

視覚と制御

橋本浩一

koichi@k2.t.u-tokyo.ac.jp

東京大学大学院
情報理工学系研究科
システム情報学専攻

予定変更

- 1月17日:橋本浩一助教授 視覚と制御
- 1月24日は講義がありません。
- 1月31日:木村英紀教授 生命現象と制御
 - 同じ教室、同じ時間
- レポート提出期限を2月7日に延期する。
 - 2件 駒場教務課の提出箱へ



注意事項

- 出席点について
 - 出席表には本人が署名すること(筆跡チェックをしている)
 - 意図:
 - 欠席してもかまわないが、不正をしてまで点数をほしがらな
 - 出席点の割合は少ない。よいレポートを書けば取り戻せる。
- 履修届を出していないが履修を希望する人
 - 連絡事項: 科類・学年、学生証番号、氏名、連絡先
 - 連絡先:
嵯峨山茂樹教授(工学部計数工学科システム情報コース幹事)
E-mail: sagayama@hil.t.u-tokyo.ac.jp /電話: 03-5841-6900



レポート様式と内容

- 4件のレポートの**それぞれ**についてA4版のレポート用紙を使用し、一枚目は右の例の形式で作成する。
- 連絡先はE-mailまたは電話など、レポートに関する問い合わせができる連絡先を記載する。
- 「他レポート」提出状況は、自分が提出したほかのレポートの担当教官名を**すべて**書く。レポート評価集計時の確認に用いる。通常は前半期は1名、後半期は3名記載されるはず。
- ワープロ作成の場合は最下行に**自筆の署名**を行う。

総合科目(一般F)

「インテリジェントな認識と行動のシステム科学II」
担当教官:工学部計数工学科 嵯峨山茂樹教授

「生命現象と制御」
木村英紀教授(1月31日)

理科I類1年 098765 前渡多南

E-mail: tanan@zento.ne.jp

Tel: 03-5841-6900

(他レポート:嵯峨山、舘、橋本)

前渡多南



Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

レポート様式と内容

- 2枚目からレポート本文を書く。
- 長さは特に定めないが、1講義あたり3～4枚を一応の目安とする。
- 内容は以下の3部分を含むこと
 - 講義内容の要約をわかりやすく手短に行い
 - その後講義を聴いて考えたこと、講義に触発されて調べてみてわかったことなどを中心に論述し、(参考文献・情報源を明記)
 - 最後に講義に関する感想も添えること
- 注意:4件のレポート(2回に分けて提出)は、すべて別々に作って提出。複数をいっしょに綴じないこと。



視覚

人間の感覚器

- 視覚、聴覚、触覚などの感覚器の中で最も重要
- 外部情報の87%

工学的背景

- コンピュータ技術の進展
- ビジョンセンサ、画像処理システムの低価格化

さまざまな応用

- 自動車運転支援
- 監視システム
- 機械制御
- 外観調査



視覚と制御

脳研究からのアナロジー

- **脳を創る** (CREST: 戦略的基礎研究推進事業、科技庁)
 - 脳機能の解明のうち、脳型情報処理システムの構築を目標とする研究を対象とする領域。
 - 具体的には、「脳型デバイス・アーキテクチャ(学習、連想記憶等)」「情報生成処理(認知認識、運動計画、思考、言語、評価、記憶等)システム」の構築を目標とする。

ビジョンシステム概説

- ビジョンの基礎、ステレオ

ビジュアルサーボ

- 位置ベースと特徴ベース
- 感度

視覚と制御

- 動的システム、Structure from Motion
- スネイク



脳研究からのアナロジー



Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

サイバネティクス (N. Wiener, 1948)

動物と機械の制御と通信に関する学問

- 動機: 神経生理学者との共同研究

動物の随意運動の制御の原理はフィードバック

- 目的:

- 動物の行動を機械の原理との類推を用いて解明
- 動物に似た行動をする機械を作る

脳のモータ制御、ロボティクス、人工知能

サイバネティクスの観点から制御工学を展望した解説:

木村(計測と制御 2001、システム／制御／情報 2001)



脳の情報処理 (D. Marr, 1982)

計算理論:

- 計算の目標は何か、なぜそれが適切なものか、そしてその実行可能な方略の論理は何か。
- 脳は何を入力として、何を出力として、どのような計算をしているのか。

表現とアルゴリズム:

- 計算理論はどのようにして実現することができるのか。
- 計算にあたっての入力と出力の表現は何か。
- 変換のためのアルゴリズムは何か。

ハードウェアによる実現:

- 表現とアルゴリズムがどのようにして物理的に実現されるか。
- アルゴリズムを実行するハードウェアは？
- 神経回路モデル



計算理論の実証

心理実験、行動実験、生理学的実験

- ブラックボックスの同定
 - 錯視(視覚認識のモデル)
 - 視覚フィードバックを断った指差し運動(運動制御モデル)
 - 脳の一部の機能に損傷のある患者や四肢機能に損傷のある患者に対するコンサルティング(脳機能モデル)
- 内部状態の観察
 - ニューロンの電氣的活動の計測(神経回路モデル)
 - MEG, FMRI, PET... (高次脳機能モデル)



空間知覚

脳の空間知覚能力

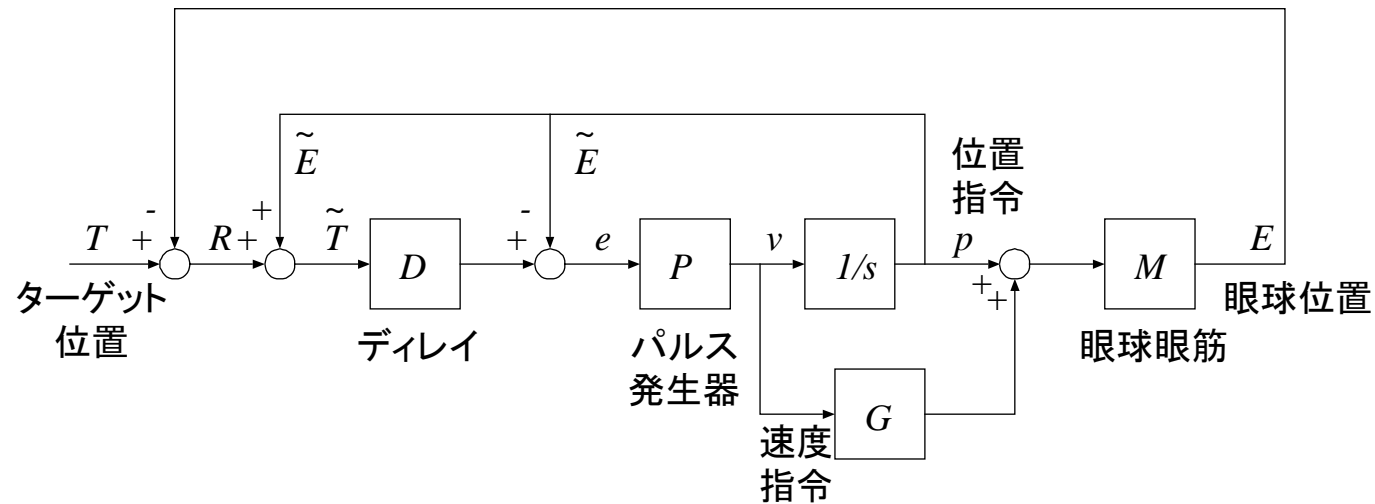
- 頭・体を動かしながらでも周りの空間を正確に認識可能
 - 体操、飛び込み、テニス... (オープンループかも?)
 - ビデオカメラを振り回して撮影した映像を見ただけでは周囲の構造を理解できない。
 - 情報処理の過程で感覚情報と運動情報が統合されている。
 - 視覚をはじめとする各種の体性感覚の統合に加え、運動指令情報が利用されているのではないか。
-
- 網膜に映る像と、眼の動き、頭の動き、体の動きなどのセンサフィードバックと、運動ニューロンに向けて出される動きの指令信号(のコピー)の統合の形態のモデルを知りたい。



視覚性空間

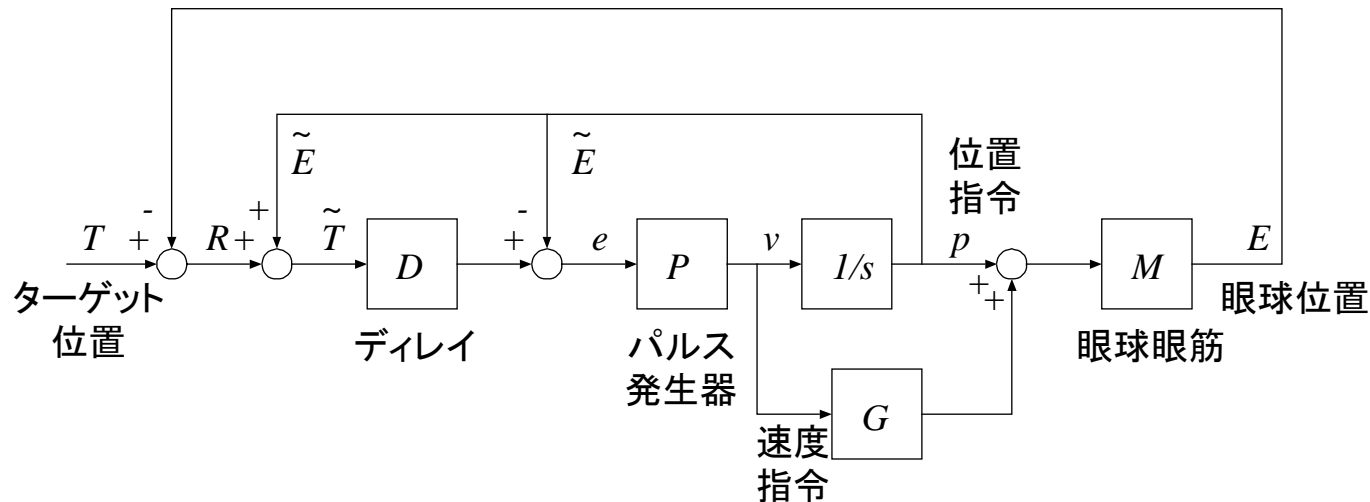
サッケード

- 次々と視線の方向を切り替える眼球運動。
- 網膜上の像と網膜中心との位置のずれを解消するための眼球運動（画像ベースビジュアルサーボ）
- 対象物の空間的位置をまず把握し、その位置に対する眼球運動（位置ベースビジュアルサーボ）



