

FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO
UNIVERZE V LJUBLJANI

Eva Stergaršek Kuzmič

Odkrivanje novih predmetov in grajenje njihovih
modelov z aktivnim robotskim sistemom

MAGISTRSKO DELO

Mentor: prof.dr. Aleš Leonardis
doc.dr. Aleš Ude

Ljubljana, 2010

Št.: 133-MAG-RI/2010
Datum: 17. 6. 2010



Eva STERGARŠEK KUZMIČ, univ. dipl. inž. el.

Ljubljana

Fakulteta za računalništvo in informatiko Univerze v Ljubljani izdaja naslednjo magistrsko nalogo

Naslov naloge: **Odkrivanje novih predmetov in grajenje njihovih modelov z aktivnim robotskim sistemom**

Detecting new objects and building models with active robot system

Tematika naloge:

Sposobnost odkrivanja novih predmetov je težko poustvariti s popolnoma pasivnimi vizualnimi sistemi brez predznanja. Ena od možnosti reševanja tega problema so aktivni robotski sistemi, ki poleg sistema za zajemanje in obdelavo slik vsebujejo elemente za manipulacijo delov prizorišča.

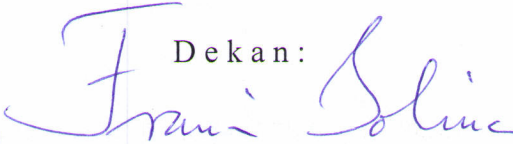
V okviru realnega robotskega sistema razvijte robusten pristop za odkrivanje neznanih predmetov v poljubnih prizorih, ki temelji na vizualnih lokalnih značilnicah. Izločene značilnice naj bodo vhod v sistem za robustno oblikovanje hipotez o obstoju predmetov. Način manipulacije določite na podlagi izbranih hipotez o obstoju predmeta, posledice manipulacije pa uporabite za robustno vrednotenje hipotez. Pokažite možnost uporabe tega pristopa za gradnjo modelov novih predmetov.

Mentor:


prof. dr. Aleš Leonardis



Dekan:


prof. dr. Franc Solina

Somentor:


doc. dr. Aleš Ude

IZJAVA O AVTORSTVU

magistrskega dela

Spodaj podpisana Eva Stergaršek Kuzmič,

z vpisno številko 63070478

sem avtorica magistrskega dela z naslovom:

Odkrivanje novih predmetov in grajenje njihovih modelov z aktivnim robotskim sistemom.

S svojim podpisom zagotavljam, da:

- sem magistrsko delo izdelala samostojno pod vodstvom mentorja prof.dr. Aleša Leonardisa in somentorja doc.dr. Aleša Udeta.
- so elektronska oblika magistrskega dela, naslova (slov., angl.), povzetka (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko magistrskega dela
- in soglašam z javno objavo elektronske oblike magistrskega dela v zbirki »Dela FRI«.

V Ljubljani, dne 23.9.2010

Zahvala

Zahvaljujem se mentorjema dr. Alešu Leonardisu in dr. Alešu Udetu. Zahvala gre tudi sodelavcem z odseka E-1 na inštitutu "Jožef Stefan", prijateljem in družini.

Kazalo

Povzetek	1
Abstract	3
1. Uvod.....	5
1.1. Struktura naloge.....	10
2. Pregled področja.....	11
3. Detekcija in učenje predmetov.....	15
3.1. Oblikovanje hipotez o obstoju predmeta.....	17
3.1.1. Pridobivanje 3-D informacij.....	18
3.1.2. Iskanje planarnih delov v prizoru	20
3.2. Potiskanje na osnovi izbrane hipoteze	23
3.3. Vrednotenje hipotez	26
3.4. Grajenje modelov.....	32
4. Metodologija	33
4.1. Iskanje in opis vizualnih značilnic.....	33
4.1.1. Odkrivanje ekstremov.....	34
4.1.2. Določanje lokacije	35
4.1.3. Določanje orientacije	36
4.1.4. Izgradnja opisnikov.....	36
4.2. Algoritem RANSAC	37
4.3. Gručenje točk.....	39
4.3.1. Algoritem K -means.....	39
4.3.2. Algoritem X -means.....	40
4.4. Razcep s singularnimi vrednostmi.....	43
4.5. Kalibracija kamere.....	43

5.	Eksperimenti	46
5.1.	Realni robotski sistem	46
5.2.	Preizkusi.....	48
5.2.1.	Oblikovanje hipotez	49
5.2.2.	Potiskanje.....	50
5.2.3.	Vrednotenje hipotez	50
5.2.4.	Grajenje modelov	55
6.	Zaključek	58
7.	Viri.....	61

Kazalo slik

Slika 1: Roboti za domača okolja.	6
Slika 2: Segmentacija neznanih predmetov.	7
Slika 3: Trije križi.	9
Slika 4: Nadzorovano učenje s sintetičnimi slikami z označenimi mesti za prijemanje.	12
Slika 5: Prijemanje na podlagi primitivov.....	13
Slika 6: Prvi dotik robotske roke in predmeta.	14
Slika 7: Oris našega pristopa.	16
Slika 8: Pridobivanje 3-D informacij.	20
Slika 10: Določanje začetne točke in smeri potiskanja.	25
Slika 11: Postopek izgradnje SIFT opisnikov.	37
Slika 12: Delovanje X-means.	42
Slika 13: Eksperimentalna robotska celica.....	46
Slika 14: Vsi testni predmeti.	48
Slika 15: Tri izmed hipotez za obstoj predmeta.	49
Slika 16: Potiskanje izbrane hipoteze za obstoj predmeta.	51
Slika 17: Primeri eksperimentov.	53
Slika 18: Grajenje modela s kopičenjem značilnic.	55
Slika 19: Model predmeta na osnovi 38 različnih pogledov.	56

Kazalo tabel

Tabela 1: Podmnožice testnih predmetov.....	49
Tabela 2: Rezultati potiskanja za 50 eksperimentov.....	55

Kazalo algoritmov

Algoritem 1: Oblikovanje hipotez.	23
Algoritem 2: Potiskanje ravnine.	26
Algoritem 3: Vrednotenje hipoteze.....	31

Povzetek

Bistven element kognitivnih robotskih sistemov je zmožnost spoznavanja predmetov, s katerimi robot pride v stik prvič, in učenja predstavitev za takšne predmete, ki so uporabne za kasnejše razpoznavanje. Naša predpostavka je, da je odkrivanje novih predmetov izvedljivo na učinkovit način z aktivnim robotskim sistemom, ki poleg vizualnega sistema za zajemanje slik prizorov vsebuje elemente za manipulacijo delov prizora. Manipulacija služi za ločitev predmeta iz okolice in za akumulacijo objektnih značilnic, na osnovi teh pa je mogoče modeliranje predmeta in ugotavljanje njegove funkcionalnosti.

V tem delu podajamo postopek odkrivanja in segmentacije predmetov s pomočjo kombinirane rabe vizualnih in manipulativnih funkcij kognitivnega robotskega sistema. Postopek smo preverili z več eksperimenti z realnim robotskim sistemom.

V predlaganem pristopu je postopek spoznavanja predmeta sestavljen iz treh procesov. Prvi proces predstavlja oblikovanje hipotez o obstoju predmetov. Oblikovanje hipotez sestavlja več podprocesov: izločanje in gručenje točk vizualnih značilnic, določanje 3-D lokacij in iskanje ravnin.

Drugi proces definira optimalne manipulativne akcije za verifikacijo predhodno določenih hipotez o obstoju predmeta. Z več zaporednimi manipulativnimi akcijami je mogoče znanje o predmetih kopičiti, na podlagi teh pa določati modele predmeta. V okviru tega dela manipulativne akcije predstavlja potiskanje z robotsko roko.

Tretji proces hipoteze o obstoju predmeta dejansko verificira na podlagi novih informacij, pridobljenih z manipulacijo. Verifikacija temelji na ugotavljanja konsistentnosti detektiranih značilnic predmeta glede na gibanje togega telesa. V primeru, da je kateri izmed procesov neuspešen, lahko sistem oblikuje in ovrednoti

dodatne hipoteze, zato predpostavka, da so procesi popolni, ni potrebna. Za zagotovitev robustnosti pristopa smo za več podprocesov uporabili verjetnostne metode, kot je na primer RANSAC (RANdom Sample Consensus).

Pridobljeni rezultati so pokazali, da je predlagan pristop uspešen pri odkrivanju in segmentaciji novih predmetov in da je znanje, ki ga pridobimo z več zaporednimi manipulativnimi akcijami, mogoče zbirati v popolnejše modele predmeta.

Ključne besede: detekcija predmetov, segmentacija predmetov, učenje z raziskovanjem, učenje konceptov, RANSAC.

Abstract

An important element of a cognitive robotic system is the ability to detect novel objects and learn their representations, which are suitable for later recognition and manipulation. The basic assumption of our work is that the detection and segmentation of new objects can be facilitated by an active robotic system, which can not only observe the objects but can also manipulate them. Manipulation supports object segmentation and the accumulation of object features, which provides the basis for building object models and for determining their functionality.

In this thesis we propose a new approach for object detection and segmentation based on the integrated use of visual and manipulative functions of a robotic system. The developed system was tested in a number of experiments on a real robot.

In the proposed approach, the object detection process consists of three consecutive procedures. The first procedure deals with the generation of hypotheses about the existence of an object. It comprises several sub-processes: extraction of visual features, calculation and clustering of 3-D points, and discovery of planar surfaces.

The second process defines manipulative actions which the robot needs to perform to validate the previously calculated object hypotheses. By employing these manipulative actions, the existence of an object can be confirmed and additional object knowledge can be accumulated. In our system, suitable manipulative actions are realized as pushing movements.

After the completion of the pushing movement, the third procedure verifies the underlying object hypothesis based on the newly acquired information. The verification procedure evaluates the consistency of the detected object features with respect to the assumption of rigid body motion. Additional hypotheses are evaluated if the verification

process fails. To ensure the robustness of the system, probabilistic methods such as RANSAC (RANdom SAmple Consensus) have been applied in several computational stages.

Our experimental results show that the system provides successful object detection and segmentation in complex scenes. Our work also demonstrates the robot can acquire more complete object models by performing several consecutive manipulative actions.

Keywords: object detection, object segmentation, learning by exploration, concept learning, RANSAC.

1. Uvod

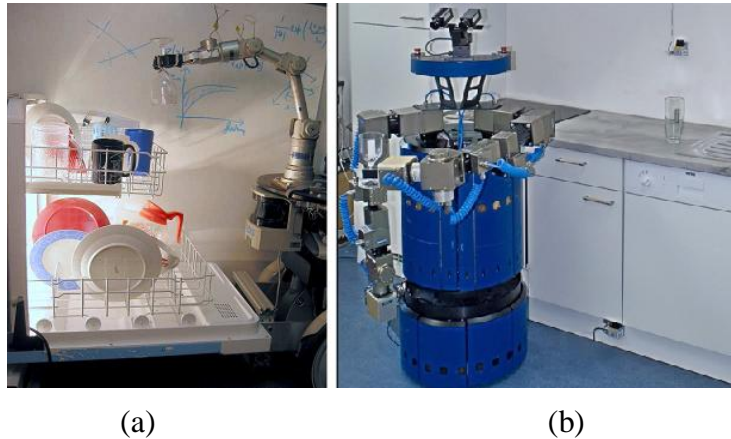
Veliko raziskav je usmerjenih v razvoj inteligentnih robotov, ki bi lahko delovali v običajnih človeških okoljih, na primer v kuhinji, in izvajali običajne naloge, kot so na primer zlaganje posode iz pomivalnega stroja. Slika 1 prikazuje dva primera takšnih robotov. Roboti za opravljanje takšnih nalog potrebujejo sposobnost interpretacije vizualnih informacij.

Ljudje smo zelo uspešni pri prepoznavanju na podlagi vizualnih informacij. Zgolj na podlagi vizualne informacije, tudi če samo pasivno opazujemo, lahko brez večjih naporov prepoznavamo različne predmete, osebe, kraje in preostale informacije, ki jih pridobimo iz prizorov, na primer gibanja in celo občutke. Uspešno prepoznavamo tudi predmete, ki jih vidimo prvič, čeprav so ti umeščeni v vizualno zelo zapletena prizorišča. To uspešnost je težko ponoviti s popolnoma pasivnimi vizualnimi sistemi. Takšni sistemi so sicer sposobni pravilnega ločevanja preprostih predmetov in prizorišč brez prekrivanj, pasivni vizualni sistemi pa lahko odpovejo pri zanesljivemu in pravilnemu ločevanju gruč neznanih predmetov [36]. Robot brez določenega nivoja predznanja se pri razstavljanju slik v smiselne dele pogosto znajde pred nemogočo nalogo. Tak primer prikazuje Slika 2.

Vizualni sistemi brez predznanja ne morejo ločiti med, na primer, dvema predmetoma, ki se prekrivata, in enim samim predmetom. Podobno težavna je naloga segmentacije predmetov z vizualno zapletenih ozadij.

Ena izmed rešitev je, da robota opremimo z modeli vseh predmetov, ki bi se lahko znašli v njegovem okolju. Ta pristop ni praktičen. Če na primer vzamemo robota za domača opravila, je zelo verjetno, da bo pri opravljanju svojih nalog (čiščenje, pospravljanje)

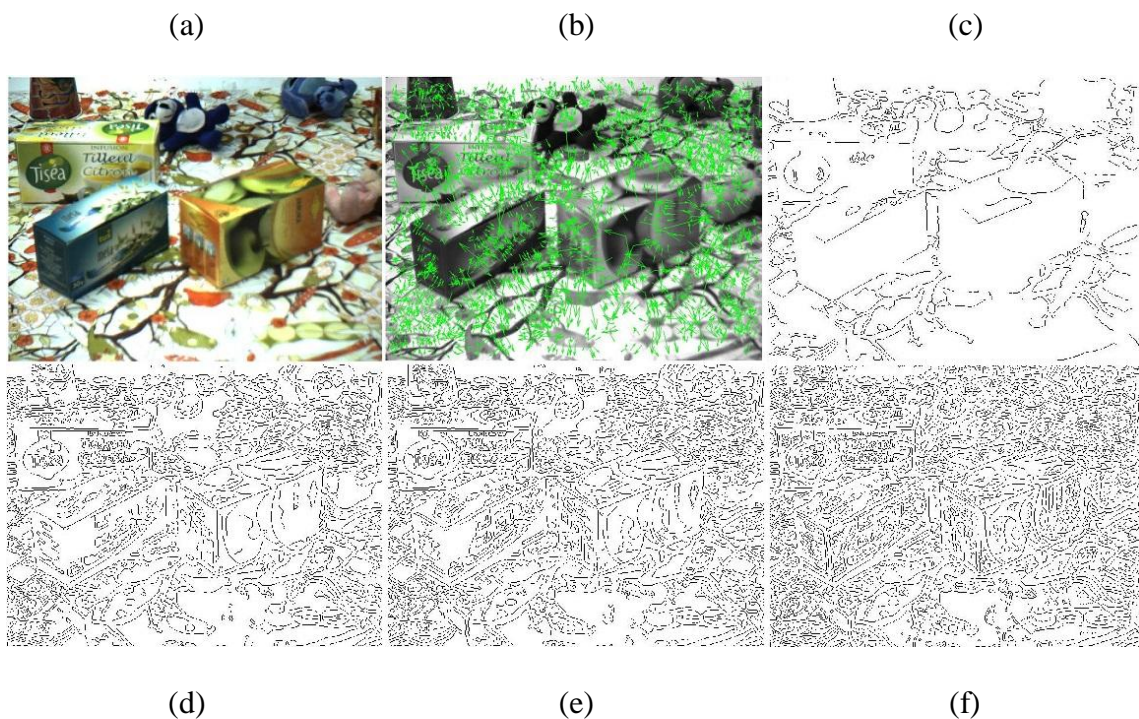
naletel na nove predmete (na primer pločevinka pijače), za katere ne bo imel modelov. Če želimo prilagodljivega robota, bo ta moral biti opremljen z drugačne vrste znanjem, in sicer z znanjem, ki mu bo omogočalo, da bo lahko odkrival, ločeval in sledil novim predmetom, za katere (še) nima modelov. Resnično avtonomen robot mora biti sposoben raziskovanja in učenja na podlagi informacij iz okolja, ker ne more biti odvisen od prejemanja vseh informacij na pasiven način [22]. Avtonomnim robotom je zato smiselno podati predznanja, ki mu pomagajo pri aktivnem raziskovanju novih okolij.



Slika 1: Roboti za domača okolja. (a) Robot pri nalogi zlaganja posode iz pomivalnega stroja. Povzeto po [32]. (b) Robot za delo v kuhinji. Povzeto po [31].

V splošnem je težko definirati, kaj pomeni pojem 'predmet'. Koncept 'predmet' ima pomembno vlogo v kognitivni znanosti, čeprav ustrezna formalna definicija ne obstaja. Definicija pojma objekt je odvisna od načina, kako razdelimo zaznani svet na smiselne enote [7]. Obstaja veliko principov, ki jih lahko uporabimo pri definiciji objektov, na primer zaprtje (closure), povezanost (connectedness), dvostranska simetrija [18], konveksnost, koplanarnost in kolinearnost obrisov. Čeprav vsak od teh principov pomaga pri zaznavanju predmetov, ni nobeden od njih vedno ustrezen. Z lahkoto poiščemo stvar, ki ustreza nekemu principu, a ni samostojen predmet. Prekrižane paličice, na primer, ustrezajo principu povezanosti, a jih dojemamo kot dva predmeta.

Zato je mogoče takšne osnovne principe uporabiti le za oblikovanje hipotez o obstoju predmetov, za potrditev hipotez pa je potrebno uvesti dodatne paradigme.



Slika 2: Segmentacija neznanih predmetov. (a) Izvirna slika. (b) Značilnice SIFT. (c)-(f) Izločeni robovi - detektor robov Canny - prag 0.5 (c), 0.1 (d), 0.05 (e) in 0.001 (f).

Glede na aktivnega agenta, na primer robota, lahko, glede na [12], objekte opišemo z naslednjimi lastnostmi:

- določena minimalna on maksimalna velikost,
- časovna stabilnost in
- upravljivost.

Določanje, kaj je predmet glede na te značilnosti, ni preprosta naloga. S pomočjo teh lastnosti lahko definiramo procese, ki jih lahko uporabimo za odločanje o obstoju predmetov. Zadnjo lastnost lahko preverimo tako, da na predmete vplivamo in jih

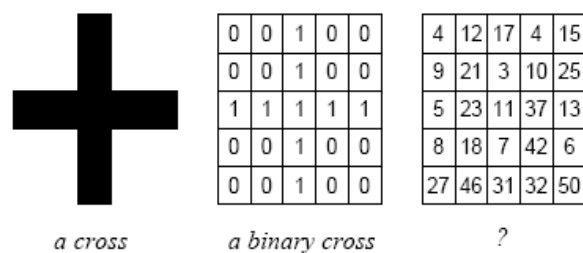
premikamo. Kriterij časovne stabilnosti pa lahko uporabimo pri primerjavah izločenih vizualnih značilnic pred in po izvedeni akciji.

Pojem povezave predmet-akcija (object-action complex, OAC) [11] povezuje predmete in dejanja, ki jih na njih izvedemo. Predmeti imajo po eni strani različne zmožnosti (angl. affordances), na primer drsenje in valjenje, po drugi strani pa jih te akcije določajo. Posledice operacije, ki jo izvedemo na predmetu, pripomorejo pri odločanju o obstoju predmeta in določanju njegove oblike. Čeprav je vizualno zaznavanje polno iluzij, je mogoče večino nejasnosti odpraviti, če lahko robot premika predmete. V prisotnosti manipulacije postane vid močnejši, in večina iluzij izgine. Poleg vida lahko pri tem pomagajo tudi taktilne informacije.

Ni povsem jasno, kakšno je predznanje človeškega sistema. Res pa je, da človeška sposobnost ločevanja predmetov ni univerzalna in se izboljšuje z izkušnjami. Primer so trije križi na naslednji sliki (Slika 3). Iz levega dela slike z lahkoto prepoznamo križ, mogoče zato, ker že imamo kriterij za razločevanje – razliko v intenziteti slikovnih elementov. Srednji del slike, malo manj jasno, a še vseeno, tudi predstavlja križ. Kriterij (ničle proti enicam) nam je poznan. Pri desnem delu slike pa, razen če nimamo dodatnih podatkov, lahko sklepamo, da gre za naključna števila. Šele ko je kriterij (praštevila proti številom) znan, lahko v sliki zopet odkrijemo križ. Primer na desnem delu slike predstavlja nov način ločevanja, na katerega verjetno do sedaj še nismo naleteli. Lahko si predstavljamo, da smo na podoben način od otroštva naprej pridobivali kriterije za ločevanje in identificiranje predmetov. Odrasel, izkušen človek si lahko pogosto razlaga vizualne prizore brez delovanja na prizorišča. Povezovanje aktivnosti in zaznavanja pa je ključna pri razvijanju te sposobnosti [22].

Naš cilj je opremiti robotske sisteme s takšnimi sposobnostmi, ki jim omogočajo aktivno raziskovanje in s tem odkrivanje novih predmetov in učenje njihovih modelov. Aktivno raziskovanje pri tem pomeni izvajanje akcij na 'stvareh'. V naši terminologiji je 'stvar'

hipoteza za predmet in se spremeni v predmet šele, ko se skozi delovanje robota izkaže, da gre resnično za predmet s konsistentnimi lastnostmi.



Slika 3: Trije križi. Povzeto po [22].

Osnovni gradniki našega pristopa so:

- Oblikovanje hipotez o obstoju predmetov. Iz stereo slik prizora izločimo vizualne značilnice in jih opišemo s pripadajočimi vektorji in 2-D lokacijo. Značilnice uredimo glede na ujemanja, ki jih ovrednotimo s pomočjo evklidske razdalje med vektorji značilnic in epipolarne geometrije. Urejeni pari značilnic in njihove 2-D lokacije nam omogočajo izračun 3-D lokacij. Med množico 3-D točk poiščemo pojavljanja skupkov točk, ki opisujejo ravne površine. Te ravnine uporabimo za določanje manipulativnih akcij.
- Manipulacija izbrane hipoteze. Na površino vplivamo z določeno akcijo, kot je na primer potiskanje, in tako pridobimo dodatne informacije, s katerimi lahko gradimo popolne in preverjene predstavitve predmetov.
- Vrednotenje hipoteze. Med 3-D točkami določimo ujemanja in ocenimo parametre povzročene gibanja predmetov. Nato preverimo, ali se točke hipotetičnega predmeta premikajo kot točke, ki naj bi pripadale predmetu. Pri tem smo predpostavili, da robot potiska toga telesa, zato se morajo izločene značilnice premikati kot toga telesa. Nekatere točke in pripadajoče značilnice lahko zavržemo ali dodamo glede na proces vrednotenja hipoteze.

- Grajenje modelov. Na podlagi več zaporednih potiskov lahko zberemo še več značilnic in točk, ki pripadajo objektu, in tako počasi zgradimo bolj kompleten model objekta.

1.1. Struktura naloge

V drugem poglavju predstavimo pomembna dela in pristope na tem področju. V tretjem poglavju opišemo naš pristop oblikovanja in vrednotenja hipotez o obstoju predmetov s pomočjo manipulativnih akcij. Sledi mu poglavje, v katerem predstavimo metode, ki smo jih uporabili pri realizaciji predlaganih procesov. To so algoritmi za odkrivanje in opisovanje vizualnih značilnic, iskanje parametrov modelov in gručenje. Peto poglavje opisuje eksperimentalno delo, kjer predstavimo naš realni robotski sistem in rezultate. Naši sklepi so podani v šestem poglavju.

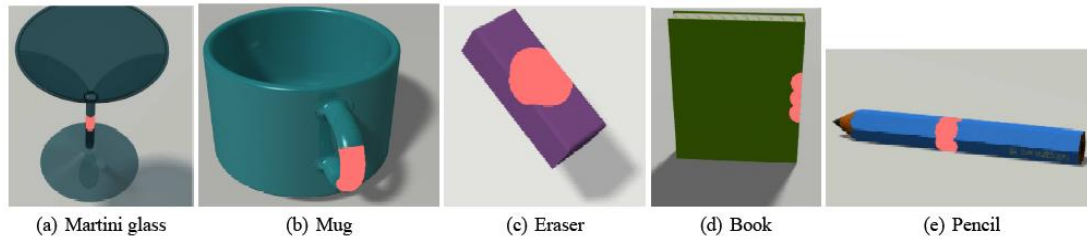
2. Pregled področja

Rezultati raziskovalnih del kažejo, da lahko robot raziskuje in se uči iz učinkov, ki jih ima njegovo delovanje na zunanji svet. Aktivno raziskovanje omogoča ločevanje predmetov od nepreglednega ozadja, kar ni preprosta naloga za pasivni vizualni sistem. Robot lahko spremeni pogled na predmet s premikanjem svojega telesa, eden od takšnih sistemov je opisan v [14], ali pa z izvajanjem akcij, ki premaknejo opazovane predmete. Predmete lahko premaknemo na več načinov, na primer s potiskanjem [22],[23] ali pa s prenašanjem v roki [16],[37]. Slednje je mogoče le, če je predmet mogoče oziroma ga znamo prijeti. Več raziskovalnih del je usmerjenih v pridobivanje novih znanj o predmetih s pomočjo prijemanja prvotno neznanih predmetov.

Saxena in sodelavci [32] so predstavili algoritem, ki za določanje načina prijemanja novih predmetov ne potrebuje in ne gradi modelov predmetov. Predlagani učni algoritem je po obdobju nadzorovanega učenja sposoben predvideti dobre točke za prijemanje. Učenje določanja točk za prijemanje poteka s pomočjo testnih slik, ki imajo označena 2-D mesta za prijemanje. Če je mogoče mesta za prijemanje določiti na več slikah predmeta (dve sliki ali več), je s triangulacijo mogoče določiti 3-D koordinate primernih mest za prijemanje, kjer lahko robot poskusi prijeti hipotetični predmet. Z nadzorovanim učenjem je v sistem vnesen določen nivo predznanja o prijemanju predmetov. Da zagotovimo neobčutljivost na spremembe v videzu predmetov (barva, merilo, tekstura, besedilo na knjigah - na te lastnosti vpliva osvetlitev, lokacija in orientacija kamer), je potrebno uporabiti ogromen nabor testnih slik.

Slika 4 prikazuje nekaj primerov testnih slik. Za predstavljenih 5 razredov predmetov (kozarec za martini, skodelica, radirka, knjiga in svinčnik) je bilo oblikovanih 2500 primerov sintetičnih slik z označenimi mesti za prijemanje. Za sposobnost posploševanja

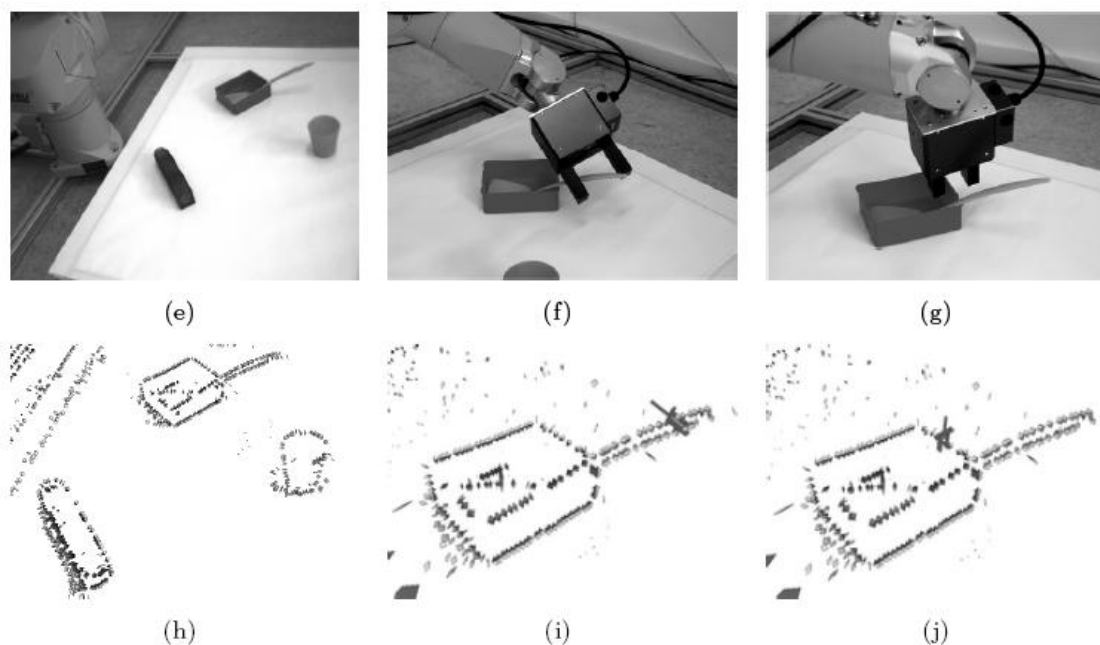
na nove predmete je poleg obsežnega nabora testnih slik potrebno uporabiti veliko različnih vrst vizualnih značilnic.



Slika 4: Nadzorovano učenje s sintetičnimi slikami z označenimi mesti za prijemanje. Povzeto po [32].

Podoben pristop za prijemanje nepoznanih predmetov opisujejo Kraft in sodelavci [16]. Njihov pristop temelji na strukturiranem predznanju. Prizore opisujejo lokalni simbolični opisniki (primitivi), ki zajemajo več vizualnih modalitet, ter 2-D in 3-D vidike vizualnih informacij. Slika 5 prikazuje realno okolje in izločene opisnike, hipoteze za prijemanje in dve uspešni prijemanji. Kot so pokazale druge raziskave, je po uspešnem prijemanju mogoče opazovati, upravljati in ločiti predmet od ozadja in sistematično pridobiti slike predmeta iz različnih zornih kotov [38].

Čeprav imamo s prijemanjem predmeta boljši nadzor nad predmetom, pa je prijemanje v primerjavi s potiskanjem veliko bolj zapletena in občutljiva operacija. Prijemanje novih predmetov, za katere nimamo modelov, je še toliko bolj zahtevna operacija. V veliko primerih je težavno prijeti nepoznan predmet, bodisi zaradi velikosti bodisi zaradi posebnosti v obliki. V takšnih primerih potiskanje predstavlja primerno alternativo, saj je preprostejše za izvedbo, kljub temu pa, če je uspešno, premakne predmet.



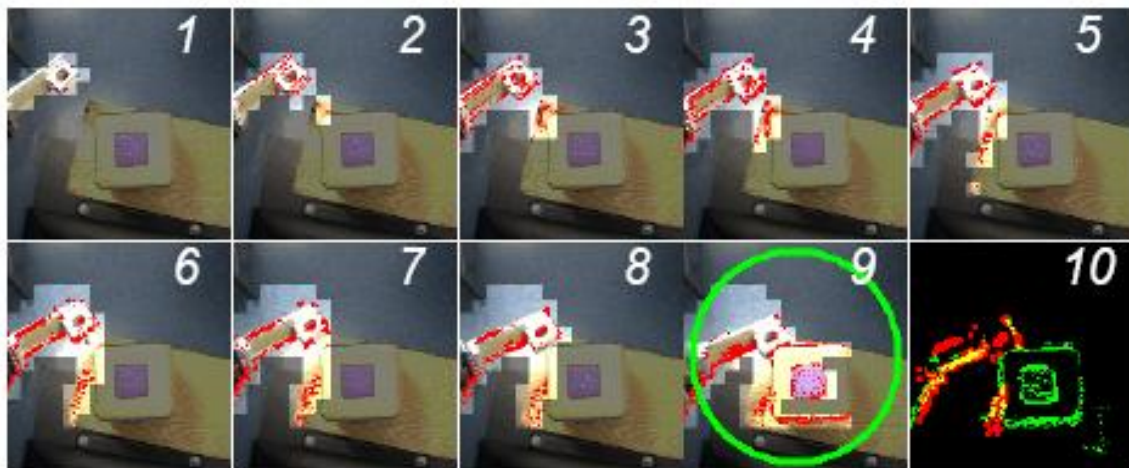
Slika 5: Prijemanje na podlagi opisnikov. Pogled na realno okolje (e) in pripadajoči 3-D opisniki (h). Uspešni prijemanji (f-g) prikazani v 3-D ponazoritvenem okolju (i-j). Povzeto po [16].

Metta in Fitzpatrick [22],[23] sta pokazala, da je mogoče s preprostimi gibanji pridobiti dovolj informacij za ločitev neznanih predmetov od ozadja in določiti zmožnosti (angl. affordances) predmetov. Eksperimenti so bili izvedeni s humanoidnim robotom z dvema rokama. Robot ni pasiven opazovalec, pač pa lahko svojo roko premika, kar omogoča ločitev roke od okolja. Z dovolj znanja o svoji roki lahko robot začne z raziskovanjem okolja. Pri dotiku roke in predmeta je možnih več scenarijev. Če je predmet velik in težak, ne bo prišlo do vizualno opaznega premika. Takšni predmeti so označeni kot ovire, saj jih robot ne more premakniti, in niso zanimivi za izvajanje poskusov. Če je predmet manjši, pa je precej bolj verjetno, da se bo premaknil v odgovor na dotik robotske roke. Slika 6 prikazuje tak primer.

