

Senzo-motorický prístup k zameriavaniu pozornosti

Andrej Lúčny

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

Abstrakt

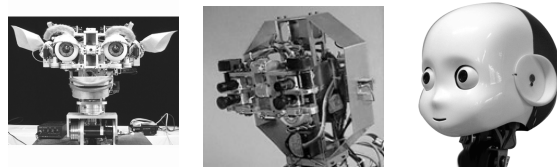
Zaoberáme sa významom očných pohybov pre zameriavanie pozornosti. Na rozdiel od väčšiny robotických hláv, ktoré majú oči vsadené pevne a otáčajú celou hlavou, skúmame, či by malo prínos pohybovať očami, či už z hľadiska ich orientácie na pozorovaný objekt, alebo ich zaostrenia. Naším cieľom nie je konkurovať prepracovaným metódam počítačového videnia, ktoré v istých ohľadoch prevyšujú i ľudské schopnosti, ale skúšať jednoduchšie, rýchlejšie a biologicky inšpirované metódy. Za cieľ zamerania pozornosti považujeme vyčlenenie objektu zo scény, ktoré je potrebné pre jeho rozpoznanie a to v prvom rade na získanie jeho reprezentácie.

1 Úvod

Väčšina humanoidných robotov má oči pevne vsadené do hlavy. Ešte i taký Kismet (obr. 1a), ktorý má vyjadrovať emócie pozerá len uprene pred seba. Je to preto, že geometria takto upretého pohľadu je jednoduchšia, a dobre sa pri nej odhadujú vzdialenosti. Niektorí z týchto robotov dokážu modelovať veľmi dobrý 3D model priestoru, v ktorom sa nachádzajú. Používajú však na to pre ľudí neviditeľný trik, že infračerveným svetlom nasvecujú scénu prúčkami rôznej veľkosti a rekonštrukciu vykonávajú z informácie o deformácii týchto prúčkov na obraze. Pri tomto procese je vhodné, ak sa ani nepohnú. Netreba dodávať, že takýto spôsob vnímania prostredia – hocako účinný – je veľmi vzdialený spôsobu, akým sa dívame my, ľudské bytosti.

Pravdepodobne prvým robotom, ktorý otáčal očami zo strany na stranu bol COG (obr. 1b), z dostupných videí sa však zdá, že smery pohľadu jeho „očí“ boli vždy rovnobežné. Z modernejších robotov, ktorí majú samostatne ovládateľné „oči“, spomeňme iCub (obr. 1c). V oboch prípadoch bolo zámerom vyvíjať riadenie robota postupne prostredníctvom napodobňovania fáz psychologického vývinu u detí, vychádzajúc pritom hlavne z prác J. Piageta [10]. Konečným cieľom pre robota COG bolo zvládnuť imitáciu [1]. Tento robot vedel sledovať pohybujúce sa postavy, zamerať pozornosť na ich oči, nájsť polohu zrenice v očnej bulve, zistiť kam sa dotýčná osoba díva

a zamerať pozornosť na dané miesto. Ako je vidno je obrázku, mal miesto jedného „oka“ dve – jedným bola kamera so širokým uhlom záberu a druhým kamera s veľmi malým uhlom záberu. Takže keď prvým okom niečo uvidel, otočil hlavou a očami daným smerom, aby sa mohlo aj druhé oko pozrieť a nasnímať zaujímavé detaily [11]. Pôsobivý opis interakcie s COGom zanechal vo svojej knihe „Typy myslí“ filozof D. Dennet [3]. Veľmi jasne z tohto prípadu vidno, že vhodná sada trikov, hierarchicky štruktúrovaných do riadiaceho systému. Čo sa týka iCub-u, väčšinou sa používa na experimenty s impulznými neuronovými sieťami [8], naše pokusy s ním teda ťažko porovnávať.



