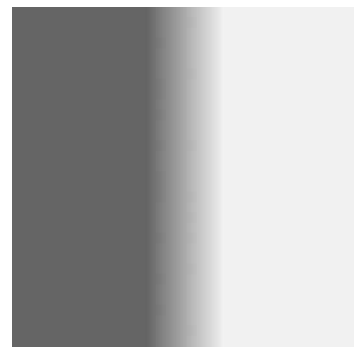
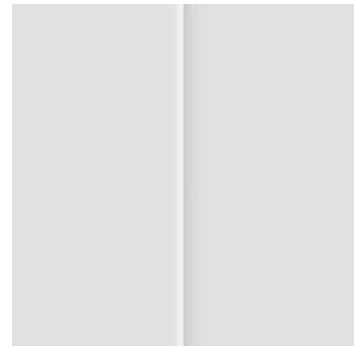
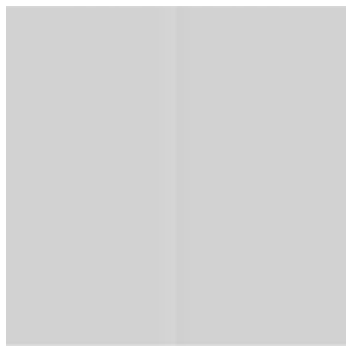
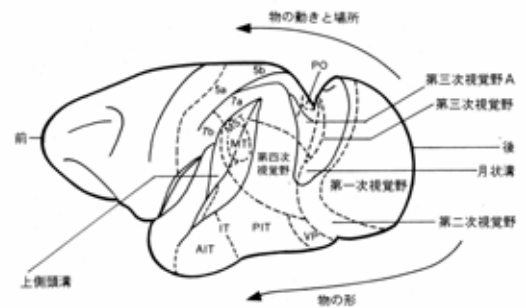
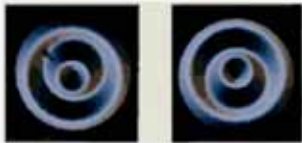
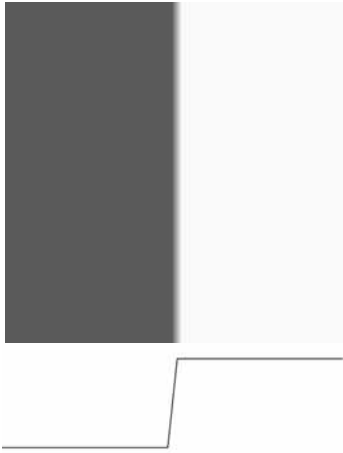


脳情報学4:
視覚の計算論

ATR脳情報研究所

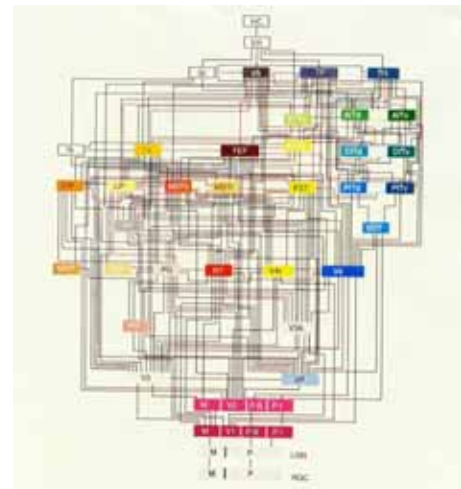
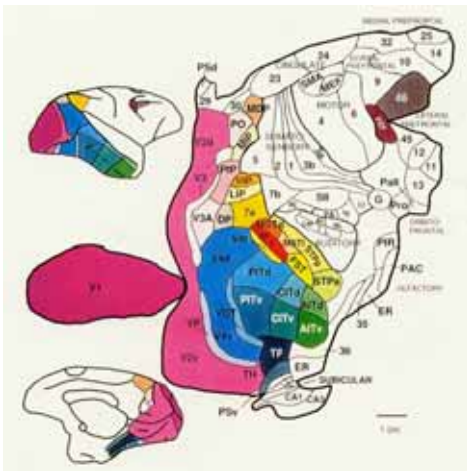
川人光男





