

UPORABA UMETNEGA SPOZNAVNEGA VIDA

Danijel Skočaj, Dušan Omerčević, Aleš Leonardis

Laboratorij za umetne vizualne spoznavne sisteme

Fakulteta za računalništvo in informatiko, Univerza v Ljubljani

E-pošta: {danijel.skocaj,dusan.omercevic,ales.leonardis}@fri.uni-lj.si

URL: <http://vicos.fri.uni-lj.si>

POVZETEK: *Povečano povpraševanje po razvoju na področju visokonivojskega računalniškega vida je v zadnjih letih pripeljalo do oblikovanja nove znanstvene discipline - umetnega spoznavnega vida. Za sisteme umetnega spoznavnega vida je značilno, da so bolj robustni, prožni in prilagodljivi. Za razliko od že uveljavljenih sistemov računalniškega vida so sposobni delovati tudi v nepredvidenih pogojih in so sposobni iz vizualnih podatkov izluščiti tudi visokonivojsko semantično informacijo. V tem članku bomo na kratko orisali glavne značilnosti umetnih vizualnih spoznavnih sistemov ter predstavili nekaj potencialnih aplikacij. Pri tem se bomo osredotočili na aplikacijo za avtomatsko lokalizacijo v urbanem okolju na osnovi vizualne informacije.*

1. UVOD

Računalniški vid se je v preteklih desetletjih že dodobra uveljavil kot pomembna znanstvena disciplina tako v teoriji kot v praksi [1]. Narejen je bil velik napredek na raznih področjih in veliko problemov je bilo zadovoljivo rešenih. Mnoga področja računalniškega vida so prišla v zrelo fazo, ko so trenutne zmogljivosti poznane, razvoj posameznih komponent dokaj predvidljiv, pričakovanja pa realna. To velja predvsem za specifična področja t.i. *nizkonivojskega računalniškega vida*¹ s točno določenim problemom ter omejenim področjem delovanja. Ob predpostavki, da lahko do določene mere nadziramo okolje, v katerem sistem računalniškega vida deluje, in je nabor možnih vhodnih signalov omejen ter poznan, lahko takšni sistemi zelo zanesljivo delujejo.

Razvoj računalniškega vida in strojne opreme pa je v zadnjih letih naredil še korak naprej. Usmeril se je tudi k razvoju bolj splošnih sistemov, ki naj bi delovali v manj predvidljivih pogojih in bi se znali odzvati na nenadne nepredvidene spremembe v okolici. V takšnih primerih je potrebna nadgradnja klasičnih rešitev in razvoj *visokonivojskih sistemov računalniškega vida*². V ta namen se pogosto uporabljajo tudi

¹ Procesiranje slik, izločanje lokalnih značilnic, izračun osnovnih značilnosti (barva, oblika, tekstura,...), ipd.

² Razumevanje ter interpretacija slik, naprednejše predstavitve, semantični opisi, ipd.

metode umetne inteligence, ki pa ponavadi na vhodu zahtevajo čiste simbolne opise. Teh pa računalniški vid ne more v popolnosti zagotavljati, zato mora za učinkovito kombinacijo teh metod priti do ustreznih prilagoditev. Tako se je v zadnjih letih začela razvijati znanstvena disciplina poimenovana *umetni spoznavni vid* (umetni kognitivni vid, ang. cognitive vision)³. Lahko bi rekli, da gre pri tem za povezovanje računalniškega vida z umetno inteligenco in kognitivnimi znanostmi. Cilj ni samo videti, ampak tudi razumeti kar vidimo (oz. vidi umetni agent). Želja je torej približati sposobnosti umetnih agentov človeškim sposobnostim vizualnega zaznavanja in spoznavanja.

Te želje in cilji niso novi in tudi raziskave na tem področju se niso začele šele s pojavom termina umetni spoznavni vid. Klasični računalniški vid in umetni spoznavni vid se v marsičem prekrivata in veliko raziskav na področju visokonivojskega računalniškega vida je bilo opravljenih že v preteklosti. Razlika je ta, da se je v zadnjih letih povečala potreba po intenzivnejših raziskavah na tem področju, zato se je tudi izoblikovala nova disciplina, ki je bolj interdisciplinarno naravnana in ki skuša s sodelovanjem računalniškega vida, umetne inteligence, robotike, kognitivnih znanosti in drugih sorodnih področij, narediti konkreten korak naprej pri razvoju sistemov spoznavnega vida in umetnih spoznavnih sistemov nasploh.