Ko je tarča (hipoteza o obstoju predmeta) določena, robotska roka seže proti njej in 'pomete' področje okoli tarče. Pristop ne vsebuje algoritmov za ugotavljanje prvotnih

hipotez o obstoju predmetov. V opisanem primeru je tarča zgolj 'zanimivi' del predmeta, v tem primeru temnejši kvadrat na zgornji ploskvi predmetu. Gibanje robotske roke ni usmerjeno k optimalnemu načinu premikanja izbranega dela prizora, pač pa v vsakem primeru le k 'pometanju' okolice v bližini tarč.

Postopek ločevanja predmeta od okolice temelji na prvem dotiku robota in predmeta, ki je zaznan vizualno. Izrazite spremembe vrednosti točkovnih elementov v slikah so označene kot hipotetično gibanje. Ko se robotska roka premika, sistem sledi njenemu gibanju, a ga izloči skupaj s senco in gibanjem v ozadju. Ko se roka premakne preko nekega področja, je to področje označeno kot 'področje brez predmetov' za kratek čas. Ko se gibanje roke razširi preko veliko večjega področja, kot bi bilo mogoče v danem času, to nakazuje prvi dotik robotske roke in predmeta, ki ju sedaj mogoče ločiti.



Slika 6: Prvi dotik robotske roke in predmeta. (1) Roka 'pomete' okolico tarče. (2-8) Sledenje gibanju roke. (9) Prvi dotik. (10) Ločitev roke in predmeta. Povzeto po [10].

3. Detekcija in učenje predmetov

Razvili smo metodologijo, po kateri na podlagi oblikovanih hipotez o obstoju predmetov ustvarimo manipulativne akcije, s katerimi hipoteze potrdimo ali ovržemo. Za generacijo hipotez predlagamo bolj zmogljiv vizualni sistem, ki ne temelji na točkovnih elementih kot v [22],[23], ampak na vizualnih značilnicah in njihovih 3-D vidikih, in upošteva tudi razmerja med značilnicami. V razvitem sistemu oblikujemo začetne hipoteze, ki jih preverimo z manipulacijo. To je pomembna prednost pred sistemi, ki ne morejo aktivno delovati na predmete.

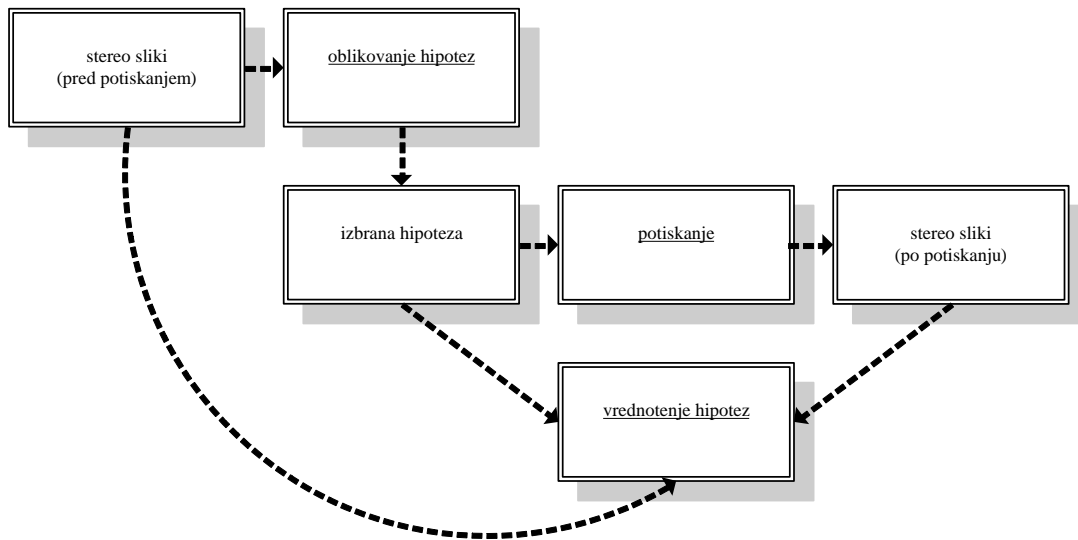
V tem delu smo se usmerili predvsem na domača okolja. Takšna okolja vsebujejo veliko predmetov z ravnimi površinami, zato smo pri razvoju metod za generacijo hipotez o predmetih obravnavali predmete, ki vsebujejo ravne površine. Ravne površine lahko predstavimo le s štirimi parametri. To je pomembno za nekatere algoritme, ki jih pri našem pristopu uporabljamo, na primer algoritem RANSAC [9]. S tem algoritmom hipoteze oblikujemo z oceno parametrov hipotetičnih ravnin. Sistem lahko oblikuje in vrednoti dodatne hipoteze, zato ni potrebno predpostaviti, da so procesi popolni. Vendar pa naš pristop ne deluje samo za ravne površine. Iz literature poznamo matematične strukture kot so geoni [3] in superkvadriki [17]. Obe metodologiji sta bili predlagani za rekonstrukcijo površin 3-D modelov. Takšne ploskve bi prav tako lahko uporabili za generacijo hipotez o obstoju predmetov, vendar pa bi pri tem morali oceniti večje število parametrov kot pri ravninah.

Določanje optimalnih premikov, ki potrdijo ali ovržejo trenutno hipotezo, je osnovano na izločenih in razvrščenih hipotetičnih ravninah. Po potiskanju lahko ocenimo parametre povzročene gibanja, in ugotovimo, ali se hipotetične točke predmeta gibajo tako, kot bi se gibale točke, ki pripadajo togemu predmetu. S pomočjo delovanja na

detektirano ravnino torej pridobimo dodatne informacije, ki jih lahko uporabimo za grajenje popolnih in preverjenih opisov predmetov.

Oblikovanje hipotez temelji na oceni parametrov ravnin, vrednotenje hipotez pa na oceni parametrov gibanja togega telesa. Pri obeh procesih, tako pri oblikovanju kot pri vrednotenju, je v množici značilnic prisotnih veliko odstopajočih točk (angl. outliers). Da za oba procesa zagotovimo robustnost, smo uporabili algoritem RANSAC [9], ki je znan zaradi svoje sposobnosti robustne ocene parametrov modela.

Oris našega pristopa prikazuje Slika 7. Iz stereo slik začetnega prizora s pomočjo primerjanja izločenih vizualnih značilnic pridobimo 3-D informacijo, ki je podlaga za oblikovanje hipotez o obstoju predmetov. Izbrana hipoteza je temelj za določanje optimalnega premikanja. S potiskanjem pridobimo dodatne informacije, s katerimi hipotezo potrdimo ali ovržemo. Proces vrednotenja hipoteze poteka 1) z oceno gibanja točk, ki so vsebovane v izbrani hipotezi in 2) s primerjanjem značilnic iz slik pred in po potiskanju.



Slika 7: Oris našega pristopa.

Načeloma bi bilo mogoče zvezno slediti značilnicam predmeta, ki ga potiskamo, kar zmanjša razlike med zaporednimi slikami in zmanjša možnost izgubljanja značilnic zaradi prevelikih razlik v glediščih. Vendar pa zvezno sledenje v našem pristopu ne bi bilo uspešno, saj je možnost, da bo robotska roka zakrila predmet med potiskanjem, zelo velika. Zato v našem sistemu primerjamo značilnice pred in po potisku šele po umiku robotske roke iz vidnega polja.

3.1. Oblikovanje hipotez o obstoju predmeta

Predmete in prizorišče je mogoče opisati na več načinov. Študije kažejo, da lahko sistemi, ki temeljijo na lokalnih, nespremenljivih značilnicah, prepoznajo predmete iz različnih zornih kotov in kasneje predmete odkrijejo v kompleksnih prizorih [27],[8],[30]. Z uporabo vizualnih značilnic lahko problem razpoznavanja predmetov prevedemo na problem, ki ga v osnovi sestavljajo tri stopnje: detekcija (angl. detection), opisovanje (angl. description) in ujemanje (angl. matching) vizualnih značilnic.

Na področju postopkov detekcije značilnic je bila razvita cela kopica detektorjev, ki temeljijo na: afini normalizaciji okolice Harrisovih in Hessejevih točk [25], razlikah Gaussovih jeder (angl. Difference of Gaussian, DoG) [17], robovih [37], ekstremih intenzitete [37], maksimalno stabilnih ekstremalnih področjih (Maximally Stable Extremal Regions, MSER) [21] ali na izrazitih področjih (angl. Salient Regions) [14].

Ocena detektorjev značilnic, predstavljena v [26], kaže na to, da je uspešnost izločanja značilnic odvisna tako od vsebine slike kot tudi od zahtevane robustnosti, točnosti in gostote značilnic.

Ocena postopkov za opisovanje značilnic [24] kaže, da pogosto najboljše rezultate doseže deskriptor SIFT (Scale Invariant Feature Transform) [17]. Uspešnost SIFT deskriptorjev potrjuje tudi raziskava delovanja deskriptorjev v primeru 3-D predmetov [27].

V naši metodologiji smo izbrali postopek za odkrivanje značilnic, ki temelji na razlikah Gaussovih jeder, in postopek SIFT za opisovanje izločenih vizualnih značilnic. Več o implementaciji v [40]. Izbira načina odkrivanja in opisovanja značilnic sicer ne vpliva na izvedbo metod, opisanih v nadaljevanju. Izberemo lahko tudi drugačno kombinacijo detektorja in deskriptorja, če tako zahteva uporaba. Naš izbrani algoritem za izločanje in opisovanje značilnic je predstavljen v poglavju 4.1.

Ključni korak pri oblikovanju hipotez je iskanje točk, ki pripadajo planarnim delom prizora. Obstaja veliko algoritmov za odkrivanje ravnin. Ravnine je mogoče iskati s pomočjo neskladja v stereo slikah (angl. disparity maps) [35]. Ta pristop pa ne zagotavlja vrste 3-D informacij, ki jih pri našem pristopu potrebujemo. Po oblikovanju hipotez sledi potiskanje množice točk, ki pripadajo izbrani hipotezi, za kar potrebujemo 3-D informacije, ki jih lahko uporabimo za manipulacijo z robotsko roko. Zato je smiselno, da si tudi pri iskanju ravnin pomagamo s 3-D informacijami, ki jih pridobimo s pomočjo izločenih vizualnih značilnic iz stereo slik prizora.

3.1.1. Pridobivanje 3-D informacij

Na začetku iz slike izločimo množico zanimivih točk. Ta postopek je neodvisen od trenutnega merila. Zanimive točke opišemo z vektorji števil s pripadajočimi 2-D lokacijami. Za dani prizor tako pridobimo množice opisnikov in njihovih 2-D lokacij: $\{ \mathbf{u}_l, \mathbf{d}_l \}$ in $\{ \mathbf{u}_d, \mathbf{d}_d \}$; $l = 1, \dots, N_L$, $d = 1, \dots, N_D$; kjer sta N_L in N_D števili izločenih značilnic iz leve in desne slike, $\mathbf{u} = (x, y)$ 2-D lokacija opisnika in \mathbf{d} opisnik oziroma deskriptor.

Ko smo zanimive točke odkrili in opisali, poskušamo poiskati iste dele prizora v obeh slikah. Ujemanje značilnic poteka na osnovi evklidske razdalje med opisniki. V našem sistemu privzamemo, da se dve značilnici ujemata, če sta njuna opisnika najmanj oddaljena med vsemi izločenimi opisniki, oziroma če je evklidska razdalja med njunima vektorjema najmanjša (najbližji sosed).

Rezultat ujemanja so pari značilnic, ena iz leve in druga iz desne slike prizora. Del ujemanj, ki jih določimo na ta način, ne povezuje dveh značilnic, ki določata isti del prizora. Razlogi za to so dvoumne značilnice ali značilnice, ki niso bile izločene v obeh slikah.

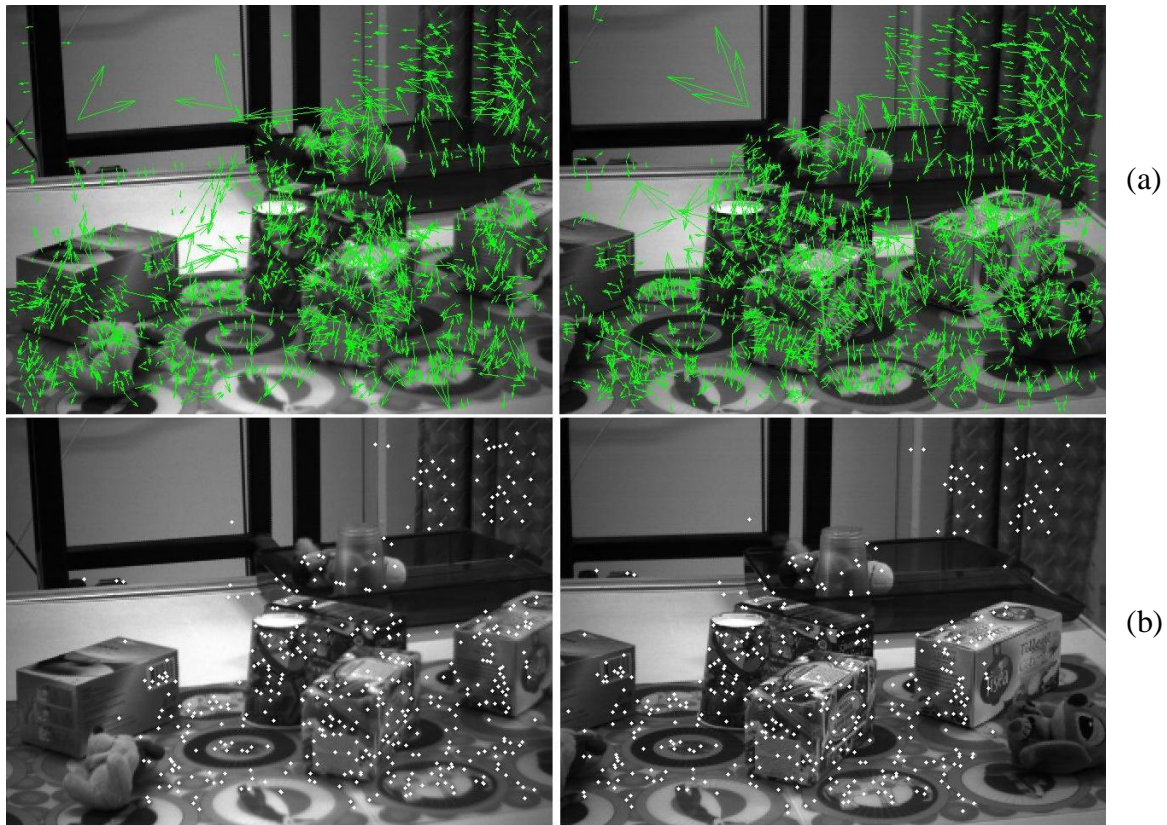
Pravilnost ujemanja lahko izboljšamo z upoštevanjem epipolarne geometrije [6]. S standardno kalibracijo (več o tem v poglavju 5.1) je mogoče pridobiti notranje in zunanje parametre obeh kamer, na podlagi katerih lahko izračunamo osnovno matriko (angl. essential matrix). Iz epipolarne geometrije dobimo zahtevo

$$\tilde{\mathbf{u}}_I^T E_{II} \tilde{\mathbf{u}}_J = 0 \quad (1)$$

kjer $\tilde{\mathbf{u}}_I = [x_I \quad y_I \quad 1]^T$ in $\tilde{\mathbf{u}}_J = [x_J \quad y_J \quad 1]^T$ označujeta lokacijo značilnic v slikah in E osnovno matriko. Osnovna matrika, ob predpostavki da model kamere ustreza modelu kamere na luknjico (angl. pinhole camera), povezuje skladne točke v stereo slikah.

Z znano postavitvijo in orientacijo kamer v prostoru in množicama 2-D lokacij, ki ustrezata množicama ujemajočih značilnic iz leve in desne slike, je mogoče določiti lokacije značilnic v prostoru. Več o tem v poglavju 5.1.

Slika 8 v zgornji vrsti prikazuje značilnice, ki smo jih izločili iz leve in desne slike. V spodnji vrsti so z 2-D lokacijami označena mesta parov značilnic, za katere smo lahko določili 3-D lokacije z ujemanjem opisnikov na podlagi Evklidske razdalje in epipolarne geometrije. 3-D točke smo predstavili s pripadajočimi 2-D lokacijami v levi in desni sliki.



Slika 8: Pridobivanje 3-D informacij. (a) Značilnice SIFT, izločene iz leve in desne slike prizora. (b) 3-D točke, označene na levi in desni sliki s pripadajočimi 2-D lokacijami.

3.1.2. Iskanje planarnih delov v prizoru

Iz množice 3-D točk želimo poiskati točke, ki opisujejo ravne površine v prizoru. Ravne površine opišemo z ravninami. V enačbi ravnine

$$ax + by + cz + d = 0 \quad (2)$$

nastopajo štirje parametri, ki jih določimo s pomočjo treh točk (x_i, y_i, z_i) ; $i = 1, 2, 3$:

$$a = \begin{vmatrix} 1 & y_1 & z_1 \\ 1 & y_2 & z_2 \\ 1 & y_3 & z_3 \end{vmatrix}, b = \begin{vmatrix} x_1 & 1 & z_1 \\ x_2 & 1 & z_2 \\ x_3 & 1 & z_3 \end{vmatrix}, c = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}, d = - \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}. \quad (3)$$

Predpostavka je, da množica 3-D točk, ki opisuje prizor, vsebuje več ravnin. Ker lahko predvidevamo, da bo v množici 3-D točk veliko odstopajočih točk (outliers), za robustnost procesa iskanja parametrov ravnin, ki opisujejo ravne površine v prizoru, uporabimo algoritem RANSAC (RANDOM Sample Consensus), ki je opisan spodaj in v poglavju 4.2.

Algoritem RANSAC za iskanje parametrov ravnin

Vhod v računski proces je množica 3-D točk. Ker želimo oblikovati več hipotez o obstoju predmetov, algoritem uporabimo večkrat. Vhodna množica v prvem koraku je množica vseh 3-D točk, ki smo jih določili na način, opisan v prejšnjem poglavju. V naslednjih korakih se lahko vhodna množica 3-D točk spreminja. Vsakič, ko določimo parametre ravnine, ki ustreza našim pogojem za ravnino, točke na ravnini izločimo iz vhodne množice točk. Ravnine iščemo, dokler ne dosežemo zeleno število ravnin k , ali pa ko vhodna množica postane premajhna oziroma ne vsebuje dovolj točk, ki bi zadostile našim pogojem za ravnino. V vsakem koraku N_p -krat naključno izberemo množico treh točk iz vhodne množice, določimo parametre ravnine in število skladnih točk F . Naš pogoj za obstoj ravnine je, da vsebuje večje ali enako število skladnih točk, kot ga določa prag $F_p > 3$. Skladnost točk s predlaganim modelom določa toleranca (t_p), ki predstavlja največjo dovoljeno oddaljenost točke od ravnine.

Proces oblikovanja hipotez o obstoju predmetov prikazuje Algoritem 1.

Za robotsko roko je potrebno določiti posebne gibe, ki bodo z visoko verjetnostjo povzročili premik množice točk, ki pripadajo hipotezi. Brez predpostavk o okolju pa ni mogoče vedeti, ali pridobljene ravnine združujejo točke, ki pripadajo enemu ali več predmetom, in ali vključujejo tudi točke, ki pripadajo ozadju. V primeru podobnega

prizora kot na prejšnji sliki (Slika 8) lahko ravnino sestavljajo točke iz ploskev ene ali več škatel, priključijo pa se lahko še točke iz ozadja. Če na primer poskusimo potisniti ravnino, ki pripada zgornjim ploskvam dveh škatel, je možnost uspešnega premika precej manjša. Robotska roka se lahko na primer premakne med škatlami, ki zato obstaneta na istem mestu.

Za boljšo verjetnost uspešnega potiskanja je potrebno preveriti razmerja med točkami ravnin. Če ravnina združuje točke, ki jih je mogoče razdeliti v gruče, je precej verjetno, da gruče opisujejo različne predmete ali ozadje. V našem sistemu ne moremo vnaprej določiti števila gruč, saj ne vemo, katerim delom prizora pripadajo točke ravnin, zato potrebujemo algoritem, ki za delovanje ne potrebuje števila gruč. Zato smo uporabili algoritem za gručenje *X-means*, ki zna sam oceniti optimalno število gruč. Algoritem je opisan v poglavju 4.3.

Hipoteze o obstoju predmeta so osnovane na množici točk, ki jih pridobimo z gručenjem točk, ki pripadajo detektiranim ravninam. Hipoteze razvrstimo glede na število točk, ki jim pripadajo. Manipulacija hipotez z večjim številom točk bo bolj verjetno uspešna, zato najprej poskusimo premakniti takšne množice. Vsako potiskanje ne bo uspešno. Potiskanje predmetov, ki so na primer pritrjeni, ne bo uspešno. V takšnih primerih lahko izberemo naslednjo hipotezo, ki vsebuje manj točk, določimo optimalen način potiskanja glede na točke, in izvedemo potiskanje.

Algoritem 1: Oblikovanje hipotez.

Vhod: množica 3-D točk

dokler več kot F_p točk in število korakov k ni doseženo

ponovi N_p krat

naključno izberi m_p točk

oceni parametre ravnine

določi točke, ki se prilegajo modelu s toleranco t_p

izberi parametre ravnine, ki ji pripada največje število točk F

če je F večje od F_p

shrani parametre ravnine in izloči točke F iz množice 3-D točk

uredi točke ravnine v gruče z algoritmom X-means

Izhod: parametri ravnin in pripadajoče točke in gruče

Izbrane eksperimentalne vrednosti, ki smo jih uporabili pri eksperimentih (glej poglavje 5.2.1):

- toleranca $t_p = 0.005\text{m}$
- število testnih podmnožic $N_p = 1000$
- prag $F_p = 20$
- število korakov $k = 6$

3.2. Potiskanje na osnovi izbrane hipoteze

V prejšnjem poglavju smo opisali proces oblikovanja hipotez, ki ga sestavljajo 1) iskanje točk, ki pripadajo ravninam, 2) gručenje točk v ravninah in 3) razvrščanje hipotez o obstoju predmeta glede na število točk, ki jih vsebujejo.

Naš cilj je potisniti izbrano hipotezo na način, ki bo s čim večjo verjetnostjo premaknil točke, ki pripadajo izbrani ravnini. Gibanje robotske roke določimo glede na parametre

ravnine, ob tem pa upoštevamo, da je gibanje odvisno tudi od podlage. Pri določanju smeri potiskanja nas vodi zahteva, da mora gibanje robotske roke potekati na ploskvi, ki je vzporedna talni površini, v našem primeru je to miza, sicer bo robot hitro trčil v mizo ali pa izgubil stik z objektom.

Določanje gibanja konca robotske roke prikazuje Algoritem 2.

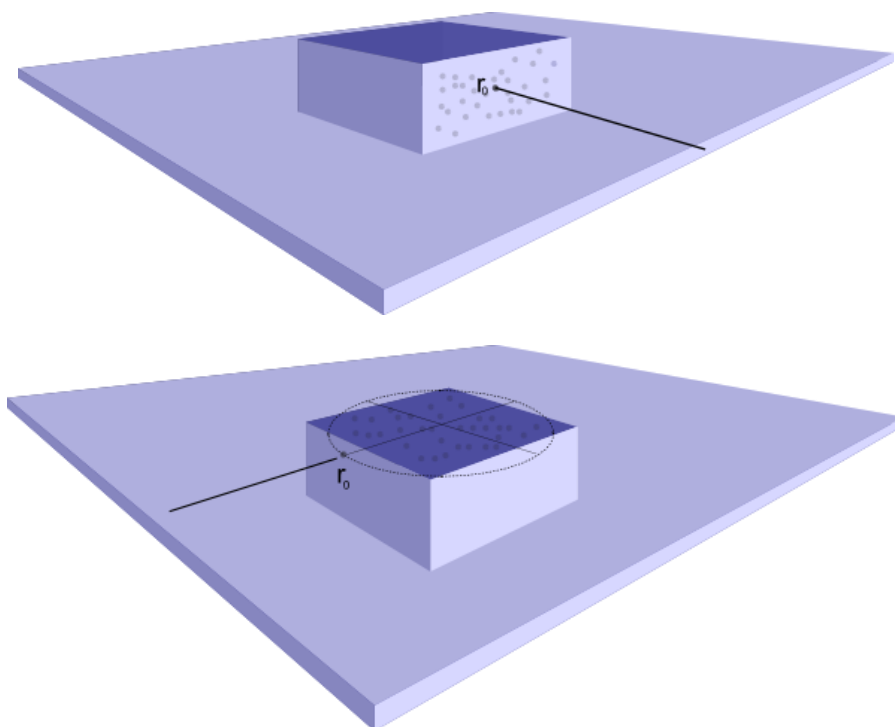
Sistem je umerjen tako, da os x robotskega sistema kaže v nasprotni smeri gravitacije. Osnova za določanje robotskih trajektorij za potiskanje je kot ϕ med talno površino in izbrano hipotetično ravnino. Trajektorije določimo s točko \mathbf{r}_0 in vektorjem \mathbf{n} , ki določata začetno točko in smer potiskanja.

Pri tem ločimo dva primera (oba primera po vrsti prikazuje Slika 9):

- $\phi \geq 45^\circ$. V tem primeru obravnavamo ravnino kot pravokotno na talno ravnino. Začetno točko \mathbf{r}_0 določimo kot povprečje vrednosti vseh točk hipotetične ravnine. Pri določanju smeri potiskanja rotiramo hipotetično ravnino okrog osi preseka s talno ravnino. Smer potiska \mathbf{n} je podana z negativno smerjo normale zarotirane ravnine. Na ta način zagotovimo, da je smer potiskanja vedno vzporedna s talno ravnino.
- $\phi > 45^\circ$. Značilnice projiciramo na ravnino, vzporedno talni ravnini. Višino ravnine določimo z enačbo (4). S pomočjo statistike drugega reda nato izračunamo elipso, ki aproksimira projicirane točke, in izračunamo točko na koncu krajše glavne osi, ki je bližje koordinatnemu sistemu obeh kamer. To je točka \mathbf{r}_0 . Potiskanje poteka vzporedno mali osi z začetkom v \mathbf{r}_0 . Smer potiskanja je podana z vektorjem, ki kaže od začetka do središča elipse \mathbf{r}_c .

$$x' = \max(x_r - \Delta, \min(x_t + \Delta, \frac{x_r + x_t}{2})), \quad (4)$$

kjer je x_r povprečna višina vseh točk ravnine in x_t višina talne ravnine. Pri eksperimentih smo uporabili vrednost $\Delta = 0.02\text{m}$.



Slika 9: Določanje začetne točke in smeri potiskanja.

Algoritem 2: Potiskanje ravnine.

Vhod: hipotetična ravnina $p = (a, b, c, d)$ in pripadajoče točke
določi talno ravnino

izračunaj kot ϕ med hipotetično in talno ravnino

če ($\phi \leq 45^\circ$)

\mathbf{r}_0 - srednja vrednost točk ravnine

rotiraj hipotetično ravnino p do navpične lege p'

$\mathbf{n} = (a', b', c')$

drugače

projiciraj točke na horizontalno ravnino višine x'

izračunaj elipso, ki aproksimira vse točke

določi \mathbf{r}_0

$\mathbf{n} = \mathbf{r}_0 - \mathbf{r}_c$

Izhod: začetna točka \mathbf{r}_0 in smer potiskanja \mathbf{n}

* potiskanje: začetek pri $\mathbf{r}_0 + \Delta\mathbf{r}$ v smeri \mathbf{n} , potiskanje $2\Delta\mathbf{r}$ v smeri $-\mathbf{n}$

3.3. Vrednotenje hipotez

Ko je potiskanje izbrane množice značilnic izvedeno, lahko z dodatnimi informacijami, ki jih pridobimo s sproženim gibanjem predmeta, hipoteze potrdimo ali ovržemo. V ta namen moramo sprejeti predpostavke o tem, kako naj bi se predmeti v robotovem okolju gibali.

V tem delu predpostavimo, da so iskani predmeti toga telesa. Potisnjene značilnice se morajo torej premakniti tako, kot se premikajo točke, ki pripadajo togemu telesu. To lahko preizkusimo s pomočjo analize stereo slike, ki jih zajamemo pred in po potisku. Značilnice, ki se nahajajo v bližini hipotetičnih točk predmeta, in se gibljejo kot hipotetične točke, lahko dodamo k modelu.

Vrednotenje hipotez poteka na podlagi ujemanj med značilnicami v stereo slikah pred in po potisku. Prvi par slik s pripadajočimi 3-D lokacijami in opisniki iz leve in desne slike $\{\mathbf{v}^1, \mathbf{d}_l^1, \mathbf{d}_d^1\}$ je osnova za oblikovanje hipotez o predmetih, izbrana hipoteza $\{\mathbf{v}^{h'}, \mathbf{d}_l^{h'}, \mathbf{d}_d^{h'}\}$ iz nabora oblikovanih hipotez pa temelj za določanje načina potiskanja. S pomočjo 3-D lokacij in opisnikov iz leve in desne slike $\{\mathbf{v}^2, \mathbf{d}_l^2, \mathbf{d}_d^2\}$ po premiku pa lahko preverimo, ali so se točke premaknile v skladu z načelom gibanja togega telesa.

Začetni nabor ujemanj pridobimo s primerjanjem opisnikov značilnic iz 1) obeh levih in 2) obeh desnih slik. Vsako izmed ujemanj nam da množico parov ujemajočih se značilnic. To je ujemanje 2-D vizualnih značilnic, zato ni nujno, da se oba opisnika iz slike pred potiskom ujemata z obema opisnikoma iz slike po potiskanju.

Vzemimo eno izmed 3-D točk iz izbrane hipoteze, s pripadajočima opisnikoma iz leve in desne slike $\mathbf{d}_l^1, \mathbf{d}_d^1$, in točko na drugi sliki, ki prikazuje premik točke v skladu z načelom gibanja togega telesa $\mathbf{d}_l^2, \mathbf{d}_d^2$. Če se, na primer, ujemata opisnika iz levih slik $\mathbf{d}_l^1, \mathbf{d}_l^2$, ni nujno, da se ujemata tudi opisnika iz desnih slik $\mathbf{d}_d^1, \mathbf{d}_d^2$. Naš pristop je, da v množico parov ujemajočih 3-D točk vključimo vse 3-D točke, katerih ujemanje kažejo pari značilnic iz levih in desnih slik.

Na primer: 3-D točko iz prve slike, ki jo določa lokacija \mathbf{v}^1 ter levi in desni opisnik $\mathbf{d}_l^1, \mathbf{d}_d^1$, določimo ujemanje za levi opisnik \mathbf{d}_l^2 in ujemanje za desni opisnik \mathbf{d}_d^2 . Par levemu opisniku je opisnik na drugi sliki \mathbf{d}_l^2 , ki določa 3-D lokacijo \mathbf{v}_1^2 , par desnemu pa opisnik \mathbf{d}_d^2 , ki določa lokacijo \mathbf{v}_2^2 . V množico parov 3-D točk vključimo oba para 3-D točk, $(\mathbf{v}^1, \mathbf{v}_1^2)$ in $(\mathbf{v}^1, \mathbf{v}_2^2)$. Ta pristop smemo uporabiti, saj je za vse značilnice mogoče določiti 3-D lokacije.

Na podlagi parov iz množic $\{\mathbf{v}_m^1\}$ in $\{\mathbf{v}_m^2\}$ poiščemo parametre transformacije togega telesa. Transformacijo togega telesa opisuje enačba

$$\mathbf{v}^2 = \mathbf{R}\mathbf{v}^1 + \mathbf{t} \quad (5)$$

kjer je \mathbf{R} rotacijska matrika in \mathbf{t} translacijski vektor.

Parametra \mathbf{R} in \mathbf{t} lahko ocenimo s prilagajanjem na osnovi najmanjših kvadratov z razcepom s singularnimi vrednostmi [1]. Za robustnost procesa iskanja parametrov gibanja togega telesa uporabimo algoritem RANSAC.

Vzemimo dve množici 3-D točk $\{\mathbf{v}_i\}$ in $\{\mathbf{v}'_i\}$; $i=1,2,\dots,N$ množici 3-D točk, N je število parov 3-D točk, ki ju povezuje enačba

$$\mathbf{v}'_i = \mathbf{R}\mathbf{v}_i + \mathbf{t} + \mathbf{e}_i \quad (6)$$

kjer je \mathbf{R} rotacijska matrika in \mathbf{t} translacijska vektor in \mathbf{e}_i vektor šuma. Prisotnost zelo velikega šuma obravnavamo v nadaljevanju besedila.

Poiskali bi radi \mathbf{R} in \mathbf{t} , ki minimizirata naslednji kriterij:

$$\sum_{i=1}^N \|\mathbf{v}'_i - (\mathbf{R}\mathbf{v}_i + \mathbf{t})\|^2. \quad (7)$$

Tu gre za problem najmanjših kvadratov, ki ga je glede na ugotovitve iz [1] mogoče razdeliti na dva dela: 1) iskanje \mathbf{R} in 2) iskanje \mathbf{t} .

