

Senzo-motorický prístup k poznávaniu scény

Andrej Lúčny

Katedra aplikovanej informatiky, Fakulta matematiky, fyziky a informatiky, univerzita Komenského
Mlynská dolina, 842 48 Bratislava
lucny@fmph.uniba.sk

Abstrakt

Nevýhody tradičného prístupu percepcia-kognícia-akcia sa už dlhé roky snaží prekonať prístup založený na predstave, že kognícia je neoddeliteľná od percepcie a akcie, prekrýva sa s nimi. Následkom toho nemožno oddeliť ani samotnú percepciu od akcie. V ostatnom čase tento prístup poskytol určitý prínos pre metódy poznávania scény, ktoré uviazli v snahe vyriešiť tento problém výlučne metódami spracovania obrazu. Živé bytosti však disponujú schopnosťou upraviť scénu a hlavne svoje postavenie v nej, takže niektoré ťažké problémy môžu vyriešiť tak, že ich obídu. Ukazuje sa, že pri správnej postupnosti senzomotorických krokov dostane kognitívny modul poznávania scény správne parametre.

1 Úvod

Čo chýba panáčikovi z Fig. 1, aby sa dostal ku svojej kamarátke? Nič, stačí mu kráčať tak, ako bežne zvyknutý vo svojej komnate, len musí vedieť, kade má ísť. Len to „len“ tu robí problém.

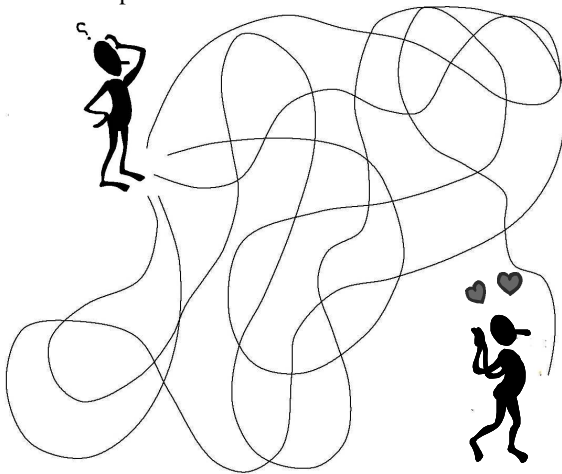


Fig. 1. Problém vývoja umelého kognitívneho systému

Myslím si, že predmet výskumu v oblasti tvorby umelých kognitívnych systémov má obdobný charakter. Pokiaľ

vieme kráčať, treba vedieť už „len“ postupnosť správnych krokov. Samozrejme „kráčanie“ tu nie je maličkosťou, zodpovedá určitej architektúre, ktorá umožňuje inkrementálny vývoj umelého systému a to dáva viac možností, podobne ako sa panáčik môže plaziť, chodiť po štyroch, kráčať alebo utekať. Ale v princípe nie je odkázaný na vynájdenie nejakého špeciálneho princípu, ako je napríklad schopnosť preletieť. Tak mi to pripadá, že väčšina pokusov o tvorbu umelých kognitívnych systémov mala práve túto ambíciu – nie nerealizovateľnú, ale ignorujúcu prízemnejšie možnosti.

Keď sa ale rozhodneme miesto lietania kráčať – ktorou cestou ísť? Tu našťastie máme jednu významnú nápovedu a tou je možnosť pozorovať vývin prírodných kognitívnych systémov. A z tohto pozorovania si môžeme vziať určité ponaučenie. Aj to sa však dá urobiť dobre a zle. Každé ponaučenie musí byť nevyhnutne zjednodušujúce a každé zjednodušenie so sebou nesie riziko, že odignorujeme niečo podstatné.

Tento príspevok poukazuje práve na jedno z takých zjednodušení a ponúka alternatívu. Konkrétne ide o to, že percepciu a na ňu nadväzujúce kognitívne procesy sa často snažíme chápať oddelenie od akcií ako samostatný modul. To by samo o sebe bolo v poriadku – koniec koncov aj pri živých bytostiach možno určit anatomicky samostatné jednotky ktoré tieto procesy realizujú – chyba však nastáva pokiaľ sa snažíme tento pohľad preniesť aj do postupného obohacovania a zlepšovania tohto modulu. Tá chyba spočíva v tom, že postupný vývin tohto modulu možno ani nie je možný bez ostatných modulov, ktoré sa v prírode vyvíjali k dokonalejším formám spolu s ním.