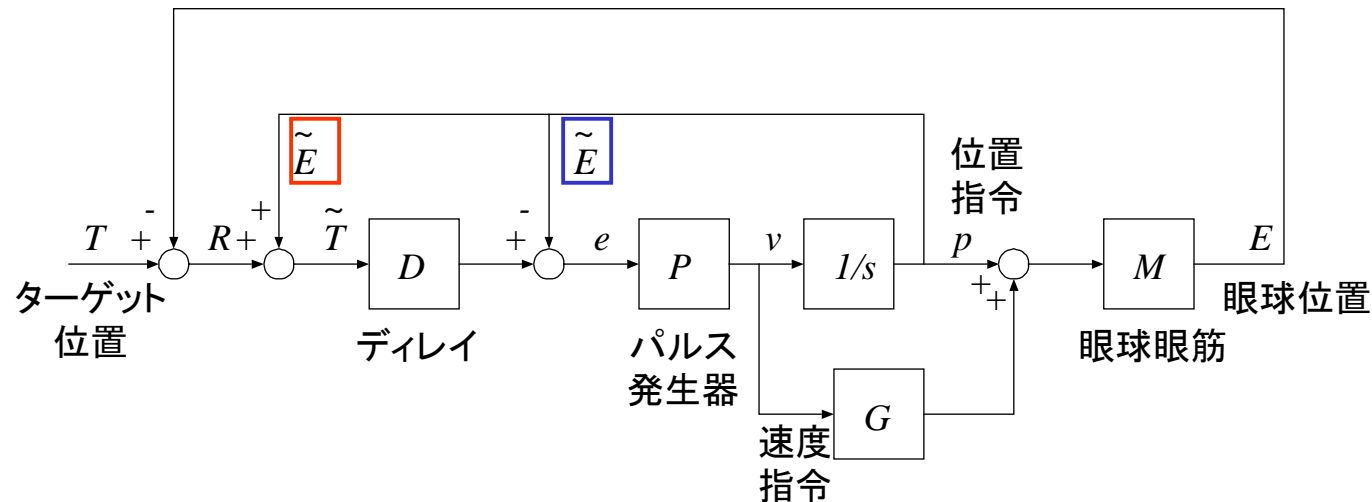
サッケードのモデル (D. Zee, 1976)

- 目標位置: T
- 眼球位置: E
- 網膜像: R
- 目標位置の脳内表現: \tilde{T}
- 眼球方向の脳内推定信号: \tilde{E}
- 速度指令: v
- 位置指令: p
- 眼球運動命令は速度指令と位置指令から構成されている。
- 眼球方向の脳内推定信号(\tilde{E})は位置指令の一部と眼筋の自己受容器の信号からなる。
- \tilde{E} を用いて目標位置の脳内表現(\tilde{T})をつくり、 \tilde{T} と \tilde{E} を一致させるようサッケードの指令を発生する。

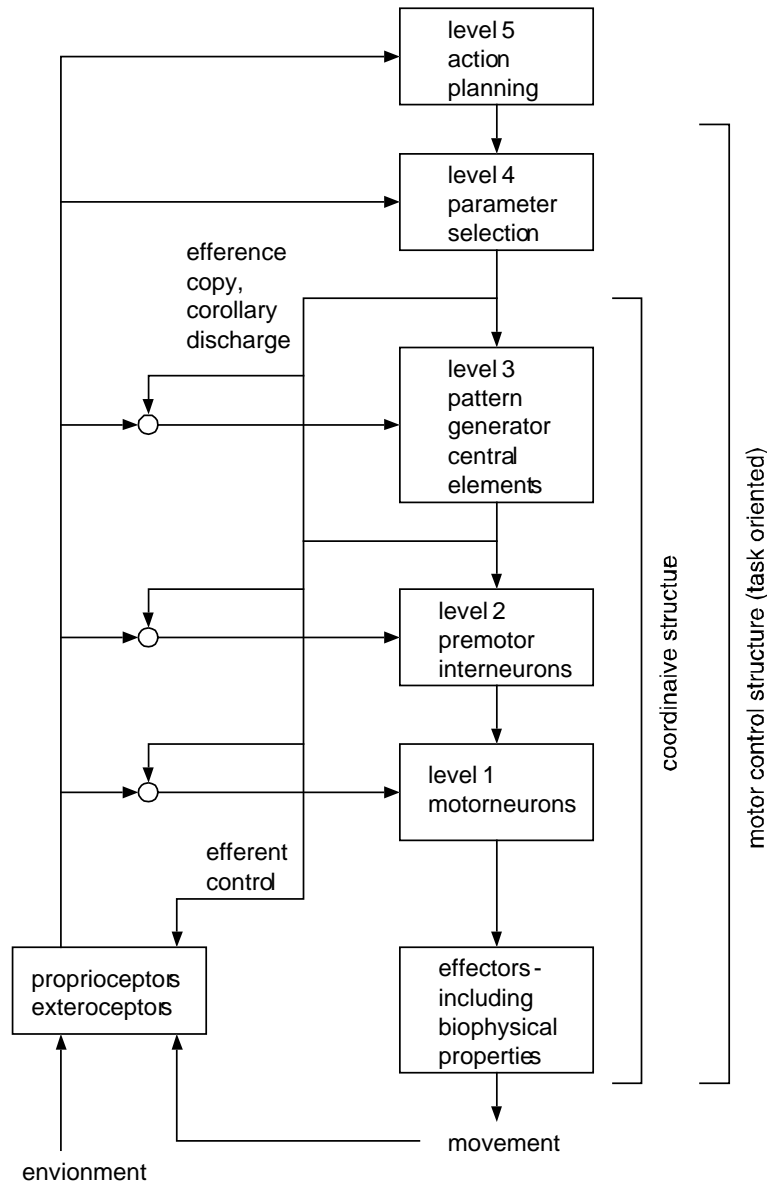


サックードモデル仮説の工学的解釈

- 脳が眼の瞬時瞬時の方向をモニタしていて、ゴール方向と一致するかどうか比較しながら運動命令を作る。
 - 時々刻々の眼球方向を表す信号(\tilde{E})
 - ターゲット位置の認識のための信号
 - 眼球位置制御のための信号



感覚情報と運動情報の相互作用モデル



- 脳からの運動指令
- 外受容器、自己受容器からのフィードバック
- 運動指令信号のコピー
- 脳から運動器官への動きの信号 (遠心性信号: efferent signal)
- 遠心性信号のコピー (遠心性コピー: efference copy)
- 自己受容器 (proprioceptor)
- 外受容器 (exteroceptor)
- 受容器から脳へ向かう信号 (求心性信号: afferent signal)



眼の位置信号

遠心性コピー

- 眼球運動命令の一部であり、自己受容器信号が眼球運動後に発火するのに比べ遅れがない。
- 両眼の共役な動きの成分のみを表す。
- 積分器を介して作られているため、眼の頭に対する絶対的な方向ではなく、現在の方向から次の方向までの相対的角度を表す。
- 遠心性コピーの信号が両眼の動きを正しく表現できるためには、眼筋の自己受容器信号を参照して常時学習する必要がある。

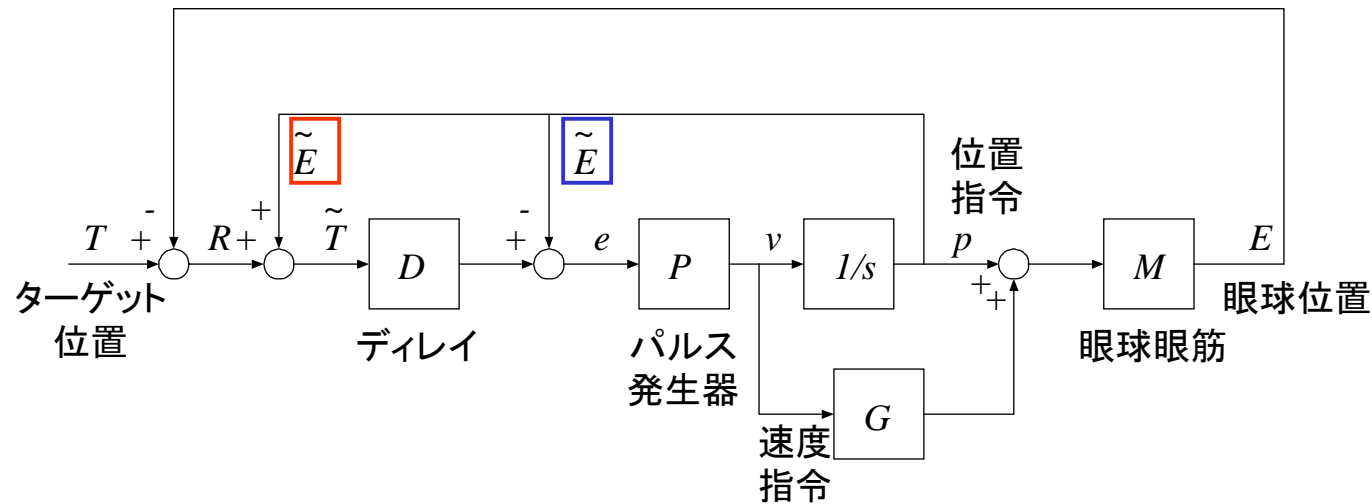
自己受容器信号

- 眼球運動後に発火するため、10ms前後の遅れが生じる。
- 各眼の頭に対する絶対的な角度を算出できる情報を含む。
- 両眼の輻輳角、したがってターゲットまでの距離を算出できる情報を含む。



視覚性空間のモデル

- 視覚目標の位置認識には自己受容器信号
- 眼球の運動制御には遠心性コピー
- サervoレートの異なる2重フィードバックループ
- 位置ベースビジュアルサーボ



運動性空間

自分の身体の一部に手を触れる動作

- かゆいところを搔く。
- 視覚を用いなくても正確にできる。

視覚と運動の協調

- 視覚目標に対して手を伸ばす (reaching)
- 対象物をつかむ (grasping)
- 対象物形状に合わせた手の形状制御 (preshaping)
- 必ずしも手の位置を見なくても行える。
- 目標位置が一度脳の中で表現されている。