1. Najprej izračunamo srednjo vrednost vseh točk in jo odštejemo od začetnih točk \mathbf{v}_i .

$$\mathbf{v} = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_i \quad (8)$$

$$\mathbf{q}_i = \mathbf{v}_i - \mathbf{v} \quad (9)$$

podobno izračunamo še \mathbf{v}' in $\mathbf{q}'_i = \mathbf{v}'_i - \mathbf{v}'$ ter določimo matriko \mathbf{H}

$$\mathbf{H} = \sum_{i=1}^N \mathbf{q}_i \mathbf{q}'_i{}^T \quad (10)$$

in poiščemo njen razcep s singularnimi vrednostmi (poglavje 4.4):

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \quad (11)$$

Nato določimo matriko \mathbf{X} :

$$\mathbf{X} = \mathbf{V} \mathbf{U}^T. \quad (12)$$

Če je determinanta $\det(\mathbf{X}) = 1$ določimo rotacijsko matriko $\mathbf{R} = \mathbf{X}$.

V primeru, da je $\det(\mathbf{X}) = -1$, matrika \mathbf{X} označuje zrcaljenje (angl. reflection):

- a) ena izmed singularnih vrednosti enaka nič, na primer $\lambda_k = 0$. Enačbo (12) preoblikujemo v $\mathbf{X}' = \mathbf{V}' \mathbf{U}'^T$, kjer je \mathbf{V}' matrika \mathbf{V} z zamenjanimi predznaki v k -tem stolpcu.
- b) nobena izmed singularnih vrednosti ni enaka nič. To je mogoče le v prisotnosti zelo velikega šuma. V takšnih primerih avtorji [1] predlagajo pristop na osnovi algoritma RANSAC.

2. Translacijski vektor določa enačba

$$\mathbf{t} = \mathbf{v}' - \mathbf{R} \mathbf{v}. \quad (13)$$

Algoritem RANSAC za iskanje rotacijske matrike in translacijskega vektorja

Iz množice vseh mogočih parov ujemajočih se 3-D točk iz $\{\mathbf{v}_m^1\}$ in $\{\mathbf{v}_m^2\}$, naključno izberemo m_r parov ($m_r = 3$) ujemajočih se točk. Za vsako testno množico parov z uporabo zgoraj opisanega algoritma ocenimo vrednosti za rotacijsko in translacijsko matriko, in določimo število točk, ki se modelu prilegajo s toleranco t_r . Za oceno kvalitete prilagajanja uporabimo razdaljo med premaknjenimi in detektiranimi točkami.

$$\|R\mathbf{v}_m^1 + \mathbf{t} - \mathbf{v}_m^2\| \quad (14)$$

Algoritem 3 prikazuje procese določanja začetne množice ujemajočih se parov točk iz slik pred in po potiskanju, ki so izhodišče za oceno parametrov gibanja togega telesa s pomočjo algoritma RANSAC.

Dodatne informacije, ki jih potrebujemo za potrjevanje hipotez o obstoju predmetov, lahko pridobimo le, če je do gibanja zares prišlo. Zato je pomembno ugotoviti, ali so se točke zares premaknile. Gibanje togega telesa je sestavljeno iz rotacijskega in translacijskega dela. Če ocenjena translacija in rotacija kažeta na to, da do gibanja ni prišlo, moramo izbrati novo hipotezo iz množice hipotez, saj v takšnem primeru nimamo dovolj informacij za nadaljnje delo. V ta namen določimo normi za rotacijsko in translacijsko matriko.

Normo rotacije določimo z izračunom osi vrtenja in kota vrtenja φ , ki ju lahko izračunamo iz rotacijske matrike. Za normo vzamemo manjšo od vrednosti $|\varphi|$ in $2\pi - |\varphi|$. Da se dokazati, da na ta način izračunana vrednost določa razdaljo v prostoru vseh rotacijskih matrik [28]. Do premika je prišlo, če je vsaj ena od obeh norm večja od mejne vrednosti za kot rotacije φ_{\min} oziroma translacije t_{\min} .

S pomočjo algoritma RANSAC določimo \mathbf{R} in \mathbf{t} in vsa ujemanja, ki ustrezajo izračunanemu togemu gibanju. Če je število točk ujemanj manjše od vnaprej določenega najmanjšega števila točk, potem nimamo na voljo dovolj informacij za gradnjo objektnih modelov. Vendar pa v tem primeru ne moremo izbrati nove hipoteze, izvesti potiskanja in ovrednotiti hipoteze v upanju, da bo število značilnic naslednjič večje. Če se je postopek končal z določenim številom točk, potem je prišlo do bodisi minimalne rotacije bodisi do minimalne translacije, torej parametri gibanja togega telesa kažejo na to, da je do gibanja prišlo. V tem primeru moramo zaradi spremembe prizora oblikovati nov nabor hipotez na osnovi novo zajetih stereo slik prizora.

Algoritem 3: Vrednotenje hipoteze.

Vhod: $\{ \mathbf{v}^1, \mathbf{d}_l^1, \mathbf{d}_d^1 \}, \{ \mathbf{v}^{hl}, \mathbf{d}_l^{hl}, \mathbf{d}_d^{hl} \}; i = 1, \dots, \text{število hipotez}$

dokler obstajajo hipoteze in ni določenih točk predmeta

izberi hipotezo in potisni ravnino (Algoritem 2)

pridobi $\{ \mathbf{v}^2, \mathbf{d}_l^2, \mathbf{d}_d^2 \}$ iz slik po potisku

določi množici točk z ujemajočimi se opisniki $\{ \mathbf{v}_m^1 \}, \{ \mathbf{v}_m^2 \}$

ponovi N_r krat

naključno izberi m_r pare iz $\{ \mathbf{v}_m^1 \}, \{ \mathbf{v}_m^2 \}$

oceni parametre R in t transformacije togega telesa

določi pare točk iz $\{ \mathbf{v}^1 \}, \{ \mathbf{v}^2 \}$, ki se prilegajo modelu s toleranco t_r in

njihovo število F

izberi parametre transformacije, ki ji pripada največje število parov F

če ($\|\varphi\| \leq \varphi_{\min}$ in $\|t\| \leq t_{\min}$)

izberi naslednjo hipotezo

drugače

če je F večje od F_R

Izhod: točke, ki ustrezajo gibanju togega telesa

drugače

Izhod: oblikuj nove hipoteze (Algoritem 1)

Izbrane eksperimentalne vrednosti, ki smo jih uporabili pri eksperimentih (glej poglavje 5.2.3):

- toleranca $t_r = 0.005m$,
- število podmnožic $N_r = 100$,
- prag $F_r = 10$,
- minimalna rotacija in translacija $\varphi_{\min} = 5^\circ$, $t_{\min} = 0.05m$.

3.4. Grajenje modelov

Če želimo zgraditi izčrpen model predmeta, moramo robotu omogočiti različne poglede na predmet. Točke, za katere smo ugotovili, da pripadajo predmetu, je potrebno kopičiti preko različnih pogledov in jih sestaviti v model predmeta.

Pri gradnji modelov predmetov moramo določiti način potiskanja, ki bo imelo za posledico rotacijo predmetov. Na ta način bomo odkrili nove, prej nevidne značilnice in jih vključili v model predmeta. Naš algoritem, ki služi pridobivanju različnih pogledov na predmet, je v osnovi spremenjen algoritem za potiskanje, opisan v poglavju 3.2. Začetno točko r_0 in smer potiskanja n določimo na način, ki ga opisuje Algoritem 2. Nato vektor n zavrtimo okrog začetne točke v ravnini, vzporedni s smerjo gravitacije, za prej določen kot α . (Pri eksperimentalnem delu smo uporabili $\alpha = 20^\circ$). Tako hipotetično ravnino potisnemo pod kotom α , in ne pravokotno, kot določa osnovni algoritem za potiskanje. Druga možnost spremembe algoritma za potiskanje je premik začetne točke r_0 .

Uspešnost kopičenja točk predmeta preko različnih pogledov je odvisna od rotacije predmeta. Rotacija predmeta mora biti po eni strani dovolj velika, da odkrije nove, prej nevidne značilnice in po drugi strani dovolj majhna, da je izračun ujemanj med značilnicami v zaporednih slikah sploh mogoč. Glede na rezultate eksperimentov delovanja različnih algoritmov za izločanje in opisovanje značilnic [27], se delovanje poljubne kombinacije detektorja in deskriptorja zelo poslabšuje s spremembami v zornem kotu, pri spremembah večjih od 30° pa nobena izmed kombinacij detektorjev in deskriptorjev ne deluje več dobro.

Po vsakem uspešnem potisku, ki predmet premakne v skladu s prej omenjenimi omejitvami v rotaciji, pridobimo nove točke predmeta in pripadajoče parametre transformacije togega telesa. Model predmeta lahko nadgradimo z novo določenimi točkami. Novo določene točke dodamo v model, lokacije starih, prej določenih točk, pa preslikamo na nove lokacije s pomočjo parametrov gibanja togega telesa.

4. Metodologija

Naš pristop temelji na izločanju, opisovanju in ujemanju vizualnih značilnic. Izbrali smo postopek za odkrivanje značilnic, ki temelji na razlikah Gaussovih jeder, in postopek SIFT za opisovanje izločenih vizualnih značilnic.

Oblikovanje hipotez temelji na oceni parametrov hipotetičnih ravnin, vrednotenje hipotez pa na oceni parametrov gibanja togega telesa. Pri obeh procesih, tako oblikovanju kot vrednotenju, je v množici značilnic prisotnih veliko odstopajočih točk (outliers). Da za oba procesa zagotovimo robustnost, smo uporabili algoritem RANSAC, ki je znan zaradi svoje sposobnosti, da robustno oceni parametre modela.

S pomočjo algoritma RANSAC pridobimo ravnine, ki jih uporabimo kot namige za obstoj predmetov. Še vedno pa lahko točke, ki so vsebovane v eni ravni, pripadajo enemu ali več predmetom in tudi ozadju. Ker je pravilnost izbrane hipoteze zelo pomembna za izračun gibanja robotske roke, ki bo s čim večjo verjetnostjo premaknilo hipotetično ravnino, točke najprej razdelimo v gruče točk, ki med seboj niso preveč oddaljene. V ta name uporabimo algoritem za gručenje *X-means*, ki sam oceni število gruč, ki vnaprej ni znano.

4.1. Iskanje in opis vizualnih značilnic

SIFT (Scale-invariant feature transform) značilnice je leta 1999 razvil David Lowe [17]. Za naše delo je pomembno, da so te značilnice neodvisne od ravninske rotacije in merila ter deloma neodvisne od spremembe osvetlitve in pogleda.

Pot do značilnic SIFT pelje skozi štiri faze:

- odkrivanje ekstremov v prostoru meril (scale-space extrema detection),
- določanje lokacij (keypoint localization),
- določanje orientacije (orientation assignment) in
- izgradnja opisnikov pomembnih točk (keypoint descriptor).

4.1.1. Odkrivanje ekstremov

Področja, ki so neodvisna od spremembe merila, poiščemo z iskanjem točk, ki so stabilne preko vseh meril slike. Osnova za to je konvolucija več Gaussovih funkcij (16) z vhodno sliko:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (15)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (16)$$

Določanje stabilnih točk pri različnih merilih poteka s konvolucijo razlike Gaussovih funkcij (difference of Gaussian) in vhodne slike:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (17)$$

Razlogi za izbiro te funkcije je več: filtrirane slike L potrebujemo za opis značilnic, D pa lahko izračunamo s preprosto operacijo odštevanja. Po vsaki takšni konvoluciji je izračunana slika podvzorčena in proces ponovljen.

Ekstreme $D(x, y, \sigma)$ poiščemo s primerjanjem vsake vzorčne točke (sample point) z osmimi sosedi pri istem merilu in pripadajočimi devetimi sosedi pri sosednjih merilih, višjemu in nižjemu. Če je vrednost točke največja ali najmanjša med vsemi sosedi, postane kandidat za pomembno točko (keypoint).

4.1.2. Določanje lokacije

Rezultat prejšnjega koraka je v večini primerov preveč kandidatov za pomembne točke. V tem koraku poleg določanja lokacije pomembnih točk tudi izločimo točke s premajhnim kontrastom in točke, ki ležijo na robovih.

Lokacijo pomembne točke določimo z interpolacijo podatkov [4]. Funkcijo $D(x, y, \sigma)$ razvijemo v Taylorjevo vrsto do kvadratnega člana in jo premaknemo v vzorčno točko. Velja, da D in njene odvode ocenimo pri vzorčni točki, $\mathbf{x} = (x, y, \sigma)^T$ pa označuje odmik od te točke.

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} \quad (18)$$

Lokacijo ekstrema $\hat{\mathbf{x}}$ določimo z naslednjo enačbo:

$$\hat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D}{\partial \mathbf{x}} \quad (19)$$

Če je vrednost $\hat{\mathbf{x}}$ pri kateri koli dimenziji večja od 0.5, ekstrem leži bližje drugi točki v sliki, zato interpolacijo ponovimo pri tej točki. V nasprotnem primeru pri določanju lokacije upoštevamo odmik in tako dobimo interpoliran približek lokacije pomembne točke.

Če navzdol omejimo vrednost $D(\hat{\mathbf{x}})$ lahko izločimo točke s premajhnim kontrastom. Lowe pri vrednostih točkovnih elementov med 0 in 1, izloči točke, kjer je $|D(\hat{\mathbf{x}})|$ manjše od 0.03.

Točke, ki so posledica odziva na robove, izločimo s pomočjo glavnih ukrivljenosti. Posledica roba bo namreč velik odziv v smeri roba in majhen odziv v pravokotni smeri. Glavne ukrivljenosti izračunamo iz Hessejeve matrike

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (20)$$

pri lokaciji in merilu pomembne točke. Lastne vrednosti te matrike so proporcionalne glavnim ukrivljenostim D .

4.1.3. Določanje orientacije

Prvi korak pri določanju orientacije pomembni točki je izbira slike, katere merilo je najbližje merilu pomembne točke. Nato pri vsaki točki slike $L(x, y)$ izračunamo velikost (21) in orientacijo gradienta (22) z razlikami med točkovnimi elementi po naslednjih enačbah:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (21)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (22)$$

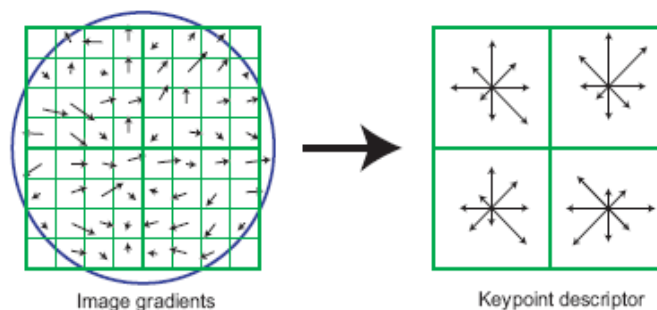
Orientacijski histogram pomembne točke sestavlja 36 razdelkov, ki pokrivajo vseh 360 stopinj. Prispevki v histogram so gradienti v okolici pomembne točke. Vsak prispevek je utežen z velikostjo in z Gaussovimi krožnim oknom. Vrhovi orientacijskega histograma ustrezajo glavnim smerem lokalnih gradientov. Upoštevamo najvišji vrh in vrhove, ki niso manjši od 80% velikosti najvišjega. Ko pomembni točki določimo orientacijo, je mogoče opisnik pomembne točke rotirati glede na orientacijo in tako doseči neodvisnost od rotacije slike.

4.1.4. Izgradnja opisnikov

V prejšnjih korakih smo ključnim točkam določili lokacijo, merilo in orientacijo. Slika 11 prikazuje postopek izgradnje opisnika na podlagi 8x8 okolice ključne točke.

Na levi strani slike so prikazane velikosti in orientacije gradientov. Krog opisuje utežitev z Gaussovimi oknom. Dolžina puščice predstavlja seštevek velikosti gradientov v bližini

določene smeri. Orientacije gradientov razvrstimo v orientacijske histograme, ki imajo 8 razdelkov, kar je prikazano na desni strani slike (Slika 11).



Slika 11: Postopek izgradnje SIFT opisnikov. Opisnik sestavljajo vsi orientacijski histogrami (2 x 2), vsak ima po 8 razdelkov. Povzeto po [17].

V našem primeru izgradnja opisnika poteka v 16 x 16 okolici ključne točke. Pri izgradnji opisnika upoštevamo vrednosti vseh razdelkov vseh orientacijskih histogramov. Na podlagi 16 x 16 okolice pridobimo šestnajst (4 x 4) orientacijskih histogramov, vsak ima 8 možnih orientacij. Opisnik torej sestavlja vektor s 128 vrednostmi. Sledi še normalizacija vektorja, nastavitev pragov in ponovna normalizacija vektorja.

4.2. Algoritem RANSAC

V mnogih znanstvenih problemih moramo zaznane podatke razložiti v okviru predhodno določenih modelov. Takšno razlago lahko razdelimo na dva podproblema. Prvi je iskanje najboljšega ujemanja med podatki in enim od modelov (problem razvrščanja), drugi pa iskanje najboljših vrednosti parametrov (v smislu odstopanja) za izbrani model (problem ocene parametrov). V praksi sta oba problema povezana - rešitev drugega po navadi zahteva rešitev prvega problema.

Pri oceni parametrov je mogoče uporabiti veliko metod: na primer metodo najmanjših kvadratov (Least Mean Square, LMS) ali oceno največje zanesljivosti (Maximum Likelihood Estimation, MLE), itd. Takšne metode pri oceni parametrov uporabijo vse podatke in nimajo notranjih mehanizmov za ugotavljanje večjih napak. Temeljijo namreč na predpostavki, da bo, ne glede na velikost množice podatkov, vedno dovolj 'dobrih' vrednosti, ki bodo zgladile večja odstopanja.

V mnogih praktičnih primerih ta predpostavka ne drži. Pri reševanju tega problema si lahko pomagamo z več hevrističnimi pristopi. Najpogosteje uporabljen pristop je variacija naslednjega postopka. Parametre modela izračunamo na podlagi vseh podatkov. Nato poiščemo podatek, ki se najmanj ujema s predlaganim modelom in ga odstranimo. Postopek iterativno ponavljamo, dokler ni skupno odstopanje manjše od predhodno določenega nivoja ali pa izločimo vse podatke. Na žalost pa je lahko predlagana hevristika že zaradi ene same 'okužene točke' neuspešna, zato ni primerna za uporabo na nepreverjenih podatkih.

Algoritem RANSAC (RANdom Sample And Concensus), ki sta ga leta 1981 v svojem delu predstavila Fischler and Bolles [9], je uspešen kot metoda za oceno parametrov modela v prisotnosti velikega števila odstopajočih točk (angl. outliers). Zaradi tega je zelo uveljavljen pri analizi prizorišč, kjer je pogosto veliko napak zaradi načina izločanja vizualnih značilnic, ki so vir podatkov. RANSAC procedura deluje v nasprotni smeri kot običajne tehnike glajenja. Običajno z glajenjem izločamo točke iz prvotne rešitve, ki temelji na čim več točkah. RANSAC pa začne z majhno prvotno množico, ki ji dodaja točke.

Bolj formalno lahko RANSAC paradigmo opišemo na naslednji način. Če vzamemo model, za katerega potrebujemo minimalno število točk m , da lahko določimo proste parametre, in množico točk P (velja, da je P večje od m). Iz P naključno izberemo podmnožico S_I , ki vsebuje m točk, in določimo parametre modela M_I . S pomočjo

modela določimo razširjeno množico S_1' , ki vsebuje točke, ki se prilegajo modelu s predhodno določeno toleranco.

Če je število točk v S_1' večje od predhodno določenega praga, ki je funkcija ocena števila odstopanj v množici P , izračunamo nov model M_1' na podlagi točk iz S_1' . Če je število točk manjše od praga, naključno izberemo novo podmnožico S_2 , in ponovimo postopek. Če po predhodno določenem številu poskusov i ne najdemo množice S_i' , ki bi vsebovala število točk večje od predhodno določenega praga, bodisi zaključimo z največjim S_i' ali pa zaključimo brez rešitve.

RANSAC ima torej tri parametre:

- toleranca, ki določa skladnost točke s predlaganim modelom,
- število testnih podmnožic in
- prag, ki določa pravilnost modela.

4.3. Gručenje točk

V našem sistemu ne moremo vnaprej vedeti, ali 3-D točke, ki smo jih razdelili v ravnine z algoritmom RANSAC, pripadajo enemu ali več predmetom in če mogoče vključujejo točke iz ozadja. Potiskanje bo prineslo dodatne informacije, na podlagi katerih bomo hipotezo o obstoju predmeta bodisi sprejeli bodisi zavrgli. V primerih potiskanja ravnin, ki vsebujejo točke, ki pripadajo različnim predmetom in so med seboj zelo oddaljene, je možnost uspešnega potiskanja precej manjša.

4.3.1. Algoritem K -means

Algoritem K -means [20] je najbolj poznan in pogosto uporabljan algoritem za gručenje, predvsem zaradi svoje preprostosti in konvergenčnih lastnosti. K -means ima nekaj slabosti, na primer počasnost. Poleg tega za svoje delovanje potrebuje število gruč K .

Zanimivo je, da lahko algoritem pri fiksnem številu gruč poišče slabše rešitve kot če se število gruč lahko spreminja [29].

Procedura se začne z določitvijo K gruč, vsako na začetku sestavlja ena naključno izbrana podatkovna točka, in nadaljuje z dodajanjem točk k gruči, katere srednja vrednost je najbližja novi točki. Ko je točka dodana gruči, je srednja vrednost gruče ponovno izračunana tako, da upošteva tudi novo točko. Algoritem se konča, ko se srednje vrednosti gruč ne spreminjajo več.

4.3.2. Algoritem X-means

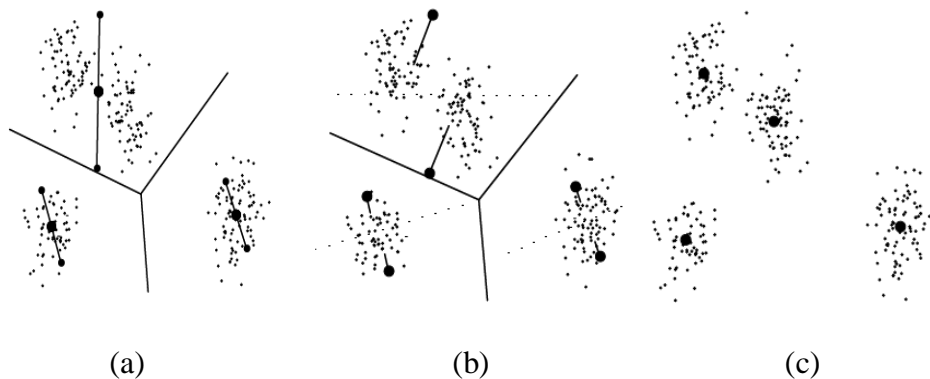
V našem pristopu potrebujemo gručenje točk ravnin, ne moremo pa podati števila gruč, saj nimamo predpostavk o predmetih v prizorih, zato algoritma K -means ne moremo uporabiti. X -means [29] je izboljšava algoritma K -means. Za nas najpomembnejša izboljšava je, da za svoje delovanje ne potrebuje števila gruč, pač pa območje vrednosti za število gruč. X -means potem sam določi najboljšo vrednost. Pri danem območju algoritem začne s K , ki je enaki spodnji meji območja. Pri svojem delovanju povečuje število centroidov, če je potrebno, dokler ne doseže zgornje meje območja. Med tem procesom algoritem beleži ocene skupin centroidov, najboljše ocenjena skupina je rezultat algoritma.

Algoritem X -means je sestavljen iz dveh operacij:

- *Izboljšava parametrov.* Operacijo sestavlja delovanje običajnega algoritma K -means.
- *Izboljšava strukture.* S to operacijo poiščemo dodatne centroide. To je mogoče z razcepom centroidov. Razcep temelji na skupku dveh idej.

- Prva ideja - eden po eden. Izberemo centroid, v bližini katerega določimo nov centroid. Za novo skupino centroidov uporabimo algoritem *K-means*. Če je novi model boljši od prejšnjega, dodamo centroid, drugače se vrnemo k prejšnji strukturi. Pomembno je odgovoriti na vprašanje, katere centroide bi veljalo razcepiti. Če bi poskusili razcepiti vse centroide in izbrali najboljšega, bi bila računaska zahtevnost prevelika.
- Druga ideja - preizkusimo del centroidov. Izberemo (na primer) polovico centroidov, ki so po nekem hevrističnem kriteriju najbolj ugodni za razcep. Izbrane centroide razcepimo in uporabimo algoritem *K-means*. Če je novi model boljši od prejšnjega, dodamo centroide, drugače se vrnemo k prejšnji strukturi. Vprašanje pa je, kaj uporabiti za hevristični kriterij, ki označuje smiselnost razcepa.

Slika 12 prikazuje delovanje algoritma *X-means* in operacij "Izboljšava parametrov" in "Izboljšava strukture". Prvotne centroide, ki smo jih pridobili z operacijo Izboljšava parametrov, razcepimo na dva naslednika. Premaknemo jih iz lokacije prvotnega centroida v nasprotnih smereh poljubno izbranega vektorja, za razdaljo, ki je proporcionalna velikosti območja. V naslednjem koraku v vsakem področju uporabimo *K-means*, kjer se otroci borijo za točke svojih staršev. Zadnji korak je preveritev modela, kjer se odločimo, ali obdržimo starša ali pa njegove naslednike. Upamo, da bodo centroidi, ki že vsebujejo točke, ki tvorijo resnične gruče, preživeli svoje otroke in jih proces ne bo spremenil. Po drugi strani pa upamo, da bodo področja, ki trenutno niso dobro predstavljena s centroidi, deležna večje pozornosti in da se bo v njih število centroidov povečalo.



Slika 12: Delovanje X-means. (a) Razcep izhodiščnih centroidov. (b) Lokalni K-means. (c) Končne gruč po izbiri modela. Povzeto po [29].

Korak, ki odloča o tem, ali obdržimo starše ali otroke, je izbira modela. Cilj izbire modela je uravnoteženje kvalitete prilagajanja k podatkom in kompleksnosti ali sposobnosti modela, da razloži vhodne podatke.

V našem primeru imamo družino alternativnih modelov M_j , kjer različni modeli pripadajo rešitvam z različnimi vrednostmi K (števili gruč). Najboljši model izberemo s pomočjo posteriornih verjetnosti. V našem primeru so vsi modeli oblike, ki jo predvideva K-means algoritem (sferični Gaussovi). Za približek posteriornih verjetnosti uporabimo naslednjo formulo, znano tudi kot Schwarzov kriterij [33] ali kriterij BIC (Bayesian Information Criterion):

$$BIC = -2 \ln L_{max} + k \ln N \quad (23)$$

kjer je L_{max} največja verjetnost (Maximum Likelihood Estimation, MLE) dosegljiva v okviru modela, k število parametrov modela in N število uporabljenih podatkovnih točk.

Izbira kriterija BIC za določanje razcepa je ena od možnosti. BIC je relativno preprosta metoda, ker potrebuje le največjo verjetnost, ki jo je mogoče doseči v okviru modela, in ne preko celotnega prostora parametrov. Izvedeni eksperimenti opisani v [29] so pokazali dobro delovanje kriterija BIC.

4.4. Razcep s singularnimi vrednostmi

Razcep s singularnimi vrednostmi (Singular Value Decomposition, SVD) [13] je pomemben matrični razcep, ki ga uporabljamo na več področjih, na primer za prilagajanje podatkov na osnovi najmanjših kvadratov in za računanje pseudoinverza.

Realno matriko A z dimenzijami $m \times n$ in $m \geq n$ lahko razcepimo na tri matrike:

$$A = U \Sigma V^T \quad (24)$$

$$U^T U = V^T V = I_n \quad (25)$$

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n) \quad (26)$$

Enačba (24) predstavlja razcep s singularnimi vrednostmi matrike A .

Matriko U sestavljajo ortonormalizirani lastni vektorji, povezani z n največjimi lastnimi vrednostmi matrike AA^T , matriko V pa ortonormalizirani lastni vektorji matrike $A^T A$. Diagonalne elementi matrike Σ so nenegativni kvadratni koreni lastnih vrednosti matrike $A^T A$. Imenujemo jih singularne vrednosti.

4.5. Kalibracija kamere

Razmerje med 3-D točko $M = [X, Y, Z]^T$ in njeno projekcijo v sliko $m = [u, v]^T$ podaja naslednja enačba:

$$s\tilde{m} = A[\mathbf{R} \quad \mathbf{t}] \tilde{M} \quad (27)$$

kjer sta $\tilde{M} = [X, Y, Z, 1]^T$, $\tilde{m} = [u, v, 1]^T$, s poljubni skalirni faktor, (\mathbf{R}, \mathbf{t}) zunanji parametri kamere (extrinsic parameters) in A notranja matrika (intrinsic matrix):

$$\mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (28)$$

(u_0, v_0) sta koordinati središčne točke (principal point), α, β skalirna faktorja za osi u in v , ter γ popačenje obeh osi slike.

Če predpostavimo $Z=0$ (vse točke so na kalibracijski plošči), sledi

$$s\tilde{\mathbf{m}} = \mathbf{H}\tilde{\mathbf{M}} \quad (29)$$

$$\mathbf{H} = \mathbf{A}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}] \quad (30)$$

in $\tilde{\mathbf{M}} = [X \quad Y \quad 1]^T$.

Na podlagi slike ravnine lahko ocenimo homografijo $\mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3]$.

Sledi:

$$[\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3] = \lambda \mathbf{A}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}], \quad (31)$$

λ je poljuben skalar.

Zaradi ortonormalnosti \mathbf{r}_1 in \mathbf{r}_2 dobimo dve glavni omejitvi za notranje parametre na podlagi ene homografije:

$$\begin{aligned} \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 &= 0 \\ \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_1 &= \mathbf{h}_2^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 \end{aligned} \quad (32)$$

Naj bo

$$\mathbf{B} = \mathbf{A}^{-T} \mathbf{A}^{-1} = \begin{bmatrix} B_{11} & B_{21} & B_{31} \\ B_{12} & B_{22} & B_{32} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \quad (33)$$

$$\mathbf{b} = [B_{11} \quad B_{12} \quad B_{22} \quad B_{13} \quad B_{23} \quad B_{33}]^T \quad (34)$$

Če je \mathbf{h}_i i -ti stolpec \mathbf{H} (30) velja:

$$\mathbf{h}_i^T \mathbf{B} \mathbf{h}_j = \mathbf{v}_{ij}^T \mathbf{b}. \quad (35)$$

Omejitve, ki sledita iz dane homografije (32), lahko zapišemo kot:

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{b} = 0. \quad (36)$$

Na podlagi n slik imamo n enačb (36) in lahko določimo

$$\mathbf{V} \mathbf{b} = 0, \quad (37)$$

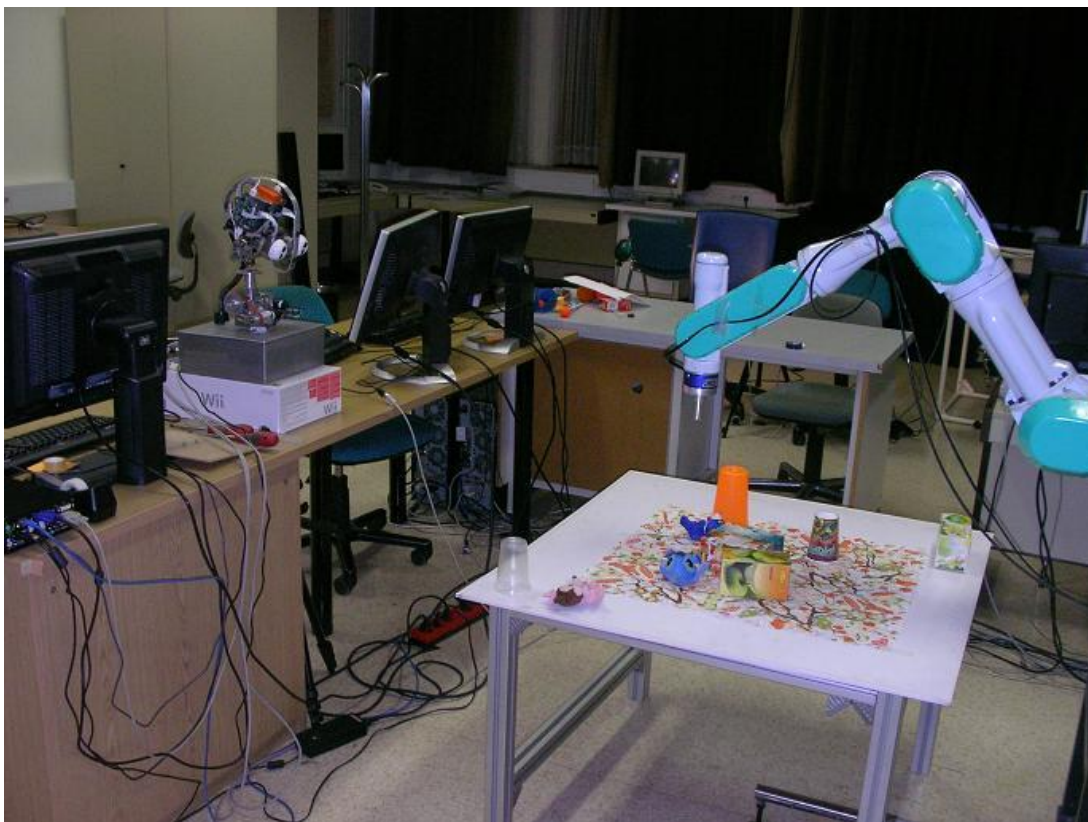
kjer je \mathbf{V} matrika $2n \times 6$.