Cilj tega članka je osvetliti ta korak in željo po kvalitetnem preskoku ter prikazati možnosti uporabe umetnega spoznavnega vida v praksi. Tako bomo v drugem razdelku bolj podrobno opredelili področje umetnega spoznavnega vida in umetnih spoznavnih sistemov, v tretjem poglavju pa prikazali nekaj že razvitih in potencialnih aplikacij. Nato se bomo v četrtem poglavju osredotočili na aplikacijo za avtomatsko lokalizacijo v urbanem okolju na osnovi vizualne informacije. Članek bomo zaključili s sklepom v katerem bomo poskušali predvideti nadaljnji razvoj na področju umetnih spoznavnih sistemov.

2. UMETNI SPOZNAVNI (VIZUALNI) SISTEMI

Termin umetni spoznavni vid je bil uveden, da bi zajel poskuse za doseganje bolj robustnih, prožnih in prilagodljivih sistemov računalniškega vida z njihovo nadgradnjo s spoznavnimi zmožnostmi: sposobnostjo učiti se, prilagajati se, primerjati alternativne rešitve, in celo razvijati nove strategije za analizo in interpretacijo [3]. Ključna lastnost sistema umetnega spoznavnega vida je sposobnost robustnega in zanesljivega delovanja tudi v nepredvidljivih okoliščinah. Tak sistem mora biti sposoben predvideti dogodke in temu ustrezno prilagajati njegovo delovanje. Cilj sistema ni le enostavna obdelava slike, ampak iz vizualne informacije izluščiti tudi višjenivojsko semantično informacijo in sliko (ali sekvenco slik) temu primerno interpretirati. Pri vsem tem igra zelo pomembno vlogo vizualno učenje. Obseg vizualnih podatkov, ki so umetnemu agentu na voljo (in jih mora obdelovati), je namreč odločno prevelik, da bi se znanje o teh podatkih zajelo v

³ Iz SSKJ: *spoznaven, kognitiven*: nanašajoč se na spoznavanje, spoznanje; *spoznavati*: na osnovi zaznav, podatkov in umske dejavnosti prihajati do poznavanja česa in vedenja [2].

ustrezne modele zgolj s programiranjem in ročnim specificiranjem določenih lastnosti. Umetni vizualni spoznavni sistem si mora veliko večino znanja čimbolj samodejno pridobiti z učenjem.

Umetni spoznavni vid pa lahko nastopa tudi v širšem kontekstu – kot del *umetnega spoznavnega sistema* [4]. Za učinkovito in zanesljivo delovanje avtonomnega umetnega agenta je namreč potrebno zagotoviti učinkovito povezavo med različnimi komponentami spoznavnega sistema. Za povečanje robustnosti ima tako tak sistem lahko več senzorjev, s katerimi zaznava okolico (vizualni, haptični, infrardeči, akustični, globinski senzorji, itn.). Pri utelešenih (ang. embodied) umetnih agentih je zelo pomemben robotski del (aktuatorji, robotske roke, mobilne platforme). Veliko vlogo pri načrtovanju in sklepanju igrajo tudi metode umetne inteligence. Nenazadnje, za učinkovito komunikacijo (verbalno in neverbalno) s človekom (pa tudi med agenti samimi) je potreben tudi prijazen in zmogljiv način komuniciranja, ki je čim bližje človeškemu naravnemu jeziku. Tu igrajo pomembno vlogo lingvistične metode. Ker pa je delovanje kateregakoli agenta v realnem okolju zelo odvisno od uspešnega procesiranja informacij, ki jih agent zaznava, in je med vsemi vrstami informacij daleč najbolj pomembna ravno vizualna (70 % informacij iz okolja dobi človek v vizualni obliki), igra umetni spoznavni vid ključno vlogo pri razvoju umetnih spoznavnih sistemov.