運動性空間

- 目標位置対する開ループ運動
- 2重ループの位置ベースビジュアルサーボ



視覚性空間と運動性空間の関係

対象物の位置情報

- 視覚・眼球運動系が獲得
- 手の運動発現
- 手の運動系
 - 自己受容器情報に基づいて、身体の一部に触れる非視覚的対象物の位置情報が得られる。

視覚系と運動系の位置情報の関係

- 脳の領域
- 表現形式
 - 共通の座標系を用いて統一的に表現？
 - 各系ごとに独自の座標系を用いて表現？
 - 系相互間で情報のやり取りがなされる？



行動実験による見解

視覚・運動協調動作

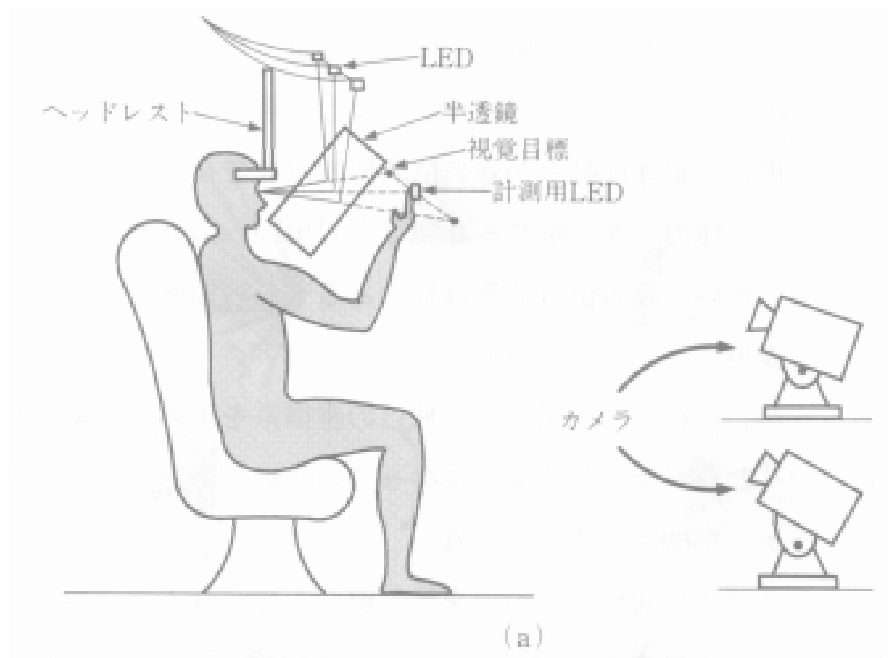
- 長時間自分の手を見る機会を奪われる
- 視覚目標を開ループ的に(自分の手を見ずに)指先で指示するときの位置精度が明らかに低下
- 視覚フィードバックがかかれば精度回復

視覚・運動各系の独立動作

- サッケードの精度は手の動作に影響されない
- 体に手を触れる動作も視覚フィードバックの影響を受けない



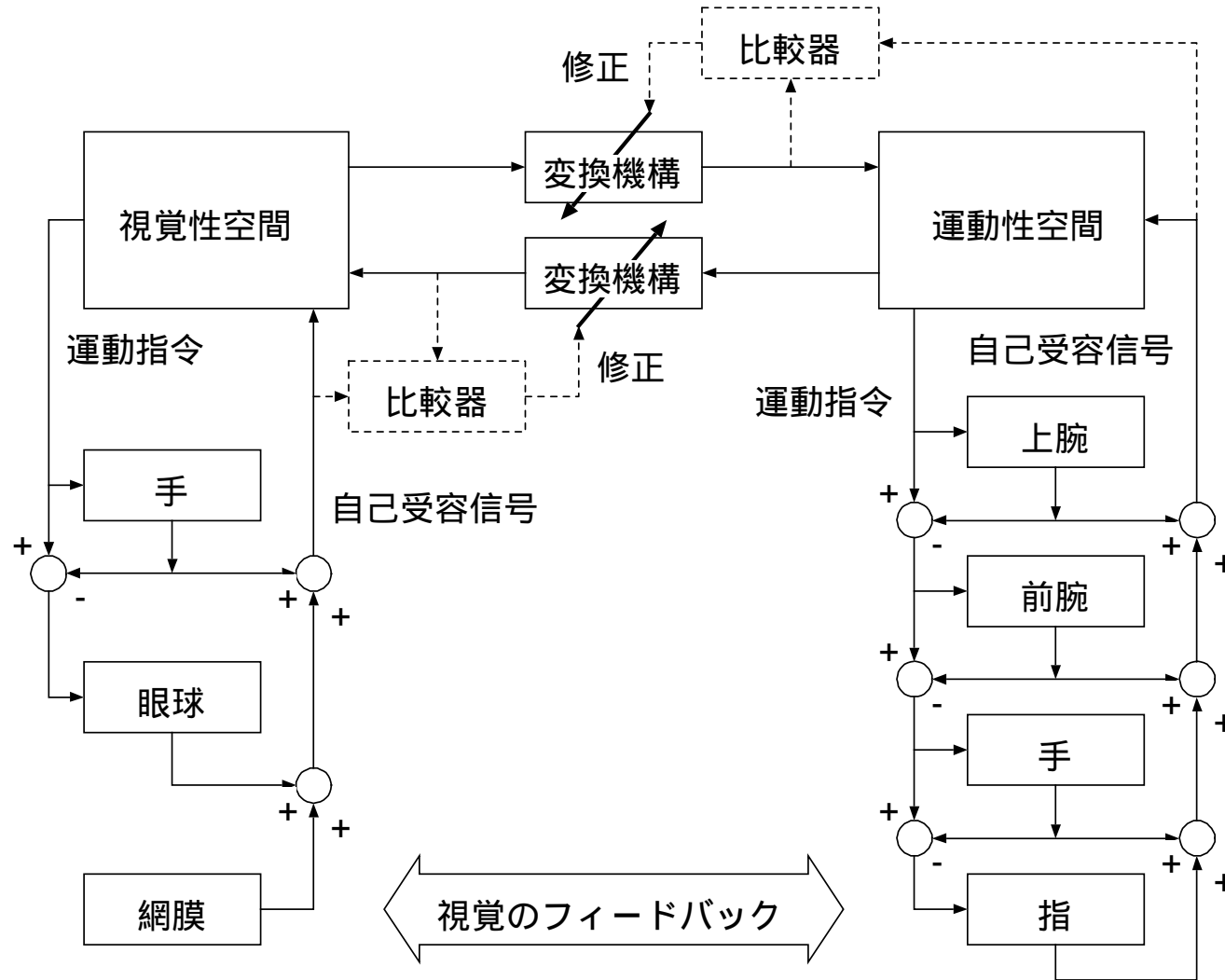
視覚目標の開ループ指示実験



- 視覚目標(LEDで提示)を指先で指示。
- 手は見えない。
- 徐々に位置誤差が増え、指先を目視すると急速に誤差は減少。
- プリズム適応でも同様
- 視覚系と手動作系の間では位置の対応関係が可塑的であり、ドリフトする。
- 手を目視確認することによって、相互間の位置対応関係が校正される。



空間知覚の統合モデル



ビジョンシステム概説



Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

視覚の階層化 (D. Marr, 1982)

初期視覚:

- 網膜に投影された2次元画像から外界の3次元空間の表面を推測する。

中間視覚:

- さまざまな視覚情報を統合して 2-1/2 次元スケッチを作る。

高次視覚:

- 3次元のモデルに基づいてパターン認識を行う。



初期視覚

2次元画像から3次元構造を推測

- 不良設定問題
- 元の世界に関する何らかの情報が必要
 - 連続性
 - 注視対象の形状・大きさ
- 人間は膨大なデータベースと照合
 - 不可能物体
 - だまし絵
 - 錯視図形

視覚のモジュール化

- 明るさ、色、形、動き、テクスチャ



中間視覚

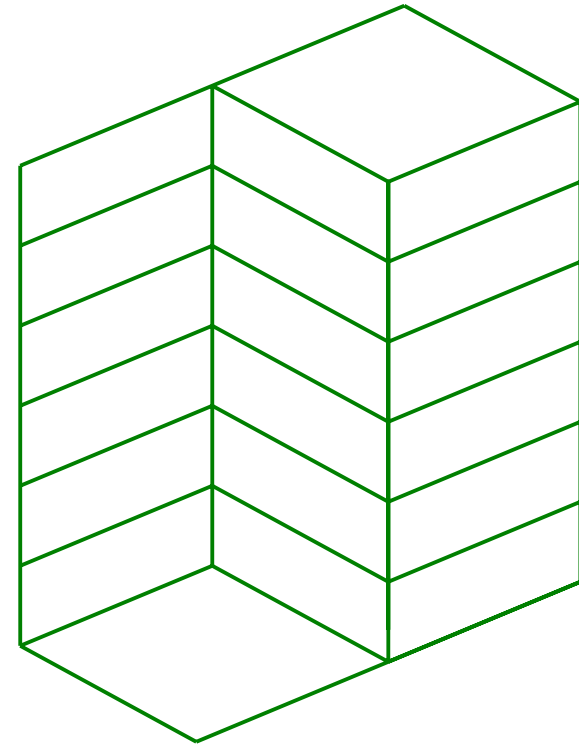
- 各モジュールの出力を統合した2次元画像と3次元立体モデルの中間的表現
 - 線画
 - 解釈が容易になるわけではない。
 - 奥行き分布図(depth map)
 - 距離計測装置の出力の自然な表現であり、有用な情報表現形態であるが、回転に弱い。
 - 法線分布図(needle diagram)
 - 照度差ステレオから容易に得られる表現であるが、回転に弱い。
 - 拡張ガウス像
 - 奥行き分布図や法線分布図から容易に得られる表現。
 - 視線方向に対する面の法線の角度に面の面積情報を付与したものであり、一種の2-1/2スケッチ。



高次視覚

パターン認識

- ランダムドットステレオにおける対応付け
 - 連続性が崩れるので対応付けが困難
- 陰影からの構造復元
 - 光源情報が欠落すると複数解が存在する。
- 連続性の仮定や光源情報があったとしても難しい。



座標変換

- ワールド座標系: O_w
- カメラ座標系: O_c

$${}^c v = \begin{bmatrix} {}^c X \\ {}^c Y \\ {}^c Z \end{bmatrix} = {}^c X r_x + {}^c Y r_y + {}^c Z r_z$$