Če imamo $n \geq 3$ slik, dobimo eno rešitev za \mathbf{b} , ki je določena do skalirnega faktorja. Ko ocenimo \mathbf{b} , lahko izračunamo notranje in zunanje parametre kamere z enačbami, ki so navedene v [41].

5. Eksperimenti

5.1. Realni robotski sistem

Pri eksperimentalnem delu je bilo potrebno združiti sistem za vid s sistemom, s katerim lahko izvajamo manipulativne akcije. Našo eksperimentalno celico, ki jo prikazuje Slika 13, sestavlja robotski manipulator Mitsubishi PA-10 in sistem za stereo vid – humanoidna glava z Univerze v Karlsruheju [2].



Slika 13: Eksperimentalna robotska celica.

Preden začnemo z eksperimenti, moramo sistem usposobiti za delovanje v realnem okolju. Kalibracija kamer je potreben korak, če želimo pretvoriti 2-D informacije, ki jih zajamemo s kamerami, v 3-D informacije, ki so uporabne za manipulacijo objektov v prostoru. Za kalibracijo kamer smo uporabili fleksibilno in robustno metodo [41], več o tem v poglavju 4.5.

Uporabljena metoda zahteva, da kamere opazujejo planarni vzorec pri nekaj (vsaj dveh) različnih orientacijah. Kalibracijski vzorec je lahko natisnjen z laserskim tiskalnikom in pritrjen na primerno površino, na primer knjigo. Premikamo lahko bodisi kamere bodisi planarni vzorec. Računalniške simulacije in preizkusi z realnimi podatki so v obeh primerih pokazali zelo dobre rezultate.

Pri postopku kalibracije smo uporabili $n=30$ slik kalibracijskega vzorca. Vzorec je šahovnica s 7×5 kvadrati. Slike smo zajeli pri različnih oddaljenostih in orientacijah vzorca. Na podlagi slik smo določili parametre kamere, ki smo jih v nadaljevanju uporabili pri določanju 3-D lokacij ujemajočih značilnic.

Za oceno transformacije iz koordinatnega sistema robotske glave v koordinatni sistem robotske roke smo kalibracijski vzorec pritrčili na robotsko roko. Na ta način lahko iz slik kalibracijskega vzorca določimo množico 3-D točk v koordinatnem sistemu robotske roke. Z znanimi parametri kamere lahko isto množico 3-D točk predstavimo v koordinatnem sistemu robotske glave. Pri oceni transformacije smo uporabili 5 slik kalibracijskega vzorca pri 5 različnih oddaljenostih, kar ustreza premiku po osi z koordinatnega sistema robotske roke, in 2 rotacijah – okrog osi x in okrog osi y koordinatnega sistema robotske roke.

5.2. Preizkusi

V naših eksperimentih smo želeli preveriti in oceniti delovanje treh procesov: proces oblikovanja hipotez o predmetih, potiskanje, ki potrdi ali ovrže hipotezo, in proces vrednotenja hipotez glede na morebitne spremembe prizora.

Pri preizkušanju procesa oblikovanja hipotez smo predpostavili dobro delovanje v primeru, da testni predmeti vsebujejo ravne površine. Zato smo poleg predmetov, ki vsebujejo ravne dele (škatle), uporabili še predmete, ki nimajo ravnih površin (igračke, lončki). Testni predmeti (prikazuje jih Slika 14) so 3 škatle, 3 lončki in 3 igrače.



Slika 14: Vsi testni predmeti.

Izvedli smo 50 poskusov, in sicer po 10 poskusov za 5 različnih podmnožic testnih predmetov. Podmnožice so označene s pikami in jih prikazuje Tabela 1. Škatle označimo po vrsti: 'jagoda', 'jabolko', 'limona'. Lončki: 'pisan', 'oranžen', 'prozoren'. Igrače: 'temnomodra', 'svetlomodra', 'roza'. Točke označujejo predmete, ki smo jih uporabili pri poskusih. Na primer: v prvih 10 poskusih smo uporabili škatli za jagodni in jabolčni čaj, pisan in prozoren lonček ter temnomodro in roza igračko.

Tabela 1: Podmnožice testnih predmetov.

Poskusi	Testni predmeti								
	škatla			lonček			igračka		
	1	2	3	1	2	3	1	2	3
01-10	×	×		×		×	×		×
11-20	×	×	×	×	×			×	×
21-30	×			×	×	×	×	×	×
31-40		×	×	×			×		
41-50	×		×		×			×	×

5.2.1. Oblikovanje hipotez

Pri vseh poskusih je bilo oblikovano vnaprej določeno število hipotez k ($k = 6$). Končno število hipotez o obstoju predmetov se je spreminjalo glede na rezultat algoritma za gručenje. Slika 15 prikazuje 3 izmed izračunanih hipotez. V tem primeru vsaka od prikazanih hipotez opisuje eno ploskev na enem predmetu.

V naših eksperimentih smo uporabili predmete z ravnimi površinami, ki se nahajajo na podobnih višinah. Pri takšnih predmetih obstaja velika verjetnost, da bodo oblikovane hipoteze povezale več predmetov. Ker smo takšne predmete pogosto postavili skupaj, je 42% hipotez vključevalo točke, ki pripadajo več predmetom.



Slika 15: Tri izmed hipotez za obstoj predmeta.

5.2.2. Potiskanje

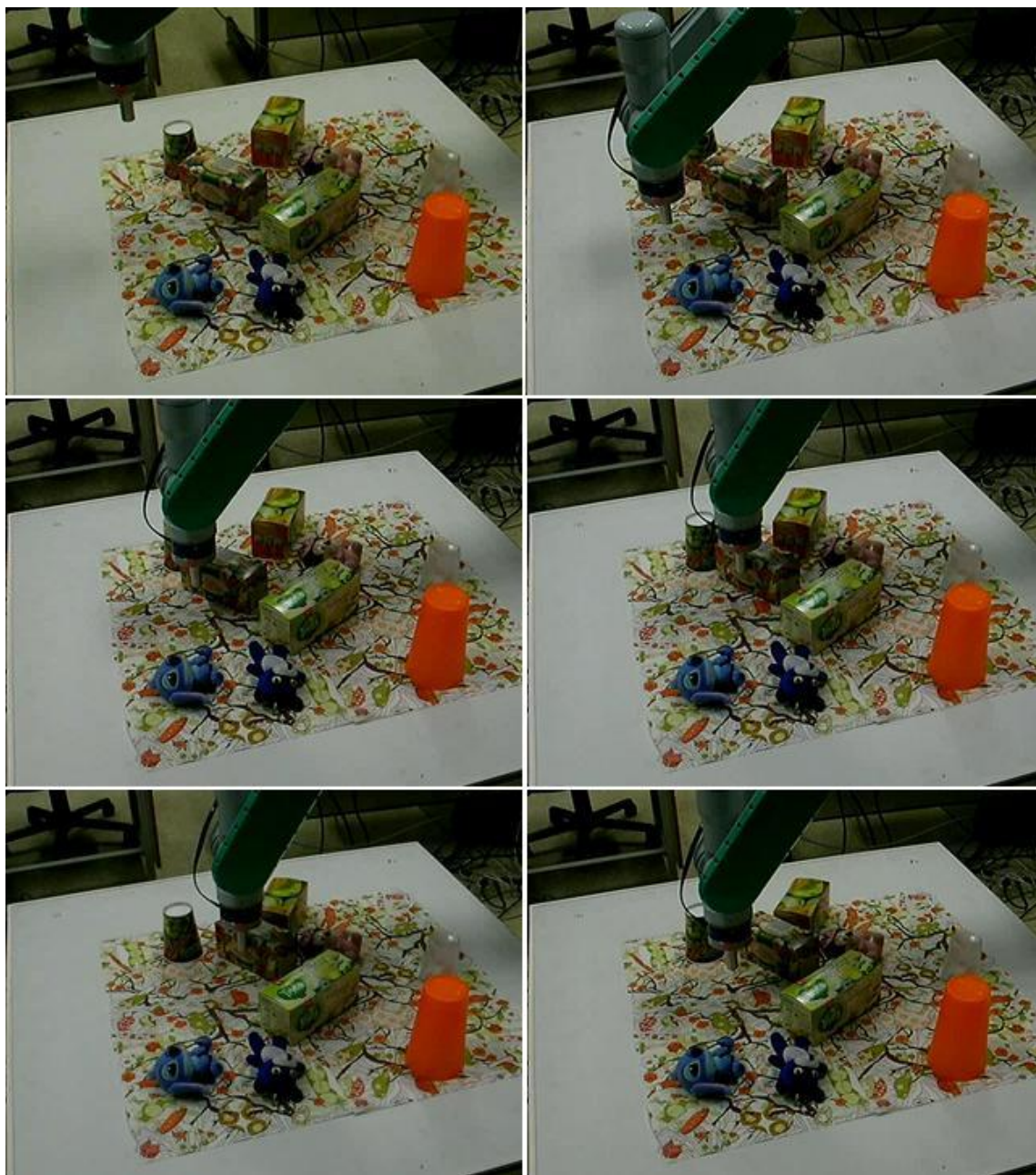
Pri 96% eksperimentov je bilo potiskanje izvedeno, pri ostalih eksperimentih določene akcije niso bile mogoče zaradi omejitev delovnega prostora. Slika 16 prikazuje enega od poskusov potiskanja izbrane hipotetične ravnine, ki poteka po vnaprej določeni trajektoriji.

5.2.3. Vrednotenje hipotez

Pri vrednotenju hipotez smo upoštevali dve možnosti. Prva možnost je, da točke, ki jih vsebuje hipoteza o predmetu, pripadajo enemu predmetu. V takšnih primerih bi moral robot točke prepoznati kot točke predmeta in (dodatno) poiskati več točk, ki pripadajo predmetu – seveda če se točke premaknejo kot posledica potiskanja.

Druga možnost je, da točke v hipotezi pripadajo več predmetom in/ali ozadju. Podobno kot v prvem primeru želimo premaknjene točke dodati modelu predmeta, ki ga robot premakne s potiskanjem. Dodatna naloga je odvzemanje točk, ki sicer so v hipotezi, a ne pripadajo predmetu, ki ga robot potisne. Predpostavka je, da se te točke ne premaknejo oziroma njihov premik ni v skladu s hipotezo o gibanju togega telesa.

Obe možnosti sta dokaj enakovredno zastopani, saj 42% hipotez vključuje točke, ki pripadajo več predmetom.



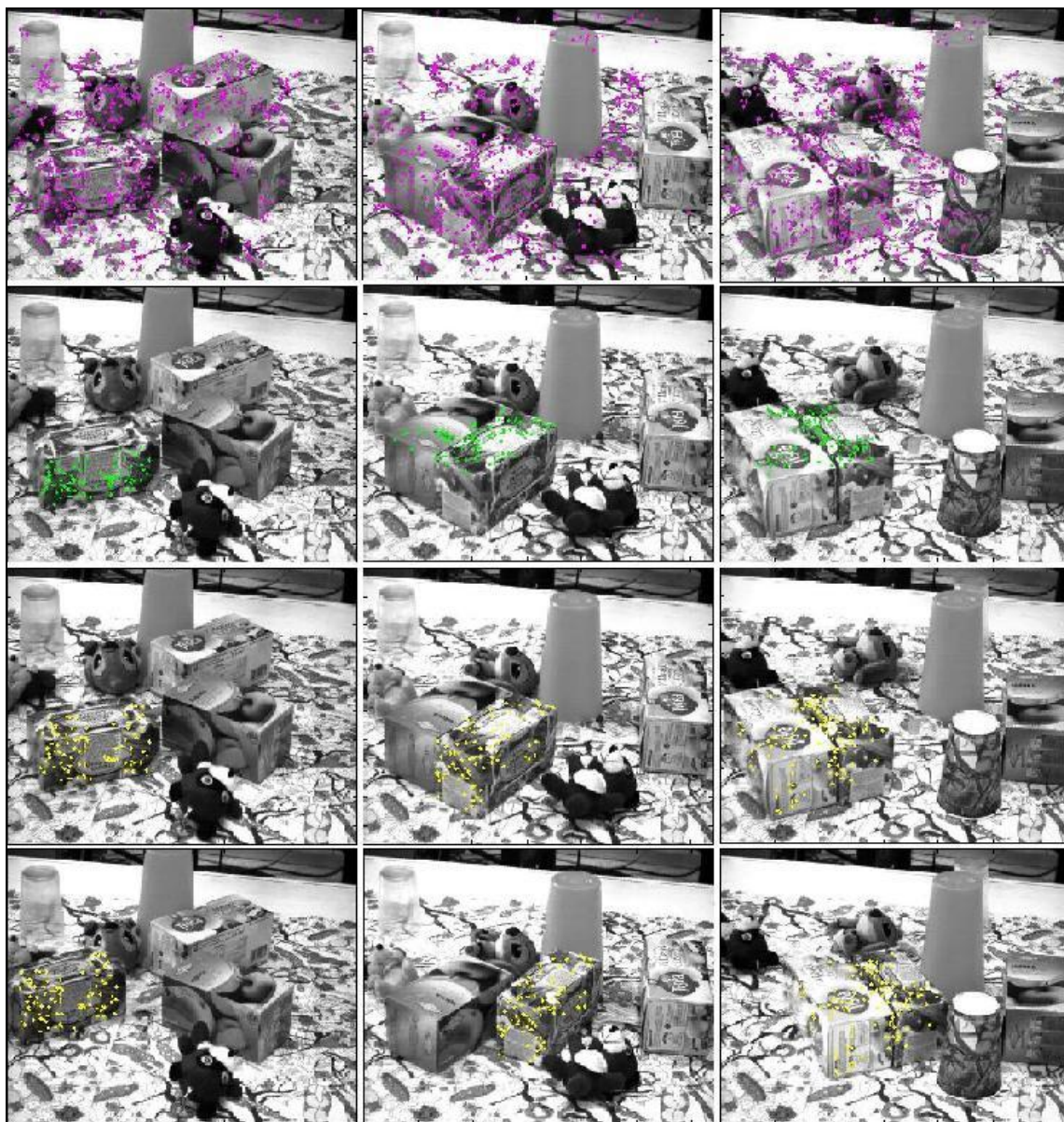
Slika 16: Potiskanje izbrane hipoteze za obstoj predmeta.

Slika 17 v stolpcih prikazuje rezultate nekaterih od izvedenih eksperimentov. Struktura slike je naslednja: prva vrstica prikazuje lokacije vseh točk, za katere smo lahko določili 3-D lokacije, druga vrstica pa točke, ki pripadajo izbrani hipotezi za obstoj predmeta. Tretja in četrta vrstica kažeta rezultate. Prikazane so točke, ki se premaknejo v skladu s hipotezo o togem gibanju predmeta. Prikazane so v slikah pred in po potiskanju.

Prvi stolpec demonstrira primer dodajanja točk. Hipoteza vsebuje točke enega predmeta. Robot ta predmet potisne in točke predmeta se zato premaknejo v skladu z načelom gibanja togega telesa. Tako pridobimo dovolj informacij za dodajanje točk modelu predmeta. Rezultat je nov model predmeta, ki poleg točk v začetni hipotezi vsebuje tudi dodatne točke.

Drugi stolpec prikazuje dodajanje in odzemanje točk. Hipoteza vsebuje točke več predmetov. Robot potisne enega izmed predmetov, katerega točke so vsebovane v hipotezi. Točke tega predmeta se premaknejo. Točke predmetov, ki se niso premaknili, algoritem uspešno zavrže.

Tretji stolpec, podobno kot drugi, prikazuje dodajanje in odzemanje točk. Hipoteza zopet vsebuje točke več predmetov. Razlika je v tem, da se oba predmeta, ko robot izvede potiskanje, premakneta zelo podobno. Točke obeh predmetov se premaknejo skladno z načelom gibanja togega telesa, zato jih algoritem označi kot točke predmeta. Med točkami različnih predmetov je torej mogoče razlikovati le, če se premaknejo različno.



Slika 17: Primeri eksperimentov.

Množico točk, ki jih naš sistem vključi v model objekta, lahko robot poveča z rotacijo objekta. Na ta način robot spremeni zorni kot, iz katerega opazuje objekt. Rotacija predmeta mora biti po eni strani dovolj velika, da odkrije nove, prej nevidne značilnice, in po drugi strani dovolj majhna, da omogoči določanje ujemanja značilnic v zaporednih slikah. Ker s potiskanjem pogosto ni mogoče natančno nadzorovati gibanja neznanih objektov, lahko pride pri potiskanju do takšnih premikov, ki niso primerni za dodajanje novih točk. Do tega lahko pride 1) zaradi prevelike rotacija predmeta, ki onemogoči ujemanje med zaporednima slikama, ali pa 2) ker do premika predmeta sploh ne pride.

Pri prevelikih rotacijah moramo ponovno potisniti opazovani predmet. Ker se je prizorišče spremenilo, mora biti potiskanje osnovano na na novo oblikovanih hipotezah o predmetih. Kadar je premik objekta premajhen, da bi bilo mogoče ločiti točke na podlagi principa togega gibanja, pa naš sistem izbere naslednjo hipotezo iz množice prej izračunanih hipotez in izvede nov potisk.

Tabela 2 prikazuje statistično oceno 50 potiskanj. Prva vrstica prikazuje delež eksperimentov, kjer je bila posledica potiska premik predmeta, in določanje točk predmeta. Delež takšnih eksperimentov, ki jih označimo kot 'Uspeh', je 58%. Druga vrstica označuje eksperimente, kjer je bil premik (oziroma rotacija) predmeta prevelika, zato ni bilo mogoče ujemanje značilnic v sosednjih slikah. Delež takšnih eksperimentov je 36%. Ker je prišlo do spremembe prizorišča, je potrebno oblikovati novo množico hipotez. Tretja vrstica opisuje eksperimente, kjer do spremembe prizorišča ni prišlo. Vzrok je neizvedeno potiskanje zaradi omejitev delovnega prostora (4% eksperimentov) in potiskanje, ki je sicer bilo izvedeno, a ni prispevalo k spremembi prizora (2% eksperimentov). V takšnih primerih ni potrebno izračunati novih hipotez, ker lahko uporabimo naslednjo izmed oblikovanih hipotez.

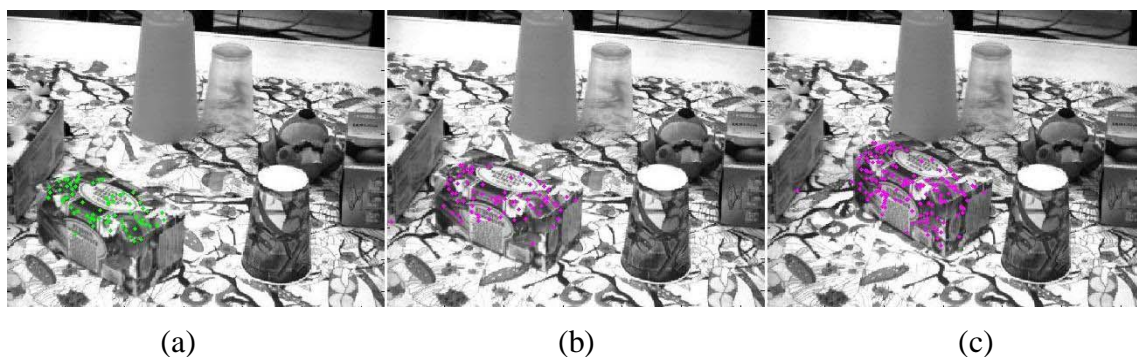
Tabela 2: Rezultati potiskanja za 50 eksperimentov

Točke predmeta	Rezultat	Delež eksperimentov
$\geq F_r$	Uspeh	58%
$< F_r$	Oblikuj nove hipoteze	36%
/	Izberi naslednjo hipotezo	6%

* $F_r = 10$ (prag)

5.2.4. Grajenje modelov

Informacije o predmetu zbiramo s spreminjanjem pogleda na predmet. Z vsakim novim pogledom želimo odkriti nove, prej nevidne značilnice, za kar potrebujemo določeno minimalno rotacijo telesa. Rotacija pa mora biti po drugi strani dovolj majhna, da je določanje ujemanja značilnic v zaporednih slikah mogoče. Slika 18 predstavlja uspešno kopičenje značilnic po dveh zaporednih potiskih, ki sta vodila v primerno rotacijo predmeta, in s tem uspešno grajenje modela predmeta.



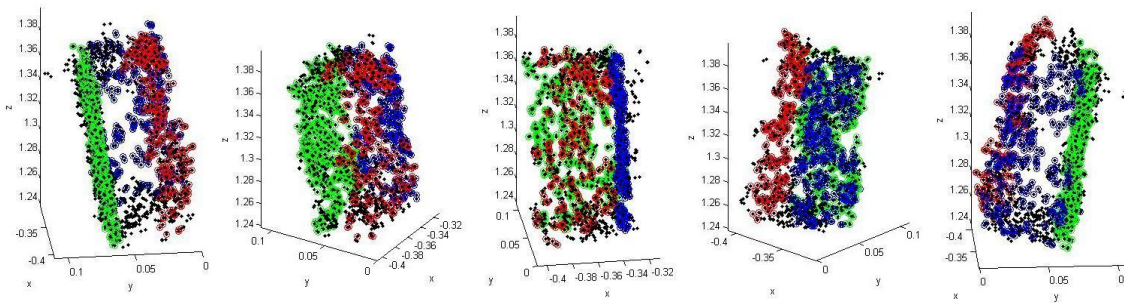
Slika 18: Grajenje modela s kopičenjem značilnic. (a) Hipoteza o obstoju predmeta. Določene točke predmeta po prvem potiskanju (b) in po drugem potiskanju (c).

S potiskanjem mora robot zagotoviti primeren premik objekta. Vendar pa lahko odziv predmeta na potisk natančno izračunamo le, če poznamo celoten model objekta. Dokler takšnega modela nimamo, lahko potiske določimo samo na hevrističen način. Zato se

lahko zgodi, da robot izvede potisk, ki ne omogoča nadaljnje gradnje modela.

V primeru velikih rotacij izračun ujemanja med vizualnimi značilnicami v sosednjih slikah ni mogoč. Zagotavljanje primerne rotacije preko več zaporednih slik je v celoti mogoče le z boljšo kontrolo nad predmetom. Večjo kontrolo nad predmetom lahko zagotovimo s prijemanjem predmeta ali pa v sodelovanju s človekom.

Zato smo izvedli eksperiment, v katerem smo preizkusili grajenje objektnega modela v sodelovanju s človekom. Pri tem človek potiska predmet na način, s katerim zagotovi, da je rotacija v zaporednih slikah primerna. Slika 19 prikazuje rezultat - model predmeta, ki vsebuje značilnice, izločene iz 38 pogledov na predmet.



Slika 19: Model predmeta na osnovi 38 različnih pogledov.

Ocenjene in dejanske mere predmeta:

- 0.060m in 0.060m,
- 0.070m in 0.067m,
- 0.131m in 0.132m.

Ocenjeni in dejanski koti med ravninami:

- 95.8° in 90° (modra in rdeča ravnina),
- 103.6° in 90° (zelena in rdeča ravnina).

Preizkusi grajenja modelov s potiskanjem robota so se zaradi prej naštetih dejstev končali preden bi lahko zajeli dovolj slik z različnimi pogledi na predmet.

Naš pristop pa je uspešen pri odkrivanju novih predmetov. Eksperimentalni rezultati kažejo, da je naš sistem uspešen pri oblikovanju hipotez, pri potiskanju izbrane detektirane ravnine in določanju točk, ki pripadajo predmetu.

6. Zaključek

Če predmet razumemo kot množico točk, ki pripadajo togemu telesu in se zato predvidljivo gibljejo, lahko nove predmete odkrivamo z analizo gibanja točk v zaporednih prizorih. V našem delu smo obravnavali prizore, ki vsebujejo objekte sestavljene iz planarnih ploskev. Ravnine torej namigujejo na obstoj predmeta, zato smo iskanje premetov smo osnovali na iskanju ravnin v slikah. Za potrditev takšne hipoteze o obstoju predmeta robot poizkusi detektirano ravnino premakniti s potiskanjem. Po uspešnem potisku lahko hipotezo o obstoju predmeta potrdimo ali ovržemo. Po več uspešnih potiskih je mogoče znanje o predmetu kopičiti preko zbranih značilnic.

Eksperimentalni rezultati kažejo, da je naš sistem uspešen pri oblikovanju hipotez. Pri vseh eksperimentih je bilo oblikovano vnaprej določeno število hipotez. Sistem je uspešen tudi pri potiskanju izbrane hipotetične ravnine. Pri določanju točk predmeta je sistem sposoben tako dodajanja točk, ki se premaknejo v skladu z ocenjenimi parametri gibanja togega telesa, kot tudi odvzemanja točk, ki ne pripadajo predmetu, iz začetne množice točk, ki sestavljajo hipotezo o obstoju predmeta. Prav tako ima sistem primeren odziv v primerih, ko je sprememba na prizorišču premajhna (ali pa je ni) in v primerih, ko je rotacija potisnjene predmeta prevelika. Lahko zaključimo, da je naš pristop uspešen pri odkrivanju novih predmetov.

Pri gradnji modelov predmetov kopičimo značilnice s spreminjanjem pogleda na predmet. V našem primeru to izvedemo s spremembo algoritma za potiskanje. Za uspešno grajenje modelov mora biti rotacija predmeta v določenih mejah. Rotacija predmeta mora biti po eni strani dovolj velika, da sistem odkrije nove, prej nevidne značilnice, in po drugi strani dovolj majhna, da je določanje ujemanja značilnic v zaporednih slikah mogoče. Ko določimo nove točke predmeta, torej točke, ki smo jih

določili iz novega pogleda, jih dodamo k obstoječim točkam predmeta. Lokacije ostalih točk predmeta preslikamo v skladu z izračunano rotacijo in translacijo.

Pridobljeni rezultati so pokazali, da je predlagan pristop uspešen pri odkrivanju novih predmetov in da je znanje, ki ga pridobimo z več zaporednimi manipulativnimi akcijami, mogoče zbirati v ustrezno verificirane modele predmeta.

Za uspešno gradnjo modelov včasih potrebujemo boljši nadzor nad gibanjem predmeta. Delež eksperimentov, kjer je bilo potrebno potiskanje ponoviti zaradi prevelike rotacije, je dovolj visok (36%), da je na mestu razmislek o uporabnosti zgolj potiskanja za gradnjo modelov. Boljši nadzor nad gibanjem predmeta je mogoče doseči s prijemanjem predmeta ali pa v sodelovanju s človekom. Na ta način lahko zagotovimo primerne premike opazovanega predmeta. Eksperimentalno smo preizkusili, da lahko v sodelovanju s človekom uspešno gradimo objektne modele.

Delo v prihodnosti bo usmerjeno predvsem v združitev predlaganega pristopa, ki je uspešen pri odkrivanju novih predmetov, in prijemanja, kar bo povečalo natančnost gibanja najdenih predmetov. V nadaljnjih raziskavah bomo odkrivanju predmetov in gradnji njihovih modelov dodali še določanje zmožnosti (angl. affordances) predmetov.

7. Viri

- [1] K. S. Arun, T. S. Huang, S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, št. 9, zv. 5, str. 698-700, 1987.
- [2] T. Asfour, K. Welke, P. Azad, A. Ude, R. Dillmann, "The Karlsruhe Humanoid Head", v zborniku *Humanoid Robots 2008*, str. 447-453, 2009.
- [3] I. Biederman, "Recognition by components: A theory of human image understanding," *Psychological Review*, št. 2, zv. 94, str. 115-147, 1987.
- [4] M. Brown, D. G. Lowe, "Invariant features from interest point groups," v zborniku *British Machine Vision Conference*, Cardiff, Wales, 2002, str. 656-665.
- [5] S. J. Dickinson, *What is cognitive science*, Basil Blackwell publishers, 1999, str. 172-207.
- [6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2004, pogl. 9.
- [7] J. Feldman, "What is a visual object?", *Trends in Cognitive Sciences*, št. 7, zv. 6, str. 252–256, 2003.
- [8] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Simultaneous object recognition and segmentation from single or multiple model views," *International Journal of Computer Vision*, št. 67, zv. 2, str. 159-188, 2006.
- [9] M. A. Fischler, R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, št. 24, zv. 6, str. 381-395, 1981.
- [10] P. Fitzpatrick, "First contact: An active vision approach to segmentation," v zborniku *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, Las Vegas, Nevada, 2003, str. 2161-2166.

- [11] Ch. Geib, K. Mourao, R. Petrick, N. Pugeault, M. Steedman, N. Krüger, F. Wörgötter, "Object Action Complexes as an Interface for Planning and Robot Control", v zborniku *International Conference on Humanoid Robots*, 2006.
- [12] J. Gibson, *The ecological approach to Visual Perception*, Houghton Mifflin, Boston, MA, 1979.
- [13] G. Golub, C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Mathematik*, št. 5., zv. 14, str. 403-420, 1970.
- [14] T. Kadir, A. Zisserman, M. Brady, "An affine invariant salient region detector," v zborniku *European Conference on Computer Vision*, Prague, Czech Republic, 2004, str. 228-241.
- [15] G. Kootstra, J. Ypma, B. de Boer, "Active exploration and keypoint clustering for object recognition," v zborniku *International Conference on In Robotics and Automation*, Pasadena, California, USA, 2008, str. 1005-1010.
- [16] D. Kraft, N. Pugeault, E. Baseski, M. Popovic, D. Kragić, S. Kalkan, F. Wörgötter, N. Krüger, "Birth of the object: detection of objectness and extraction of object shape through object–action complexes," *International Journal of Humanoid Robotics*, št. 5, zv. 2, str. 247-256, 2008.
- [17] Leonardis, A. Jaklič, F. Solina, "Superquadrics for segmentation and modeling range data," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, št. 19, zv. 11, str. 1289-1295, 1997.
- [18] W. H. Li, A. M. Zhang, and L. Kleeman, "Bilateral Symmetry Detection for Real-time Robotics Applications," *The International Journal of Robotics Research*, št. 7., Vol. 27, str. 785-814, 2008.
- [19] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, št. 2, zv. 60, str. 91-110, 2004.
- [20] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," v zborniku *5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, zv. 1, str. 281-297, 1967.
- [21] J. Matas, O. Chum, M. Urban, T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," v zborniku *13th British Machine Vision Conference*, Cardiff, UK, 2002, str. 384-393.

- [22] G. Metta and P. Fitzpatrick, "Grounding vision through experimental manipulation," *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, št. 361, zv. 1811, str. 2165-2185, 2003.
- [23] G. Metta and P. Fitzpatrick, "Early integration of vision and manipulation," *Adaptive Behavior*, št. 11, zv. 2, str. 109-128, 2003.
- [24] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, št. 27, zv. 10, str. 1615-1630, 2005.
- [25] K. Mikolajczyk, C. Schmid, "Scale & Affine Invariant Interest Point Detectors," *International Journal of Computer Vision*, št. 60, zv. 1, str. 63-86, 2004.
- [26] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, št. 65, zv. 1-2, str. 43-72, 2005.
- [27] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3-D objects," v zborniku *10th IEEE International Conference on Computer Vision*, Beijing, China, 2005, zv. 1, str. 800-807.
- [28] R. M. Murray, Z. Li., S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation* (1st Edition), CRC Press, 1994, pogl. 2.
- [29] D. Pelleg, A. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," v zborniku *Seventeenth International Conference on Machine Learning*, San Francisco, California, 2000, str. 727-734.
- [30] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, "3-D object modeling and recognition using affine-invariant patches and multi-view spatial constraints," v zborniku *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, 2003, zv. 2, str. 272-277.
- [31] R.B. Rusu, B. Gerkey, M. Beetz, "Robots in the kitchen: Exploiting ubiquitous sensing and actuation," *Robotics and Autonomous Systems Journal (Special Issue on Network Robot Systems)*, št. 10, zv. 56, str. 844-856, 2008.
- [32] A. Saxena, J. Driemeyer, A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, št. 2., zv. 27, str. 157-173, 2008.

- [33] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, št. 2, zv. 6, str. 461-464, 1978.
- [34] E. Trucco, A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.
- [35] E. Trucco, F. Isgro and F. Bracchi, "Plane detection in disparity space," v zborniku *International Conference on Visual Information Engineering*, Guildford, UK, 2003, str. 73–76.
- [36] C. J. Tsikos and R. K. Bajcsy, "Segmentation via manipulation," *IEEE Transactions on Robotics and Automation*, št. 7, zv. 3, str. 306-319, 1991.
- [37] Tuytelaars, L. Van Gool, "Matching Widely Separated Views Based on Affine Invariant Regions," *International Journal of Computer Vision*, št. 59, zv. 1, str. 61-85, 2004.
- [38] A. Ude, D. Omrčen, and G. Cheng, "Making object learning and recognition an active process," *International Journal of Humanoid Robotics*, št. 5, zv. 2, str. 267-286, 2008.
- [39] A. Ude and E. Oztop, "Active 3-D vision on a humanoid head," v zborniku *Proc. 14th Int. Conf. Advanced Robotics*, Munich, Germany, 2009.
- [40] (2009) A. Vedaldi, SIFT for Matlab. Dostopno na:
<http://www.vlfeat.org/~vedaldi/code/sift.html>
- [41] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, št. 22, zv. 11, str. 1330-1334, 2000.