Že iz kratkega opisa lastnosti umetnih spoznavnih sistemov je razvidno dvoje – da so umetni spoznavni sistemi potencialno zelo uporabni na mnogih področjih, ter da je njihov razvoj precej kompleksno in zahtevno delo. Dosedanji razvoj na področju računalniškega vida in drugih zgoraj omenjenih področjih je dejansko omogočil resnejše raziskave in aplikacije v smeri splošnih spoznavnih sistemov. Ravno ta potencialna uporabnost (in posledično korist) je botrovala precejšnjemu povečanju vlaganj na to področje, tako v zasebnem kot tudi v javnem sektorju. Tako so v okviru Generalnega direktorata za informacijsko družbo in medije pri Evropski komisiji ustanovili posebno enoto *Cognition*, ki je zadolžena za podporo raziskavam na področju umetnih spoznavnih sistemov [5]. Samo v okviru te enote je bilo od leta 2003 na treh razpisih šestega okvirnega raziskovalnega programa namenjenih za razvoj kognitivnih sistemov in naprednih robotov preko 100 milijonov evrov. Glede na splošnost in robustnost umetnih spoznavnih sistemov, pa tudi glede na njihovo prijaznost do uporabnikov in prilagodljivost, lahko pričakujemo vstop te tehnologije na mnoga področja, od industrijskih obratov do domačega okolja.

3. APLIKACIJE

Iz ohlapnega in precej ambicioznega opisa umetnih vizualnih spoznavnih sistemov in spoznavnih sistemov nasploh lahko kmalu identificiramo kopico konkretnih aplikacij letih. Te aplikacije lahko razumemo kot obogatitev že obstoječih aplikacij klasičnega računalniškega vida, ki izkoriščajo že obstoječo infrastrukturo in jim dodajajo novo vsebino, ali pa kot razvoj popolnoma novih sistemov. V nadaljevanju bomo prikazali

nekaj obstoječih in potencialnih aplikacij ter pokazali, kako se le-te razlikujejo od klasičnih aplikacij računalniškega vida.

Industrijske aplikacije. V industrijskih obratih je že nameščenih zelo veliko sistemov strojnega vida, ki v omejenih pogojih učinkovito opravljajo predvsem kvantitativne meritve pri optičnem nadzoru kakovosti in sorodnih nalogah. Metode spoznavnega vida lahko nabor nalog razširijo tudi na kvalitativne meritve in estetske preglede ter zmanjšajo potrebo po kontroliranih pogojih. S povečevanjem splošnosti, robustnosti in prilagodljivosti sistemov lahko le-ti postanejo bolj prijazni za upravljanje, bolj prilagojeni željam posameznih uporabnikov, večkrat uporabljivi ter bolj fleksibilni, kar omogoča, da se jih da na enostaven način (z učenjem ter povratno zanko) učinkovito rekonfigurirati.

Video nadzor. Trg s sistemi za video nadzor se iz leta v leto povečuje in samodejni sistemi za video nadzor pridobivajo na pomembnosti. Že zelo veliko takšnih sistemov je nameščenih po raznih javnih in zasebnih stavbah in zunanjih površinah. Ti sistemi pa so skoraj brez izjeme zelo enostavni - sposobni so le zaznati gibanje, mogoče tudi slediti objektom oz. osebam v vidnem polju. Te sisteme je mogoče zelo lepo nadgraditi s prvinami spoznavnega vida. Nadgrajeni sistemi, temelječi na že obstoječi infrastrukturi, bi lahko bili sposobni samodejnega razpoznavanja neobičajnega obnašanja, učinkovitega razpoznavanja konkretnih oseb na slikah, lahko bi bili bolj splošni, prilagodljivi, s samodejnim učenjem bi se lahko hitro in enostavno samo-konfigurirali in prilagodili na nove situacije.

Promet. Zelo pomemben segment za uporabo umetnih vizualnih sistemov so tudi aplikacije v prometu. Na cesti pogoji ne bodo nikoli dovolj predvidljivi in vizualna informacija bo vedno odločilna pri odločitvah. Izdelanih je bilo že precej klasičnih aplikacij računalniškega vida, ki znajo samodejno slediti avtomobile na sliki, šteti promet in delati razne druge statistike. Aplikacije obogatene s prvinami umetnega spoznavnega vida pa poleg tega tudi poskušajo razumeti dogajanje na cesti in predvideti prihodnje dogodke. Tako lahko s spremljanjem vozil, ki vozijo v bližini nekega avtomobila, uspešno predvidijo njegovo prihodnje ravnanje ter na ta način tudi zaznajo nenavadno obnašanje na cesti. Sistemi, ki so nameščeni v notranjosti vozila, pa poskušajo zaznati neobičajno obnašanje voznika (npr. zaspanost), omogočajo varnejšo uporabo zračnih blazin (glede na starost oz. velikost osebe), ipd.