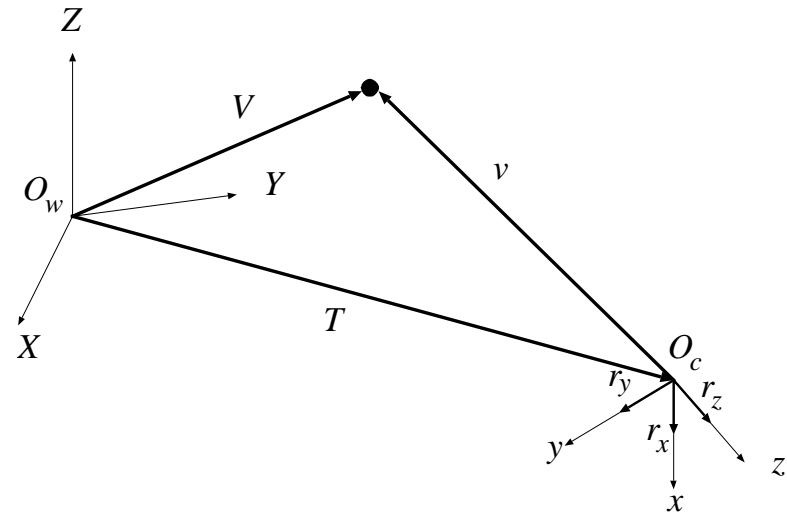
$${}^w R_c = [r_x \ r_y \ r_z]$$

$$V = {}^w R_c {}^c v + T$$

$$\begin{bmatrix} V \\ 1 \end{bmatrix} = \begin{bmatrix} {}^w R_c & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} {}^c v \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} {}^c v \\ 1 \end{bmatrix} = \begin{bmatrix} {}^w R_c^T & -{}^w R_c^T T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V \\ 1 \end{bmatrix}$$

外部パラメータ行列



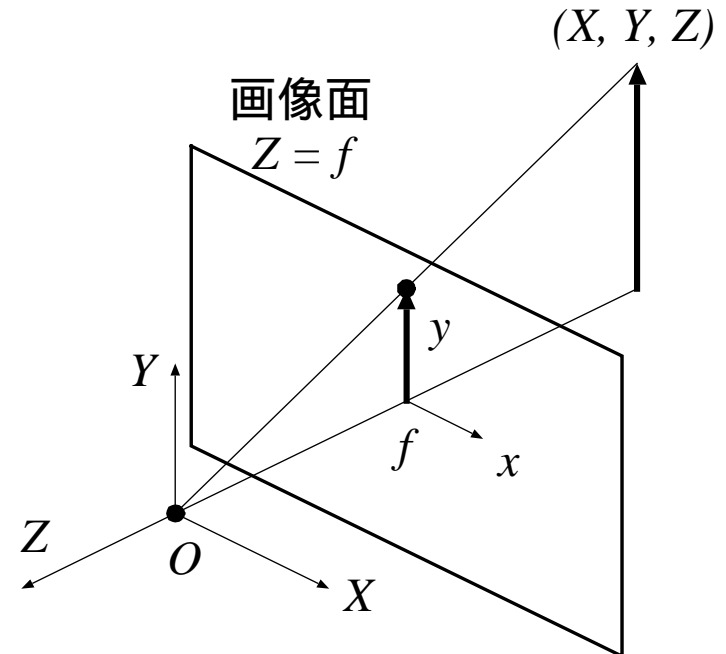
同次変換

- 透視変換

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{cZ} \begin{bmatrix} cX \\ cY \end{bmatrix}$$

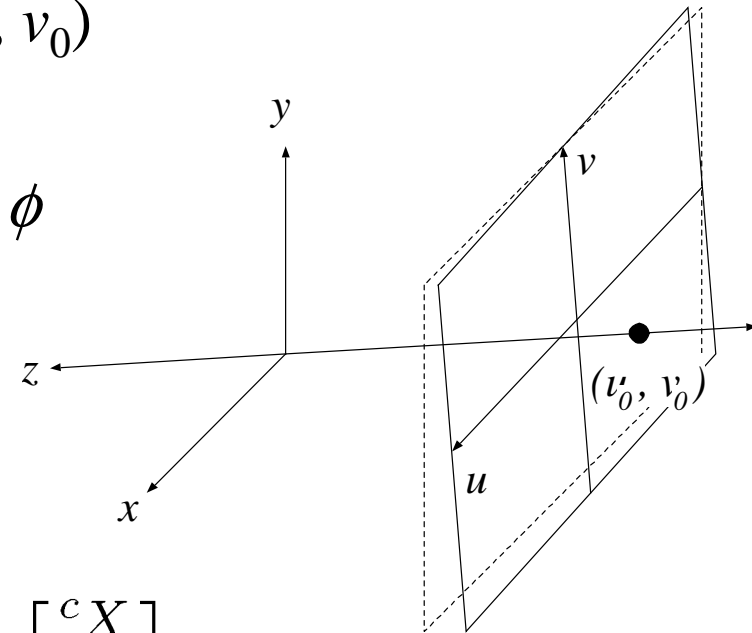
- 同次変換

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} cX \\ cY \\ cZ \end{bmatrix}$$



画像座標

- フレーム系と CCD 系のずれ
 - 光軸と CCD 面の交点: (u_0, v_0)
 - 縦横のスケール: (k_u, k_v)
 - u - v 平面と x - y 平面のずれ: ϕ



$$w \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f k_u & f k_u \cot \phi & u_0 \\ 0 & f k_v / \sin \phi & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} {}^c X \\ {}^c Y \\ {}^c Z \end{bmatrix}$$

内部パラメータ行列: A



透視投影行列

- 座標変換

$${}^c v = \begin{bmatrix} {}^c X \\ {}^c Y \\ {}^c Z \end{bmatrix} = \underbrace{[{}^w R_c^T \quad -{}^w R_c^T T]}_B \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- 透視変換・画像座標

$$w \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f k_u & f k_u \cot \phi & u_0 \\ 0 & f k_v / \sin \phi & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_A \begin{bmatrix} {}^c X \\ {}^c Y \\ {}^c Z \end{bmatrix}$$

- 透視投影変換

$$w \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = AB \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$P = AB$: 透視投影行列



キャリブレーション

- 透視投影行列

$$P = \begin{bmatrix} P_{11} & \cdots & P_{14} \\ \cdots & & \\ P_{31} & \cdots & P_{34} \end{bmatrix}$$

- 透視投影変換はスケール不変ゆえ、11自由度
 - 外部パラメータ:6
 - 内部パラメータ:5
- 6点を観測すれば最小2乗推定が可能



ステレオ

$$w_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = P_i \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \begin{array}{l} P_i = A_i [R_i^T \quad -R_i^T T_i] \\ t = R_1^T (T_2 - T_1) \\ R = R_1^T R_2 \end{array}$$

m_i M

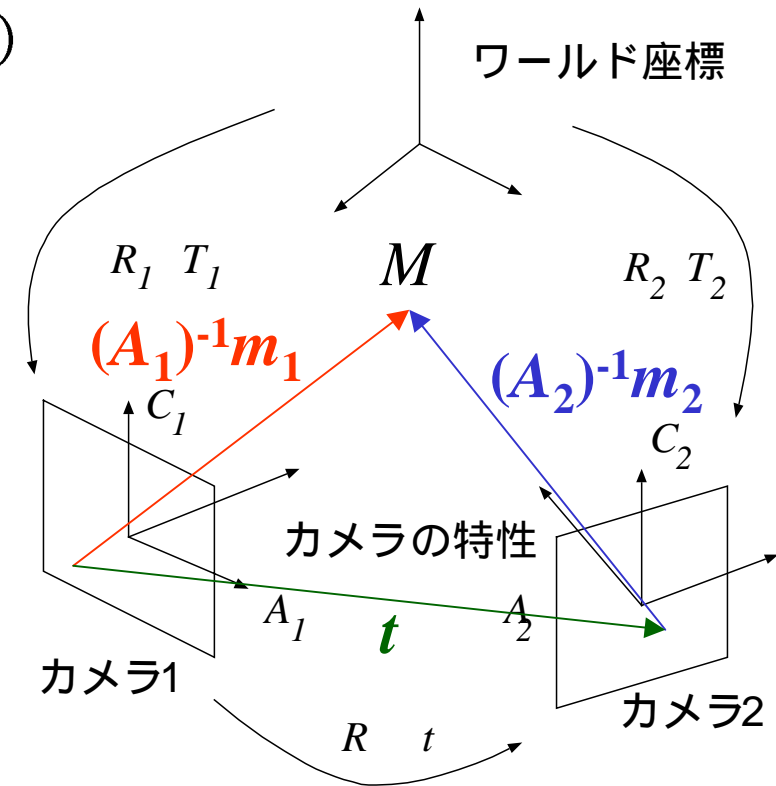
$$w_i R_i (A_i)^{-1} m_i = [I \quad -T_i] M = M - T_i$$

$$w_1 (A_1)^{-1} m_1 - w_2 R (A_2)^{-1} m_2 = t$$

$$T = t \wedge = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$$

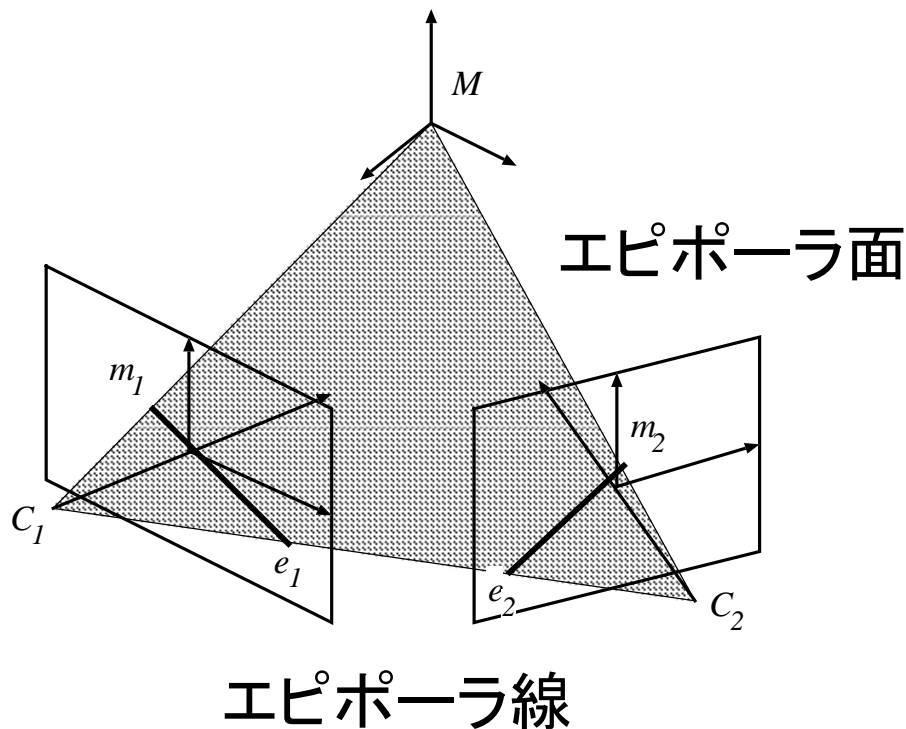
$$m_1^T (A_1^{-1})^T T R (A_2)^{-1} m_2 = 0$$