Pritom ťažiť z poznatkov o vývine detí, môžu v oblasti tvorby software pre umelé kognitívne systémy v prvom rade také architektúry, ktoré podporujú inkrementálny vývoj. Inkrementálny vývoj prvý vyzdvihol Rodney Brooks vo svojej subsumpcnej architektúre [2] v polovici 80-tych rokov minulého storočia. O niečo modernejšiu inkrementalitu mi umožnil prístup na báze multi-agentových systémov [5], ktorý som zhrnul v tzv. architektúre Agent-Space [7]¹.

¹ www.agentspace.org

2 Senzomotorický prístup

Prvý, kto prišiel s myšlienkou, že myslenie nepovstáva len z vnímania, ale aj zo senzomotorickej činnosti, bol zrejme Jean Piaget [11]. Odvodil ju z experimentov ako je napríklad „The A not B error“ (Fig. 2): Dieťaťu je prezentovaná scéna s dvomi skryšami. Do jednej zo skryš schovávame pre dieťa zaujímavý predmet a demonštrujeme, že odkrytím skryše sa vieme predmetu opäť zmocniť. Opakujeme to, pokiaľ sa dieťa nenaučí predmet odkryť. Potom predmet schováme do druhej skryše. Napriek tomu, že dieťa vidí kde predmet schovávame, dieťa mladšie ako 10 mesiacov, nebude hľadať schovaný predmet tam, kde sme ho schovali, ale tam, kde ho predtým nachádzalo. Piaget z tohto pokusu urobil záver, že dieťa nie je schopné predmet reprezentovať nezávisle od motorickej akcie, ktorou sa ho zmocňuje. Obdobie vývinu dieťaťa, v ktorom sú jeho reprezentačné schopnosti takto obmedzené, nazval senzomotorickým obdobím.



Fig. 2. Experiment „The A not B error“, (prevzaté z [13])

Podľa Lindy Smithovej [12] mal síce Piaget pravdu v tom, že prvé reprezentácie u dieťaťa sú senzomotorické, ale mýlil sa v predstave, že v neskorších vývinových fázach sú tieto reprezentácie nahradené dokonalejšími. Smithová predpokladá, že experimenty ako je „The A not B error“ odhaľujú najvzrušujúcejšiu črtu našich kognitívnych schopností, ktorá nás sprevádza celým životom. Senzomotorické obdobie dieťaťa je podľa nej špecifické len tým, že táto črta ľahko vypláva na povrch a môžeme ju tak zvonku pozorovať. Dokonca i v tomto období môžeme zabrániť vonkajšiemu prejavu chyby, pokiaľ medzi zmenou skryše zmeníme polohu dieťaťa (napríklad sediace postavíme, alebo stojace posadíme – pritom stále sediace alebo stále stojace dieťa chybu demonštruje) alebo zmeníme čokoľvek čo má súvis

s „rečou jeho tela“ (napríklad mu dáme na zápästie závažie).

Tento názor veľmi dobre zapadá do mojej predstavy (vychádzajúcej z prác Brooks [2] a Minského [9]), že myseľ je dokonalý systém poskladaný z nedokonalých vecí, ktoré vyplávajú na povrch pri rôznych klamoch [8]. Takisto sa mi pozdáva predstava, že základné stavebné jednotky mysle majú senzomotorický charakter (priekopníkom tohto prístupu v je našej domácej komunite Peter Kostelník, viď [6]) – mojimi slovami ide o tzv. reaktívnych agentov. Najdôležitejšou motiváciou je však pre mňa poukázanie na to, že akcie nie sú len následkom vnímania, ale že sa aktívne podieľajú na konštitúcii vnímania. Charron, Labbani-Igbida a Mouaddib v [3] získavali základ tvaru telesa v scéne z nekolízneho pohybu okolo tohto telesa, čo je výhodné nakoľko výstupy na motory sú výrazne jednoduchšie než vstupy z kamery.