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ある。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

初期視覚 (early vision)



中間視覚 (middle vision)



高次視覚 (high-level vision)

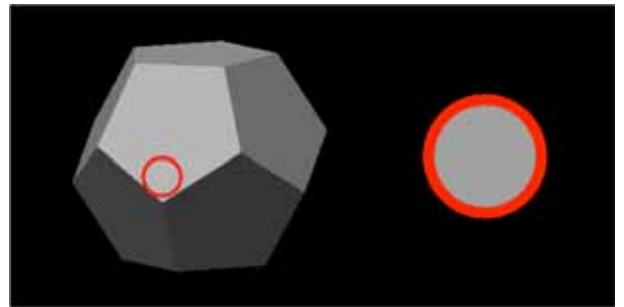
視覚の不良設定性

良設定問題

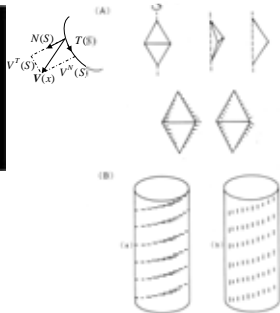
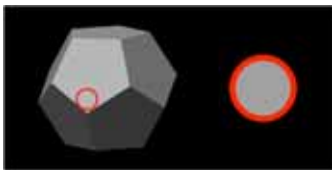
- (1) 問題の解が存在し
- (2) 解が一意であり
- (3) 解がデータに連続に依存する

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



Contour Based Optical Flow (Hildreth)



Within the aperture, only motion component perpendicular to the contour can be measured.

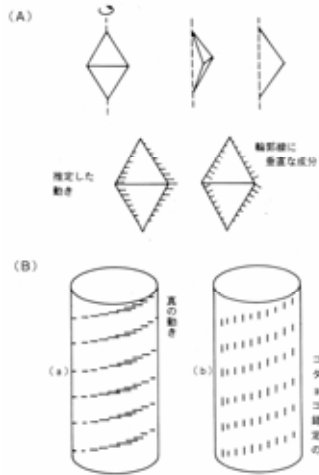
How to determine velocity vector?

V

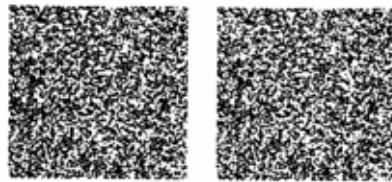
$$\|V \cdot N - V^N\|^2 + \lambda \int \left(\frac{\partial V}{\partial s} \right)^2 ds$$

Smoothness of the velocity field

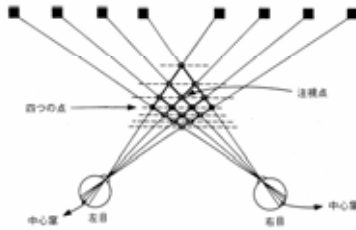




深層のマークはなぜよる
 動きが輪郭線に沿ってあまり変化しないという前提知識に基づいて、輪郭線に垂直な動きの成分から、垂直の動きの成分を推定した
 シンセティックな生成。(A) 図像面を2つ取り合わせたような世界
 図像を垂直な軸のまわりに回転させた場合に、輪郭線に垂直な動き
 成分はつかぬ。この前提知識に基づいて、垂直に回転させた
 ときの検出条件を求めて、垂直の動きを推定する。図(B)のように正
 しい前提知識がある。(B) 図像のマークについて、図(A)を推定す
 る。垂直の動き成分(a)のようなもの。図(B)のように回転させ
 られたマークという結果を得る。