F: Fundamental Matrix



エピポーラ面、エピポーラ線

- 2台のカメラ中心と注目点を含む面をエピポーラ面という。
- エピポーラ面の画像平面への投影は線になり、これをエピポーラ線という。
- 対応点はかならずエピポーラ線上に存在する。



$$m_1^T F m_2 = 0$$

$$m_1^T F = [a_1 \ b_1 \ c_1]$$

$$a_1 u_2 + b_1 v_2 + c_1 = 0$$

エピポーラ線方程式



ステレオ画像の遷移

- カメラ1、カメラ2で見えていた画像はカメラ3ではどのように見えるか？
- ステレオにより空間座標を計算して逆算するのではなく、基礎行列のみから推定

$$m_1^T F_{13} m_3 = 0, \quad m_2^T F_{23} m_3 = 0$$

$$m_3 = F_{13}^T m_1 \wedge F_{23}^T m_2$$

- R, T がわかると F は計算できるので、4, 5, 6... 台目のカメラ画像も容易に計算できる。
- さまざまな視点からの像が3次元情報を再構成せずに容易に作成できる。



基礎行列の決定

- F は 3x3 行列で、定数倍の自由度: 未知数は8

$$m_1^T F m_2 = 0$$
$$[u_1 \ v_1 \ 1] \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = 0$$

- 8点の対応が取れば F は計算できる。

$$u^T f = 0$$
$$u^T = [u_1 u_2 \ v_1 u_2 \ u_2 \ u_1 v_2 \ v_1 v_2 \ v_2 \ u_1 \ v_1 \ 1]$$
$$f^T = [F_{11} \ F_{21} \ F_{31} \ \cdots \ F_{13} \ F_{23} \ F_{33}]$$
$$U f = 0, \quad U = \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix}$$

基本行列

- 基礎行列

$$F = (A_1^{-1})^T T R (A_2)^{-1}$$

$$m_1^T F m_2 = 0, \quad m_i = [u_i \ v_i \ 1]^T$$

- 基本行列

$$E = T R$$

$$\xi_1^T E \xi_2 = 0, \quad \xi_i = [x_i \ y_i \ 1]^T = (A_i)^{-1} m_i$$

- カメラの内部パラメータ(A_i)が既知であれば基礎行列から基本行列が計算できる。



基本行列の性質

- 基本行列の集合: E
- 任意の 3x3 行列: Q

$$Q \in E \Leftrightarrow \Sigma = \text{diag}\{\lambda, \lambda, 0\} \text{ where } Q = U\Sigma V^T$$

- 基本行列から外部パラメータ(t, R)の抽出

$$E = TR, \quad T = t \wedge$$

$$t = \|Q\|U_{.3}, \quad R = UR_zV^T$$

$$U_{.3} = U \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad R_z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



動画像

- 解析の原理はステレオと同じ
 - カメラが動くかオブジェクトが動くかの違い
- 画像の系列
 - 対象の追跡が容易
 - 対応付けが容易
 - 動きのモデルが与えられれば精度が向上
- 基本行列に基づく運動復元が可能



オプティカルフロー

- カメラ(またはオブジェクト)の動きと画像の動きの対応
 - カメラの動き: (V, Ω)
 - 注目点の位置: $\xi = [x \ y]^T = \left[\frac{X}{Z} \ \frac{Y}{Z} \right]^T$
 - オプティカルフロー:

$$\dot{\xi} = \begin{bmatrix} -\frac{1}{Z} & 0 & \frac{x}{Z} & xy & -1 - x^2 & y \\ 0 & -\frac{1}{Z} & \frac{y}{Z} & 1 + y^2 & -xy & -x \end{bmatrix} \begin{bmatrix} V \\ \Omega \end{bmatrix}$$

- オプティカルフローと深さがわかれば運動がわかる
- 運動とオプティカルフローがわかれば深さがわかる
- 並進速度と深さを分離して求めることはできない



特徴ベースビジュアルサーボ

- 見え方の違いによりカメラの動きを制御
 - 目標画像: $\xi_d = [x_d \ y_d]^T$
 - 現在画像: $\xi = [x \ y]^T$
 - オプティカルフロー: $\dot{\xi} = J \begin{bmatrix} V \\ \Omega \end{bmatrix}$
 - カメラの動き:

$$\begin{bmatrix} V \\ \Omega \end{bmatrix} = J^\dagger(\xi_d - \xi), \quad J^\dagger = (J^T J)^{-1} J^T$$

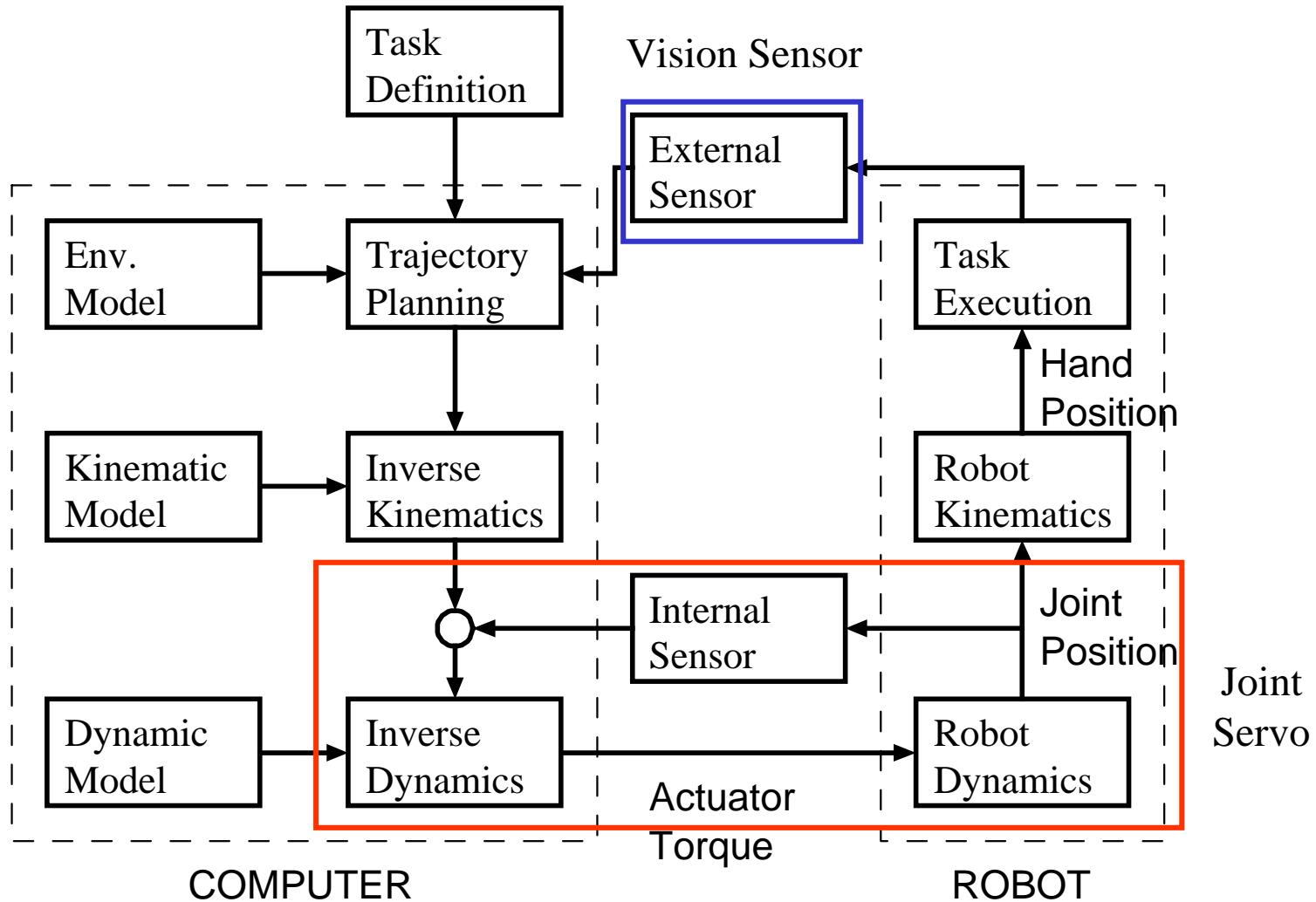


ビジュアルサーボ

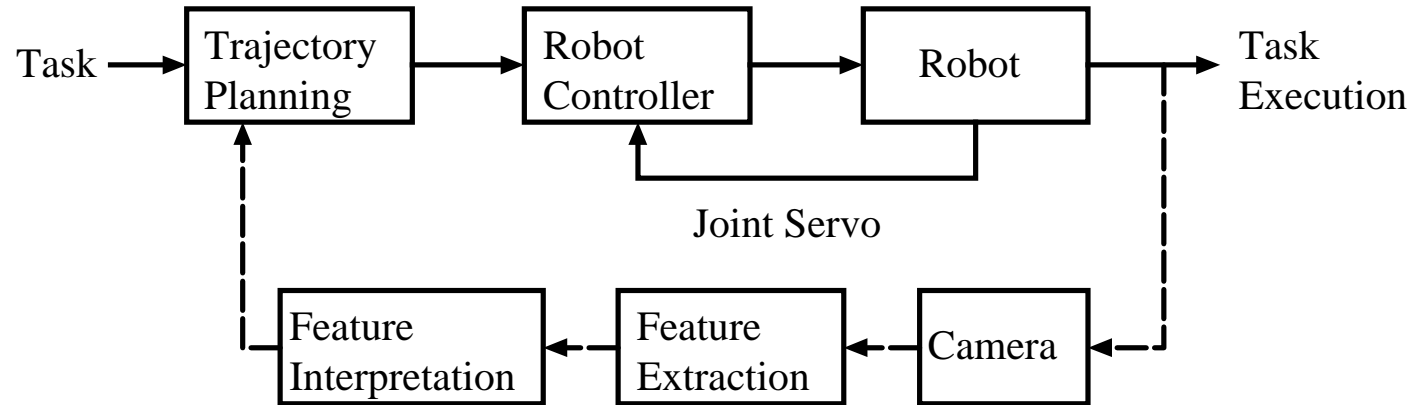


Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

階層的ロボット制御



静的な視覚フィードバックとその歴史



- 最も初期の文献: Y. Shirai and H. Inoue: Guiding a robot by visual feedback in assembly tasks. *Pattern Recognition*, 5: 99-108, 1973.
- 積み木の世界: 完全な自動化
- 背景からの物体の切り出しと位置認識のために入力画像を線画に変換し、解釈
- 1枚の画像に基づいて軌道を計画
- アームが動き終わるまでに少しでも環境が変化すると作業は失敗
- 画像入力、特徴抽出、特徴理解の処理は逐次的
- 視覚フィードバック周期は10秒
- 「見てから動く」静的な方法