Št.: 133-MAG-RI/2010
Datum: 17. 6. 2010



Eva STERGARŠEK KUZMIČ, univ. dipl. inž. el.

Ljubljana

Fakulteta za računalništvo in informatiko Univerze v Ljubljani izdaja naslednjo magistrsko nalogo

Naslov naloge: **Odkrivanje novih predmetov in grajenje njihovih modelov z aktivnim robotskim sistemom**

Detecting new objects and building models with active robot system

Tematika naloge:

Sposobnost odkrivanja novih predmetov je težko poustvariti s popolnoma pasivnimi vizualnimi sistemi brez predznanja. Ena od možnosti reševanja tega problema so aktivni robotski sistemi, ki poleg sistema za zajemanje in obdelavo slik vsebujejo elemente za manipulacijo delov prizorišča.

V okviru realnega robotskega sistema razvijte robusten pristop za odkrivanje neznanih predmetov v poljubnih prizorih, ki temelji na vizualnih lokalnih značilnicah. Izločene značilnice naj bodo vhod v sistem za robustno oblikovanje hipotez o obstoju predmetov. Način manipulacije določite na podlagi izbranih hipotez o obstoju predmeta, posledice manipulacije pa uporabite za robustno vrednotenje hipotez. Pokažite možnost uporabe tega pristopa za gradnjo modelov novih predmetov.

Mentor:


prof. dr. Aleš Leonardis



Dekan:


prof. dr. Franc Solina

Somentor:


doc. dr. Aleš Ude

FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO
UNIVERZE V LJUBLJANI

Eva Stergaršek Kuzmič

Odkrivanje novih predmetov in grajenje njihovih
modelov z aktivnim robotskim sistemom

MAGISTRSKO DELO

Mentor: prof.dr. Aleš Leonardis
doc.dr. Aleš Ude

Ljubljana, 2010

IZJAVA O AVTORSTVU

magistrskega dela

Spodaj podpisana Eva Stergaršek Kuzmič,

z vpisno številko 63070478

sem avtorica magistrskega dela z naslovom:

Odkrivanje novih predmetov in grajenje njihovih modelov z aktivnim robotskim sistemom.

S svojim podpisom zagotavljam, da:

- sem magistrsko delo izdelala samostojno pod vodstvom mentorja prof.dr. Aleša Leonardisa in somentorja doc.dr. Aleša Udeta.
- so elektronska oblika magistrskega dela, naslova (slov., angl.), povzetka (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko magistrskega dela
- in soglašam z javno objavo elektronske oblike magistrskega dela v zbirki »Dela FRI«.

V Ljubljani, dne 23.9.2010

Zahvala

Zahvaljujem se mentorjema dr. Alešu Leonardisu in dr. Alešu Udetu. Zahvala gre tudi sodelavcem z odseka E-1 na inštitutu "Jožef Stefan", prijateljem in družini.

Kazalo

Povzetek	1
Abstract	3
1. Uvod.....	5
1.1. Struktura naloge.....	10
2. Pregled področja.....	11
3. Detekcija in učenje predmetov.....	15
3.1. Oblikovanje hipotez o obstoju predmeta.....	17
3.1.1. Pridobivanje 3-D informacij.....	18
3.1.2. Iskanje planarnih delov v prizoru	20
3.2. Potiskanje na osnovi izbrane hipoteze	23
3.3. Vrednotenje hipotez	26
3.4. Grajenje modelov.....	32
4. Metodologija	33
4.1. Iskanje in opis vizualnih značilnic.....	33
4.1.1. Odkrivanje ekstremov.....	34
4.1.2. Določanje lokacije	35
4.1.3. Določanje orientacije	36
4.1.4. Izgradnja opisnikov.....	36
4.2. Algoritem RANSAC	37
4.3. Gručenje točk.....	39
4.3.1. Algoritem K -means.....	39
4.3.2. Algoritem X -means.....	40
4.4. Razcep s singularnimi vrednostmi.....	43
4.5. Kalibracija kamere.....	43

5.	Eksperimenti	46
5.1.	Realni robotski sistem	46
5.2.	Preizkusi.....	48
5.2.1.	Oblikovanje hipotez	49
5.2.2.	Potiskanje.....	50
5.2.3.	Vrednotenje hipotez	50
5.2.4.	Grajenje modelov	55
6.	Zaključek	58
7.	Viri.....	61

Kazalo slik

Slika 1: Roboti za domača okolja.	6
Slika 2: Segmentacija neznanih predmetov.	7
Slika 3: Trije križi.	9
Slika 4: Nadzorovano učenje s sintetičnimi slikami z označenimi mesti za prijemanje.	12
Slika 5: Prijemanje na podlagi primitivov.	13
Slika 6: Prvi dotik robotske roke in predmeta.	14
Slika 7: Oris našega pristopa.	16
Slika 8: Pridobivanje 3-D informacij.	20
Slika 10: Določanje začetne točke in smeri potiskanja.	25
Slika 11: Postopek izgradnje SIFT opisnikov.	37
Slika 12: Delovanje X-means.	42
Slika 13: Eksperimentalna robotska celica.	46
Slika 14: Vsi testni predmeti.	48
Slika 15: Tri izmed hipotez za obstoj predmeta.	49
Slika 16: Potiskanje izbrane hipoteze za obstoj predmeta.	51
Slika 17: Primeri eksperimentov.	53
Slika 18: Grajenje modela s kopičenjem značilnic.	55
Slika 19: Model predmeta na osnovi 38 različnih pogledov.	56

Kazalo tabel

Tabela 1: Podmnožice testnih predmetov.....	49
Tabela 2: Rezultati potiskanja za 50 eksperimentov.....	55

Kazalo algoritmov

Algoritem 1: Oblikovanje hipotez.	23
Algoritem 2: Potiskanje ravnine.	26
Algoritem 3: Vrednotenje hipoteze.....	31

Povzetek

Bistven element kognitivnih robotskih sistemov je zmožnost spoznavanja predmetov, s katerimi robot pride v stik prvič, in učenja predstavitev za takšne predmete, ki so uporabne za kasnejše razpoznavanje. Naša predpostavka je, da je odkrivanje novih predmetov izvedljivo na učinkovit način z aktivnim robotskim sistemom, ki poleg vizualnega sistema za zajemanje slik prizorov vsebuje elemente za manipulacijo delov prizora. Manipulacija služi za ločitev predmeta iz okolice in za akumulacijo objektnih značilnic, na osnovi teh pa je mogoče modeliranje predmeta in ugotavljanje njegove funkcionalnosti.

V tem delu podajamo postopek odkrivanja in segmentacije predmetov s pomočjo kombinirane rabe vizualnih in manipulativnih funkcij kognitivnega robotskega sistema. Postopek smo preverili z več eksperimenti z realnim robotskim sistemom.

V predlaganem pristopu je postopek spoznavanja predmeta sestavljen iz treh procesov. Prvi proces predstavlja oblikovanje hipotez o obstoju predmetov. Oblikovanje hipotez sestavlja več podprocesov: izločanje in gručenje točk vizualnih značilnic, določanje 3-D lokacij in iskanje ravnin.

Drugi proces definira optimalne manipulativne akcije za verifikacijo predhodno določenih hipotez o obstoju predmeta. Z več zaporednimi manipulativnimi akcijami je mogoče znanje o predmetih kopičiti, na podlagi teh pa določati modele predmeta. V okviru tega dela manipulativne akcije predstavlja potiskanje z robotsko roko.

Tretji proces hipoteze o obstoju predmeta dejansko verificira na podlagi novih informacij, pridobljenih z manipulacijo. Verifikacija temelji na ugotavljanja konsistentnosti detektiranih značilnic predmeta glede na gibanje togega telesa. V primeru, da je kateri izmed procesov neuspešen, lahko sistem oblikuje in ovrednoti

dodatne hipoteze, zato predpostavka, da so procesi popolni, ni potrebna. Za zagotovitev robustnosti pristopa smo za več podprocesov uporabili verjetnostne metode, kot je na primer RANSAC (RANdom Sample Consensus).

Pridobljeni rezultati so pokazali, da je predlagan pristop uspešen pri odkrivanju in segmentaciji novih predmetov in da je znanje, ki ga pridobimo z več zaporednimi manipulativnimi akcijami, mogoče zbirati v popolnejše modele predmeta.

Ključne besede: detekcija predmetov, segmentacija predmetov, učenje z raziskovanjem, učenje konceptov, RANSAC.

Abstract

An important element of a cognitive robotic system is the ability to detect novel objects and learn their representations, which are suitable for later recognition and manipulation. The basic assumption of our work is that the detection and segmentation of new objects can be facilitated by an active robotic system, which can not only observe the objects but can also manipulate them. Manipulation supports object segmentation and the accumulation of object features, which provides the basis for building object models and for determining their functionality.

In this thesis we propose a new approach for object detection and segmentation based on the integrated use of visual and manipulative functions of a robotic system. The developed system was tested in a number of experiments on a real robot.

In the proposed approach, the object detection process consists of three consecutive procedures. The first procedure deals with the generation of hypotheses about the existence of an object. It comprises several sub-processes: extraction of visual features, calculation and clustering of 3-D points, and discovery of planar surfaces.

The second process defines manipulative actions which the robot needs to perform to validate the previously calculated object hypotheses. By employing these manipulative actions, the existence of an object can be confirmed and additional object knowledge can be accumulated. In our system, suitable manipulative actions are realized as pushing movements.

After the completion of the pushing movement, the third procedure verifies the underlying object hypothesis based on the newly acquired information. The verification procedure evaluates the consistency of the detected object features with respect to the assumption of rigid body motion. Additional hypotheses are evaluated if the verification

process fails. To ensure the robustness of the system, probabilistic methods such as RANSAC (RANdom SAmple Consensus) have been applied in several computational stages.

Our experimental results show that the system provides successful object detection and segmentation in complex scenes. Our work also demonstrates the robot can acquire more complete object models by performing several consecutive manipulative actions.

Keywords: object detection, object segmentation, learning by exploration, concept learning, RANSAC.

1. Uvod

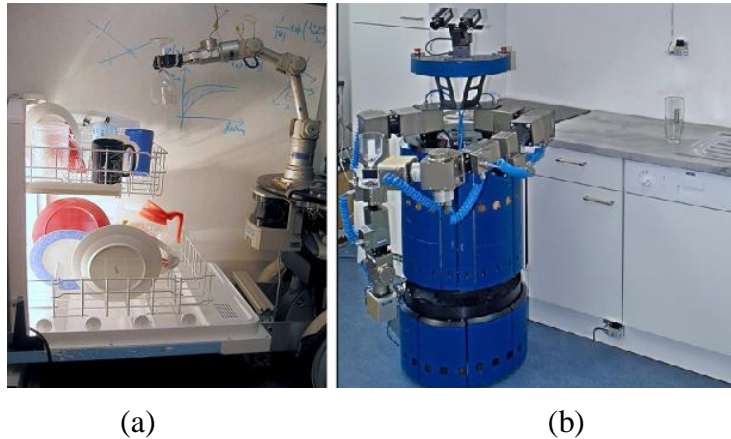
Veliko raziskav je usmerjenih v razvoj inteligentnih robotov, ki bi lahko delovali v običajnih človeških okoljih, na primer v kuhinji, in izvajali običajne naloge, kot so na primer zlaganje posode iz pomivalnega stroja. Slika 1 prikazuje dva primera takšnih robotov. Roboti za opravljanje takšnih nalog potrebujejo sposobnost interpretacije vizualnih informacij.

Ljudje smo zelo uspešni pri prepoznavanju na podlagi vizualnih informacij. Zgolj na podlagi vizualne informacije, tudi če samo pasivno opazujemo, lahko brez večjih naporov prepoznavamo različne predmete, osebe, kraje in preostale informacije, ki jih pridobimo iz prizorov, na primer gibanja in celo občutke. Uspešno prepoznamo tudi predmete, ki jih vidimo prvič, čeprav so ti umeščeni v vizualno zelo zapletena prizorišča. To uspešnost je težko ponoviti s popolnoma pasivnimi vizualnimi sistemi. Takšni sistemi so sicer sposobni pravilnega ločevanja preprostih predmetov in prizorišč brez prekrivanj, pasivni vizualni sistemi pa lahko odpovejo pri zanesljivemu in pravilnemu ločevanju gruč neznanih predmetov [36]. Robot brez določenega nivoja predznanja se pri razstavljanju slik v smiselne dele pogosto znajde pred nemogočo nalogo. Tak primer prikazuje Slika 2.

Vizualni sistemi brez predznanja ne morejo ločiti med, na primer, dvema predmetoma, ki se prekrivata, in enim samim predmetom. Podobno težavna je naloga segmentacije predmetov z vizualno zapletenih ozadij.

Ena izmed rešitev je, da robota opremimo z modeli vseh predmetov, ki bi se lahko znašli v njegovem okolju. Ta pristop ni praktičen. Če na primer vzamemo robota za domača opravila, je zelo verjetno, da bo pri opravljanju svojih nalog (čiščenje, pospravljanje)

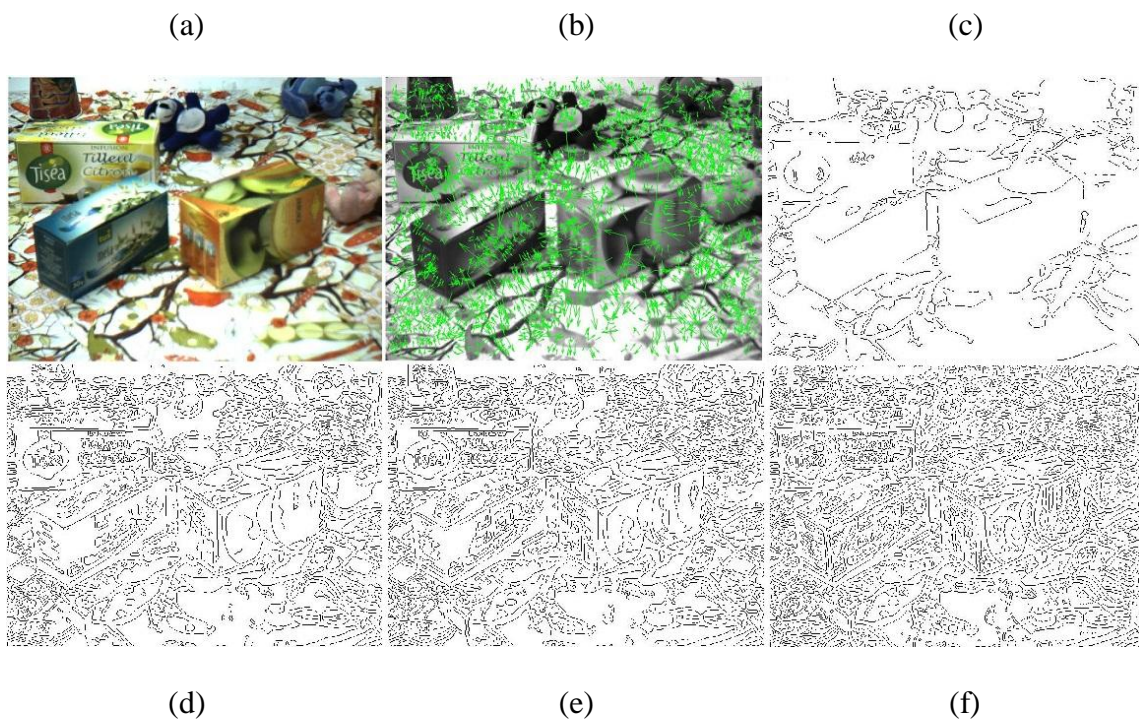
naletel na nove predmete (na primer pločevinka pijače), za katere ne bo imel modelov. Če želimo prilagodljivega robota, bo ta moral biti opremljen z drugačne vrste znanjem, in sicer z znanjem, ki mu bo omogočalo, da bo lahko odkrival, ločeval in sledil novim predmetom, za katere (še) nima modelov. Resnično avtonomen robot mora biti sposoben raziskovanja in učenja na podlagi informacij iz okolja, ker ne more biti odvisen od prejemanja vseh informacij na pasiven način [22]. Avtonomnim robotom je zato smiselno podati predznanja, ki mu pomagajo pri aktivnem raziskovanju novih okolij.



Slika 1: Roboti za domača okolja. (a) Robot pri nalogi zlaganja posode iz pomivalnega stroja. Povzeto po [32]. (b) Robot za delo v kuhinji. Povzeto po [31].

V splošnem je težko definirati, kaj pomeni pojem 'predmet'. Koncept 'predmet' ima pomembno vlogo v kognitivni znanosti, čeprav ustrezna formalna definicija ne obstaja. Definicija pojma objekt je odvisna od načina, kako razdelimo zaznani svet na smiselne enote [7]. Obstaja veliko principov, ki jih lahko uporabimo pri definiciji objektov, na primer zaprtje (closure), povezanost (connectedness), dvostranska simetrija [18], konveksnost, koplanarnost in kolinearnost obrisov. Čeprav vsak od teh principov pomaga pri zaznavanju predmetov, ni nobeden od njih vedno ustrezen. Z lahkoto poiščemo stvar, ki ustreza nekemu principu, a ni samostojen predmet. Prekrižane paličice, na primer, ustrezajo principu povezanosti, a jih dojemamo kot dva predmeta.

Zato je mogoče takšne osnovne principe uporabiti le za oblikovanje hipotez o obstoju predmetov, za potrditev hipotez pa je potrebno uvesti dodatne paradigme.



Slika 2: Segmentacija neznanih predmetov. (a) Izvirna slika. (b) Značilnice SIFT. (c)-(f) Izločeni robovi - detektor robov Canny - prag 0.5 (c), 0.1 (d), 0.05 (e) in 0.001 (f).

Glede na aktivnega agenta, na primer robota, lahko, glede na [12], objekte opišemo z naslednjimi lastnostmi:

- določena minimalna on maksimalna velikost,
- časovna stabilnost in
- upravljivost.

Določanje, kaj je predmet glede na te značilnosti, ni preprosta naloga. S pomočjo teh lastnosti lahko definiramo procese, ki jih lahko uporabimo za odločanje o obstoju predmetov. Zadnjo lastnost lahko preverimo tako, da na predmete vplivamo in jih

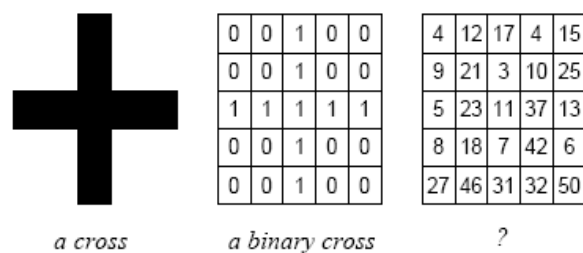
premikamo. Kriterij časovne stabilnosti pa lahko uporabimo pri primerjavah izločenih vizualnih značilnic pred in po izvedeni akciji.

Pojem povezave predmet-akcija (object-action complex, OAC) [11] povezuje predmete in dejanja, ki jih na njih izvedemo. Predmeti imajo po eni strani različne zmožnosti (angl. affordances), na primer drsenje in valjenje, po drugi strani pa jih te akcije določajo. Posledice operacije, ki jo izvedemo na predmetu, pripomorejo pri odločanju o obstoju predmeta in določanju njegove oblike. Čeprav je vizualno zaznavanje polno iluzij, je mogoče večino nejasnosti odpraviti, če lahko robot premika predmete. V prisotnosti manipulacije postane vid močnejši, in večina iluzij izgine. Poleg vida lahko pri tem pomagajo tudi taktilne informacije.

Ni povsem jasno, kakšno je predznanje človeškega sistema. Res pa je, da človeška sposobnost ločevanja predmetov ni univerzalna in se izboljšuje z izkušnjami. Primer so trije križi na naslednji sliki (Slika 3). Iz levega dela slike z lahkoto prepoznamo križ, mogoče zato, ker že imamo kriterij za razločevanje – razliko v intenziteti slikovnih elementov. Srednji del slike, malo manj jasno, a še vseeno, tudi predstavlja križ. Kriterij (ničle proti enicam) nam je poznan. Pri desnem delu slike pa, razen če nimamo dodatnih podatkov, lahko sklepamo, da gre za naključna števila. Šele ko je kriterij (praštevila proti številom) znan, lahko v sliki zopet odkrijemo križ. Primer na desnem delu slike predstavlja nov način ločevanja, na katerega verjetno do sedaj še nismo naleteli. Lahko si predstavljamo, da smo na podoben način od otroštva naprej pridobivali kriterije za ločevanje in identificiranje predmetov. Odrasel, izkušen človek si lahko pogosto razlaga vizualne prizore brez delovanja na prizorišča. Povezovanje aktivnosti in zaznavanja pa je ključna pri razvijanju te sposobnosti [22].

Naš cilj je opremiti robotske sisteme s takšnimi sposobnostmi, ki jim omogočajo aktivno raziskovanje in s tem odkrivanje novih predmetov in učenje njihovih modelov. Aktivno raziskovanje pri tem pomeni izvajanje akcij na 'stvareh'. V naši terminologiji je 'stvar'

hipoteza za predmet in se spremeni v predmet šele, ko se skozi delovanje robota izkaže, da gre resnično za predmet s konsistentnimi lastnostmi.



Slika 3: Trije križi. Povzeto po [22].

Osnovni gradniki našega pristopa so:

- Oblikovanje hipotez o obstoju predmetov. Iz stereo slik prizora izločimo vizualne značilnice in jih opišemo s pripadajočimi vektorji in 2-D lokacijo. Značilnice uredimo glede na ujemanja, ki jih ovrednotimo s pomočjo evklidske razdalje med vektorji značilnic in epipolarne geometrije. Urejeni pari značilnic in njihove 2-D lokacije nam omogočajo izračun 3-D lokacij. Med množico 3-D točk poiščemo pojavljanja skupkov točk, ki opisujejo ravne površine. Te ravnine uporabimo za določanje manipulativnih akcij.
- Manipulacija izbrane hipoteze. Na površino vplivamo z določeno akcijo, kot je na primer potiskanje, in tako pridobimo dodatne informacije, s katerimi lahko gradimo popolne in preverjene predstavitve predmetov.
- Vrednotenje hipoteze. Med 3-D točkami določimo ujemanja in ocenimo parametre povzročene gibanja predmetov. Nato preverimo, ali se točke hipotetičnega predmeta premikajo kot točke, ki naj bi pripadale predmetu. Pri tem smo predpostavili, da robot potiska toga telesa, zato se morajo izločene značilnice premikati kot toga telesa. Nekatere točke in pripadajoče značilnice lahko zavržemo ali dodamo glede na proces vrednotenja hipoteze.

- Grajenje modelov. Na podlagi več zaporednih potiskov lahko zberemo še več značilnic in točk, ki pripadajo objektu, in tako počasi zgradimo bolj kompleten model objekta.

1.1. Struktura naloge

V drugem poglavju predstavimo pomembna dela in pristope na tem področju. V tretjem poglavju opišemo naš pristop oblikovanja in vrednotenja hipotez o obstoju predmetov s pomočjo manipulativnih akcij. Sledi mu poglavje, v katerem predstavimo metode, ki smo jih uporabili pri realizaciji predlaganih procesov. To so algoritmi za odkrivanje in opisovanje vizualnih značilnic, iskanje parametrov modelov in gručenje. Peto poglavje opisuje eksperimentalno delo, kjer predstavimo naš realni robotski sistem in rezultate. Naši sklepi so podani v šestem poglavju.

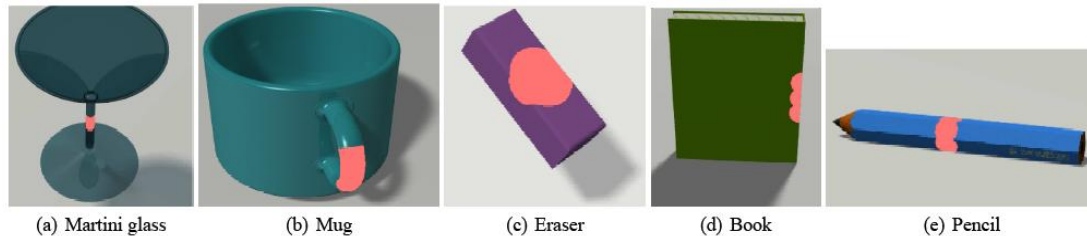
2. Pregled področja

Rezultati raziskovalnih del kažejo, da lahko robot raziskuje in se uči iz učinkov, ki jih ima njegovo delovanje na zunanji svet. Aktivno raziskovanje omogoča ločevanje predmetov od nepreglednega ozadja, kar ni preprosta naloga za pasivni vizualni sistem. Robot lahko spremeni pogled na predmet s premikanjem svojega telesa, eden od takšnih sistemov je opisan v [14], ali pa z izvajanjem akcij, ki premaknejo opazovane predmete. Predmete lahko premaknemo na več načinov, na primer s potiskanjem [22],[23] ali pa s prenašanjem v roki [16],[37]. Slednje je mogoče le, če je predmet mogoče oziroma ga znamo prijeti. Več raziskovalnih del je usmerjenih v pridobivanje novih znanj o predmetih s pomočjo prijemanja prvotno neznanih predmetov.

Saxena in sodelavci [32] so predstavili algoritem, ki za določanje načina prijemanja novih predmetov ne potrebuje in ne gradi modelov predmetov. Predlagani učni algoritem je po obdobju nadzorovanega učenja sposoben predvideti dobre točke za prijemanje. Učenje določanja točk za prijemanje poteka s pomočjo testnih slik, ki imajo označena 2-D mesta za prijemanje. Če je mogoče mesta za prijemanje določiti na več slikah predmeta (dve sliki ali več), je s triangulacijo mogoče določiti 3-D koordinate primernih mest za prijemanje, kjer lahko robot poskusi prijeti hipotetični predmet. Z nadzorovanim učenjem je v sistem vnesen določen nivo predznanja o prijemanju predmetov. Da zagotovimo neobčutljivost na spremembe v videzu predmetov (barva, merilo, tekstura, besedilo na knjigah - na te lastnosti vpliva osvetlitev, lokacija in orientacija kamer), je potrebno uporabiti ogromen nabor testnih slik.

Slika 4 prikazuje nekaj primerov testnih slik. Za predstavljenih 5 razredov predmetov (kozarec za martini, skodelica, radirka, knjiga in svinčnik) je bilo oblikovanih 2500 primerov sintetičnih slik z označenimi mesti za prijemanje. Za sposobnost posploševanja

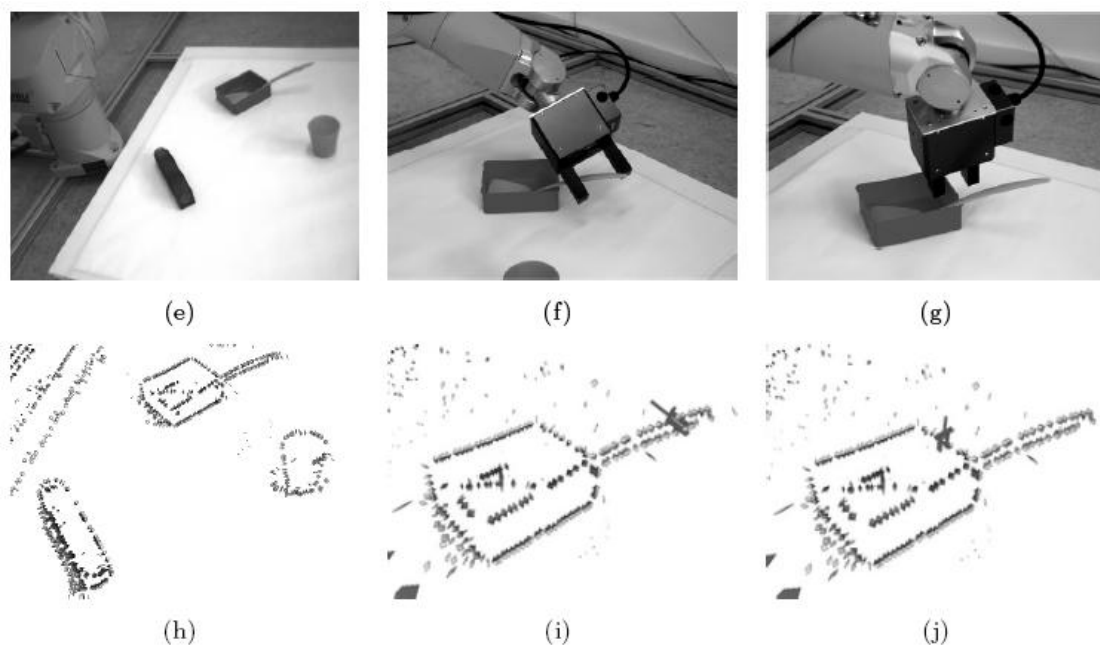
na nove predmete je poleg obsežnega nabora testnih slik potrebno uporabiti veliko različnih vrst vizualnih značilnic.



Slika 4: Nadzorovano učenje s sintetičnimi slikami z označenimi mesti za prijemanje. Povzeto po [32].

Podoben pristop za prijemanje nepoznanih predmetov opisujejo Kraft in sodelavci [16]. Njihov pristop temelji na strukturiranem predznanju. Prizore opisujejo lokalni simbolični opisniki (primitivi), ki zajemajo več vizualnih modalitet, ter 2-D in 3-D vidike vizualnih informacij. Slika 5 prikazuje realno okolje in izločene opisnike, hipoteze za prijemanje in dve uspešni prijemanji. Kot so pokazale druge raziskave, je po uspešnem prijemanju mogoče opazovati, upravljati in ločiti predmet od ozadja in sistematično pridobiti slike predmeta iz različnih zornih kotov [38].

Čeprav imamo s prijemanjem predmeta boljši nadzor nad predmetom, pa je prijemanje v primerjavi s potiskanjem veliko bolj zapletena in občutljiva operacija. Prijemanje novih predmetov, za katere nimamo modelov, je še toliko bolj zahtevna operacija. V veliko primerih je težavno prijeti nepoznan predmet, bodisi zaradi velikosti bodisi zaradi posebnosti v obliki. V takšnih primerih potiskanje predstavlja primerno alternativo, saj je preprostejše za izvedbo, kljub temu pa, če je uspešno, premakne predmet.



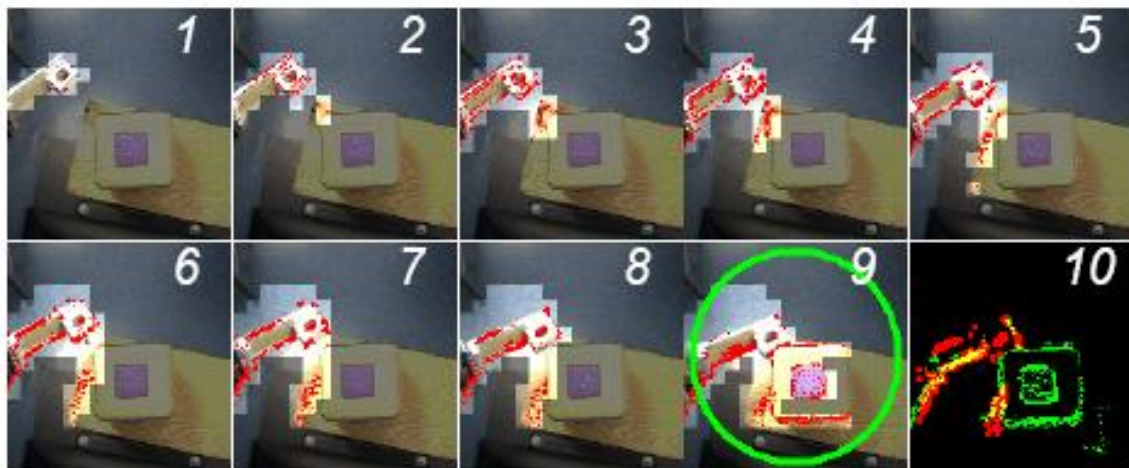
Slika 5: Prijemanje na podlagi opisnikov. Pogled na realno okolje (e) in pripadajoči 3-D opisniki (h). Uspešni prijemanji (f-g) prikazani v 3-D ponazoritvenem okolju (i-j). Povzeto po [16].

Metta in Fitzpatrick [22],[23] sta pokazala, da je mogoče s preprostimi gibanji pridobiti dovolj informacij za ločitev neznanih predmetov od ozadja in določiti zmožnosti (angl. affordances) predmetov. Eksperimenti so bili izvedeni s humanoidnim robotom z dvema rokama. Robot ni pasiven opazovalec, pač pa lahko svojo roko premika, kar omogoča ločitev roke od okolja. Z dovolj znanja o svoji roki lahko robot začne z raziskovanjem okolja. Pri dotiku roke in predmeta je možnih več scenarijev. Če je predmet velik in težak, ne bo prišlo do vizualno opaznega premika. Takšni predmeti so označeni kot ovire, saj jih robot ne more premakniti, in niso zanimivi za izvajanje poskusov. Če je predmet manjši, pa je precej bolj verjetno, da se bo premaknil v odgovor na dotik robotske roke. Slika 6 prikazuje tak primer.

