http://ilab.usc.edu/publications/doc/Elazary_Itti10vss.pdf

Framework and implementation for perception

Lior Elazary and Laurent Itti
Computer Science, Neuroscience
University of Southern California

Motivation

The main problem with AI and robotics is perception.

Kasparov playing against Deep Blue

Many algorithms exist for optimal decision making (MDPs,POMDPs, min-max, etc.), but all require that the state of the world is known (precisely or probabilistically).

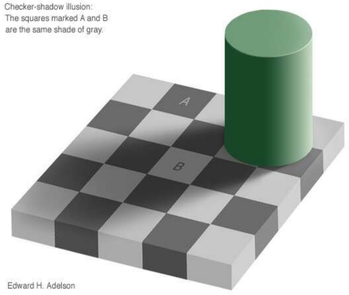
If the perception problem were solved, many of the machines promised in the past would be able to materialize.

Perception

To perceive the world (W) is to know p(W).

We do not need sensors (vision or others) to find p(W), but it helps.

Visual sensors often have problems with occlusions, missing data and noise.



We use the concept of hallucinations based on priors to fill in the missing data, which results in better recognition.

Vision is nothing but controlled hallucinations.

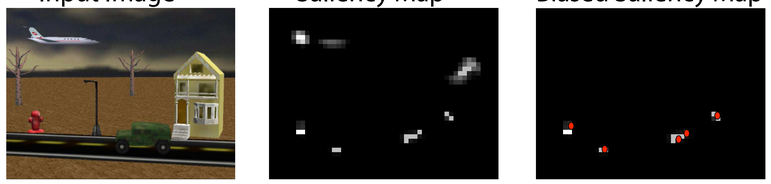
Finding p(W)

Use bottom-up methods to narrow down the possible solutions and use top-down methods to evaluate the solutions.

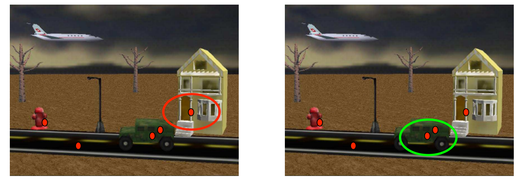
Huttenlocher and Ullman 87
DDMCMC (Z. Tu and S. C. Zhu 2001).
FastSlam 2.0 (Montemerlo et al. 03).

Finding p(W)

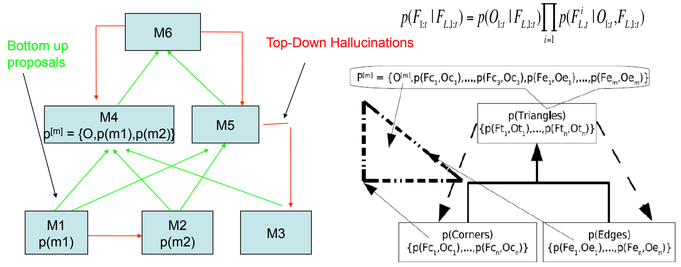
Input Image | Saliency Map | Biased Saliency Map



Model



Hierarchical Model



$$p(F_{L,t} \mid F_{L,1:t}) = p(O_{1:t} \mid F_{L,1:t}) \prod_{i=1}^{n} p(F_{L,t}^{i} \mid O_{1:t}, F_{L,1:t})$$

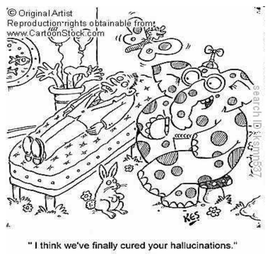
Hallucination

$$p(F \mid O_{1:t}, F_{L,1:t}) = \eta p(F_L \mid O, F)\, p(F \mid O_{1:t-1}, F_{L,1:t-1})$$

$$p(F_L \mid O, F) = p(F_L \mid O, F) + p(F \mid O_{h,1:t-1}, F_{1:t-1})$$

Can be seen as taking the likelihood or prior

Use Surprise to control hallucinations (Itti&Baldi 05)

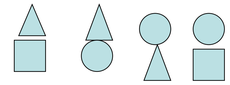


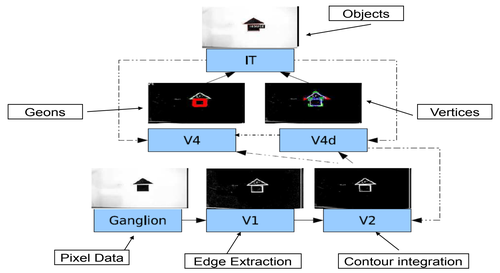
$$\text{Surpise} = KL(p(O,F \mid F_L), p(O,F)) = \int p(O,F \mid F_L) \log\left(\frac{p(O,F \mid F_L)}{p(O,F)}\right) dO,F$$

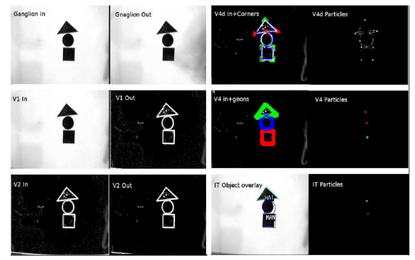
Implementation

Structure is inspired from Hummel & Biederman (1992) Geon work.

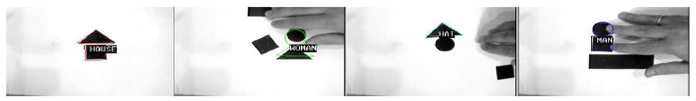
Geons: 

Objects: 





Results

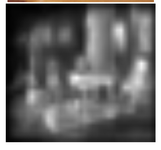


Method	No Clutter n=342	Clutter n=279	Total n=621
SIFT	41.81%	33.33%	38.00%
Proposed System	100%	97.13%	98.71%

Moving to 3D

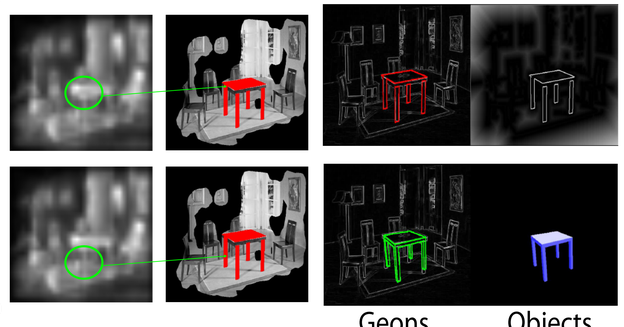
Edge Distance

$$p(I \mid C) = \prod_{i}^{N} G(e_i; 0, \sigma_i)$$



Regions

$$p(I \mid R) = \prod_{i}^{R} G(I_i; \mu_i, \sigma_i)$$



Geons | Objects

The framework demonstrates the ability to integrate bottom-up and top-down information, in order to achieve perception from noisy sensors in the presence of occlusions.

This work was supported by the Army Research Office (ARO), the Human Frontier Science Program (HFSP), the National Science Foundation (NSF), the Defense Advanced Research Pro jects Agency (DARPA), and the National Geospatial-Intelligence Agency (NGA).