Motorická zložka môže byť podstatná aj vtedy, keď sa nám na prvý pohľad zdá, že sa pozorovaný subjekt vôbec nehýbe. Aline Bompas a Kevin O'Regan v [1] napríklad ukázali, že motorika oko-hybných svalov má vplyv na vnímanie farieb. Testovaným osobám predkladali do pravej časti zorného poľa zelenú škvrnu a do ľavej červenú v rôznom poradí. Potom im ukazovali v oboch častiach žltú rôznych odtieňov. Pri normálnom pohybe očí testovanej osoby dochádzalo k tomu, že rovnaký vnem vyvolávali rôzne odtiene žltej – na pravej strane musela byť bližšie k zelenej a na ľavej k červenej. Pokiaľ boli oči fixované, k tomuto efektu nedochádzalo. Autori testu predpokladajú, že motorické dáta z oko-hybných svalov sú použité na kalibráciu sýtosti farieb.

Smithová v [12] sa zamerala na spôsob akým deti v nízkom veku manipulujú s objektmi v ich okolí v porovnaní s dospelými. Na základe analýzy videa z kamery umiestnenej na čele dieťaťa a dospelého objektívne preukázala, že dieťa si na rozdiel od dospelého aktívne umiestňuje predmety do zorného poľa tak, aby ho prakticky celé pokrývali. To mu má umožniť lepšie oddeliť daný predmet od ostatnej scény, uložiť si do pamäti jeho opis a následne tak eliminovať potrebu si ho prikladať k očiam, lebo zapamätané parametre vie použiť na to, aby ho rozpoznalo aj diaľky a v rámci zložitejšej scény. Toto pozorovanie je zaujímavé z toho hľadiska, že varuje inžinierov, ktorí sa zaoberajú rozpoznávaním scény, že snímanie obrazu z kamery a jeho následné spracovanie nedáva dostatočný materiálny základ na napodobnenie analogického rozpoznávacieho mechanizmu u človeka. Na správnu kalibráciu takéhoto systému môže byť nevyhnutný aj aktívny pohyb kamery, či dokonca manipulátor na umiestňovanie predmetov do zorného uhla kamery. Zo Smithovej experimentu by mohlo dokonca vyplývať, že je veľmi podstatné mať dve kamery s možnosťou ovládania ich zaostrenia. Pokiaľ si

totiž dieťa umiestni predmet blízko k očiam, ľahko ho odseparuje od zvyšku scény aj tým, že keď naň zaostrí zrak, je to jediný predmet, ktorý vidí ostro. Používanie zdanlivo dokonalejšieho vizuálneho systému na simuláciu zraku môže takto zmať simuláciu zrakového vnímania v zárodku – keďže aktívnou motorikou sa niektoré zdanlivé nevýhody dajú využiť. Motorické dáta okohybných svalov sú opäť zjavne jednoduchšie než obraz a môžu hrať nezastupiteľnú úlohu v povstávaní rozpoznávacieho mechanizmu.

3 Príklad: rozpoznávanie scény

Tento prístup demonštrujeme na jednoduchom príklade mobilného robota rozpoznávajúceho scénu (Fig. 3). Uvedený príklad nebude biologicky relevantný – na to je tento robot príliš primitívny – ale dúfam, že osvetlí myšlienku senzomotorického prístupu.

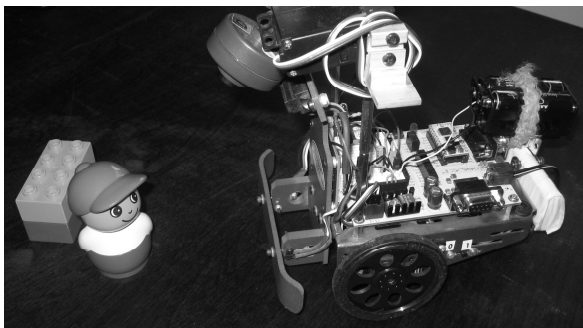


Fig. 3. Mobilný robot a objekty v scéne

Robot je vybavený (jednou) kamerou a dvojkolesovým podvozkom. Kamera je bezdrôtová a podvozok na diaľkové ovládanie cez bluetooth, takže obe máme pripojené k PC, na ktorom dokážeme bežať aj zložitejšie riadenie, založené v našom prípade na architektúre Agent-Space. Scéna je upravená tak, že obsahuje len svetlé objekty na čiernom podklade, aby pri kvalite použitej kamery bolo vôbec niečo možné oddemonštrovať.

Predstavme si, že robot vidí Fig. 4. Otázka je čo vlastne vidí: jeden objekt alebo dva? My vidíme dva, ale ako na to príde náš robot? Tradičný prístup v počítačovom videní vychádza z Guzmanovej metódy [4] a v princípe sa snaží určiť čo je na obraze vypuklé, čo duté a pod., následne spočítať hĺbku (vzdialenosť od kamery) pre každý pixel obrazu a na základe určitej logiky – preferujúc určité štandardné tvary objektov – rozhodnúť, či ide o jeden, či o dva objekty.