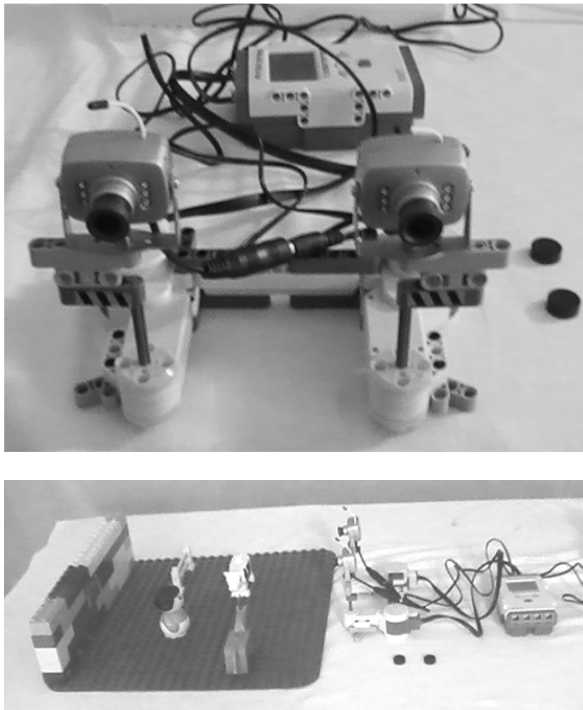
Obr. 1. zľava doprava: a) KISMET, b) COG, c) iCub

Biologický význam motorickej zložky vnímania, vyzdvihujú práce Piagetových nasledovníkov. L. 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. Snímala len priestor pred očami, takže žiaľ neskúmala, aké sú rozdiely sú v zameriavaní pozornosti na predmet medzi dospelým a dieťaťom. Nie je jasné, či napríklad zaostrovanie na predmet hrá významnú úlohu (t.j. či dieťa vyextrahuje objekt z prostredia na základe toho, že keď ho má pred očami, tak ten jediný vidí ostro), alebo stačí, že blízky predmet okupuje prevládajúcu časť obrazu. Bolo by zaujímavé, keby nejaký vývinový psychológ podrobil túto otázku skúmaniu.

2 Experimenty

Pre naše účely sme zostrojili pomerne jednoduchý experimentálny prostriedok, keďže stavať robotický

hlavu je nad naše schopnosti i možnosti. Umožňuje však nasnímať sériu obrázkov, s ktorými sa dá ďalej hrať. Pozostáva z dvojice kamier, ktoré sa môžu otáčať vo vodorovnom smere pomocou motorov ovládaných z LEGO kocky (obr. 2).



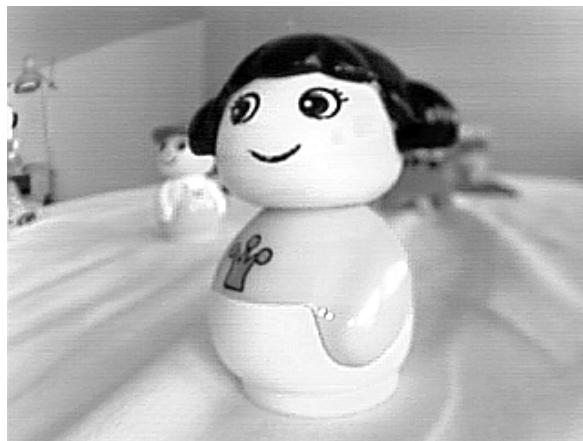
Obr. 2. Použitý experimentálny prostriedok

Z technického hľadiska sme použili OpenCV na spracovanie obrazu z kamier, remoteNXT na ovládanie motorov a DevC++ ako vývojové prostredie. Je len zhoda okolností, že bol systém bezdrôtový. Použili sme kvalitnú kameru CCTV s veľkou šírkou záberu vyvinutú pre sledovanie vtákov vo vtáčích búdkach. Zaostrenie objektívu (*gain*) sa však na nej dá robiť len ručne, takže naše experimenty neboli plne automatizované. Podobne sme nemali v systéme tretí motor, ktorý by hýbal zo strany na stranu celou sústavou, miesto sme toto natočenie simulujúce pohyb hlavy robili ručne.

Rozdiel medzi takýmto systémom a ľudským okom je samozrejme obrovský. Ale niekde začať musíme. S týmto systémom sme vykonali nasledovné experimenty:

Experiment 1. Umiestnili sme objekt do tesnej blízkosti kamery a menili sme jej zaostrenie tak, aby bol objekt ešte pomerne ostrý, ale objekty v pozadí sa zahmlievali. Vieme teraz z nasnímanej dvojice obrazov vyextrahovať objekt a akou metódou? Na obr. 3 vidíme, že pomocou

nastavovania ostrosti sa dá s obrazom výrazne pracovať. Veľmi dobre tomu napomáha algoritmus zvýraznenia hrán (*sharpen*).