Ko je tarča (hipoteza o obstoju predmeta) določena, robotska roka seže proti njej in 'pomete' področje okoli tarče. Pristop ne vsebuje algoritmov za ugotavljanje prvotnih

hipotez o obstoju predmetov. V opisanem primeru je tarča zgolj 'zanimivi' del predmeta, v tem primeru temnejši kvadrat na zgornji ploskvi predmetu. Gibanje robotske roke ni usmerjeno k optimalnemu načinu premikanja izbranega dela prizora, pač pa v vsakem primeru le k 'pometanju' okolice v bližini tarč.

Postopek ločevanja predmeta od okolice temelji na prvem dotiku robota in predmeta, ki je zaznan vizualno. Izrazite spremembe vrednosti točkovnih elementov v slikah so označene kot hipotetično gibanje. Ko se robotska roka premika, sistem sledi njenemu gibanju, a ga izloči skupaj s senco in gibanjem v ozadju. Ko se roka premakne preko nekega področja, je to področje označeno kot 'področje brez predmetov' za kratek čas. Ko se gibanje roke razširi preko veliko večjega področja, kot bi bilo mogoče v danem času, to nakazuje prvi dotik robotske roke in predmeta, ki ju sedaj mogoče ločiti.



Slika 6: Prvi dotik robotske roke in predmeta. (1) Roka 'pometa' okolico tarče. (2-8) Sledenje gibanju roke. (9) Prvi dotik. (10) Ločitev roke in predmeta. Povzeto po [10].

3. Detekcija in učenje predmetov

Razvili smo metodologijo, po kateri na podlagi oblikovanih hipotez o obstoju predmetov ustvarimo manipulativne akcije, s katerimi hipoteze potrdimo ali ovržemo. Za generacijo hipotez predlagamo bolj zmogljiv vizualni sistem, ki ne temelji na točkovnih elementih kot v [22],[23], ampak na vizualnih značilnicah in njihovih 3-D vidikih, in upošteva tudi razmerja med značilnicami. V razvitem sistemu oblikujemo začetne hipoteze, ki jih preverimo z manipulacijo. To je pomembna prednost pred sistemi, ki ne morejo aktivno delovati na predmete.

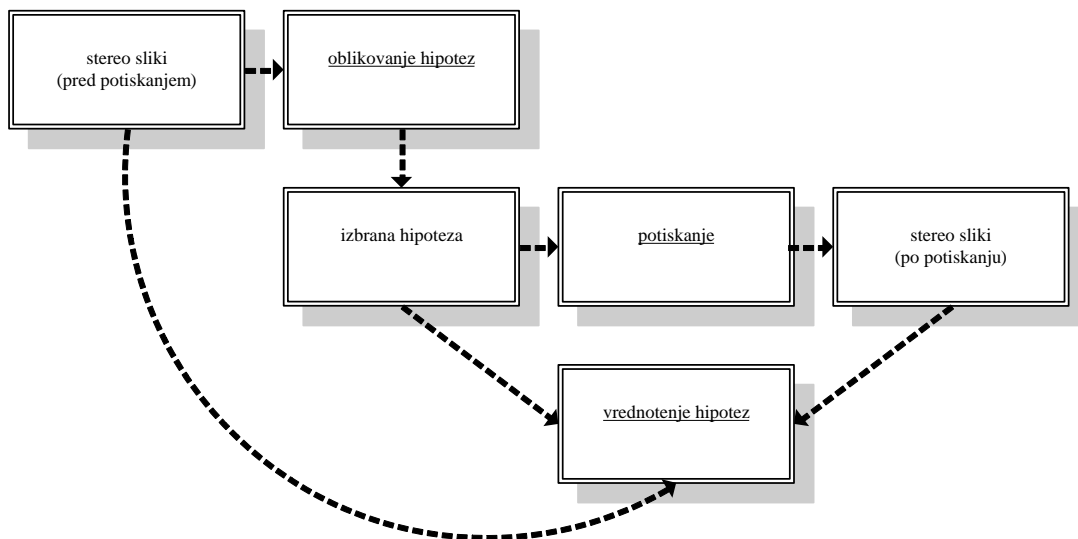
V tem delu smo se usmerili predvsem na domača okolja. Takšna okolja vsebujejo veliko predmetov z ravnimi površinami, zato smo pri razvoju metod za generacijo hipotez o predmetih obravnavali predmete, ki vsebujejo ravne površine. Ravne površine lahko predstavimo le s štirimi parametri. To je pomembno za nekatere algoritme, ki jih pri našem pristopu uporabljamo, na primer algoritem RANSAC [9]. S tem algoritmom hipoteze oblikujemo z oceno parametrov hipotetičnih ravnin. Sistem lahko oblikuje in vrednoti dodatne hipoteze, zato ni potrebno predpostaviti, da so procesi popolni. Vendar pa naš pristop ne deluje samo za ravne površine. Iz literature poznamo matematične strukture kot so geoni [3] in superkvadriki [17]. Obe metodologiji sta bili predlagani za rekonstrukcijo površin 3-D modelov. Takšne ploskve bi prav tako lahko uporabili za generacijo hipotez o obstoju predmetov, vendar pa bi pri tem morali oceniti večje število parametrov kot pri ravninah.

Določanje optimalnih premikov, ki potrdijo ali ovržejo trenutno hipotezo, je osnovano na izločenih in razvrščenih hipotetičnih ravninah. Po potiskanju lahko ocenimo parametre povzročene gibanja, in ugotovimo, ali se hipotetične točke predmeta gibajo tako, kot bi se gibale točke, ki pripadajo togemu predmetu. S pomočjo delovanja na

detektirano ravnino torej pridobimo dodatne informacije, ki jih lahko uporabimo za grajenje popolnih in preverjenih opisov predmetov.

Oblikovanje hipotez temelji na oceni parametrov ravnin, vrednotenje hipotez pa na oceni parametrov gibanja togega telesa. Pri obeh procesih, tako pri oblikovanju kot pri vrednotenju, je v množici značilnic prisotnih veliko odstopajočih točk (angl. outliers). Da za oba procesa zagotovimo robustnost, smo uporabili algoritem RANSAC [9], ki je znan zaradi svoje sposobnosti robustne ocene parametrov modela.

Oris našega pristopa prikazuje Slika 7. Iz stereo slik začetnega prizora s pomočjo primerjanja izločenih vizualnih značilnic pridobimo 3-D informacijo, ki je podlaga za oblikovanje hipotez o obstoju predmetov. Izbrana hipoteza je temelj za določanje optimalnega premikanja. S potiskanjem pridobimo dodatne informacije, s katerimi hipotezo potrdimo ali ovržemo. Proces vrednotenja hipoteze poteka 1) z oceno gibanja točk, ki so vsebovane v izbrani hipotezi in 2) s primerjanjem značilnic iz slik pred in po potiskanju.



Slika 7: Oris našega pristopa.

Načeloma bi bilo mogoče zvezno slediti značilnicam predmeta, ki ga potiskamo, kar zmanjša razlike med zaporednimi slikami in zmanjša možnost izgubljanja značilnic zaradi prevelikih razlik v glediščih. Vendar pa zvezno sledenje v našem pristopu ne bi bilo uspešno, saj je možnost, da bo robotska roka zakrila predmet med potiskanjem, zelo velika. Zato v našem sistemu primerjamo značilnice pred in po potisku šele po umiku robotske roke iz vidnega polja.

3.1. Oblikovanje hipotez o obstoju predmeta

Predmete in prizorišče je mogoče opisati na več načinov. Študije kažejo, da lahko sistemi, ki temeljijo na lokalnih, nespremenljivih značilnicah, prepoznajo predmete iz različnih zornih kotov in kasneje predmete odkrijejo v kompleksnih prizorih [27],[8],[30]. Z uporabo vizualnih značilnic lahko problem razpoznavanja predmetov prevedemo na problem, ki ga v osnovi sestavljajo tri stopnje: detekcija (angl. detection), opisovanje (angl. description) in ujemanje (angl. matching) vizualnih značilnic.

Na področju postopkov detekcije značilnic je bila razvita cela kopica detektorjev, ki temeljijo na: afini normalizaciji okolice Harrisovih in Hessejevih točk [25], razlikah Gaussovih jeder (angl. Difference of Gaussian, DoG) [17], robovih [37], ekstremih intenzitete [37], maksimalno stabilnih ekstremalnih področjih (Maximally Stable Extremal Regions, MSER) [21] ali na izrazitih področjih (angl. Salient Regions) [14].

Ocena detektorjev značilnic, predstavljena v [26], kaže na to, da je uspešnost izločanja značilnic odvisna tako od vsebine slike kot tudi od zahtevane robustnosti, točnosti in gostote značilnic.

Ocena postopkov za opisovanje značilnic [24] kaže, da pogosto najboljše rezultate doseže deskriptor SIFT (Scale Invariant Feature Transform) [17]. Uspešnost SIFT deskriptorjev potrjuje tudi raziskava delovanja deskriptorjev v primeru 3-D predmetov [27].

V naši metodologiji smo izbrali postopek za odkrivanje značilnic, ki temelji na razlikah Gaussovih jeder, in postopek SIFT za opisovanje izločenih vizualnih značilnic. Več o implementaciji v [40]. Izbira načina odkrivanja in opisovanja značilnic sicer ne vpliva na izvedbo metod, opisanih v nadaljevanju. Izberemo lahko tudi drugačno kombinacijo detektorja in deskriptorja, če tako zahteva uporaba. Naš izbrani algoritem za izločanje in opisovanje značilnic je predstavljen v poglavju 4.1.

Ključni korak pri oblikovanju hipotez je iskanje točk, ki pripadajo planarnim delom prizora. Obstaja veliko algoritmov za odkrivanje ravnin. Ravnine je mogoče iskati s pomočjo neskladja v stereo slikah (angl. disparity maps) [35]. Ta pristop pa ne zagotavlja vrste 3-D informacij, ki jih pri našem pristopu potrebujemo. Po oblikovanju hipotez sledi potiskanje množice točk, ki pripadajo izbrani hipotezi, za kar potrebujemo 3-D informacije, ki jih lahko uporabimo za manipulacijo z robotsko roko. Zato je smiselno, da si tudi pri iskanju ravnin pomagamo s 3-D informacijami, ki jih pridobimo s pomočjo izločenih vizualnih značilnic iz stereo slik prizora.

3.1.1. Pridobivanje 3-D informacij

Na začetku iz slike izločimo množico zanimivih točk. Ta postopek je neodvisen od trenutnega merila. Zanimive točke opišemo z vektorji števil s pripadajočimi 2-D lokacijami. Za dani prizor tako pridobimo množice opisnikov in njihovih 2-D lokacij: $\{ \mathbf{u}_l, \mathbf{d}_l \}$ in $\{ \mathbf{u}_d, \mathbf{d}_d \}$; $l = 1, \dots, N_L$, $d = 1, \dots, N_D$; kjer sta N_L in N_D števili izločenih značilnic iz leve in desne slike, $\mathbf{u} = (x, y)$ 2-D lokacija opisnika in \mathbf{d} opisnik oziroma deskriptor.

Ko smo zanimive točke odkrili in opisali, poskušamo poiskati iste dele prizora v obeh slikah. Ujemanje značilnic poteka na osnovi evklidske razdalje med opisniki. V našem sistemu privzamemo, da se dve značilnici ujemata, če sta njuna opisnika najmanj oddaljena med vsemi izločenimi opisniki, oziroma če je evklidska razdalja med njunima vektorjema najmanjša (najbližji sosed).

Rezultat ujemanja so pari značilnic, ena iz leve in druga iz desne slike prizora. Del ujemanj, ki jih določimo na ta način, ne povezuje dveh značilnic, ki določata isti del prizora. Razlogi za to so dvoumne značilnice ali značilnice, ki niso bile izločene v obeh slikah.

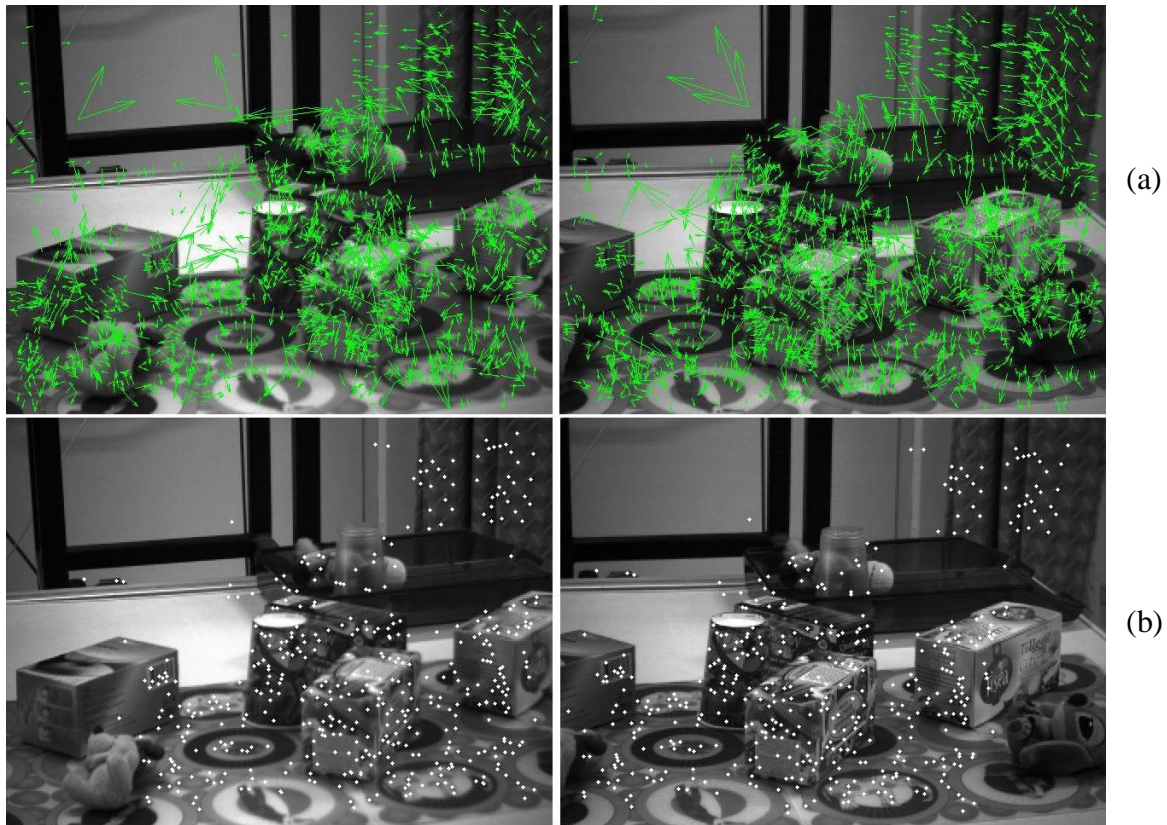
Pravilnost ujemanja lahko izboljšamo z upoštevanjem epipolarne geometrije [6]. S standardno kalibracijo (več o tem v poglavju 5.1) je mogoče pridobiti notranje in zunanje parametre obeh kamer, na podlagi katerih lahko izračunamo osnovno matriko (angl. essential matrix). Iz epipolarne geometrije dobimo zahtevo

$$\tilde{\mathbf{u}}_I^T E_{II} \tilde{\mathbf{u}}_J = 0 \quad (1)$$

kjer $\tilde{\mathbf{u}}_I = [x_I \quad y_I \quad 1]^T$ in $\tilde{\mathbf{u}}_J = [x_J \quad y_J \quad 1]^T$ označujeta lokacijo značilnic v slikah in E osnovno matriko. Osnovna matrika, ob predpostavki da model kamere ustreza modelu kamere na luknjico (angl. pinhole camera), povezuje skladne točke v stereo slikah.

Z znano postavitvijo in orientacijo kamer v prostoru in množicama 2-D lokacij, ki ustrezata množicama ujemajočih značilnic iz leve in desne slike, je mogoče določiti lokacije značilnic v prostoru. Več o tem v poglavju 5.1.

Slika 8 v zgornji vrsti prikazuje značilnice, ki smo jih izločili iz leve in desne slike. V spodnji vrsti so z 2-D lokacijami označena mesta parov značilnic, za katere smo lahko določili 3-D lokacije z ujemanjem opisnikov na podlagi Evklidske razdalje in epipolarne geometrije. 3-D točke smo predstavili s pripadajočimi 2-D lokacijami v levi in desni sliki.



Slika 8: Pridobivanje 3-D informacij. (a) Značilnice SIFT, izločene iz leve in desne slike prizora. (b) 3-D točke, označene na levi in desni sliki s pripadajočimi 2-D lokacijami.

3.1.2. Iskanje planarnih delov v prizoru

Iz množice 3-D točk želimo poiskati točke, ki opisujejo ravne površine v prizoru. Ravne površine opišemo z ravninami. V enačbi ravnine

$$ax + by + cz + d = 0 \quad (2)$$

nastopajo štirje parametri, ki jih določimo s pomočjo treh točk (x_i, y_i, z_i) ; $i = 1, 2, 3$:

$$a = \begin{vmatrix} 1 & y_1 & z_1 \\ 1 & y_2 & z_2 \\ 1 & y_3 & z_3 \end{vmatrix}, b = \begin{vmatrix} x_1 & 1 & z_1 \\ x_2 & 1 & z_2 \\ x_3 & 1 & z_3 \end{vmatrix}, c = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}, d = - \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}. \quad (3)$$

Predpostavka je, da množica 3-D točk, ki opisuje prizor, vsebuje več ravnin. Ker lahko predvidevamo, da bo v množici 3-D točk veliko odstopajočih točk (outliers), za robustnost procesa iskanja parametrov ravnin, ki opisujejo ravne površine v prizoru, uporabimo algoritem RANSAC (RANDOM Sample Consensus), ki je opisan spodaj in v poglavju 4.2.

Algoritem RANSAC za iskanje parametrov ravnin

Vhod v računski proces je množica 3-D točk. Ker želimo oblikovati več hipotez o obstoju predmetov, algoritem uporabimo večkrat. Vhodna množica v prvem koraku je množica vseh 3-D točk, ki smo jih določili na način, opisan v prejšnjem poglavju. V naslednjih korakih se lahko vhodna množica 3-D točk spreminja. Vsakič, ko določimo parametre ravnine, ki ustreza našim pogojem za ravnino, točke na ravnini izločimo iz vhodne množice točk. Ravnine iščemo, dokler ne dosežemo zeleno število ravnin k , ali pa ko vhodna množica postane premajhna oziroma ne vsebuje dovolj točk, ki bi zadostile našim pogojem za ravnino. V vsakem koraku N_p -krat naključno izberemo množico treh točk iz vhodne množice, določimo parametre ravnine in število skladnih točk F . Naš pogoj za obstoj ravnine je, da vsebuje večje ali enako število skladnih točk, kot ga določa prag $F_p > 3$. Skladnost točk s predlaganim modelom določa toleranca (t_p), ki predstavlja največjo dovoljeno oddaljenost točke od ravnine.

Proces oblikovanja hipotez o obstoju predmetov prikazuje Algoritem 1.

Za robotsko roko je potrebno določiti posebne gibe, ki bodo z visoko verjetnostjo povzročili premik množice točk, ki pripadajo hipotezi. Brez predpostavk o okolju pa ni mogoče vedeti, ali pridobljene ravnine združujejo točke, ki pripadajo enemu ali več predmetom, in ali vključujejo tudi točke, ki pripadajo ozadju. V primeru podobnega

prizora kot na prejšnji sliki (Slika 8) lahko ravnino sestavljajo točke iz ploskev ene ali več škatel, priključijo pa se lahko še točke iz ozadja. Če na primer poskusimo potisniti ravnino, ki pripada zgornjim ploskvam dveh škatel, je možnost uspešnega premika precej manjša. Robotska roka se lahko na primer premakne med škatlami, ki zato obstaneta na istem mestu.

Za boljšo verjetnost uspešnega potiskanja je potrebno preveriti razmerja med točkami ravnin. Če ravnina združuje točke, ki jih je mogoče razdeliti v gruče, je precej verjetno, da gruče opisujejo različne predmete ali ozadje. V našem sistemu ne moremo vnaprej določiti števila gruč, saj ne vemo, katerim delom prizora pripadajo točke ravnin, zato potrebujemo algoritem, ki za delovanje ne potrebuje števila gruč. Zato smo uporabili algoritem za gručenje *X-means*, ki zna sam oceniti optimalno število gruč. Algoritem je opisan v poglavju 4.3.

Hipoteze o obstoju predmeta so osnovane na množici točk, ki jih pridobimo z gručenjem točk, ki pripadajo detektiranim ravninam. Hipoteze razvrstimo glede na število točk, ki jim pripadajo. Manipulacija hipotez z večjim številom točk bo bolj verjetno uspešna, zato najprej poskusimo premakniti takšne množice. Vsako potiskanje ne bo uspešno. Potiskanje predmetov, ki so na primer pritrjeni, ne bo uspešno. V takšnih primerih lahko izberemo naslednjo hipotezo, ki vsebuje manj točk, določimo optimalen način potiskanja glede na točke, in izvedemo potiskanje.

Algoritem 1: Oblikovanje hipotez.

Vhod: množica 3-D točk

dokler več kot F_p točk in število korakov k ni doseženo

ponovi N_p krat

naključno izberi m_p točk

oceni parametre ravnine

določi točke, ki se prilegajo modelu s toleranco t_p

izberi parametre ravnine, ki ji pripada največje število točk F

če je F večje od F_p

shrani parametre ravnine in izloči točke F iz množice 3-D točk

uredi točke ravnine v gruče z algoritmom X-means

Izhod: parametri ravnin in pripadajoče točke in gruče

Izbrane eksperimentalne vrednosti, ki smo jih uporabili pri eksperimentih (glej poglavje 5.2.1):

- toleranca $t_p = 0.005\text{m}$
- število testnih podmnožic $N_p = 1000$
- prag $F_p = 20$
- število korakov $k = 6$

3.2. Potiskanje na osnovi izbrane hipoteze

V prejšnjem poglavju smo opisali proces oblikovanja hipotez, ki ga sestavljajo 1) iskanje točk, ki pripadajo ravninam, 2) gručenje točk v ravninah in 3) razvrščanje hipotez o obstoju predmeta glede na število točk, ki jih vsebujejo.

Naš cilj je potisniti izbrano hipotezo na način, ki bo s čim večjo verjetnostjo premaknil točke, ki pripadajo izbrani ravnini. Gibanje robotske roke določimo glede na parametre

ravnine, ob tem pa upoštevamo, da je gibanje odvisno tudi od podlage. Pri določanju smeri potiskanja nas vodi zahteva, da mora gibanje robotske roke potekati na ploskvi, ki je vzporedna talni površini, v našem primeru je to miza, sicer bo robot hitro trčil v mizo ali pa izgubil stik z objektom.

Določanje gibanja konca robotske roke prikazuje Algoritem 2.

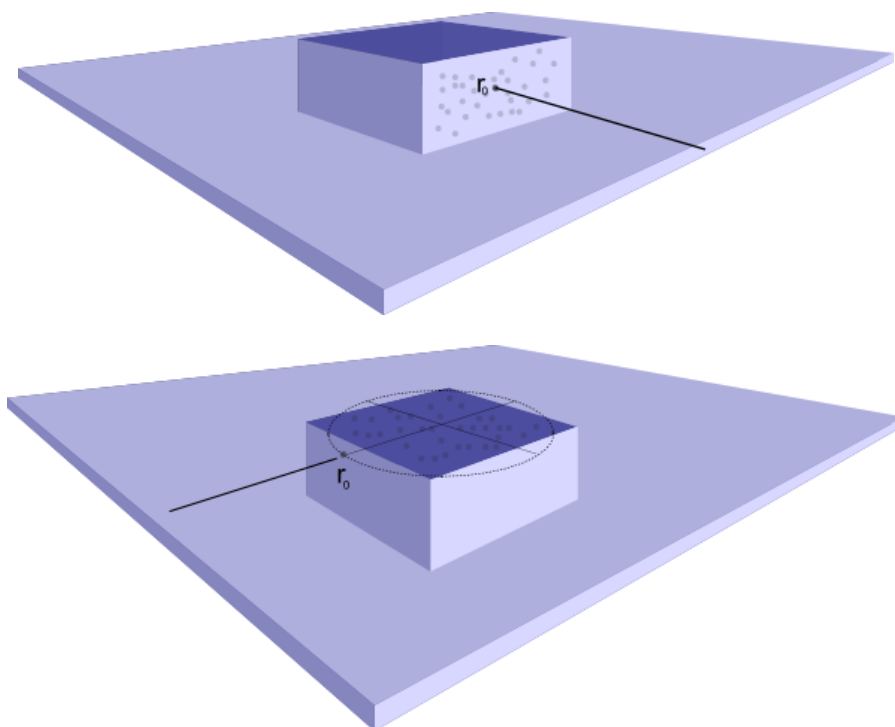
Sistem je umerjen tako, da os x robotskega sistema kaže v nasprotni smeri gravitacije. Osnova za določanje robotskih trajektorij za potiskanje je kot ϕ med talno površino in izbrano hipotetično ravnino. Trajektorije določimo s točko \mathbf{r}_0 in vektorjem \mathbf{n} , ki določata začetno točko in smer potiskanja.

Pri tem ločimo dva primera (oba primera po vrsti prikazuje Slika 9):

- $\phi \geq 45^\circ$. V tem primeru obravnavamo ravnino kot pravokotno na talno ravnino. Začetno točko \mathbf{r}_0 določimo kot povprečje vrednosti vseh točk hipotetične ravnine. Pri določanju smeri potiskanja rotiramo hipotetično ravnino okrog osi preseka s talno ravnino. Smer potiska \mathbf{n} je podana z negativno smerjo normale zarotirane ravnine. Na ta način zagotovimo, da je smer potiskanja vedno vzporedna s talno ravnino.
- $\phi > 45^\circ$. Značilnice projiciramo na ravnino, vzporedno talni ravnini. Višino ravnine določimo z enačbo (4). S pomočjo statistike drugega reda nato izračunamo elipso, ki aproksimira projicirane točke, in izračunamo točko na koncu krajše glavne osi, ki je bližje koordinatnemu sistemu obeh kamer. To je točka \mathbf{r}_0 . Potiskanje poteka vzporedno mali osi z začetkom v \mathbf{r}_0 . Smer potiskanja je podana z vektorjem, ki kaže od začetka do središča elipse \mathbf{r}_c .

$$x' = \max(x_r - \Delta, \min(x_t + \Delta, \frac{x_r + x_t}{2})), \quad (4)$$

kjer je x_r povprečna višina vseh točk ravnine in x_t višina talne ravnine. Pri eksperimentih smo uporabili vrednost $\Delta = 0.02\text{m}$.



Slika 9: Določanje začetne točke in smeri potiskanja.

Algoritem 2: Potiskanje ravnine.

Vhod: hipotetična ravnina $p = (a, b, c, d)$ in pripadajoče točke
določi talno ravnino

izračunaj kot ϕ med hipotetično in talno ravnino

če ($\phi \sigma 45^\circ$)

\mathbf{r}_0 - srednja vrednost točk ravnine

rotiraj hipotetično ravnino p do navpične lege p'

$\mathbf{n} = (a', b', c')$

drugače

projiciraj točke na horizontalno ravnino višine x'

izračunaj elipso, ki aproksimira vse točke

določi \mathbf{r}_0

$\mathbf{n} = \mathbf{r}_0 - \mathbf{r}_c$

Izhod: začetna točka \mathbf{r}_0 in smer potiskanja \mathbf{n}

* potiskanje: začetek pri $\mathbf{r}_0 + \Delta\mathbf{r}$ v smeri \mathbf{n} , potiskanje $2\Delta\mathbf{r}$ v smeri $-\mathbf{n}$

3.3. Vrednotenje hipotez

Ko je potiskanje izbrane množice značilnic izvedeno, lahko z dodatnimi informacijami, ki jih pridobimo s sproženim gibanjem predmeta, hipoteze potrdimo ali ovržemo. V ta namen moramo sprejeti predpostavke o tem, kako naj bi se predmeti v robotovem okolju gibali.

V tem delu predpostavimo, da so iskani predmeti toga telesa. Potisnjene značilnice se morajo torej premakniti tako, kot se premikajo točke, ki pripadajo togemu telesu. To lahko preizkusimo s pomočjo analize stereo slike, ki jih zajamemo pred in po potisku. Značilnice, ki se nahajajo v bližini hipotetičnih točk predmeta, in se gibljejo kot hipotetične točke, lahko dodamo k modelu.

Vrednotenje hipotez poteka na podlagi ujemanj med značilnicami v stereo slikah pred in po potisku. Prvi par slik s pripadajočimi 3-D lokacijami in opisniki iz leve in desne slike $\{\mathbf{v}^1, \mathbf{d}_l^1, \mathbf{d}_d^1\}$ je osnova za oblikovanje hipotez o predmetih, izbrana hipoteza $\{\mathbf{v}^{h'}, \mathbf{d}_l^{h'}, \mathbf{d}_d^{h'}\}$ iz nabora oblikovanih hipotez pa temelj za določanje načina potiskanja. S pomočjo 3-D lokacij in opisnikov iz leve in desne slike $\{\mathbf{v}^2, \mathbf{d}_l^2, \mathbf{d}_d^2\}$ po premiku pa lahko preverimo, ali so se točke premaknile v skladu z načelom gibanja togega telesa.

Začetni nabor ujemanj pridobimo s primerjanjem opisnikov značilnic iz 1) obeh levih in 2) obeh desnih slik. Vsako izmed ujemanj nam da množico parov ujemajočih se značilnic. To je ujemanje 2-D vizualnih značilnic, zato ni nujno, da se oba opisnika iz slike pred potiskom ujemata z obema opisnikoma iz slike po potiskanju.

Vzemimo eno izmed 3-D točk iz izbrane hipoteze, s pripadajočima opisnikoma iz leve in desne slike $\mathbf{d}_l^1, \mathbf{d}_d^1$, in točko na drugi sliki, ki prikazuje premik točke v skladu z načelom gibanja togega telesa $\mathbf{d}_l^2, \mathbf{d}_d^2$. Če se, na primer, ujemata opisnika iz levih slik $\mathbf{d}_l^1, \mathbf{d}_l^2$, ni nujno, da se ujemata tudi opisnika iz desnih slik $\mathbf{d}_d^1, \mathbf{d}_d^2$. Naš pristop je, da v množico parov ujemajočih 3-D točk vključimo vse 3-D točke, katerih ujemanje kažejo pari značilnic iz levih in desnih slik.

Na primer: 3-D točko iz prve slike, ki jo določa lokacija \mathbf{v}^1 ter levi in desni opisnik $\mathbf{d}_l^1, \mathbf{d}_d^1$, določimo ujemanje za levi opisnik \mathbf{d}_l^2 in ujemanje za desni opisnik \mathbf{d}_d^2 . Par levemu opisniku je opisnik na drugi sliki \mathbf{d}_l^2 , ki določa 3-D lokacijo \mathbf{v}_1^2 , par desnemu pa opisnik \mathbf{d}_d^2 , ki določa lokacijo \mathbf{v}_2^2 . V množico parov 3-D točk vključimo oba para 3-D točk, $(\mathbf{v}^1, \mathbf{v}_1^2)$ in $(\mathbf{v}^1, \mathbf{v}_2^2)$. Ta pristop smemo uporabiti, saj je za vse značilnice mogoče določiti 3-D lokacije.

Na podlagi parov iz množic $\{\mathbf{v}_m^1\}$ in $\{\mathbf{v}_m^2\}$ poiščemo parametre transformacije togega telesa. Transformacijo togega telesa opisuje enačba

$$\mathbf{v}^2 = \mathbf{R}\mathbf{v}^1 + \mathbf{t} \quad (5)$$

kjer je \mathbf{R} rotacijska matrika in \mathbf{t} translacijski vektor.

Parametra \mathbf{R} in \mathbf{t} lahko ocenimo s prilagajanjem na osnovi najmanjših kvadratov z razcepom s singularnimi vrednostmi [1]. Za robustnost procesa iskanja parametrov gibanja togega telesa uporabimo algoritem RANSAC.

Vzemimo dve množici 3-D točk $\{\mathbf{v}_i\}$ in $\{\mathbf{v}'_i\}$; $i=1,2,\dots,N$ množici 3-D točk, N je število parov 3-D točk, ki ju povezuje enačba

$$\mathbf{v}'_i = \mathbf{R}\mathbf{v}_i + \mathbf{t} + \mathbf{e}_i \quad (6)$$

kjer je \mathbf{R} rotacijska matrika in \mathbf{t} translacijska vektor in \mathbf{e}_i vektor šuma. Prisotnost zelo velikega šuma obravnavamo v nadaljevanju besedila.

Poiskali bi radi \mathbf{R} in \mathbf{t} , ki minimizirata naslednji kriterij:

$$\sum_{i=1}^N \|\mathbf{v}'_i - (\mathbf{R}\mathbf{v}_i + \mathbf{t})\|^2. \quad (7)$$

Tu gre za problem najmanjših kvadratov, ki ga je glede na ugotovitve iz [1] mogoče razdeliti na dva dela: 1) iskanje \mathbf{R} in 2) iskanje \mathbf{t} .