Fig. 4. Pohľad zpredu



Fig. 5. Pohľad z boku



Fig. 6. Segmentovaný pohľad zpredu

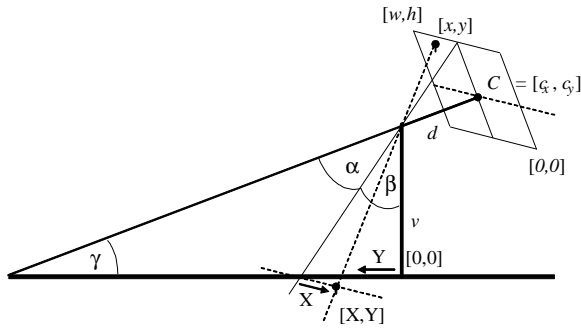


Fig. 7a. Parametre optického systému robota:

h – vertikálne rozlíšenie obrazu

w – horizontálne rozlíšenie obrazu

α - vertikálny uhol záberu

β - horizontálny uhol záberu

(tieto uhly sú v dokumentácii kamery alebo sa dajú ľahko odmerať pri vertikálnom natočení kamery)

$C = [c_x, c_y]$ - úbežný bod (zvyčajne $[w/2, h/2]$, dá sa taktiež odmerať pri horizontálnom natočení kamery ako priesečník vhodných rovnobežných priamok na obraze)

d - ohnisková vzdialenosť kamery v pixloch:

$$d = \frac{h - c_y}{\text{tg} \alpha} = \frac{w - c_x}{\text{tg} \beta}$$

γ - uhol sklopenia kamery

v – výška umiestnenia kamery

x, y – pozícia na obraze

X, Y – rovina po ktorej sa pohybuje robot (pôdorys)

Úlohu je spočítať $[X, Y]$ z $[x, y]$.

(rotáciu kamery okolo osi jej smerovania neuvažujeme.)

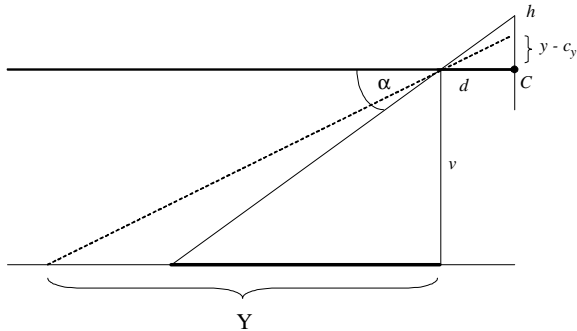


Fig. 7b. Keby kamera nebola sklopená, súradnicu Y ľahko spočítame:

$$Y = \frac{vd}{y - c_y}$$

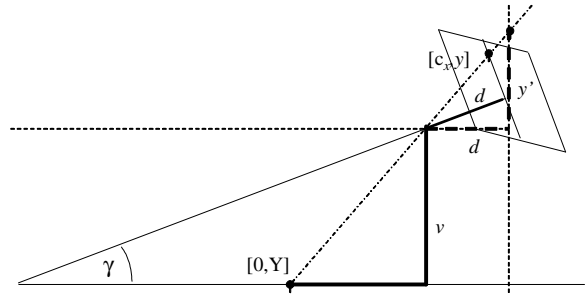


Fig. 7c. Keďže kamera je sklopená, zistíme kam by bol bod na obraze otočením zobrazený a pozrieme odtiaľ podľa 7b:

$$y' = d \frac{d \sin \gamma + (y - c_y) \cos \gamma}{d \cos \gamma - (y - c_y) \sin \gamma} \quad Y = \frac{vd}{y'}$$