Športne igre. Podobno je tudi pri sistemih za spremljanje in nadzor športnih iger, recimo nogometa. Poznamo več sistemov, ki znajo mozaičiti slike z več kamer in jih združevati v enotno sliko, ki pokriva celotno igrišče. Taki sistemi znajo tudi uspešno slediti igralcem in žogi, računati razne statistike, itn. To so klasični problemi računalniškega vida. Pred precej drugačen problem pa se postavimo, če od sistema zahtevamo tudi interpretacijo igre. Če želimo, da sistem razume, kaj se dogaja, da ve, katera ekipa napada, kdaj je priložnost za gol, kdaj je bil dosežen zadetek in kdaj storjen prekršek. Če želimo torej sistem, ki bi lahko igral vlogo avtomatskega televizijskega komentatorja. Takšni problemi so precej bolj zahtevni in spadajo v domeno umetnega spoznavnega vida.

Iskanje po slikovnih podatkovnih zbirkah. Večina obstoječih sistemov za razpoznavanje objektov je sposobnih razpoznati samo tiste objekte, ki so bili predstavljeni v fazi učenja. Tudi sistemi za iskanje po slikovnih podatkovnih zbirkah so precej omejeni in so sposobni iskanja le tistih slik, ki imajo podobne nekatere enostavne značilnosti (recimo barvo, mogoče teksturo, ipd.). Ne obstajajo pa še metode, ki bi znale učinkovito kategorizirati predmete, ki bi znale recimo poiskati vse slike, na katerih so prikazani psi (ki so lahko zelo različni). Z razvojem takšnih metod bo omogočeno dejansko razumevanje in semantično označevanje slik in iskanje po slikovnih ter video podatkovnih zbirkah.

Film, TV in zabava. Na področju zabave se pojavljajo marsikateri možnosti uporabe umetnega spoznavnega vida – od sistemov za samodejno označevanje, indeksiranje in iskanje po video podatkovnih zbirkah, preko sistemov za detekcijo določenih dogodkov ali predmetov, do nudenja dodatnih informacij gledalcu (samodejno pridobljenih v realnem času med neposrednim prenosom).

Mobilna robotika. Mobilni roboti potrebujejo za samonavigacijo dovolj dober in hiter sistem za samostojno ugotavljanje trenutnega položaja in določanje poti. V večini primerov morajo roboti pri tem upoštevati tudi (ali celo predvsem) vizualno informacijo. Učinkovito se morejo izogibati oviram, dinamično prilagajati svoj zemljevid okolja in se ustrezno odzvati na nepredvidene dogodke. Takšni roboti so zelo uporabni v človeku neprijaznih in nevarnih okoljih.

Kognitivni asistenti. Glavne karakteristike spoznavnih sistemov – robustnost, prilagodljivost in prijaznost, so idealne lastnosti za kognitivnega asistenta – umetnega agenta za pomoč pri vsakdanjih opravilih. Takšne pomoči so najbolj potrebni invalidi in ostareli. In ker populacija Evrope postaja vse starejša, lahko pričakujemo, da se bo trg s takšnimi in podobnimi osebnimi pomočniki zelo povečal.

Raziskave vesolja in vojaški sistemi. Umetni spoznavni sistemi so nujno potrebni pri raziskavah vesolja; avtonomni mobilni roboti se morajo namreč inteligentno gibati po popolnoma neznanem in nepredvidljivem okolju in se samostojno ter hitro odzvati na nepredvidljive dogodke. Seveda pa so avtonomni in decentralizirani inteligentni sistemi, sposobni samostojnega odločanje v neznanem, nepredvidljivem in spreminjajočem se okolju, izrednega pomena tudi za vojsko.

4. KONKRETNI PRIMER: NA VIDU TEMELJEČ POLOŽAJNI SISTEM

V tem razdelku predstavljamo konkretni primer uporabe metod umetnega spoznavnega vida na primeru aplikacije za avtomatsko lokalizacijo v urbanem okolju na osnovi vizualne informacije, ki ga razvijamo v okviru projekta MOBVIS [6]. Namen projekta je uporabiti tehnologije s področja umetnega spoznavnega vida in sorodnih znanosti na mobilnih napravah in jih narediti bolj uporabne in prijazne uporabniku. Iz narave mobilnih naprav sledi, da niso namenjene samo uporabi v zaprtih prostorih ampak naj bi spremljale uporabnika povsod, kjer se le-ta giblje, to je na ulici, letališču, na sestankih in

še marsikje drugje. Sistem, ki ga gradimo, bo moral odslikovati vso kompleksnost okolja, ki nas obdaja. Človek si v tem okolju pomaga tako, da združuje informacije iz različnih čutil in svojo sposobnost predvidevanja in odločanja. Podoben pristop smo uporabili tudi v okviru projekta MOBVIS, znotraj katerega združujemo tehnologije umetnega spoznavnega vida, globalni položajni sistem (GPS), položajne sisteme na osnovi radijskih oddajnikov, inercialne senzorzje, digitalne zemljevide in metode umetne inteligence.

