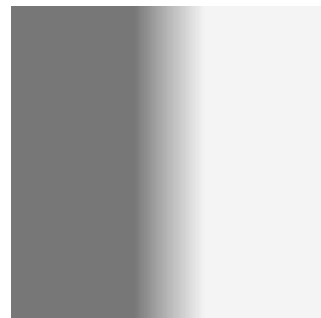
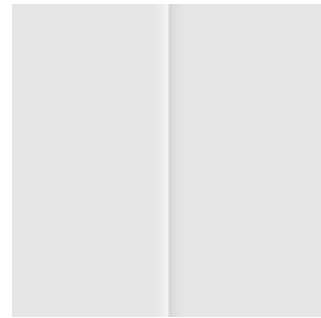
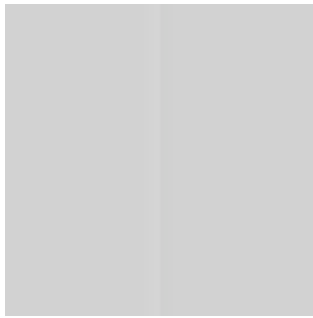


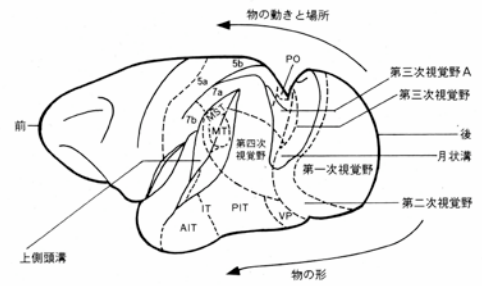
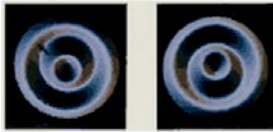
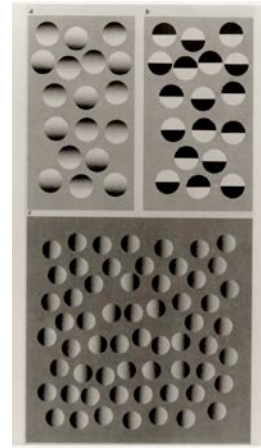
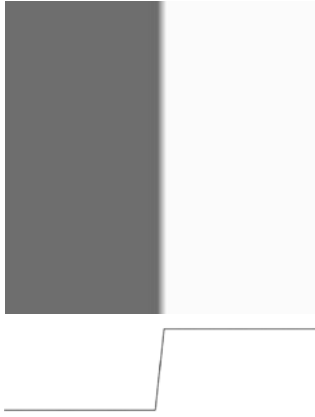
脳の計算理論

視覚の計算論

川人光男

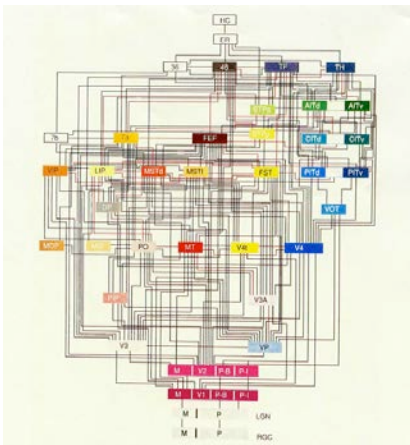
ATR脳情報通信総合研究所





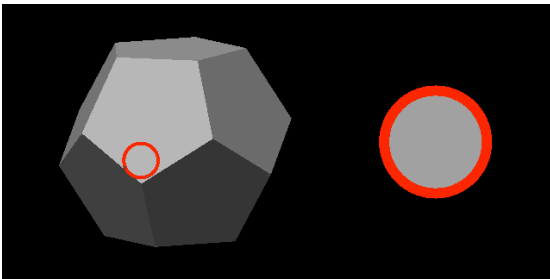
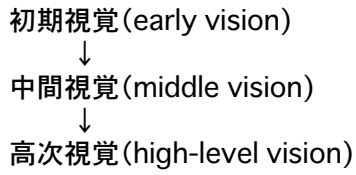
猿の視覚野

猿の複数の視覚野と頭頂連合野5野と7野。第三次視覚野、第三次視覚野A、POを見るために、月状溝を開いて広げてある。同じようにMT野、MST野を見るために上側頭溝が開いて広げている。視覚情報処理は、第一次視覚野から、第二次、第三次、MT、MST、7野、5野と進む対象物操作や歩行のための空間視の階層的流れと、第一次視覚野、第二次、第四次、PIT、AITと進むパターン認識のための形態視の階層的流れの2つに大別できる。