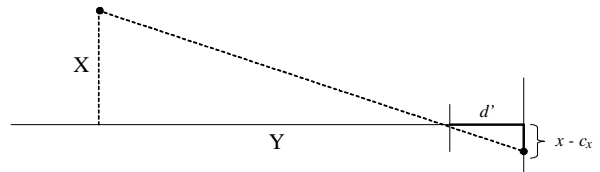


Fig. 7d. Z pohľadu zhora potom spočítame X

$$d' = d \cos \gamma \quad X = Y \frac{x - c_x}{d'}$$

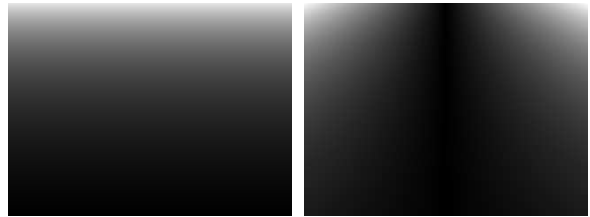


Fig. 8. Graficky znázornené súradnice Y (vľavo) a X (vpravo) reprezentujúce zložky vzdialenosti bodov na pôdoryse zobrazených na obraz kamery robota

Senzomotorický prístup na to ide inak. Predpokladá, že v momente keď robot prvý raz čelí pohľadu z Fig. 4, nie je schopný vydolovať z neho úplnú informáciu. Je však schopný na scénu aplikovať málo účinnú ale rýchlu segmentáciu (metódou 2-means clustering) (Fig. 6), ktorá mu napovie, že niečo na scéne je. Robot taktiež vie zhruba určiť vzdialenosť videného, predpokladajúc, že sa videné nachádza na zemi a vediac výšku a orientáciu kamery (Fig. 7, 8). Na základe predstavy o relatívnom umiestnení robota a objektu na pôdoryse scény sa robot postaví do určitej vzdialenosti čelne k predmetu a spustí motorickú akciu za účelom bližšieho preskúmania objektu.

Táto motorická akcia je načisto „nadrôtovaná“ a obchádza predmet z pravej strany, pričom tento pohyb je prerušovaný otáčaním sa k objektu a späť (Fig. 9). Dôležité je, že motorická akcia je dostatočne krátka na to, aby sa dalo zo sekvencie príkazov na motory s dostatočnou presnosťou viesť po akej trajektórii robot šiel.

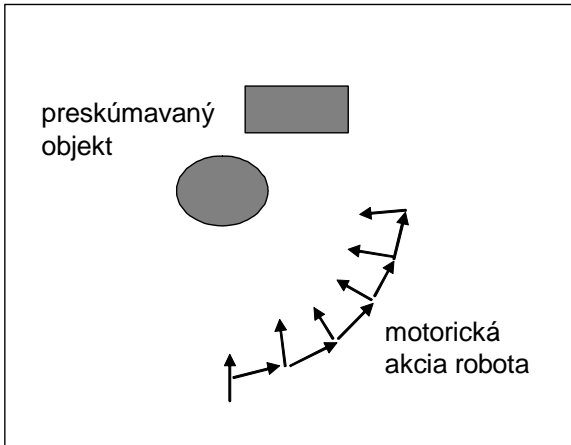


Fig. 9. „Nadrôtovaná“ motorická akcia robota za účelom preskúmania objektu



Fig. 10. Pozorovaný objekt z rôznych pohľadov

Pri tomto pohybe si robot segmentuje obraz z rôznych pohľadov (Fig. 10) – z tej istej výšky, len z rôznych uhlov. Všíma si pritom viac podlahu ako samotný objekt, lebo ak v nejakom bode obrazu vidí podlahu, vie, že po celej priamke od kamery do tohto bodu sa objekt nenachádza. Túto priamku môže teda použiť ako rezací lúč, ktorým do modelu 3D priestoru vyrezáva skúmaný objekt. (Parametre tejto priamky má vzhľadom na pevné pripevnenie kamery vypočítané dopredu podľa Fig. 8). Keďže robot predpokladá, že motorickú akciu dokáže vykonať presne – takže vie relatívne polohy miest z ktorých sa bude na objekt pozeráť, zlúčiť jednotlivé pohľady nie je ťažké (Fig. 11). Takáto 3D rekonštrukcia má pochopiteľne svoje chyby (veľké problémy napríklad robí tieň čapice rekonštruovaného LEGO panáčika – v 3D sa to prejaví tým, že má diery v hlave). (Existujú samozrejme ďaleko rafinovanejšie metódy 3D rekonštrukcie, pričom na to stačia dva obrázky [14]. Nič menej ide o nesmierne ťažký

problém – po matematickej i výpočtovej stránke.) Počas tejto rekonštrukcie robot dosiahne aj taký bočný pohľad (Fig. 5) pri ktorom je jasné, že videné predmety sú dva. Výsledkom motorickej akcie sú teda dva 3D objekty (Fig. 12).

