564–2571. IEEE, 2011.
- [8] Martín Abadi, Ashish Agarwal, and Paul Barham et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception Architecture for Computer Vision. *ArXiv e-prints*, December 2015.
- [10] Mandy Ding. Augmented reality in museums. *Arts Management & Technology Laboratory*, 12, 2017.