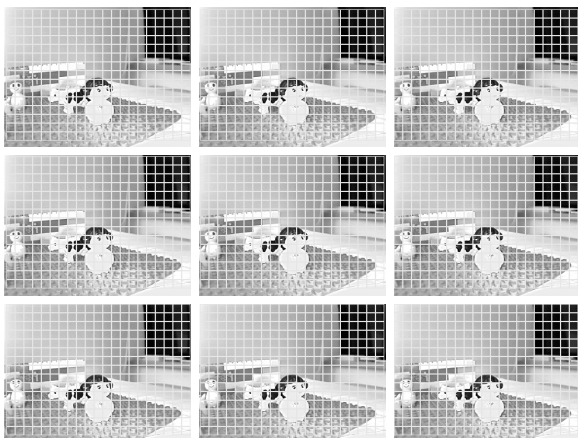


Obr. 3. Záber z jedného „oka“ s rôznym zaostrením (hore je kamera zaostrená na dievčatko, dolu na koníka)

Existuje niekoľko algoritmov na detekciu rozmazania obrazu (*blur detection*) [14] – hoci nie sú voľne dostupné. Keby sme na obraz naniesli sústavu prekrývajúcich sa regiónov (obr. 4.) a na každom z nich vyhodnotili pomocou tohto algoritmu rozmazanie a spočítali minimum pre každý bod, zrejme by to poskytlo celkom dobrý podklad na extrakciu objektov zo scény, ktorá nespojí objekty, ktoré sú na obraze blízko seba, avšak v skutočnosti sa nachádzajú v rôznej vzdialenosti.

Experiment 2. Nasmerovali sme jedno oko systému tak, aby bolo zamerané na určitý objekt v scéne a druhým okom sme plynule pohybovali okolo analogického zamerania. Dokážeme z takýchto obrázkov vyextrahovať videné objekty? Keď na seba nasnímané obrázky

položíme a vypočítame ich diferenciu (obr. 5), pre každý región na obraze (opäť uvažujeme mriežkovú sústavu prekrývajúcich sa regiónov na obraze – obr. 4) sa dá povedať pri akom smerovaní druhého oka dosahuje minimálnu diferenciu. Drobné odchýlky v polohe prikladaných obrazov pritom odstránime tým, že vyberieme polohu s minimálnou diferenciou v rámci určitej malej tolerancie. Pre každý región sa potom pozrieme, aká je preň distribúcia minimálnej diferencie podľa jednotlivých smerovaní. Zaujímajú nás také regióny, ktoré pre určitý smer obsahuje v tejto distribúcii významný výkyv. Regióny, ktoré dosahujú hodnoty blízke minimu v pre väčšinu smerovaní, zodpovedajú plochám scény. Ale tie z výkyvom zodpovedajú pozorovaným objektom. Každý objekt má pritom iné smerovanie, pri ktorom dosahuje minimum diferencie, čiže je to určitá cesta ako navzájom tieto objekty separovať. Vďaka príslušnosti každého bodu k sústave regiónov, môžeme optimálne smerovanie povedať pre každý bod a uskutočniť separáciu objektu z obrazu s presnosťou na pixle (musíme pravda opraviť nepresnosti dilatáciou a eróziou). Táto separácia je však málo účinná voči zákrytom (Obr. 6) – objekty stojacie v tom istom smere budú spravidla fantomicky spojené.



Obr. 4. Sústava prekrývajúcich sa regiónov. (mriežky sú vzájomne posunuté)

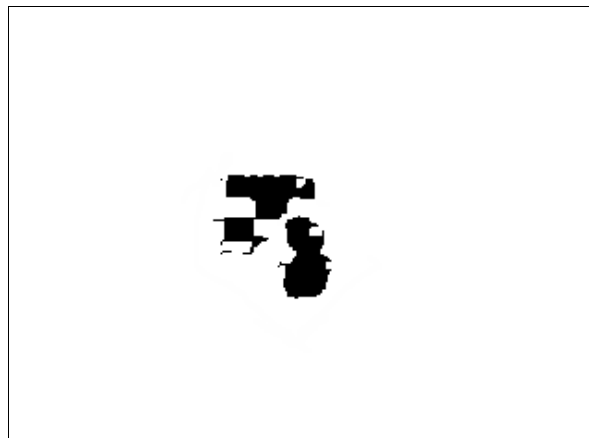
Na základe približnej identifikácie objektu na obraze by sa teraz dalo pokračovať tým, že obe oči nasmerujeme tak, aby bol objekt v strede obrazu. Zatiaľ sme to neskúšali. Domnievame sa, že by výsledok separácie z týchto dát mohol byť o niečo presnejší.