視覚の役割

2次元画像から外界の3次元構造を推測すること



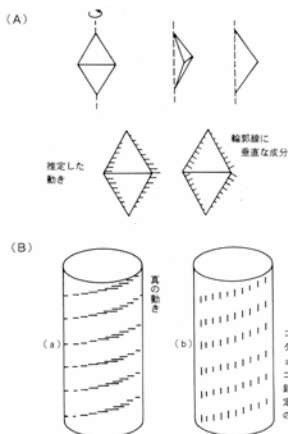
視覚の不良設定性

良設定問題

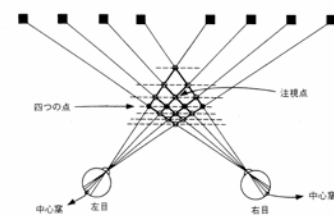
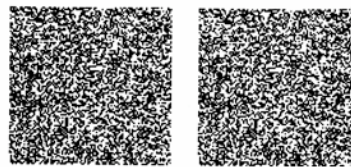
- (1) 問題の解が存在し
- (2) 解が一意であり
- (3) 解がデータに連続的に依存する

光学 3次元空間→2次元画像
視覚(逆光学) 2次元画像→3次元空間

不良設定性



深層のマークはなぜ上る
動きが輪郭線に沿ってあまり変化しないという前提知識に基づいて、輪郭線に垂直な動きの成分から、本来の動きの成分を推定したシムソン・ウェン・移動。(A) 図の動きをつくりあわせてような多面体を垂直な軸のまわりに回転させた場合に、輪郭線に垂直な動き成分しか現れない。この計算データに基づいて、さらに定めらるるの前提条件を定めて、真の動きを推定する。図(a)のように正しい回転が再現できる。(B) 深層のマークについて、同じ計算をする。真の動きは図(a)のようなのに、図(b)のように同じ全体が上へ動くという結果を記す。



ランダム・ドット・ステレオグラムと関連立体視図は、ユレスのランダム・ドット・ステレオグラムと、左目の目の像のずれで、両眼を動かして各点の像を合わせた。視線を動かして両眼の像を重なるように調整して立体像が現れるようにする。この方法はより遠くや速くは重なる物体の存在の目的像は、固定で両眼の像に応じて自動的に、このずれから両眼の像から両眼の像を計算できる。ランダム・ドット・ステレオグラムには、このずれの増減しか含まれていないが、立体的に見える。

標準正則化理論

光学

$$Ax=y$$

画像

データ y から x を解く(逆問題)

3次元世界の様子を推測

$$\|Ax-y\| + \lambda\|Px\|^2 \text{ を最小にする } x \text{ を求める}$$

画像データ拘束 世界のモデル・なめらかさ・先験的知識

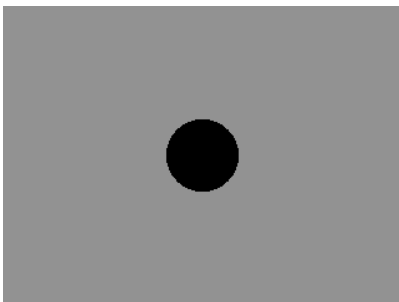
$\|Px\|$: 線形汎関数

標準正則化汎関数

問題	正則化原理
エッジ検出	$\int [(S \cdot f - I)^2 + \lambda(f_x)^2] dx$
領域に基づくオプティカルフロー	$\int [(i_x u + i_y v + i_z) + \lambda(u_x^2 + u_y^2 + v_x^2 + v_y^2)] dx dy$
輪郭に基づくオプティカルフロー	$\int [(v \cdot N - v^N)^2 + \lambda \left(\frac{\partial v}{\partial s} \right)^2] ds$
表面再構成	$\int [(S \cdot f - d)^2 + \lambda(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2)] dx dy$
時間間近似	$\int [(S \cdot f - I)^2 + \lambda(\nabla f \cdot V + f_t)^2] dx dy dt$
色	$\ I' - A\ + \lambda\ P\ ^2$
陰影からの構造復元	$\int [(E - R(f, g))^2 + \lambda(f_x^2 + f_y^2 + g_x^2 + g_y^2)] dx dy$
ステレオ	$\int \left[\left[\nabla^2 G * (L(x, y) - R(x + d(x, y), y)) \right]^2 + \lambda(\nabla d)^2 \right] dx dy$

最大事後確率推定としてみた標準正則化理論

3次元世界	x
画像	y
光学	$A \quad (Ax = y)$
観測ノイズ	$n \quad (n = y - Ax)$
3次元世界の先見分布	$P(x) = \frac{1}{Z} \exp(-U_{prior}(x)) = \frac{1}{Z} \exp(-\ Qx\ ^2)$
条件付確率	$P(y x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{n^2}{2\sigma^2}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(Ax-y)^2}{2\sigma^2}\right\}$
事後確率	$P(x y) = \frac{P(y x)P(x)}{p(y)} = \frac{1}{p(y)\sqrt{2\pi}\sigma Z} \exp\left\{-\frac{(Ax-y)^2 + 2\sigma^2\ Qx\ ^2}{2\sigma^2}\right\}$
最大事後確率推定	$P(x y)$ 最大
標準正則化	$\ Ax-y\ ^2 + \lambda\ Qx\ ^2$ 最小