左目に入る画像 右目に入る画像



ランダム・ドット・ステレオグラムと同様原理
 図は、3Dのランダム・ドット・ステレオグラム。左目の目の
 像のズレで、奥行き感が得られる仕組みを示した。図像を透視
 点で結ぶと奥行き感が得られる。図に2つの透視点があるようにす
 る。この透視点より近く奥行きが浅く遠く奥行きが深い透視点
 まで透視点に応じてズレが得られる。このズレから奥行きが3次元
 量の透視点で計算できる。ランダム・ドット・ステレオグラムは、
 このズレの透視点から奥行きが得られる。透視点のズレ。

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

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

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

$$\int [(v \cdot N - v^N)^2 + \lambda (\frac{\partial v}{\partial s})^2] ds$$

表面再構成

$$\int [(S \cdot f - d)^2 + \lambda (f_x^2 + 2f_y^2 + f_z^2)^2] dx dy$$

時空間近似

$$\int [(S \cdot f - i)^2 + \lambda (\nabla f \cdot V + f_t)^2] dx dy dt$$

色

$$\|r^* - A_c\|^2 + \lambda \|P_c\|^2$$

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

$$\int [(V^2 G * (L(x, y) - R(x + d(x, y), y)))^2 + \lambda (\nabla d)^2] dx dy$$

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 ($Ax = y$)

観測ノイズ

n ($n = y - Ax$)

3次元世界の先見分布

$$P(x) = \frac{1}{Z} \exp\{-U_{\text{prior}}(x)\} = \frac{1}{Z} \exp\{-\|Qx\|^2\}$$

条件付確率

$$P(y|x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{n^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(Ax - y)^2}{2\sigma^2}\right\}$$

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

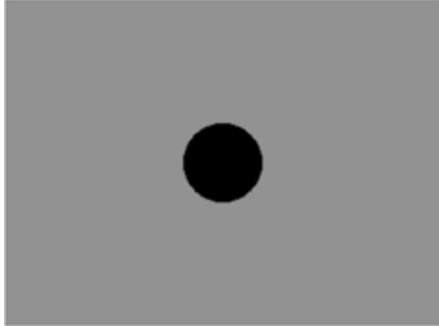
中間視覚 (middle vision)

複数の手がかりの統合



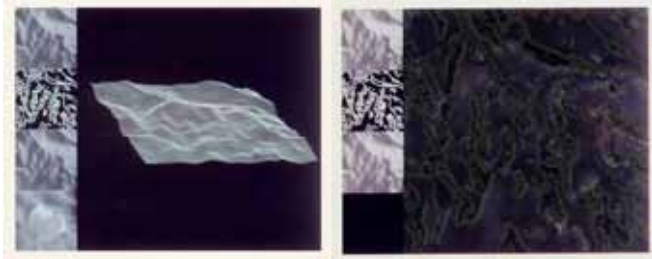
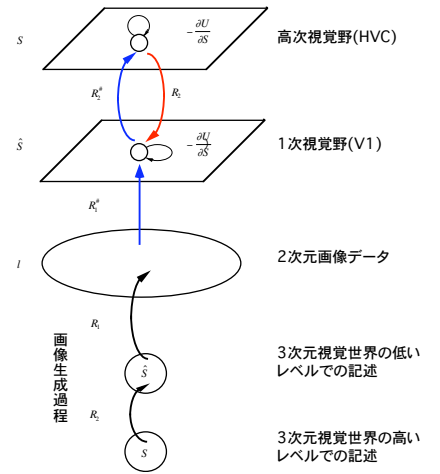
2.1/2次元スケッチ

ライン過程(不連続)が統合のかなめ



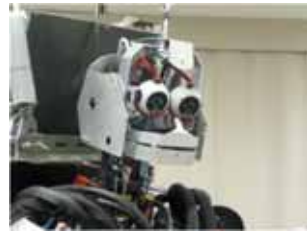
一次視覚野と高次視覚野とで構成される階層構造の基本計算モデル

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と、低いレベルでの記述から2次元画像データへのと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

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.



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