Tieto experimenty do istej miery naznačujú, že ako zaostrovanie, tak smerovanie očí je z hľadiska extrakcie objektov z obrazu dôležité. Prvé separuje objekty odlišujúce sa vzdialenosťou, druhé smerom umiestnenia. Prírodné polohovanie takýchto objektov

v 3D je tu založené na polárnych súradniciach, ktoré zodpovedajú motorickým stavom vizuálneho aparátu: smerovanie oka (potenciálne dva uhly) a zaostrenie objektívu. Našom prípade sme pritom smerovanie merali iba časom (po ktorý sa vykonáva pohyb kým sústava dospeje k danému smerovaniu), čiže v jednotkách, ktoré by sa nedali použiť pre umiestňovanie objektu v 3D modeli.



Obr. 5. Diferencia obrazov z jednotlivých „očí“ (pri pohľade na panáčika v strede)

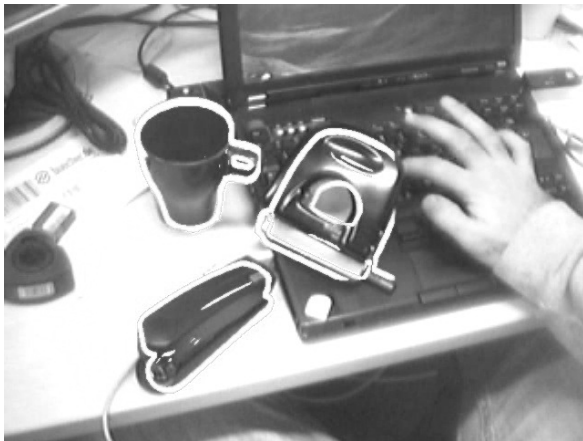


Obr. 6. Separácia objektu.

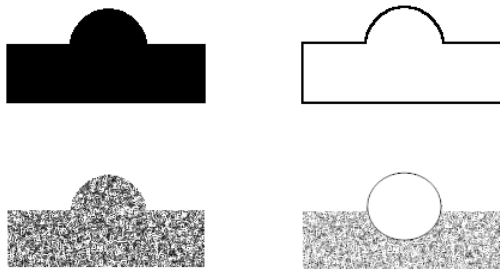
3 Rozpoznávanie objektov

Pokiaľ sa nám podarí vyextrahovať určitý objekt z prostredia – a to najlepšie z viacerých pohľadov (čo si samozrejme žiada motorickú akciu), môžeme využiť niektorú z metód rozpoznávania objektu podľa tvaru na vytvorenie jeho reprezentácie. Za obzvlášť sľubnú metódu v tomto smere považujeme DOT [5], ktorá je

založená na štatistickom spracovaní hrán. Pokiaľ táto metóda disponuje reprezentáciou objektu, je schopná rozpoznáť ho aj v rámci veľmi bohatej scény (obr. 7) a je schopná vrátiť jeho hranicu na obraze. Problémom je získanie tejto reprezentácie. Pri tradičnom použití je DOT oživená tým, že objekt je snímaný voči jednofarebnému kontrastnému pozadiu, z ktorého ide ľahko extrahovať. Pochopiteľne teda vyvoláva záujem o hľadanie prirodzenejších spôsobov extrakcie. Kvalita separácie dosiahnutá pri vyššie uvedených experimentoch nie je dostatočná na to, aby dokázala DOT naštartovať. Nič menej to neznamená, že sa principiálne mýlime (a že teda motorika nezohráva v separácii objektov významnú úlohu).