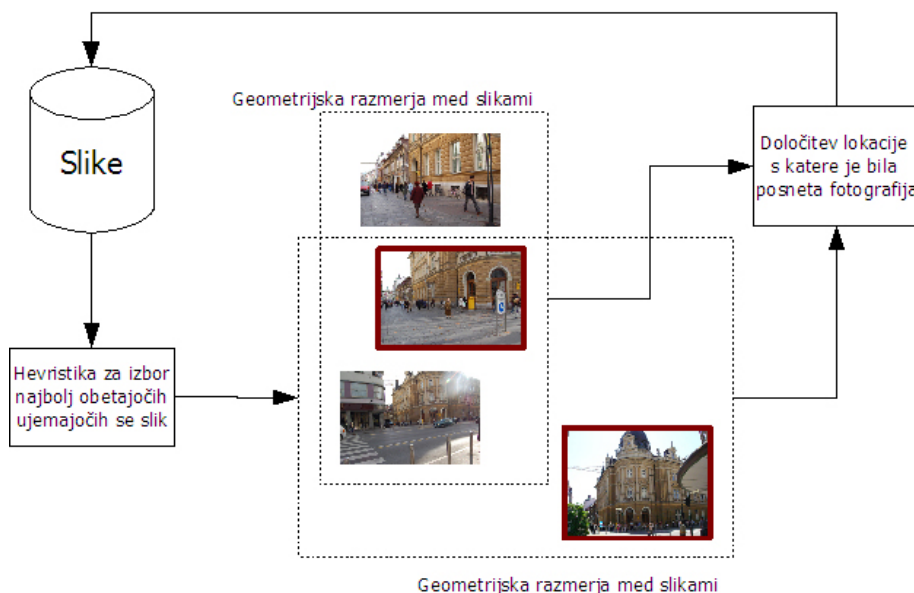
Podobno kot pri človeku bo tudi pri mobilnih sistemih, ki jih gradimo v okviru projekta MOBVIS, poglavitni vir informacij o uporabnikovi okolici vizualna informacija. Na podlagi vizualne informacije je možno ugotoviti ne samo to, kje se uporabnik nahaja, to je njegov geografski položaj, ampak tudi kaj je v njegovi okolici in posledično v kakšnem kontekstu se trenutno nahaja. To bo omogočilo gradnjo mobilnih aplikacij, ki so sedaj le stežka uresničljive, saj zahtevajo veliko informacij o uporabnikovem okolju, ki jih lahko trenutno zagotovi le uporabnik sam, preko omejenega uporabniškega vmesnika mobilnih naprav. Predstavljamo si mobilni telefon, ki avtomatično poveča glasnost zvonjenja, ko ugotovi, da je uporabnik na ulici polni ljudi in ga avtomatično utiša na koncertu klasične glasbe ali sestanku. Predstavljamo si turista, ki ga mobilna naprava vodi po zanimivem mestu, kot je prikazano na Sliki 1. Predstavljamo si spletni dnevnik, ki avtomatično beleži pomembne stvari iz življenja svojega lastnika. Vse te in še mnogo drugih aplikacij je možno razviti, v kolikor ne samo zajamemo sliko uporabnikove okolice, ampak tudi razumemo kaj se na sliki nahaja in dogaja.



Slika 1: Prototip na vidu temelječe mobilne storitve: sposobnost dojemanja turista opremljenega z dlančnikom se poveča, če videnim predmetom (npr. znamenitostim) dopolnimo pomen z dodatnimi informacijami.