研究の背景

電総研

- 線画抽出のための非階層的手法(白井、1973)
- トルク制御マニピュレータ(高瀬、1973)

日立

- プリント基板の欠陥検出、半導体組み立て(江尻、1973)

MIT

- シーンアナリシスにおける線画の解釈 (Winston, 1975)

Stanford

- ロボットアーム制御法 (Paul, 1972)

- 高性能マニピュレータの開発の遅れが原因で、マシンビジョン・マシンインテリジェンスの研究とマニピュレーションの研究に分化
- マシンインテリジェンスの研究の一分野としてマニピュレーションが研究された



ビジュアルフィードバック

SRI International

- ボルトの挿入、コンベア上の物体のつまみあげにビジュアルフィードバックを利用 (Rosen, 1976, 1978)
- 2値化を用いた平面上の物体の位置検出、物体までの距離推定アルゴリズム (Hill, 1979)
- ビジュアルフィードバック制御における動的な問題が指摘された

University of Pennsylvania

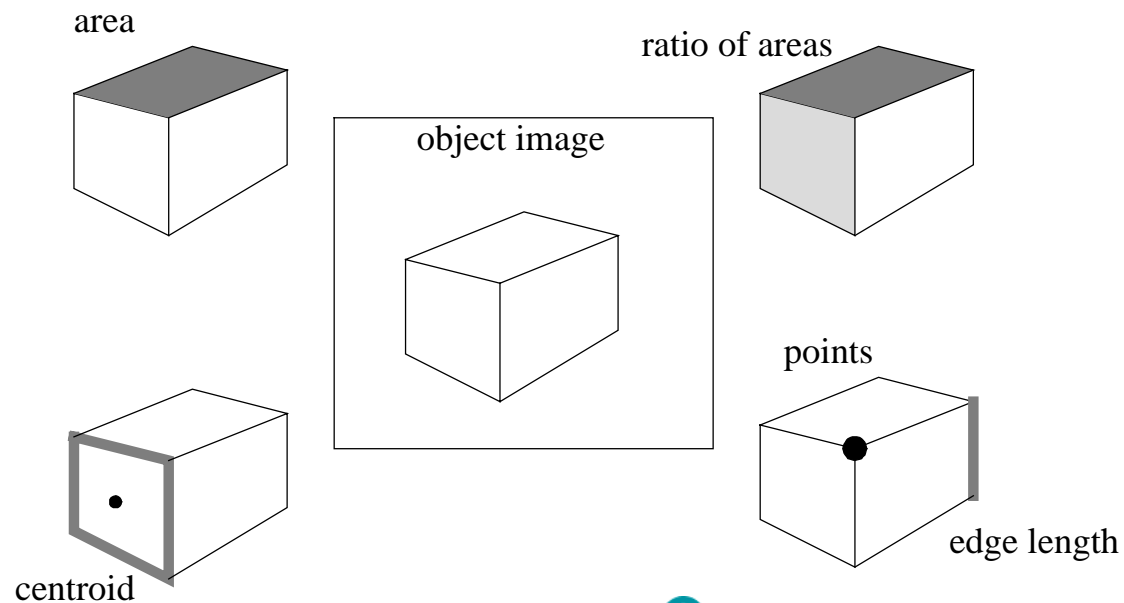
- ピンポンロボット (Andersson, 1987)
- ボールの軌道モデルを用いてボールを打つためのハンドの軌道を動的にプランニング
- 人工知能的ビジュアルフィードバックのハイライト



画像特徴量

CMU

- 特徴ベースビジュアルサーボ (Weiss, 1984, 1987)
- ロボット姿勢と画像特徴量との非線形名関係を適応的に推定
- 画像中の物体の像の特徴を表す量
- 画像の解釈を行わずに簡単な前処理で得られるもの



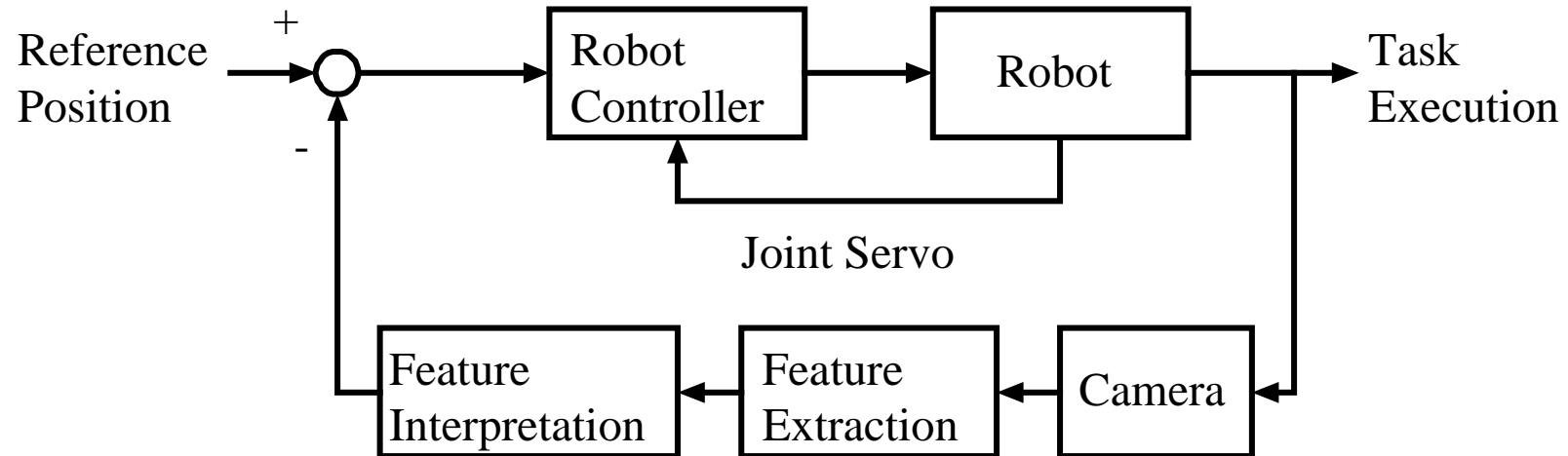
Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

ビジュアルサーボ

- 動的な視覚フィードバック制御
- 動的：「見ながら動き、動きながら見る」
- 視覚センサをフィードバックループ内に組み込み、運動する物体に対して作業することを目的として設計された制御系
- 位置ベース法と特徴ベース法



位置ベースビジュアルサーボ



- サーボ系への目標値入力: 物体とハンドの相対位置・姿勢
- 従来の静的な方法の自然な拡張
- ロボット制御、画像処理、画像認識が並行して実行される
- ロボット制御系は通常の位置制御系
- ロバストかつ高速な位置・姿勢の検出法が研究の対象



位置測定法

写真測量

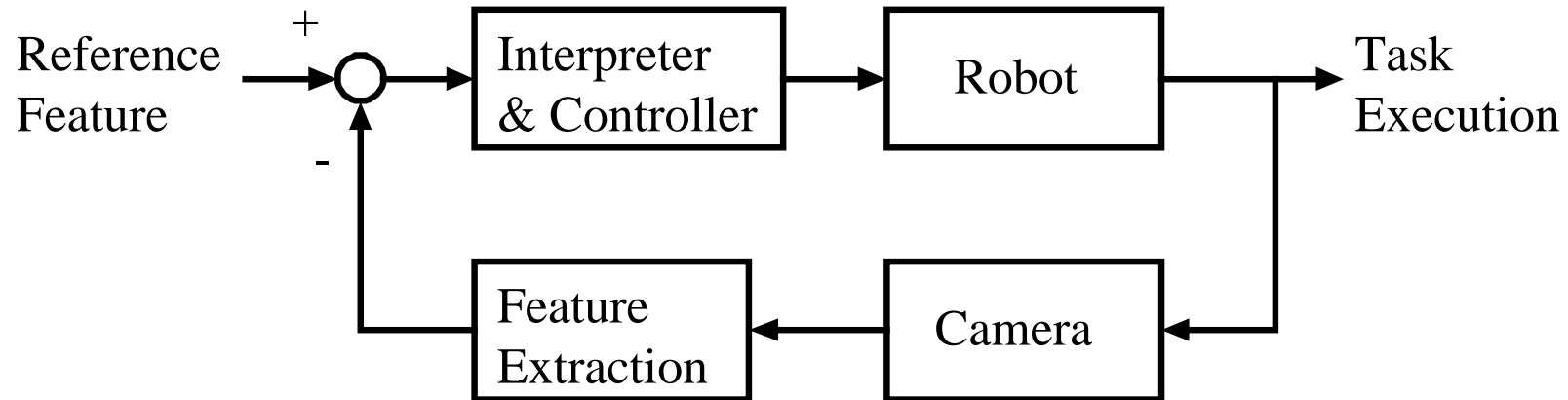
- 既知の大きさ・形状の物体の位置・姿勢を一枚の画像から求める
- ノイズの影響を受けやすく、精度がよくない
- たくさんの特徴点を用いてノイズの影響を排除(空間的)
- 特徴点の時系列を用いてノイズの影響を排除(時間的)
- AR モデルと self tuning regulator (Koivo, 1991)
- 拡張 Kalman Filter (Wilson, 1993)

ステレオ

- 画像の対応づけ
- 暗い背景の中で白いボールを追跡(ピンポン: Andersson, 1987, お手玉: Rizzi, 1992, Miyazaki, 1994)
- オプティカルフロー(おもちゃの列車: Allen, 1993)
- 高速ビジョン(動く物体の把握: Ishikawa, 1996)