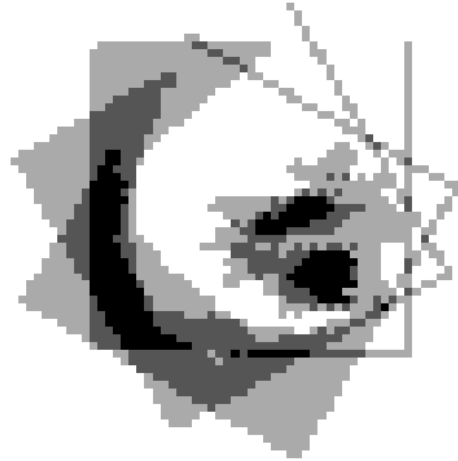


Fig. 11. Rekonštrukcia z rôznych pohľadov (jeden vodorovný rez, pohľad zhora, zobrazené sú i „duchy“: väčší – na obrázku zľava - je výsledkom úzkeho záberu kamery, menší pásik vzadu – na obrázku sprava – je dôsledkom toho, že u vyšších rezov padá horizont do obrazu a nad horizontom nemôže byť podlaha).

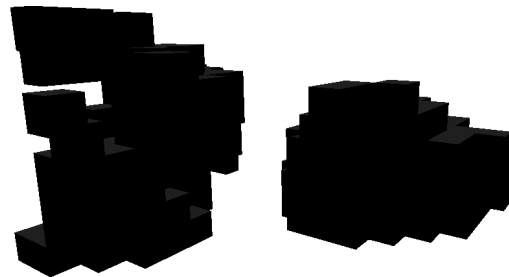


Fig. 12. Zrekonštruované 3D objekty. (Ťažko by sme spoznali v ľavom panáčika a v pravom kváder. Na lepší výsledok by boli potrebné lepšie vedomosti i hardware. Výhoda je, že tento výsledok vieme dostať v reálnom čase, čo je pre použitie na robotovi nevyhnutná podmienka.)

Zrekonštruované objekty by sme – teoreticky (keby mali lepšiu kvalitu) – vedeli použiť na renderovanie objektu z ľubovoľného pohľadu. (Nám pritom stačí uvažovať pohľady z pozície kamery na robote.) Pri ďalšom pohybe v scéne potom robot pri opätovnom pohľade na dané objekty skúsi najprv vyrenderovať objekty, ktoré pozná a prirovnať ich

k videnému. Pokiaľ sa mu podarí dosiahnuť uspokojivú zhodu, predpokladá, že vidí tie objekty, ktoré už videl. Je ho možné samozrejme ľahko oklamať tým, že mu do scény dáme jeden objekt, ktorý sa podobá na dva už videné (Fig. 13). Nič menej rovnako možno oklamať aj človeka.

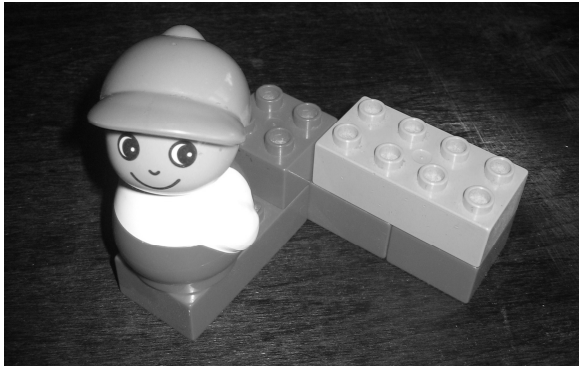


Fig. 13. Objekt, ktorý po „naučení“ robota vyvolá klam.

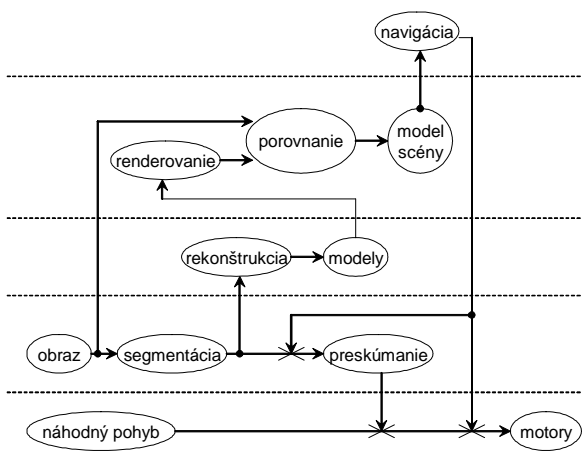


Fig. 14. Architektonický návrh riadenia robota (krížik označuje zásah do činnosti nižšej – „evolučne“ staršej – vrstvy z vyššej – „evolučne“ novej – vrstvy. Tento zásah je v zmysle [2] i [7] realizovaný ovplyvnením komunikácie medzi časťami nižšej vrstvy)