Projekt MOBVIS smo razdelili na dva dela. V prvem delu, ki je že pripeljal do delujočega prototipa, smo razvili sistem, ki omogoča določanje geografskega položaja na osnovi posnete fotografije. V drugem delu načrtujemo razviti sistem, ki bo omogočal avtomatsko prepoznavanje statičnih predmetov, kot so npr. zgradbe, spomeniki, ulično pohištvo in drugo, in dinamičnih predmetov kot so npr. pešci, vozila in živali. Na podlagi poznavanja predmetov, zgradb, vozil, oseb in ostalih stvari v uporabnikovi okolici, bo sistem lahko sklepal o kontekstu v katerem se uporabnik nahaja in uporabniku ponudil informacije in storitve, ki so v tem kontekstu najbolj primerne.

Prototip na vidu temeljčnega položajnega sistema, ki smo ga razvili, omogoča določanje geografskega položaja brez uporabe posebne opreme, saj zadostuje že navadni digitalni fotoaparati. Določanje položaja se izvede tako, da se uporabnikova fotografija posneta z neznanе lokacije primerja s shranjenimi fotografijami istega prizora posnetimi z znanih lokacij. Nato se s pomočjo matematičnih metod [7] izračuna geografski položaj s katerega je bila posneta uporabnikova fotografija. Skica sistema je prikazana na Sliki 2.



Slika 2: Skica sistema za določanje geografskega položaja. Za obrabljeni slike ne poznamo točnega geografskega položaja s katerega sta bili posneti, medtem ko smo neobrabljeni slike posneli z znanega položaja. Na podlagi geometrijskih razmerij med slikami lahko z uporabo matematičnih metod izračunamo geografski položaj s katerega sta bili posneti obrabljeni slike.

5. ZAKLJUČEK

V tem članku smo poskušali opisati novo znanstveno disciplino, ki se poraja v zadnjih letih - umetni spoznavni vid. Izpostavili smo nekatere značilnosti sistemov umetnega spoznavnega vida, po katerih se razlikujejo od klasičnih sistemov strojnega in

računalniškega vida, ter te razlike prikazali tudi na izbranih primerih aplikacij. V tem kontekstu smo tudi predstavili aplikacijo za avtomatsko lokalizacijo v urbanem okolju na osnovi vizualne informacije.

Naj ta članek zaključimo z opisom zelo ambicioznega a še precej oddaljenega cilja raziskovalcev s področja umetnega spoznavnega vida in spoznavnih sistemov nasploh - zgraditi pravega kognitivnega asistenta. Tak asistent naj bi bil sposoben samostojno raziskati okolico in zgraditi ustrezen zemljevid, po katerem bi se v prihodnje znal orientirati in načrtovano gibati po prostoru. Ravno tako bi se bil sposoben naučiti ter nato razpoznavati in kategorizirati predmete ter razumeti njihove namene in funkcije. Znal bi razpoznavati in razumeti akcije v njegovi okolici, ter se te akcije naučiti in ponoviti. Tak asistent bi znal verbalno in neverbalno komunicirati s ljudmi in drugimi roboti. Zaznaval bi nove situacije in se ustrezno odzval nanje. Neprestano bi se učil in posodabljal svoje znanje ter sposobnosti in temu ustrezno prilagajal svoje vedenje.

Čeprav ljudje vsa ta opravila (bolj ali manj) z lahkoto opravljamo, je z znanstvenega in inženirskega vidika zgraditi tak umeten sistem zelo zahtevna naloga. Raziskave na področju umetnih spoznavnih sistemov in še posebej umetnega spoznavnega vida nas peljejo v to smer. In čeprav je pot do končnega cilja še dolga, pa v vmesnem času nastajajo in se razvijajo mnoge komponente, ki so že in bodo še našle pot do takojšnje uporabe v različnih praktičnih aplikacijah.

LITERATURA

1. D. Forsyth and J. Ponce (2003), *Computer Vision A Modern Approach*, Prentice Hall.
2. (1995), *Slovar slovenskega knjižnega jezika*, DZS, Ljubljana.
3. D. Vernon et al. (2005), *A research roadmap of cognitive vision*. Technical report, ECVision, http://www.ecvision.org/research_planning/Research_Roadmap.htm .
4. <http://cognitivesystems.org>
Cognitive Systems for Cognitive Assistants – CoSy.
5. <http://www.cordis.lu/ist/cognition>
Information Society Technologies, Directorate E, Unit E5 – Cognition.
6. <http://www.mobvis.org>
Vision Technologies and Intelligent Maps for Mobile Attentive Interfaces in Urban Scenarios – MOBVIS.
7. R. Hartley and A. Zisserman (2003), *Multiple View Geometry in Computer Vision*, Cambridge University Press, Second Edition.