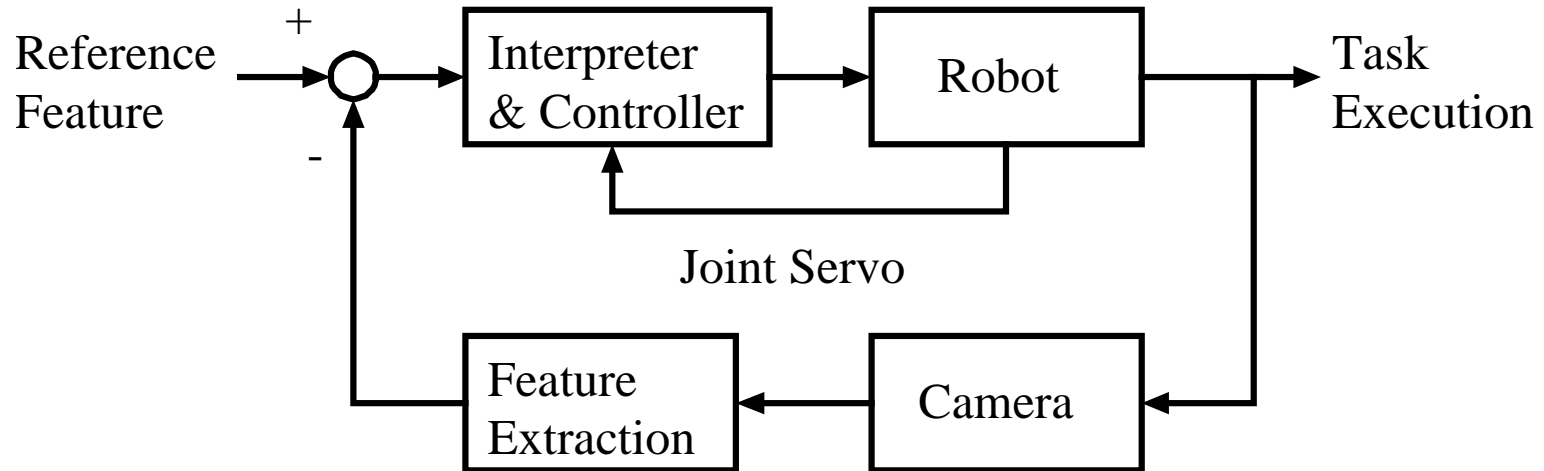
特徴ベースビジュアルサーボ



- 画像上の特徴量を直接フィードバックして制御
- すべてのブロックを並列に処理
- サーボ系への目標入力の特徴量
- 物体のモデルやロボットの機構学的モデルの不確かさに対してロバスト
- 3次元再構成をしないので、ノイズにロバスト
- 目標値の生成は teach-by-showing



特徴ベースビジュアルサーボII



- カメラのサンプリング周期の長さ: 33ms
- ロボット制御に要求されるサンプリング周期: 1ms
- 内部にジョイントサーボ系を持つ2重ループ構成
- 視覚はジョイントサーボ系に対して追従目標指令系
- ジョイントサーボ系は従来のもと同じ



特徴ベース定式化

- 特徴量ベクトル: ξ
- ハンドの位置・姿勢: p

$$\xi = \psi(p)$$

- 目標特徴量: ξ_d
- 目標ハンド位置: p_d

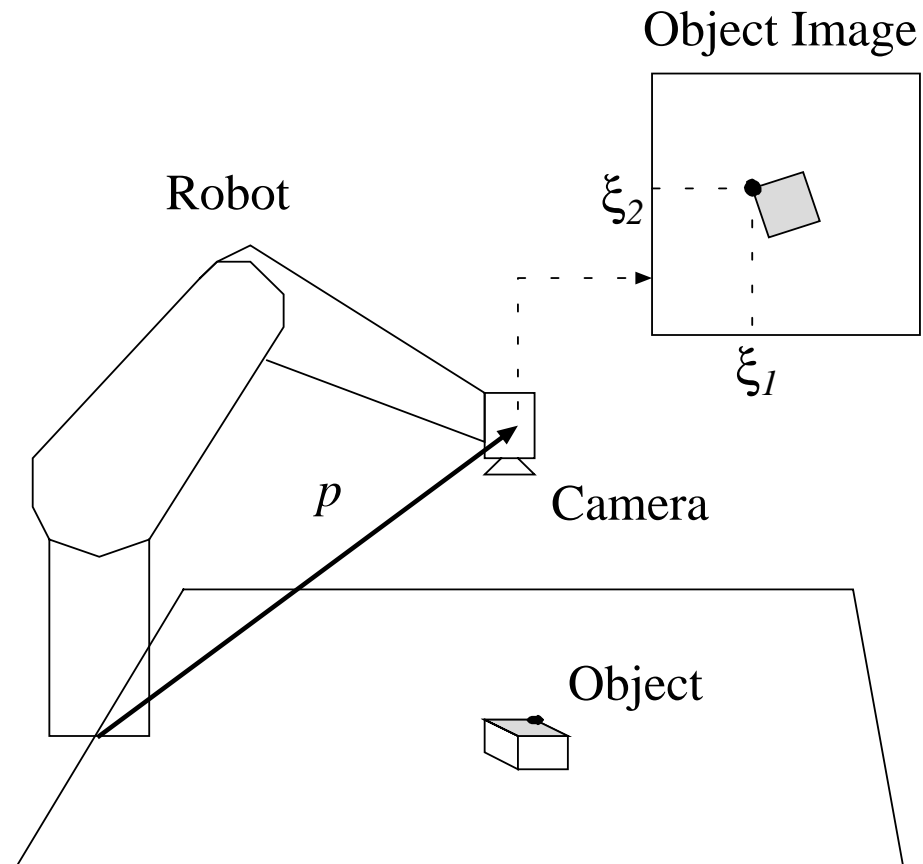
$$\xi_d = \psi(p_d)$$

- 画像ヤコビアン

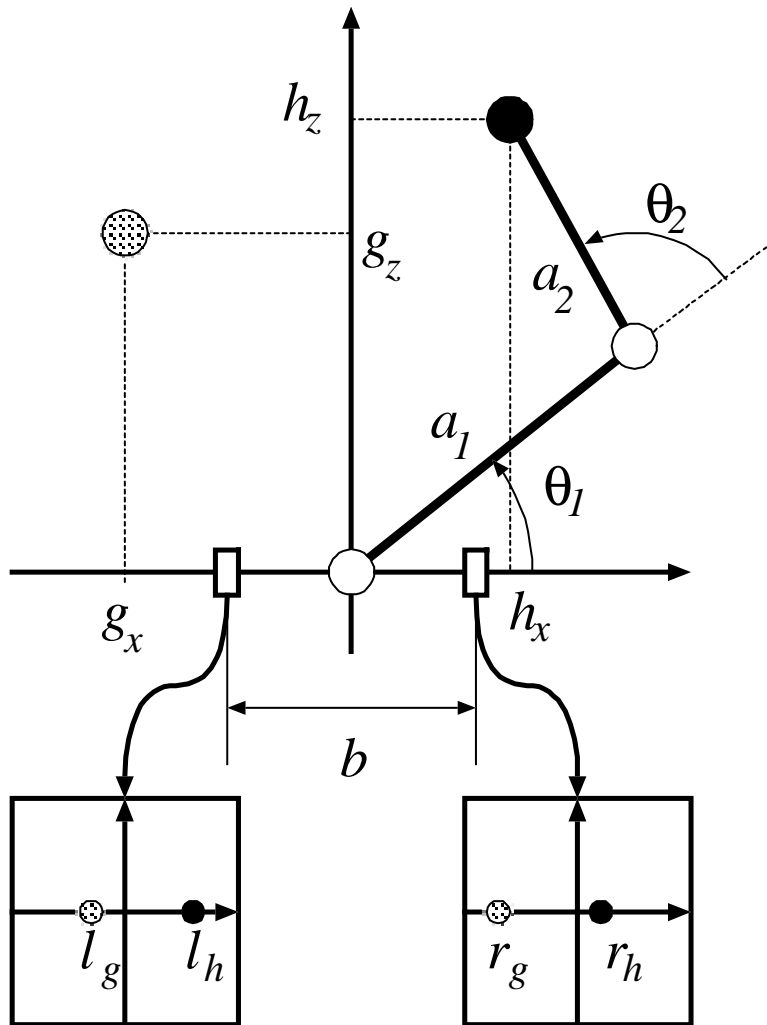
$$J = \frac{\partial \xi}{\partial p}$$

- 制御則

$$\dot{p} = \lambda J^\dagger (\xi_d - \xi)$$



位置ベースと特徴ベースの比較



- 関節角: $\theta = (\theta_1, \theta_2)$
- リンク長さ: $a = (a_1, a_2)$
- ハンド位置: $h = (h_x, h_z)$
- ハンド画像: $\xi_h = (l_h, r_h)$
- ゴール位置: $g = (g_x, g_z)$
- ゴール画像: $\xi_g = (l_g, r_g)$
- 目標: ハンド位置 h をゴール位置 g に位置決めする。

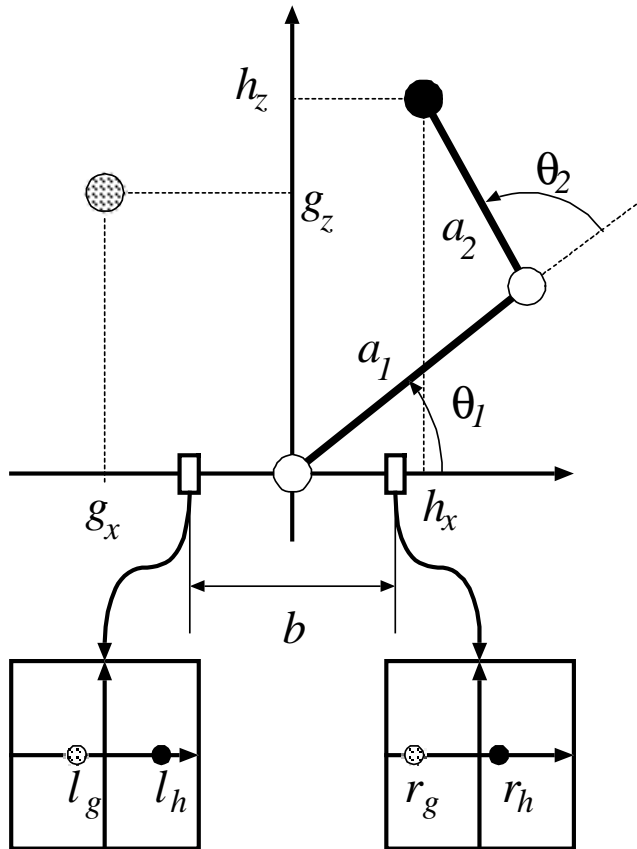


Ishikawa Hashimoto Laboratory

UNIVERSITY OF TOKYO

<http://www.k2.t.u-tokyo.ac.jp/>

3つのモデル



- 撮像モデル: $C_g : g \mapsto \xi_g$

$$l_g = f \frac{g_x + b/2}{g_z}, \quad r_g = f \frac{g_x - b/2}{g_z}$$

- ステレオモデル: $\hat{E}_g : \xi_g \mapsto \hat{g}$

$$\hat{g}_x = f \frac{b(l_g + r_g)}{d_g}, \quad \hat{g}_z = f \frac{b}{d_g}$$

$$d_g = l_g - r_g$$

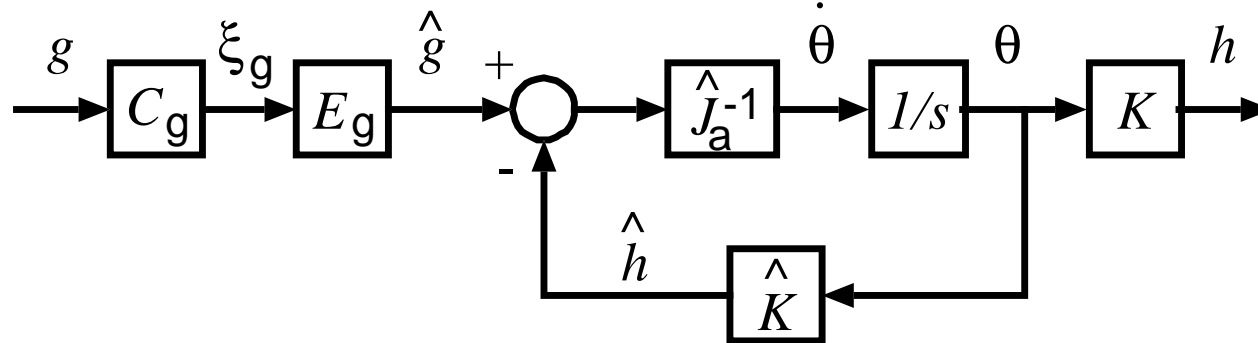
- 機構学モデル: $\hat{K} : \theta \mapsto \hat{h}$