Na záver tohto príkladu poznamenávam, že nebol reálne uvedený do plnej činnosti, odskúšané boli len jednotlivé techniky (v rástane ich vykonateľnosti v reálnom čase). Cieľom príkladu je hlavne demonštrácia myšlienky senzomotorického prístupu. Uvedenie do činnosti vyžaduje – mimo množstva „drobných“ technických problémov – implementovať tento systém

inkrementálnym spôsobom. Jeden z možných návrhov inkrementov uvádza Fig. 14.

4 Záver

V našom príspevku sme uviedli myšlienku senzomotorického prístupu k tvorbe umelých kognitívnych systémov. Základným predpokladom tohto prístupu je, že kognitívny systém sa vnútorne skladá z elementárnych jednotiek pozostávajúcich zo senzorickeho vnímania a motorickej akcie. Dôsledkom je, že vyššie úrovne senzorickeho vnímania nemožno inkrementálne vyvinúť bez súčasného vývoja motorických akcií. Toto zdánlivé obmedzenie môže byť výhodou – implementovať musíme zdánlivo viac, nič menej jednotlivé implementačné kroky sú jednoduchšie.

Senzomotorický prístup neprináša nič svetoborného okrem poukázania na skutočnosť, že pri zvolení správnej vývojovej cesty, rešpektujúc súbežný rozvoj senzorickejších a motorických aktivít, nič svetoborného – pravdepodobne – ani nepotrebujeme.

PodĎakovanie

Tento príspevok vznikol s podporou grantu VEGA 1/0280/08 *Aplikácia multi-agentovej modularity pre tvorbu riadiacich systémov a počítačových modelov.*

Literatúra

- [1] Bompas, A. – O'Regan K.: *More evidence for sensorimotor adaptation in color perception.* Journal of Vision 6, 2006, pp. 145-153
- [2] Brooks, R.: *Cambrian Intelligence*, The MIT Press, Cambridge, Mass., 1999.
- [3] Charron, C. - Labbani-Igbida, O. – Mouaddib, E. M.: *Sensory - motor approach to shape recognition.* Perception 32, ECVF Abstract Supplement, 2003
- [4] Guzman, A.: *Computer recognition of three dimensional objects in a visual scene.* TR-59. MIT. Cambridge, 1968
- [5] Kelemen, J.: *Multiagent symbol systems and behavior-based robots.* Applied Artificial Intelligence 7, (1993), 419-432.
- [6] Kostelník, P.: *Získavanie konceptov zo senzorickejších dát pri riadení mobilného robota.* Kognícia, umelý život a počítačová inteligencia (CALCI 2003), Stará lesná, 2003.
- [7] Lúčny, A.: *Building Complex Systems with Agent-Space Architecture.* Computing and Informatics, Vol. 23 (2004), pp. 1001-1036
- [8] Lúčny, A.: *Modelovanie kognitívneho zlyhania multiagentovým systémom so stigmergickou komunikáciou.* In: Kognice a umelý život II

- (Kelemen, J. – Kvasnička, V., eds.), FPF SU, Opava, 2002, pp. 119-132
- [9] Minsky, M.: *Society of Mind*. Simon & Schuster, New York, 1986
- [10] Misra, N. - Choe, Y.: *Shape Recognition Through Dynamic Motor Representations*. In: *Neurodynamics of Cognition and Consciousness* (Perlovsky, L. – Kozma, R. eds.) Springer-Verlag, Berlin Heidelberg, 2007
- [11] Piaget, J., Inhelderová, B.: *Psychológia dieťaťa*, SOFA, Bratislava 1993
- [12] Smith, L.: *From Fragments to Geometric Shape: Changes in Visual Object Recognition Between 18- and 24-Months*. *Current Directions in Psychological Science*. Vol. 18, No. 5, 290-294
- [13] Smith, L.: Piaget was right. www.iub.edu/~cogdev
- [14] Sonka, M. – Hlaváč, V. – Boyle, R.: *Image processing, analysis and machine vision*. 3rd edition, Thomson Learning, Toronto, 2007