1. Najprej izračunamo srednjo vrednost vseh točk in jo odštejemo od začetnih točk \mathbf{v}_i .

$$\mathbf{v} = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_i \quad (8)$$

$$\mathbf{q}_i = \mathbf{v}_i - \mathbf{v} \quad (9)$$

podobno izračunamo še \mathbf{v}' in $\mathbf{q}'_i = \mathbf{v}'_i - \mathbf{v}'$ ter določimo matriko \mathbf{H}

$$\mathbf{H} = \sum_{i=1}^N \mathbf{q}_i \mathbf{q}'_i{}^T \quad (10)$$

in poiščemo njen razcep s singularnimi vrednostmi (poglavje 4.4):

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \quad (11)$$

Nato določimo matriko \mathbf{X} :

$$\mathbf{X} = \mathbf{V} \mathbf{U}^T. \quad (12)$$

Če je determinanta $\det(\mathbf{X}) = 1$ določimo rotacijsko matriko $\mathbf{R} = \mathbf{X}$.

V primeru, da je $\det(\mathbf{X}) = -1$, matrika \mathbf{X} označuje zrcaljenje (angl. reflection):

- a) ena izmed singularnih vrednosti enaka nič, na primer $\lambda_k = 0$. Enačbo (12) preoblikujemo v $\mathbf{X}' = \mathbf{V}' \mathbf{U}'^T$, kjer je \mathbf{V}' matrika \mathbf{V} z zamenjanimi predznaki v k -tem stolpcu.
- b) nobena izmed singularnih vrednosti ni enaka nič. To je mogoče le v prisotnosti zelo velikega šuma. V takšnih primerih avtorji [1] predlagajo pristop na osnovi algoritma RANSAC.

2. Translacijski vektor določa enačba

$$\mathbf{t} = \mathbf{v}' - \mathbf{R} \mathbf{v}. \quad (13)$$

Algoritem RANSAC za iskanje rotacijske matrike in translacijskega vektorja

Iz množice vseh mogočih parov ujemajočih se 3-D točk iz $\{\mathbf{v}_m^1\}$ in $\{\mathbf{v}_m^2\}$, naključno izberemo m_r parov ($m_r = 3$) ujemajočih se točk. Za vsako testno množico parov z uporabo zgoraj opisanega algoritma ocenimo vrednosti za rotacijsko in translacijsko matriko, in določimo število točk, ki se modelu prilegajo s toleranco t_r . Za oceno kvalitete prilagajanja uporabimo razdaljo med premaknjenimi in detektiranimi točkami.

$$\|R\mathbf{v}_m^1 + \mathbf{t} - \mathbf{v}_m^2\| \quad (14)$$

Algoritem 3 prikazuje procese določanja začetne množice ujemajočih se parov točk iz slik pred in po potiskanju, ki so izhodišče za oceno parametrov gibanja togega telesa s pomočjo algoritma RANSAC.

Dodatne informacije, ki jih potrebujemo za potrjevanje hipotez o obstoju predmetov, lahko pridobimo le, če je do gibanja zares prišlo. Zato je pomembno ugotoviti, ali so se točke zares premaknile. Gibanje togega telesa je sestavljeno iz rotacijskega in translacijskega dela. Če ocenjena translacija in rotacija kažeta na to, da do gibanja ni prišlo, moramo izbrati novo hipotezo iz množice hipotez, saj v takšnem primeru nimamo dovolj informacij za nadaljnje delo. V ta namen določimo normi za rotacijsko in translacijsko matriko.

Normo rotacije določimo z izračunom osi vrtenja in kota vrtenja φ , ki ju lahko izračunamo iz rotacijske matrike. Za normo vzamemo manjšo od vrednosti $|\varphi|$ in $2\pi - |\varphi|$. Da se dokazati, da na ta način izračunana vrednost določa razdaljo v prostoru vseh rotacijskih matrik [28]. Do premika je prišlo, če je vsaj ena od obeh norm večja od mejne vrednosti za kot rotacije φ_{\min} oziroma translacije t_{\min} .

S pomočjo algoritma RANSAC določimo \mathbf{R} in \mathbf{t} in vsa ujemanja, ki ustrezajo izračunanemu togemu gibanju. Če je število točk ujemanj manjše od vnaprej določenega najmanjšega števila točk, potem nimamo na voljo dovolj informacij za gradnjo objektnih modelov. Vendar pa v tem primeru ne moremo izbrati nove hipoteze, izvesti potiskanja in ovrednotiti hipoteze v upanju, da bo število značilnic naslednjič večje. Če se je postopek končal z določenim številom točk, potem je prišlo do bodisi minimalne rotacije bodisi do minimalne translacije, torej parametri gibanja togega telesa kažejo na to, da je do gibanja prišlo. V tem primeru moramo zaradi spremembe prizora oblikovati nov nabor hipotez na osnovi novo zajetih stereo slik prizora.

Algoritem 3: Vrednotenje hipoteze.

Vhod: $\{ \mathbf{v}^1, \mathbf{d}_l^1, \mathbf{d}_d^1 \}, \{ \mathbf{v}^{hl}, \mathbf{d}_l^{hl}, \mathbf{d}_d^{hl} \}; i = 1, \dots, \text{število hipotez}$

dokler obstajajo hipoteze in ni določenih točk predmeta

izberi hipotezo in potisni ravnino (Algoritem 2)

pridobi $\{ \mathbf{v}^2, \mathbf{d}_l^2, \mathbf{d}_d^2 \}$ iz slik po potisku

določi množici točk z ujemajočimi se opisniki $\{ \mathbf{v}_m^1 \}, \{ \mathbf{v}_m^2 \}$

ponovi N_r krat

naključno izberi m_r pare iz $\{ \mathbf{v}_m^1 \}, \{ \mathbf{v}_m^2 \}$

oceni parametre R in t transformacije togega telesa

določi pare točk iz $\{ \mathbf{v}^1 \}, \{ \mathbf{v}^2 \}$, ki se prilegajo modelu s toleranco t_r in

njihovo število F

izberi parametre transformacije, ki ji pripada največje število parov F

če ($\|\varphi\| \leq \varphi_{\min}$ in $\|t\| \leq t_{\min}$)

izberi naslednjo hipotezo

drugače

če je F večje od F_R

Izhod: točke, ki ustrezajo gibanju togega telesa

drugače

Izhod: oblikuj nove hipoteze (Algoritem 1)

Izbrane eksperimentalne vrednosti, ki smo jih uporabili pri eksperimentih (glej poglavje 5.2.3):

- toleranca $t_r = 0.005m$,
- število podmnožic $N_r = 100$,
- prag $F_r = 10$,
- minimalna rotacija in translacija $\varphi_{\min} = 5^\circ$, $t_{\min} = 0.05m$.

3.4. Grajenje modelov

Če želimo zgraditi izčrpen model predmeta, moramo robotu omogočiti različne poglede na predmet. Točke, za katere smo ugotovili, da pripadajo predmetu, je potrebno kopičiti preko različnih pogledov in jih sestaviti v model predmeta.

Pri gradnji modelov predmetov moramo določiti način potiskanja, ki bo imelo za posledico rotacijo predmetov. Na ta način bomo odkrili nove, prej nevidne značilnice in jih vključili v model predmeta. Naš algoritem, ki služi pridobivanju različnih pogledov na predmet, je v osnovi spremenjen algoritem za potiskanje, opisan v poglavju 3.2. Začetno točko r_0 in smer potiskanja n določimo na način, ki ga opisuje Algoritem 2. Nato vektor n zavrtimo okrog začetne točke v ravnini, vzporedni s smerjo gravitacije, za prej določen kot α . (Pri eksperimentalnem delu smo uporabili $\alpha = 20^\circ$). Tako hipotetično ravnino potisnemo pod kotom α , in ne pravokotno, kot določa osnovni algoritem za potiskanje. Druga možnost spremembe algoritma za potiskanje je premik začetne točke r_0 .

Uspešnost kopičenja točk predmeta preko različnih pogledov je odvisna od rotacije predmeta. Rotacija predmeta mora biti po eni strani dovolj velika, da odkrije nove, prej nevidne značilnice in po drugi strani dovolj majhna, da je izračun ujemanj med značilnicami v zaporednih slikah sploh mogoč. Glede na rezultate eksperimentov delovanja različnih algoritmov za izločanje in opisovanje značilnic [27], se delovanje poljubne kombinacije detektorja in deskriptorja zelo poslabšuje s spremembami v zornem kotu, pri spremembah večjih od 30° pa nobena izmed kombinacij detektorjev in deskriptorjev ne deluje več dobro.

Po vsakem uspešnem potisku, ki predmet premakne v skladu s prej omenjenimi omejitvami v rotaciji, pridobimo nove točke predmeta in pripadajoče parametre transformacije togega telesa. Model predmeta lahko nadgradimo z novo določenimi točkami. Novo določene točke dodamo v model, lokacije starih, prej določenih točk, pa preslikamo na nove lokacije s pomočjo parametrov gibanja togega telesa.

4. Metodologija

Naš pristop temelji na izločanju, opisovanju in ujemanju vizualnih značilnic. Izbrali smo postopek za odkrivanje značilnic, ki temelji na razlikah Gaussovih jeder, in postopek SIFT za opisovanje izločenih vizualnih značilnic.

Oblikovanje hipotez temelji na oceni parametrov hipotetičnih ravnin, vrednotenje hipotez pa na oceni parametrov gibanja togega telesa. Pri obeh procesih, tako oblikovanju kot vrednotenju, je v množici značilnic prisotnih veliko odstopajočih točk (outliers). Da za oba procesa zagotovimo robustnost, smo uporabili algoritem RANSAC, ki je znan zaradi svoje sposobnosti, da robustno oceni parametre modela.

S pomočjo algoritma RANSAC pridobimo ravnine, ki jih uporabimo kot namige za obstoj predmetov. Še vedno pa lahko točke, ki so vsebovane v eni ravni, pripadajo enemu ali več predmetom in tudi ozadju. Ker je pravilnost izbrane hipoteze zelo pomembna za izračun gibanja robotske roke, ki bo s čim večjo verjetnostjo premaknilo hipotetično ravnino, točke najprej razdelimo v gruče točk, ki med seboj niso preveč oddaljene. V ta name uporabimo algoritem za gručenje *X-means*, ki sam oceni število gruč, ki vnaprej ni znano.

4.1. Iskanje in opis vizualnih značilnic

SIFT (Scale-invariant feature transform) značilnice je leta 1999 razvil David Lowe [17]. Za naše delo je pomembno, da so te značilnice neodvisne od ravninske rotacije in merila ter deloma neodvisne od spremembe osvetlitve in pogleda.

Pot do značilnic SIFT pelje skozi štiri faze:

- odkrivanje ekstremov v prostoru meril (scale-space extrema detection),
- določanje lokacij (keypoint localization),
- določanje orientacije (orientation assignment) in
- izgradnja opisnikov pomembnih točk (keypoint descriptor).

4.1.1. Odkrivanje ekstremov

Področja, ki so neodvisna od spremembe merila, poiščemo z iskanjem točk, ki so stabilne preko vseh meril slike. Osnova za to je konvolucija več Gaussovih funkcij (16) z vhodno sliko:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (15)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (16)$$

Določanje stabilnih točk pri različnih merilih poteka s konvolucijo razlike Gaussovih funkcij (difference of Gaussian) in vhodne slike:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (17)$$

Razlogi za izbiro te funkcije je več: filtrirane slike L potrebujemo za opis značilnic, D pa lahko izračunamo s preprosto operacijo odštevanja. Po vsaki takšni konvoluciji je izračunana slika podvzorčena in proces ponovljen.

Ekstreme $D(x, y, \sigma)$ poiščemo s primerjanjem vsake vzorčne točke (sample point) z osmimi sosedi pri istem merilu in pripadajočimi devetimi sosedi pri sosednjih merilih, višjemu in nižjemu. Če je vrednost točke največja ali najmanjša med vsemi sosedi, postane kandidat za pomembno točko (keypoint).

4.1.2. Določanje lokacije

Rezultat prejšnjega koraka je v večini primerov preveč kandidatov za pomembne točke. V tem koraku poleg določanja lokacije pomembnih točk tudi izločimo točke s premajhnim kontrastom in točke, ki ležijo na robovih.

Lokacijo pomembne točke določimo z interpolacijo podatkov [4]. Funkcijo $D(x, y, \sigma)$ razvijemo v Taylorjevo vrsto do kvadratnega člana in jo premaknemo v vzorčno točko. Velja, da D in njene odvode ocenimo pri vzorčni točki, $\mathbf{x} = (x, y, \sigma)^T$ pa označuje odmik od te točke.

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} \quad (18)$$

Lokacijo ekstrema $\hat{\mathbf{x}}$ določimo z naslednjo enačbo:

$$\hat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D}{\partial \mathbf{x}} \quad (19)$$

Če je vrednost $\hat{\mathbf{x}}$ pri kateri koli dimenziji večja od 0.5, ekstrem leži bližje drugi točki v sliki, zato interpolacijo ponovimo pri tej točki. V nasprotnem primeru pri določanju lokacije upoštevamo odmik in tako dobimo interpoliran približek lokacije pomembne točke.

Če navzdol omejimo vrednost $D(\hat{\mathbf{x}})$ lahko izločimo točke s premajhnim kontrastom. Lowe pri vrednostih točkovnih elementov med 0 in 1, izloči točke, kjer je $|D(\hat{\mathbf{x}})|$ manjše od 0.03.

Točke, ki so posledica odziva na robove, izločimo s pomočjo glavnih ukrivljenosti. Posledica roba bo namreč velik odziv v smeri roba in majhen odziv v pravokotni smeri. Glavne ukrivljenosti izračunamo iz Hessejeve matrike

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (20)$$

pri lokaciji in merilu pomembne točke. Lastne vrednosti te matrike so proporcionalne glavnim ukrivljenostim D .

4.1.3. Določanje orientacije

Prvi korak pri določanju orientacije pomembni točki je izbira slike, katere merilo je najbližje merilu pomembne točke. Nato pri vsaki točki slike $L(x, y)$ izračunamo velikost (21) in orientacijo gradienta (22) z razlikami med točkovnimi elementi po naslednjih enačbah:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (21)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (22)$$

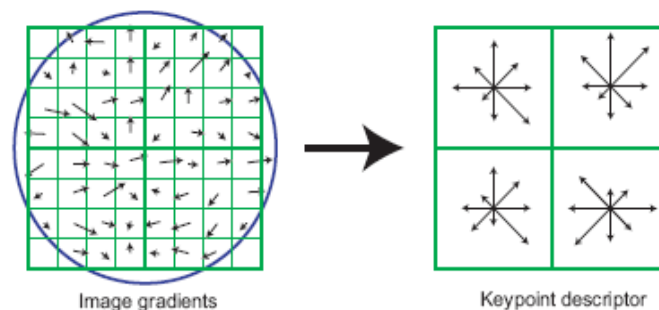
Orientacijski histogram pomembne točke sestavlja 36 razdelkov, ki pokrivajo vseh 360 stopinj. Prispevki v histogram so gradienti v okolici pomembne točke. Vsak prispevek je utežen z velikostjo in z Gaussovim krožnim oknom. Vrhovi orientacijskega histograma ustrezajo glavnim smerem lokalnih gradientov. Upoštevamo najvišji vrh in vrhove, ki niso manjši od 80% velikosti najvišjega. Ko pomembni točki določimo orientacijo, je mogoče opisnik pomembne točke rotirati glede na orientacijo in tako doseči neodvisnost od rotacije slike.

4.1.4. Izgradnja opisnikov

V prejšnjih korakih smo ključnim točkam določili lokacijo, merilo in orientacijo. Slika 11 prikazuje postopek izgradnje opisnika na podlagi 8x8 okolice ključne točke.

Na levi strani slike so prikazane velikosti in orientacije gradientov. Krog opisuje utežitev z Gaussovim oknom. Dolžina puščice predstavlja seštevek velikosti gradientov v bližini

določene smeri. Orientacije gradientov razvrstimo v orientacijske histograme, ki imajo 8 razdelkov, kar je prikazano na desni strani slike (Slika 11).



Slika 11: Postopek izgradnje SIFT opisnikov. Opisnik sestavljajo vsi orientacijski histogrami (2 x 2), vsak ima po 8 razdelkov. Povzeto po [17].

V našem primeru izgradnja opisnika poteka v 16 x 16 okolici ključne točke. Pri izgradnji opisnika upoštevamo vrednosti vseh razdelkov vseh orientacijskih histogramov. Na podlagi 16 x 16 okolice pridobimo šestnajst (4 x 4) orientacijskih histogramov, vsak ima 8 možnih orientacij. Opisnik torej sestavlja vektor s 128 vrednostmi. Sledi še normalizacija vektorja, nastavitev pragov in ponovna normalizacija vektorja.

4.2. Algoritem RANSAC

V mnogih znanstvenih problemih moramo zaznane podatke razložiti v okviru predhodno določenih modelov. Takšno razlago lahko razdelimo na dva podproblema. Prvi je iskanje najboljšega ujemanja med podatki in enim od modelov (problem razvrščanja), drugi pa iskanje najboljših vrednosti parametrov (v smislu odstopanja) za izbrani model (problem ocene parametrov). V praksi sta oba problema povezana - rešitev drugega po navadi zahteva rešitev prvega problema.

Pri oceni parametrov je mogoče uporabiti veliko metod: na primer metodo najmanjših kvadratov (Least Mean Square, LMS) ali oceno največje zanesljivosti (Maximum Likelihood Estimation, MLE), itd. Takšne metode pri oceni parametrov uporabijo vse podatke in nimajo notranjih mehanizmov za ugotavljanje večjih napak. Temeljijo namreč na predpostavki, da bo, ne glede na velikost množice podatkov, vedno dovolj 'dobrih' vrednosti, ki bodo zgladile večja odstopanja.

V mnogih praktičnih primerih ta predpostavka ne drži. Pri reševanju tega problema si lahko pomagamo z več hevrističnimi pristopi. Najpogosteje uporabljen pristop je variacija naslednjega postopka. Parametre modela izračunamo na podlagi vseh podatkov. Nato poiščemo podatek, ki se najmanj ujema s predlaganim modelom in ga odstranimo. Postopek iterativno ponavljamo, dokler ni skupno odstopanje manjše od predhodno določenega nivoja ali pa izločimo vse podatke. Na žalost pa je lahko predlagana hevristika že zaradi ene same 'okužene točke' neuspešna, zato ni primerna za uporabo na nepreverjenih podatkih.

Algoritem RANSAC (RANdom Sample And Concensus), ki sta ga leta 1981 v svojem delu predstavila Fischler and Bolles [9], je uspešen kot metoda za oceno parametrov modela v prisotnosti velikega števila odstopajočih točk (angl. outliers). Zaradi tega je zelo uveljavljen pri analizi prizorišč, kjer je pogosto veliko napak zaradi načina izločanja vizualnih značilnic, ki so vir podatkov. RANSAC procedura deluje v nasprotni smeri kot običajne tehnike glajenja. Običajno z glajenjem izločamo točke iz prvotne rešitve, ki temelji na čim več točkah. RANSAC pa začne z majhno prvotno množico, ki ji dodaja točke.

Bolj formalno lahko RANSAC paradigmo opišemo na naslednji način. Če vzamemo model, za katerega potrebujemo minimalno število točk m , da lahko določimo proste parametre, in množico točk P (velja, da je P večje od m). Iz P naključno izberemo podmnožico S_I , ki vsebuje m točk, in določimo parametre modela M_I . S pomočjo

modela določimo razširjeno množico S_1' , ki vsebuje točke, ki se prilegajo modelu s predhodno določeno toleranco.

Če je število točk v S_1' večje od predhodno določenega praga, ki je funkcija ocena števila odstopanj v množici P , izračunamo nov model M_1' na podlagi točk iz S_1' . Če je število točk manjše od praga, naključno izberemo novo podmnožico S_2 , in ponovimo postopek. Če po predhodno določenem številu poskusov i ne najdemo množice S_i' , ki bi vsebovala število točk večje od predhodno določenega praga, bodisi zaključimo z največjim S_i' ali pa zaključimo brez rešitve.

RANSAC ima torej tri parametre:

- toleranca, ki določa skladnost točke s predlaganim modelom,
- število testnih podmnožic in
- prag, ki določa pravilnost modela.

4.3. Gručenje točk

V našem sistemu ne moremo vnaprej vedeti, ali 3-D točke, ki smo jih razdelili v ravnine z algoritmom RANSAC, pripadajo enemu ali več predmetom in če mogoče vključujejo točke iz ozadja. Potiskanje bo prineslo dodatne informacije, na podlagi katerih bomo hipotezo o obstoju predmeta bodisi sprejeli bodisi zavrgli. V primerih potiskanja ravnin, ki vsebujejo točke, ki pripadajo različnim predmetom in so med seboj zelo oddaljene, je možnost uspešnega potiskanja precej manjša.

4.3.1. Algoritem K -means

Algoritem K -means [20] je najbolj poznan in pogosto uporabljan algoritem za gručenje, predvsem zaradi svoje preprostosti in konvergenčnih lastnosti. K -means ima nekaj slabosti, na primer počasnost. Poleg tega za svoje delovanje potrebuje število gruč K .

Zanimivo je, da lahko algoritem pri fiksnem številu gruč poišče slabše rešitve kot če se število gruč lahko spreminja [29].

Procedura se začne z določitvijo K gruč, vsako na začetku sestavlja ena naključno izbrana podatkovna točka, in nadaljuje z dodajanjem točk k gruči, katere srednja vrednost je najbližja novi točki. Ko je točka dodana gruči, je srednja vrednost gruče ponovno izračunana tako, da upošteva tudi novo točko. Algoritem se konča, ko se srednje vrednosti gruč ne spreminjajo več.

4.3.2. Algoritem X-means

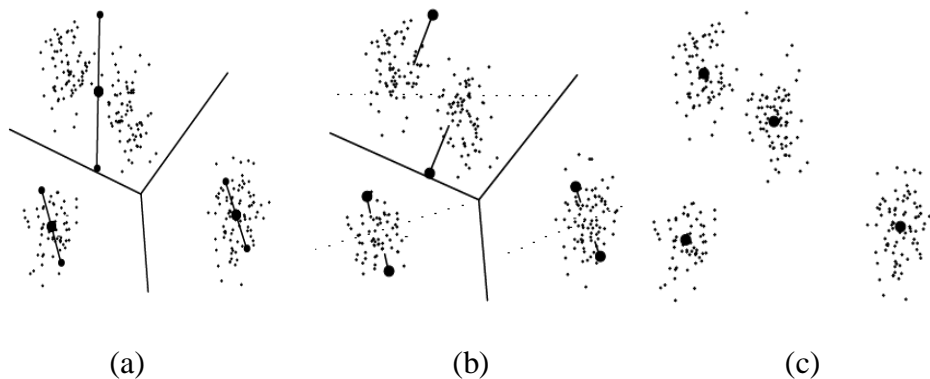
V našem pristopu potrebujemo gručenje točk ravnin, ne moremo pa podati števila gruč, saj nimamo predpostavk o predmetih v prizorih, zato algoritma K -means ne moremo uporabiti. X -means [29] je izboljšava algoritma K -means. Za nas najpomembnejša izboljšava je, da za svoje delovanje ne potrebuje števila gruč, pač pa območje vrednosti za število gruč. X -means potem sam določi najboljšo vrednost. Pri danem območju algoritem začne s K , ki je enaki spodnji meji območja. Pri svojem delovanju povečuje število centroidov, če je potrebno, dokler ne doseže zgornje meje območja. Med tem procesom algoritem beleži ocene skupin centroidov, najboljše ocenjena skupina je rezultat algoritma.

Algoritem X -means je sestavljen iz dveh operacij:

- *Izboljšava parametrov.* Operacijo sestavlja delovanje običajnega algoritma K -means.
- *Izboljšava strukture.* S to operacijo poiščemo dodatne centroide. To je mogoče z razcepom centroidov. Razcep temelji na skupku dveh idej.

- Prva ideja - eden po eden. Izberemo centroid, v bližini katerega določimo nov centroid. Za novo skupino centroidov uporabimo algoritem *K*-means. Če je novi model boljši od prejšnjega, dodamo centroid, drugače se vrnemo k prejšnji strukturi. Pomembno je odgovoriti na vprašanje, katere centroide bi veljalo razcepiti. Če bi poskusili razcepiti vse centroide in izbrali najboljšega, bi bila računaska zahtevnost prevelika.
- Druga ideja - preizkusimo del centroidov. Izberemo (na primer) polovico centroidov, ki so po nekem hevrističnem kriteriju najbolj ugodni za razcep. Izbrane centroide razcepimo in uporabimo algoritem *K*-means. Če je novi model boljši od prejšnjega, dodamo centroide, drugače se vrnemo k prejšnji strukturi. Vprašanje pa je, kaj uporabiti za hevristični kriterij, ki označuje smiselnost razcepa.

Slika 12 prikazuje delovanje algoritma *X*-means in operacij "Izboljšava parametrov" in "Izboljšava strukture". Prvotne centroide, ki smo jih pridobili z operacijo Izboljšava parametrov, razcepimo na dva naslednika. Premaknemo jih iz lokacije prvotnega centroida v nasprotnih smereh poljubno izbranega vektorja, za razdaljo, ki je proporcionalna velikosti območja. V naslednjem koraku v vsakem področju uporabimo *K*-means, kjer se otroci borijo za točke svojih staršev. Zadnji korak je preveritev modela, kjer se odločimo, ali obdržimo starša ali pa njegove naslednike. Upamo, da bodo centroidi, ki že vsebujejo točke, ki tvorijo resnične gruče, preživeli svoje otroke in jih proces ne bo spremenil. Po drugi strani pa upamo, da bodo področja, ki trenutno niso dobro predstavljena s centroidi, deležna večje pozornosti in da se bo v njih število centroidov povečalo.



Slika 12: Delovanje X-means. (a) Razcep izhodiščnih centroidov. (b) Lokalni K-means. (c) Končne gručice po izbiri modela. Povzeto po [29].

Korak, ki odloča o tem, ali obdržimo starše ali otroke, je izbira modela. Cilj izbire modela je uravnoteženje kvalitete prilagajanja k podatkom in kompleksnosti ali sposobnosti modela, da razloži vhodne podatke.

V našem primeru imamo družino alternativnih modelov M_j , kjer različni modeli pripadajo rešitvam z različnimi vrednostmi K (števili gruč). Najboljši model izberemo s pomočjo posteriornih verjetnosti. V našem primeru so vsi modeli oblike, ki jo predvideva K-means algoritem (sferični Gaussovi). Za približek posteriornih verjetnosti uporabimo naslednjo formulo, znano tudi kot Schwarzov kriterij [33] ali kriterij BIC (Bayesian Information Criterion):

$$BIC = -2 \ln L_{max} + k \ln N \quad (23)$$

kjer je L_{max} največja verjetnost (Maximum Likelihood Estimation, MLE) dosegljiva v okviru modela, k število parametrov modela in N število uporabljenih podatkovnih točk.

Izbira kriterija BIC za določanje razcepa je ena od možnosti. BIC je relativno preprosta metoda, ker potrebuje le največjo verjetnost, ki jo je mogoče doseči v okviru modela, in ne preko celotnega prostora parametrov. Izvedeni eksperimenti opisani v [29] so pokazali dobro delovanje kriterija BIC.

4.4. Razcep s singularnimi vrednostmi

Razcep s singularnimi vrednostmi (Singular Value Decomposition, SVD) [13] je pomemben matrični razcep, ki ga uporabljamo na več področjih, na primer za prilagajanje podatkov na osnovi najmanjših kvadratov in za računanje pseudoinverza.

Realno matriko A z dimenzijami $m \times n$ in $m \geq n$ lahko razcepimo na tri matrike:

$$A = U \Sigma V^T \quad (24)$$

$$U^T U = V^T V = I_n \quad (25)$$

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n) \quad (26)$$

Enačba (24) predstavlja razcep s singularnimi vrednostmi matrike A .

Matriko U sestavljajo ortonormalizirani lastni vektorji, povezani z n največjimi lastnimi vrednostmi matrike AA^T , matriko V pa ortonormalizirani lastni vektorji matrike $A^T A$. Diagonalne elementi matrike Σ so nenegativni kvadratni koreni lastnih vrednosti matrike $A^T A$. Imenujemo jih singularne vrednosti.

4.5. Kalibracija kamere

Razmerje med 3-D točko $M = [X, Y, Z]^T$ in njeno projekcijo v sliko $m = [u, v]^T$ podaja naslednja enačba:

$$s\tilde{m} = A[\mathbf{R} \quad \mathbf{t}] \tilde{M} \quad (27)$$

kjer sta $\tilde{M} = [X, Y, Z, 1]^T$, $\tilde{m} = [u, v, 1]^T$, s poljubni skalirni faktor, (\mathbf{R}, \mathbf{t}) zunanji parametri kamere (extrinsic parameters) in A notranja matrika (intrinsic matrix):

$$\mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (28)$$

(u_0, v_0) sta koordinati središčne točke (principal point), α, β skalirna faktorja za osi u in v , ter γ popačenje obeh osi slike.

Če predpostavimo $Z=0$ (vse točke so na kalibracijski plošči), sledi

$$s\tilde{\mathbf{m}} = \mathbf{H}\tilde{\mathbf{M}} \quad (29)$$

$$\mathbf{H} = \mathbf{A}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}] \quad (30)$$

in $\tilde{\mathbf{M}} = [X \quad Y \quad 1]^T$.

Na podlagi slike ravnine lahko ocenimo homografijo $\mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3]$.

Sledi:

$$[\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3] = \lambda \mathbf{A}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}], \quad (31)$$

λ je poljuben skalar.

Zaradi ortonormalnosti \mathbf{r}_1 in \mathbf{r}_2 dobimo dve glavni omejitvi za notranje parametre na podlagi ene homografije:

$$\begin{aligned} \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 &= 0 \\ \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_1 &= \mathbf{h}_2^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 \end{aligned} \quad (32)$$

Naj bo

$$\mathbf{B} = \mathbf{A}^{-T} \mathbf{A}^{-1} = \begin{bmatrix} B_{11} & B_{21} & B_{31} \\ B_{12} & B_{22} & B_{32} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \quad (33)$$

$$\mathbf{b} = [B_{11} \quad B_{12} \quad B_{13} \quad B_{21} \quad B_{22} \quad B_{23}]^T \quad (34)$$

Če je \mathbf{h}_i i -ti stolpec \mathbf{H} (30) velja:

$$\mathbf{h}_i^T \mathbf{B} \mathbf{h}_j = \mathbf{v}_{ij}^T \mathbf{b}. \quad (35)$$

Omejitve, ki sledita iz dane homografije (32), lahko zapišemo kot:

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{b} = 0. \quad (36)$$

Na podlagi n slik imamo n enačb (36) in lahko določimo

$$\mathbf{V} \mathbf{b} = 0, \quad (37)$$

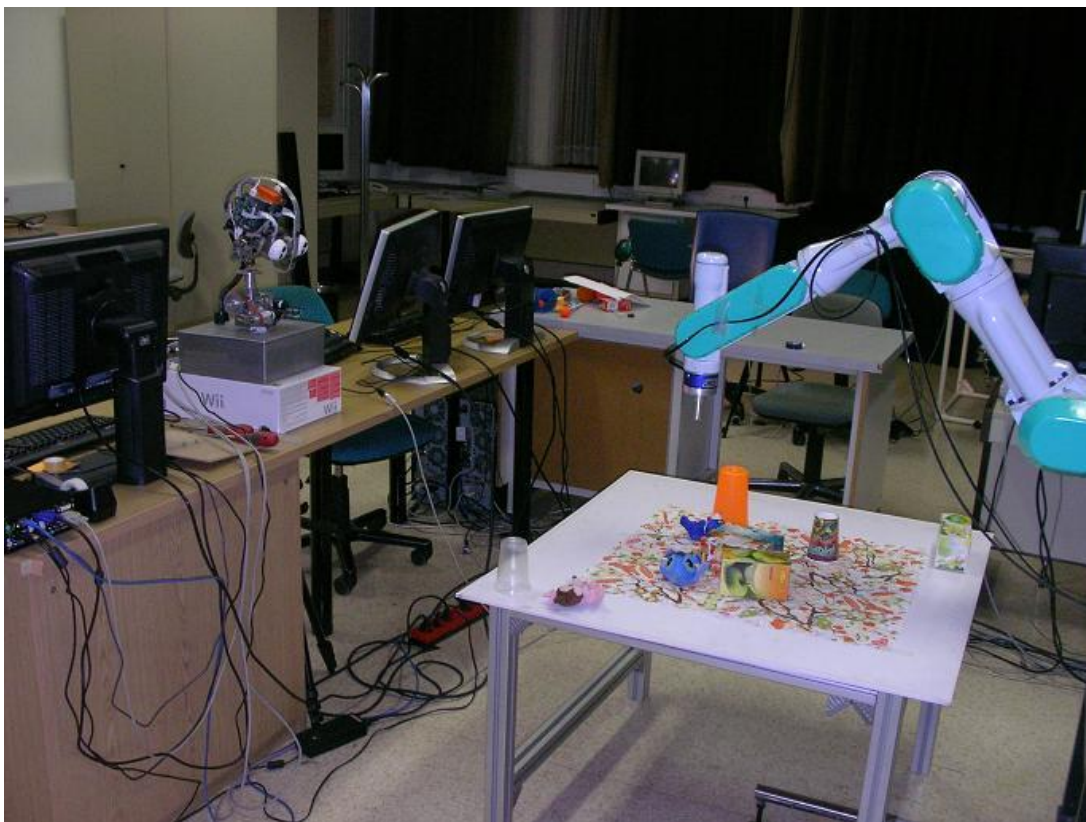
kjer je \mathbf{V} matrika $2n \times 6$.