中間視覚 (middle vision)

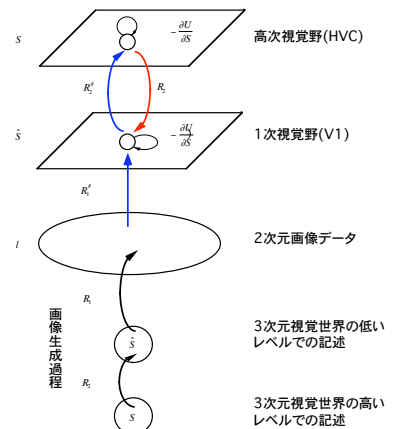
複数の手がかりの統合

↓
2・1/2次元スケッチ

ライン過程(不連続)が統合のかなめ

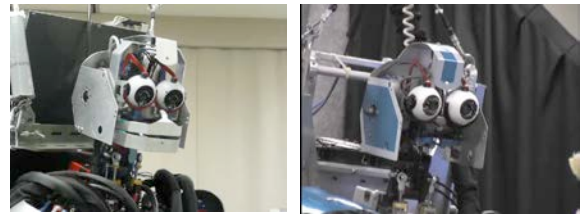
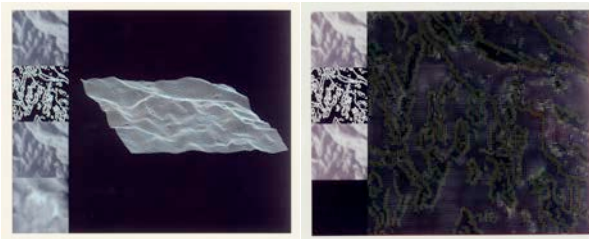
一次視覚野と高次視覚野とで構成される階層構造の基本計算モデル

一次視覚野では、高次視覚野の表現と2次元画像データとの中間的な表現が処理される。これに対応して画像生成過程は、高いレベルでの記述から低いレベルでの記述への R_1 と、低いレベルでの記述から2次元画像データへの R_2 の直列計算で表せる。



Humanoid Vision

- Optical system with foveated vision (high definition only in the center)
- Biomimetic motor control: capture targets accurately on the fovea, dealing with delays, active stereo vision



CB-i

i-1

Benefits of foveated vision



Object recognition on a humanoid head

Based on the capabilities of a humanoid visual system with foveated vision, we developed an object recognition system that integrates visuomotor processes and foveation to achieve reliable recognition. Training is done in interaction with the teacher.



Active object recognition

Training the classifier

Recognition rates with SVMs

Tr. views per object	SVM	NNC	Tr. views per object	SVM	NNC	Tr. views per object	SVM	NNC
208	97.6%	95.9%	208	94.2%	89.3%	208	91.0%	84.7%
104	96.7%	93.7%	104	92.4%	87.3%	104	87.2%	81.5%
52	95.1%	91.5%	52	90.7%	84.4%	52	82.4%	77.8%
26	91.9%	86.7%	26	86.7%	79.2%	26	77.1%	72.1%

Fovea resolution: 160x120 Fovea resolution: 80x60 Fovea resolution: 40x30

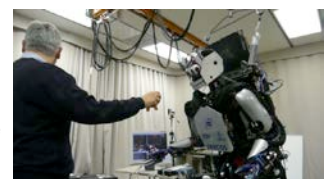
- 14 objects
- At most 2912 views were used for SVM training.
- Models invariant against 3-D depth rotations were learned.

Active 3-D vision

3-D vision is not necessary to follow the object with foveal cameras. Only 2-D information and simplified kinematics are used for this purpose.



Grasping can benefit from 3-D vision. Here the robot actively follows the object and determines the grasping configuration by continuously updating the relevant coordinate transformations.



Autonomous learning of object representations

- Machine vision has lots of problems to identify and segment objects in general scenes.
- Current vision: statistical approaches, but it is not clear how to put such knowledge into the learning system.
- The prior knowledge necessary for segmentation come from the experience we gain when we interact with the environment.
- By taking control of the object, the robot can focus on the relevant part of the image, thus bypassing potential pitfalls of pure bottom-up attention and segmentation.

Snapshot extraction

- Integration of vision and motor control
- Bayesian approach for snapshot acquisition: the position and extent of the object can be estimated by minimizing the log-likelihood (using EM-algorithm)
- No prior knowledge about objects necessary