Obr. 7. Rozpoznávanie pomocou DOT (prevzaté z [5])



Obr. 8. Rozpoznanie fantomického objektu (kruhu). (Vľavo hore je pôvodný obraz a vedľa neho jeho hrany. Vľavo dole je pôvodný obraz zašumený, vpravo dole sú jeho hrany, prekryté rozpoznávaným objektom)

4 Inteligencia verzus omyly

Je pozoruhodné, že takýto systém rozpoznávania (t.j. DOT kŕmený zatiaľ neexistujúcou metódou separácie) je v praxi použiteľný a z hľadiska reálneho času rýchlo

pracujúci, hoci z hľadiska logiky v poriadku nie je. Môže pri ňom napríklad dôjsť k rozpoznaniu fantomických objektov [4] ako vidíme na obr. 8 dolu. Zatiaľ čo pri dokonalom obraze (obr. 8. hore), má systém tendenciu chápať obrázok ako jeden kus, pri zašumení, ktoré nám vnútri objektu vygeneruje množstvo hrán rôznych orientácii, systém v obraze uvidí jeho časť ako fantomický objekt.

Pre takýto omyl vizuálneho systému však poznáme aj vznešenejšie slovo: predstavivosť. Je zaujímavé, ako v umelej inteligencii v rôznych obdobiach nachádzame vzťah, ktorý ako prvý postrehol Turing [6]: „Čo je inteligentné, sa mýli“¹.

5 Záver

V našom príspevku sme sa zaoberali významom motoriky vo vizuálnom systéme robota. Inšpirovali sme sa senzomotorickým prístupom z vývinovej psychológie. Urobili sme zopár experimentov, o ktorých si uvedomujeme, že vyvolávajú viac otázok než dávajú odpovedí. Každopádne si myslíme, že bohatá motorika musí byť súčasťou každého biomimetického vizuálneho systému. Naznačili sme, že táto by motorika mohla výrazne napomôcť zameriavaniu pozornosti s následným výstupom vhodným pre rozpoznávanie objektov.

PodĎakovanie

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

Literatúra

- [1] Brooks, R. – Stein, L.: *Building Brains for Bodies*. Journal Autonomous Robots, Volume 1 Issue 1, 1994
- [2] Davies, E. R.: *Machine Vision: Theory, Algorithms, Practicalities*. Elsevier, 2005
- [3] Dennet, D.: *Typy Myslí*. Kaligram, 1997
- [4] Fitzpatrick, P.: *Open object recognition for Humanoid Robots*, SPIE Robotics and Machine Perception newsletter, 12(2), pp. 9, September 2003
- [5] Hinterstoisser, S. - Lepetit, V. - Ilic, S. - Fua, P. - Navab, N.: *Dominant Orientation Templates for Real-Time Detection of Texture-Less Objects*. IEEE Computer Society Conference on Computer Vision

¹ Turing bol postavený pred problém spočívajúci v tom, že neomylnému stroju je možné podhodit' vstup, na ktorom sa zacyklí. Vyriešil ho tak, že tomuto stroju dovolil mýliť sa.

- and Pattern Recognition (CVPR), San Francisco, California (USA), June 2010
- [6] Kelemen, J.: *Myslenie a stroj*. Kaligram, 2010
 - [7] Lúčny, A.: *Senzomotorický prístup k poznávaniu scény*. In: *Kognice a umělý život II* (Kelemen, J. – Kvasnička, V., eds.), FPF SU, Opava, 2002, pp. 119-132
 - [8] Metta, G. - Natale L. - Nori, F. - Sandini, G. - Vernon, D. - Fadiga, L. - von Hofsten, C. - Santos-Victor, J. - Bernardino, A. - Montesano L.: *The iCub Humanoid Robot: An Open-Systems Platform for Research in Cognitive Development*, Neural Networks, special issue on Social Cognition: From Babies to Robots, Vol. 23, pp. 1125-1134, 2010
 - [9] Minsky, M.: *Society of Mind*. Simon & Schuster, New York, 1986
 - [10] Piaget, J., Inhelderová, B.: *Psychológia dieťaťa*, SOFA, Bratislava 1993
 - [11] B. Scassellati. *A binocular, foveated, active vision system*. MIT AI Memo 1628, March 1998.
 - [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] Sonka, M. – Hlaváč, V. – Boyle, R.: *Image processing, analysis and machine vision*. 3rd edition, Thomson Learning, Toronto, 2007
 - [14] Tong, H. - Li M. - Zhang, H.- Zhang, Ch.: *Blur Detection for Digital Images Using Wavelet Transform*. In In Proceedings of IEEE International Conference on Multimedia & Expo (2004), pp. 17-20