Če imamo $n \geq 3$ slik, dobimo eno rešitev za \mathbf{b} , ki je določena do skalirnega faktorja. Ko ocenimo \mathbf{b} , lahko izračunamo notranje in zunanje parametre kamere z enačbami, ki so navedene v [41].

5. Eksperimenti

5.1. Realni robotski sistem

Pri eksperimentalnem delu je bilo potrebno združiti sistem za vid s sistemom, s katerim lahko izvajamo manipulativne akcije. Našo eksperimentalno celico, ki jo prikazuje Slika 13, sestavlja robotski manipulator Mitsubishi PA-10 in sistem za stereo vid – humanoidna glava z Univerze v Karlsruheju [2].



Slika 13: Eksperimentalna robotska celica.

Preden začnemo z eksperimenti, moramo sistem usposobiti za delovanje v realnem okolju. Kalibracija kamer je potreben korak, če želimo pretvoriti 2-D informacije, ki jih zajamemo s kamerami, v 3-D informacije, ki so uporabne za manipulacijo objektov v prostoru. Za kalibracijo kamer smo uporabili fleksibilno in robustno metodo [41], več o tem v poglavju 4.5.

Uporabljena metoda zahteva, da kamere opazujejo planarni vzorec pri nekaj (vsaj dveh) različnih orientacijah. Kalibracijski vzorec je lahko natisnjen z laserskim tiskalnikom in pritrjen na primerno površino, na primer knjigo. Premikamo lahko bodisi kamere bodisi planarni vzorec. Računalniške simulacije in preizkusi z realnimi podatki so v obeh primerih pokazali zelo dobre rezultate.

Pri postopku kalibracije smo uporabili $n=30$ slik kalibracijskega vzorca. Vzorec je šahovnica s 7×5 kvadrati. Slike smo zajeli pri različnih oddaljenostih in orientacijah vzorca. Na podlagi slik smo določili parametre kamere, ki smo jih v nadaljevanju uporabili pri določanju 3-D lokacij ujemajočih značilnic.

Za oceno transformacije iz koordinatnega sistema robotske glave v koordinatni sistem robotske roke smo kalibracijski vzorec pritrčili na robotsko roko. Na ta način lahko iz slik kalibracijskega vzorca določimo množico 3-D točk v koordinatnem sistemu robotske roke. Z znanimi parametri kamere lahko isto množico 3-D točk predstavimo v koordinatnem sistemu robotske glave. Pri oceni transformacije smo uporabili 5 slik kalibracijskega vzorca pri 5 različnih oddaljenostih, kar ustreza premiku po osi z koordinatnega sistema robotske roke, in 2 rotacijah – okrog osi x in okrog osi y koordinatnega sistema robotske roke.

5.2. Preizkusi

V naših eksperimentih smo želeli preveriti in oceniti delovanje treh procesov: proces oblikovanja hipotez o predmetih, potiskanje, ki potrdi ali ovrže hipotezo, in proces vrednotenja hipotez glede na morebitne spremembe prizora.

Pri preizkušanju procesa oblikovanja hipotez smo predpostavili dobro delovanje v primeru, da testni predmeti vsebujejo ravne površine. Zato smo poleg predmetov, ki vsebujejo ravne dele (škatle), uporabili še predmete, ki nimajo ravnih površin (igračke, lončki). Testni predmeti (prikazuje jih Slika 14) so 3 škatle, 3 lončki in 3 igrače.



Slika 14: Vsi testni predmeti.

Izvedli smo 50 poskusov, in sicer po 10 poskusov za 5 različnih podmnožic testnih predmetov. Podmnožice so označene s pikami in jih prikazuje Tabela 1. Škatle označimo po vrsti: 'jagoda', 'jabolko', 'limona'. Lončki: 'pisan', 'oranžen', 'prozoren'. Igrače: 'temnomodra', 'svetlomodra', 'roza'. Točke označujejo predmete, ki smo jih uporabili pri poskusih. Na primer: v prvih 10 poskusih smo uporabili škatli za jagodni in jabolčni čaj, pisan in prozoren lonček ter temnomodro in roza igračko.

Tabela 1: Podmnožice testnih predmetov.

Poskusi	Testni predmeti								
	škatla			lonček			igračka		
	1	2	3	1	2	3	1	2	3
01-10	×	×		×		×	×		×
11-20	×	×	×	×	×			×	×
21-30	×			×	×	×	×	×	×
31-40		×	×	×			×		
41-50	×		×		×			×	×

5.2.1. Oblikovanje hipotez

Pri vseh poskusih je bilo oblikovano vnaprej določeno število hipotez k ($k = 6$). Končno število hipotez o obstoju predmetov se je spreminjalo glede na rezultat algoritma za gručenje. Slika 15 prikazuje 3 izmed izračunanih hipotez. V tem primeru vsaka od prikazanih hipotez opisuje eno ploskev na enem predmetu.

V naših eksperimentih smo uporabili predmete z ravnimi površinami, ki se nahajajo na podobnih višinah. Pri takšnih predmetih obstaja velika verjetnost, da bodo oblikovane hipoteze povezale več predmetov. Ker smo takšne predmete pogosto postavili skupaj, je 42% hipotez vključevalo točke, ki pripadajo več predmetom.



Slika 15: Tri izmed hipotez za obstoj predmeta.

5.2.2. Potiskanje

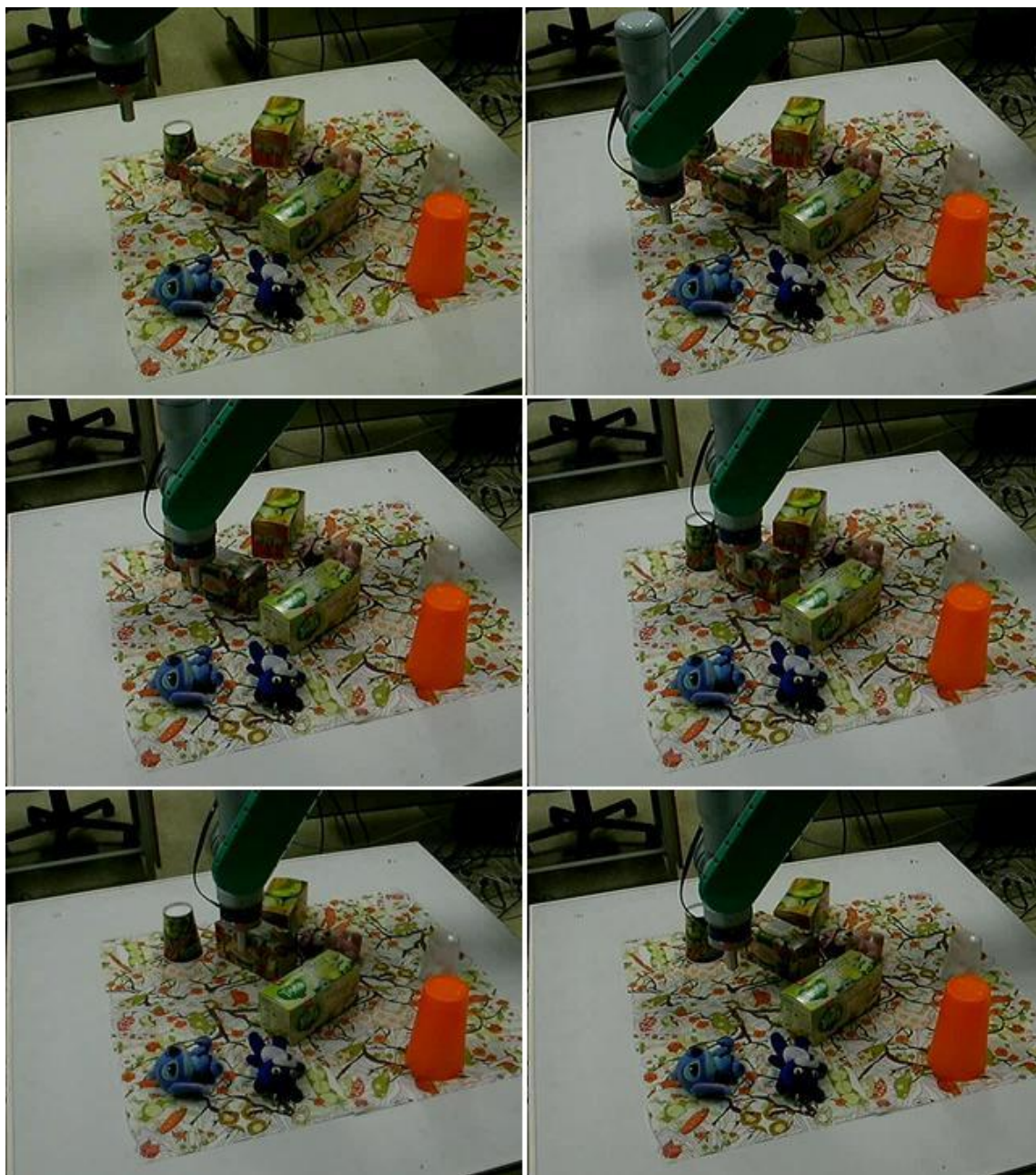
Pri 96% eksperimentov je bilo potiskanje izvedeno, pri ostalih eksperimentih določene akcije niso bile mogoče zaradi omejitev delovnega prostora. Slika 16 prikazuje enega od poskusov potiskanja izbrane hipotetične ravnine, ki poteka po vnaprej določeni trajektoriji.

5.2.3. Vrednotenje hipotez

Pri vrednotenju hipotez smo upoštevali dve možnosti. Prva možnost je, da točke, ki jih vsebuje hipoteza o predmetu, pripadajo enemu predmetu. V takšnih primerih bi moral robot točke prepoznati kot točke predmeta in (dodatno) poiskati več točk, ki pripadajo predmetu – seveda če se točke premaknejo kot posledica potiskanja.

Druga možnost je, da točke v hipotezi pripadajo več predmetom in/ali ozadju. Podobno kot v prvem primeru želimo premaknjene točke dodati modelu predmeta, ki ga robot premakne s potiskanjem. Dodatna naloga je odvzemanje točk, ki sicer so v hipotezi, a ne pripadajo predmetu, ki ga robot potisne. Predpostavka je, da se te točke ne premaknejo oziroma njihov premik ni v skladu s hipotezo o gibanju togega telesa.

Obe možnosti sta dokaj enakovredno zastopani, saj 42% hipotez vključuje točke, ki pripadajo več predmetom.



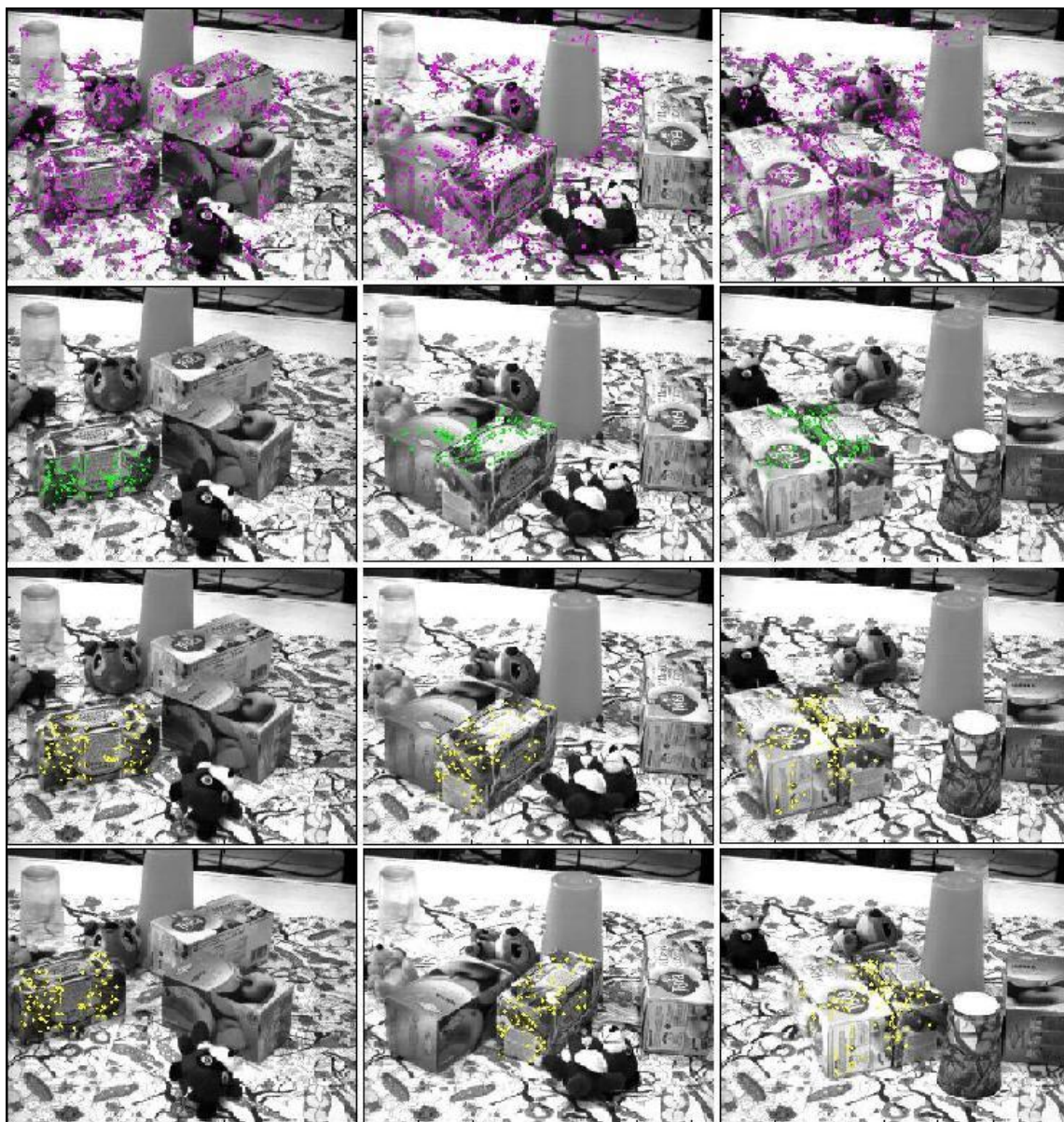
Slika 16: Potiskanje izbrane hipoteze za obstoj predmeta.

Slika 17 v stolpcih prikazuje rezultate nekaterih od izvedenih eksperimentov. Struktura slike je naslednja: prva vrstica prikazuje lokacije vseh točk, za katere smo lahko določili 3-D lokacije, druga vrstica pa točke, ki pripadajo izbrani hipotezi za obstoj predmeta. Tretja in četrta vrstica kažeta rezultate. Prikazane so točke, ki se premaknejo v skladu s hipotezo o togem gibanju predmeta. Prikazane so v slikah pred in po potiskanju.

Prvi stolpec demonstrira primer dodajanja točk. Hipoteza vsebuje točke enega predmeta. Robot ta predmet potisne in točke predmeta se zato premaknejo v skladu z načelom gibanja togega telesa. Tako pridobimo dovolj informacij za dodajanje točk modelu predmeta. Rezultat je nov model predmeta, ki poleg točk v začetni hipotezi vsebuje tudi dodatne točke.

Drugi stolpec prikazuje dodajanje in odzemanje točk. Hipoteza vsebuje točke več predmetov. Robot potisne enega izmed predmetov, katerega točke so vsebovane v hipotezi. Točke tega predmeta se premaknejo. Točke predmetov, ki se niso premaknili, algoritem uspešno zavrže.

Tretji stolpec, podobno kot drugi, prikazuje dodajanje in odzemanje točk. Hipoteza zopet vsebuje točke več predmetov. Razlika je v tem, da se oba predmeta, ko robot izvede potiskanje, premakneta zelo podobno. Točke obeh predmetov se premaknejo skladno z načelom gibanja togega telesa, zato jih algoritem označi kot točke predmeta. Med točkami različnih predmetov je torej mogoče razlikovati le, če se premaknejo različno.



Slika 17: Primeri eksperimentov.

Množico točk, ki jih naš sistem vključi v model objekta, lahko robot poveča z rotacijo objekta. Na ta način robot spremeni zorni kot, iz katerega opazuje objekt. Rotacija predmeta mora biti po eni strani dovolj velika, da odkrije nove, prej nevidne značilnice, in po drugi strani dovolj majhna, da omogoči določanje ujemanja značilnic v zaporednih slikah. Ker s potiskanjem pogosto ni mogoče natančno nadzorovati gibanja neznanih objektov, lahko pride pri potiskanju do takšnih premikov, ki niso primerni za dodajanje novih točk. Do tega lahko pride 1) zaradi prevelike rotacija predmeta, ki onemogoči ujemanje med zaporednima slikama, ali pa 2) ker do premika predmeta sploh ne pride.

Pri prevelikih rotacijah moramo ponovno potisniti opazovani predmet. Ker se je prizorišče spremenilo, mora biti potiskanje osnovano na na novo oblikovanih hipotezah o predmetih. Kadar je premik objekta premajhen, da bi bilo mogoče ločiti točke na podlagi principa togega gibanja, pa naš sistem izbere naslednjo hipotezo iz množice prej izračunanih hipotez in izvede nov potisk.

Tabela 2 prikazuje statistično oceno 50 potiskanj. Prva vrstica prikazuje delež eksperimentov, kjer je bila posledica potiska premik predmeta, in določanje točk predmeta. Delež takšnih eksperimentov, ki jih označimo kot 'Uspeh', je 58%. Druga vrstica označuje eksperimente, kjer je bil premik (oziroma rotacija) predmeta prevelika, zato ni bilo mogoče ujemanje značilnic v sosednjih slikah. Delež takšnih eksperimentov je 36%. Ker je prišlo do spremembe prizorišča, je potrebno oblikovati novo množico hipotez. Tretja vrstica opisuje eksperimente, kjer do spremembe prizorišča ni prišlo. Vzrok je neizvedeno potiskanje zaradi omejitev delovnega prostora (4% eksperimentov) in potiskanje, ki je sicer bilo izvedeno, a ni prispevalo k spremembi prizora (2% eksperimentov). V takšnih primerih ni potrebno izračunati novih hipotez, ker lahko uporabimo naslednjo izmed oblikovanih hipotez.

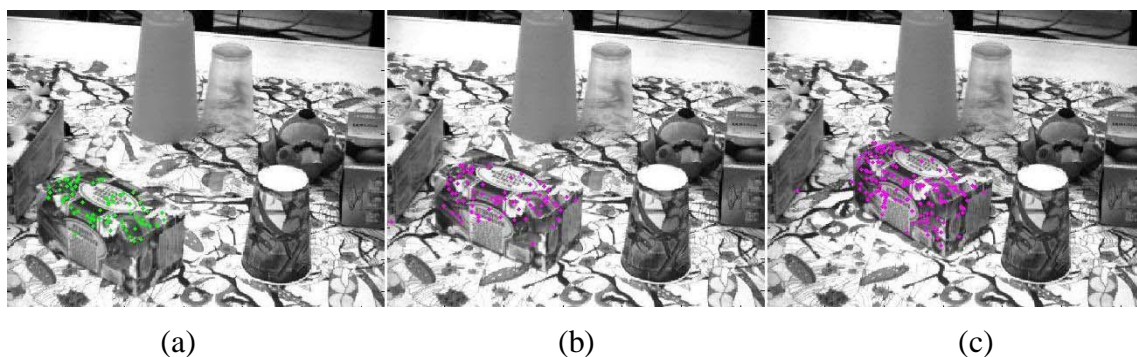
Tabela 2: Rezultati potiskanja za 50 eksperimentov

Točke predmeta	Rezultat	Delež eksperimentov
$\geq F_r$	Uspeh	58%
$< F_r$	Oblikuj nove hipoteze	36%
/	Izberi naslednjo hipotezo	6%

* $F_r = 10$ (prag)

5.2.4. Grajenje modelov

Informacije o predmetu zbiramo s spreminjanjem pogleda na predmet. Z vsakim novim pogledom želimo odkriti nove, prej nevidne značilnice, za kar potrebujemo določeno minimalno rotacijo telesa. Rotacija pa mora biti po drugi strani dovolj majhna, da je določanje ujemanja značilnic v zaporednih slikah mogoče. Slika 18 predstavlja uspešno kopičenje značilnic po dveh zaporednih potiskih, ki sta vodila v primerno rotacijo predmeta, in s tem uspešno grajenje modela predmeta.



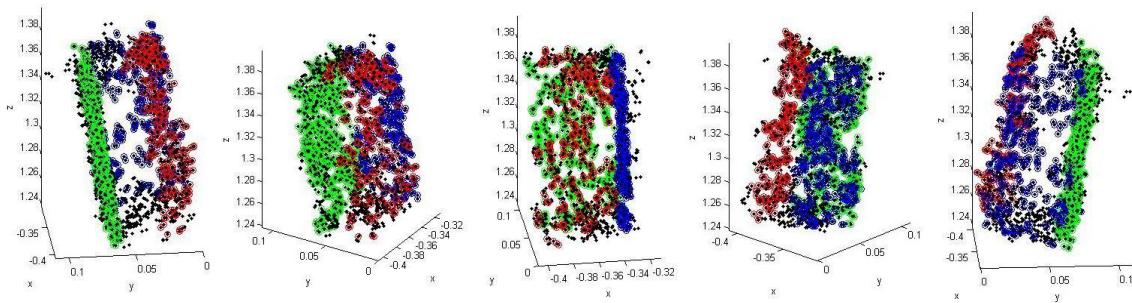
Slika 18: Grajenje modela s kopičenjem značilnic. (a) Hipoteza o obstoju predmeta. Določene točke predmeta po prvem potiskanju (b) in po drugem potiskanju (c).

S potiskanjem mora robot zagotoviti primeren premik objekta. Vendar pa lahko odziv predmeta na potisk natančno izračunamo le, če poznamo celoten model objekta. Dokler takšnega modela nimamo, lahko potiske določimo samo na hevrističen način. Zato se

lahko zgodi, da robot izvede potisk, ki ne omogoča nadaljnje gradnje modela.

V primeru velikih rotacij izračun ujemanja med vizualnimi značilnicami v sosednjih slikah ni mogoč. Zagotavljanje primerne rotacije preko več zaporednih slik je v celoti mogoče le z boljšo kontrolo nad predmetom. Večjo kontrolo nad predmetom lahko zagotovimo s prijemanjem predmeta ali pa v sodelovanju s človekom.

Zato smo izvedli eksperiment, v katerem smo preizkusili grajenje objektnega modela v sodelovanju s človekom. Pri tem človek potiska predmet na način, s katerim zagotovi, da je rotacija v zaporednih slikah primerna. Slika 19 prikazuje rezultat - model predmeta, ki vsebuje značilnice, izločene iz 38 pogledov na predmet.



Slika 19: Model predmeta na osnovi 38 različnih pogledov.

Ocenjene in dejanske mere predmeta:

- 0.060m in 0.060m,
- 0.070m in 0.067m,
- 0.131m in 0.132m.

Ocenjeni in dejanski koti med ravninami:

- 95.8° in 90° (modra in rdeča ravnina),
- 103.6° in 90° (zelena in rdeča ravnina).

Preizkusi grajenja modelov s potiskanjem robota so se zaradi prej naštetih dejstev končali preden bi lahko zajeli dovolj slik z različnimi pogledi na predmet.

Naš pristop pa je uspešen pri odkrivanju novih predmetov. Eksperimentalni rezultati kažejo, da je naš sistem uspešen pri oblikovanju hipotez, pri potiskanju izbrane detektirane ravnine in določanju točk, ki pripadajo predmetu.

6. Zaključek

Če predmet razumemo kot množico točk, ki pripadajo togemu telesu in se zato predvidljivo gibljejo, lahko nove predmete odkrivamo z analizo gibanja točk v zaporednih prizorih. V našem delu smo obravnavali prizore, ki vsebujejo objekte sestavljene iz planarnih ploskev. Ravnine torej namigujejo na obstoj predmeta, zato smo iskanje premetov smo osnovali na iskanju ravnin v slikah. Za potrditev takšne hipoteze o obstoju predmeta robot poizkusi detektirano ravnino premakniti s potiskanjem. Po uspešnem potisku lahko hipotezo o obstoju predmeta potrdimo ali ovržemo. Po več uspešnih potiskih je mogoče znanje o predmetu kopičiti preko zbranih značilnic.

Eksperimentalni rezultati kažejo, da je naš sistem uspešen pri oblikovanju hipotez. Pri vseh eksperimentih je bilo oblikovano vnaprej določeno število hipotez. Sistem je uspešen tudi pri potiskanju izbrane hipotetične ravnine. Pri določanju točk predmeta je sistem sposoben tako dodajanja točk, ki se premaknejo v skladu z ocenjenimi parametri gibanja togega telesa, kot tudi odzemanja točk, ki ne pripadajo predmetu, iz začetne množice točk, ki sestavljajo hipotezo o obstoju predmeta. Prav tako ima sistem primeren odziv v primerih, ko je sprememba na prizorišču premajhna (ali pa je ni) in v primerih, ko je rotacija potisnjene predmeta prevelika. Lahko zaključimo, da je naš pristop uspešen pri odkrivanju novih predmetov.

Pri gradnji modelov predmetov kopičimo značilnice s spreminjanjem pogleda na predmet. V našem primeru to izvedemo s spremembo algoritma za potiskanje. Za uspešno grajenje modelov mora biti rotacija predmeta v določenih mejah. Rotacija predmeta mora biti po eni strani dovolj velika, da sistem odkrije nove, prej nevidne značilnice, in po drugi strani dovolj majhna, da je določanje ujemanja značilnic v zaporednih slikah mogoče. Ko določimo nove točke predmeta, torej točke, ki smo jih

določili iz novega pogleda, jih dodamo k obstoječim točkam predmeta. Lokacije ostalih točk predmeta preslikamo v skladu z izračunano rotacijo in translacijo.

Pridobljeni rezultati so pokazali, da je predlagan pristop uspešen pri odkrivanju novih predmetov in da je znanje, ki ga pridobimo z več zaporednimi manipulativnimi akcijami, mogoče zbirati v ustrezno verificirane modele predmeta.

Za uspešno gradnjo modelov včasih potrebujemo boljši nadzor nad gibanjem predmeta. Delež eksperimentov, kjer je bilo potrebno potiskanje ponoviti zaradi prevelike rotacije, je dovolj visok (36%), da je na mestu razmislek o uporabnosti zgolj potiskanja za gradnjo modelov. Boljši nadzor nad gibanjem predmeta je mogoče doseči s prijemanjem predmeta ali pa v sodelovanju s človekom. Na ta način lahko zagotovimo primerne premike opazovanega predmeta. Eksperimentalno smo preizkusili, da lahko v sodelovanju s človekom uspešno gradimo objektne modele.

Delo v prihodnosti bo usmerjeno predvsem v združitev predlaganega pristopa, ki je uspešen pri odkrivanju novih predmetov, in prijemanja, kar bo povečalo natančnost gibanja najdenih predmetov. V nadaljnjih raziskavah bomo odkrivanju predmetov in gradnji njihovih modelov dodali še določanje zmožnosti (angl. affordances) predmetov.

7. Viri

- [1] K. S. Arun, T. S. Huang, S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, št. 9, zv. 5, str. 698-700, 1987.
- [2] T. Asfour, K. Welke, P. Azad, A. Ude, R. Dillmann, "The Karlsruhe Humanoid Head", v zborniku *Humanoid Robots 2008*, str. 447-453, 2009.
- [3] I. Biederman, "Recognition by components: A theory of human image understanding," *Psychological Review*, št. 2, zv. 94, str. 115-147, 1987.
- [4] M. Brown, D. G. Lowe, "Invariant features from interest point groups," v zborniku *British Machine Vision Conference*, Cardiff, Wales, 2002, str. 656-665.
- [5] S. J. Dickinson, *What is cognitive science*, Basil Blackwell publishers, 1999, str. 172-207.
- [6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2004, pogl. 9.
- [7] J. Feldman, "What is a visual object?", *Trends in Cognitive Sciences*, št. 7, zv. 6, str. 252–256, 2003.
- [8] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Simultaneous object recognition and segmentation from single or multiple model views," *International Journal of Computer Vision*, št. 67, zv. 2, str. 159-188, 2006.
- [9] M. A. Fischler, R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, št. 24, zv. 6, str. 381-395, 1981.
- [10] P. Fitzpatrick, "First contact: An active vision approach to segmentation," v zborniku *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, Las Vegas, Nevada, 2003, str. 2161-2166.

- [11] Ch. Geib, K. Mourao, R. Petrick, N. Pugeault, M. Steedman, N. Krüger, F. Wörgötter, "Object Action Complexes as an Interface for Planning and Robot Control", v zborniku *International Conference on Humanoid Robots*, 2006.
- [12] J. Gibson, *The ecological approach to Visual Perception*, Houghton Mifflin, Boston, MA, 1979.
- [13] G. Golub, C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Mathematik*, št. 5., zv. 14, str. 403-420, 1970.
- [14] T. Kadir, A. Zisserman, M. Brady, "An affine invariant salient region detector," v zborniku *European Conference on Computer Vision*, Prague, Czech Republic, 2004, str. 228-241.
- [15] G. Kootstra, J. Ypma, B. de Boer, "Active exploration and keypoint clustering for object recognition," v zborniku *International Conference on In Robotics and Automation*, Pasadena, California, USA, 2008, str. 1005-1010.
- [16] D. Kraft, N. Pugeault, E. Baseski, M. Popovic, D. Kragić, S. Kalkan, F. Wörgötter, N. Krüger, "Birth of the object: detection of objectness and extraction of object shape through object–action complexes," *International Journal of Humanoid Robotics*, št. 5, zv. 2, str. 247-256, 2008.
- [17] Leonardis, A. Jaklič, F. Solina, "Superquadrics for segmentation and modeling range data," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, št. 19, zv. 11, str. 1289-1295, 1997.
- [18] W. H. Li, A. M. Zhang, and L. Kleeman, "Bilateral Symmetry Detection for Real-time Robotics Applications," *The International Journal of Robotics Research*, št. 7., Vol. 27, str. 785-814, 2008.
- [19] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, št. 2, zv. 60, str. 91-110, 2004.
- [20] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," v zborniku *5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, zv. 1, str. 281-297, 1967.
- [21] J. Matas, O. Chum, M. Urban, T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," v zborniku *13th British Machine Vision Conference*, Cardiff, UK, 2002, str. 384-393.

- [22] G. Metta and P. Fitzpatrick, "Grounding vision through experimental manipulation," *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, št. 361, zv. 1811, str. 2165-2185, 2003.
- [23] G. Metta and P. Fitzpatrick, "Early integration of vision and manipulation," *Adaptive Behavior*, št. 11, zv. 2, str. 109-128, 2003.
- [24] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, št. 27, zv. 10, str. 1615-1630, 2005.
- [25] K. Mikolajczyk, C. Schmid, "Scale & Affine Invariant Interest Point Detectors," *International Journal of Computer Vision*, št. 60, zv. 1, str. 63-86, 2004.
- [26] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, št. 65, zv. 1-2, str. 43-72, 2005.
- [27] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3-D objects," v zborniku *10th IEEE International Conference on Computer Vision*, Beijing, China, 2005, zv. 1, str. 800-807.
- [28] R. M. Murray, Z. Li., S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation* (1st Edition), CRC Press, 1994, pogl. 2.
- [29] D. Pelleg, A. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," v zborniku *Seventeenth International Conference on Machine Learning*, San Francisco, California, 2000, str. 727-734.
- [30] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, "3-D object modeling and recognition using affine-invariant patches and multi-view spatial constraints," v zborniku *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, 2003, zv. 2, str. 272-277.
- [31] R.B. Rusu, B. Gerkey, M. Beetz, "Robots in the kitchen: Exploiting ubiquitous sensing and actuation," *Robotics and Autonomous Systems Journal (Special Issue on Network Robot Systems)*, št. 10, zv. 56, str. 844-856, 2008.
- [32] A. Saxena, J. Driemeyer, A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, št. 2., zv. 27, str. 157-173, 2008.

- [33] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, št. 2, zv. 6, str. 461-464, 1978.
- [34] E. Trucco, A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.
- [35] E. Trucco, F. Isgro and F. Bracchi, "Plane detection in disparity space," v zborniku *International Conference on Visual Information Engineering*, Guildford, UK, 2003, str. 73–76.
- [36] C. J. Tsikos and R. K. Bajcsy, "Segmentation via manipulation," *IEEE Transactions on Robotics and Automation*, št. 7, zv. 3, str. 306-319, 1991.
- [37] Tuytelaars, L. Van Gool, "Matching Widely Separated Views Based on Affine Invariant Regions," *International Journal of Computer Vision*, št. 59, zv. 1, str. 61-85, 2004.
- [38] A. Ude, D. Omrčen, and G. Cheng, "Making object learning and recognition an active process," *International Journal of Humanoid Robotics*, št. 5, zv. 2, str. 267-286, 2008.
- [39] A. Ude and E. Oztop, "Active 3-D vision on a humanoid head," v zborniku *Proc. 14th Int. Conf. Advanced Robotics*, Munich, Germany, 2009.
- [40] (2009) A. Vedaldi, SIFT for Matlab. Dostopno na:
<http://www.vlfeat.org/~vedaldi/code/sift.html>
- [41] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, št. 22, zv. 11, str. 1330-1334, 2000.