$$h_x = a_1 \cos \theta_1 + a_2 \cos(\theta_1 + \theta_2)$$

$$h_z = a_1 \sin \theta_1 + a_2 \sin(\theta_1 + \theta_2)$$



位置ベース法



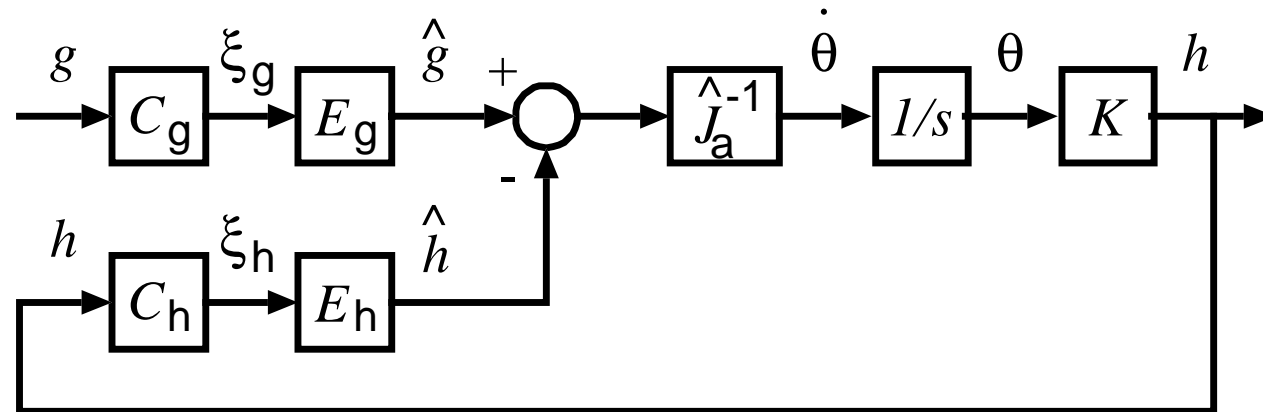
- 内部センサによるハンド位置推定値 \hat{h} と外部センサによるゴール位置推定値 \hat{g} に基づく制御
- 制御則

$$\dot{\theta} = \lambda \hat{J}_a^\dagger (\hat{g} - \hat{h}), \quad J_a = \frac{\partial h}{\partial \theta}$$

- 機構学モデル \hat{K} 、ステレオモデル \hat{E}_g に誤差があったら？
- ヤコビアンにも機構学モデル、ステレオモデルパラメータが含まれる
 - 閉ループ系が安定のための条件: $J_a \hat{J}_a^{-1} > 0$



位置ベース法II



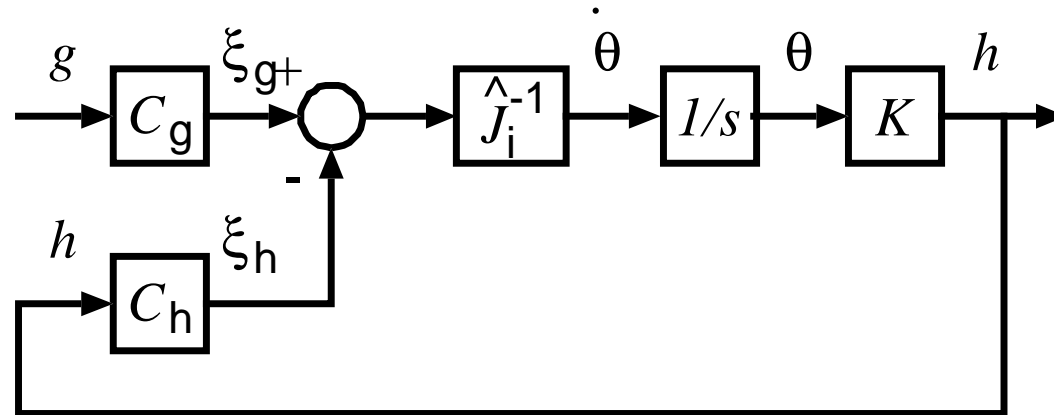
- 外部センサによるハンド位置推定値 \hat{h} とゴール位置推定値 \hat{g} に基づく制御
- 制御則

$$\dot{\theta} = \lambda \hat{J}_a^\dagger (\hat{g} - \hat{h}), \quad J_a = \frac{\partial h}{\partial \theta}$$

- 機構学モデルは用いない
- ヤコビアンには機構学パラメータも含まれる



特徴ベース法

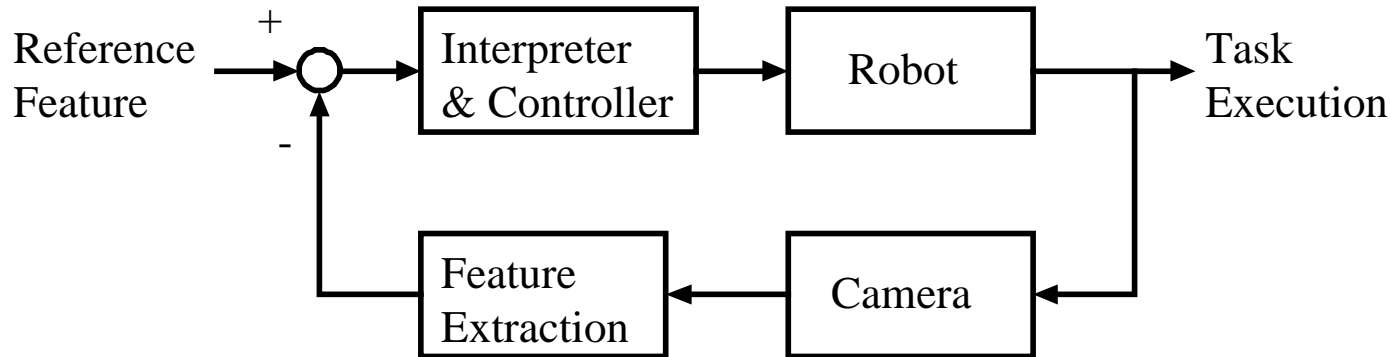


- ハンド画像 ξ_h とゴール画像 ξ_g に基づく制御
- 制御則

$$\dot{\theta} = \lambda \hat{J}_i^\dagger (\xi_g - \xi_h), \quad J_i = \frac{\partial \xi_h}{\partial h} \frac{\partial h}{\partial \theta}$$

- 機構学モデルもステレオモデルも用いない
- ヤコビアンにはこれらのパラメータは含まれる
 - 閉ループ系が安定のための条件: $J_i \hat{J}_i^{-1} > 0$

特徴ベースビジュアルサーボ系の特性



特徴量の選択基準

- 画像処理の安定性、ロバスト性
- 処理時間
- 制御系に与える影響

閉ループ系の特性

- 制御しやすさ
- ノイズに対する影響

入出力特性

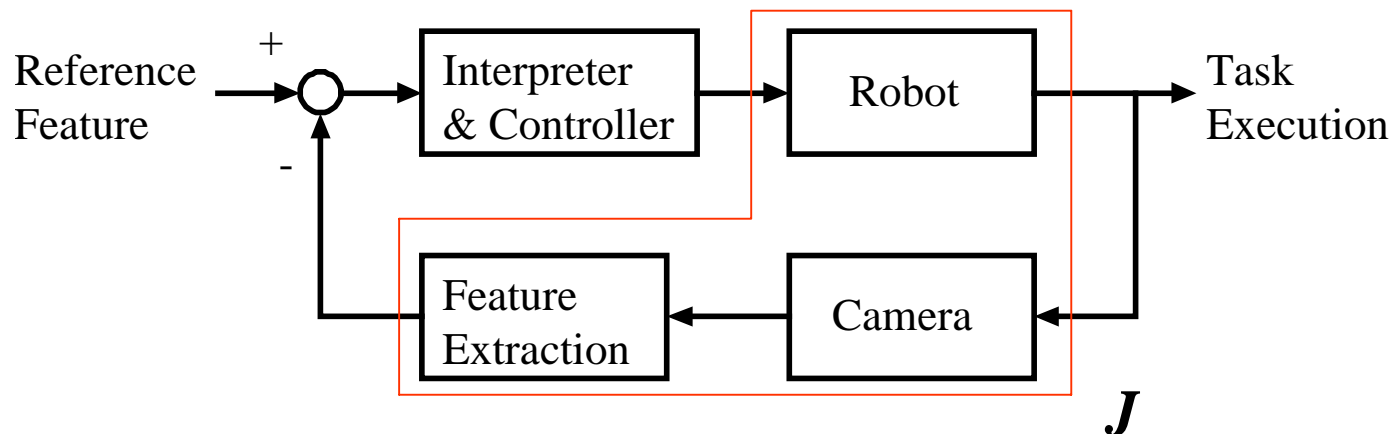
- 入力: 関節角度 θ
- 出力: 特徴量 ξ
- 入出力関係

$$J = \frac{\partial \xi}{\partial \theta}$$

画像ヤコビアン



画像ヤコビアン



- 画像ヤコビアンの特異値分解

$$J = UDV^T, \quad Jv_k = d_k u_k$$

$$U = [u_1 \cdots u_m], \quad V = [v_1 \cdots v_n], \quad |u_k| = |v_k| = 1$$

- v_k 方向のカメラの動き u_k 方向の特徴の動きに変換され、大きさは d_k 倍される。



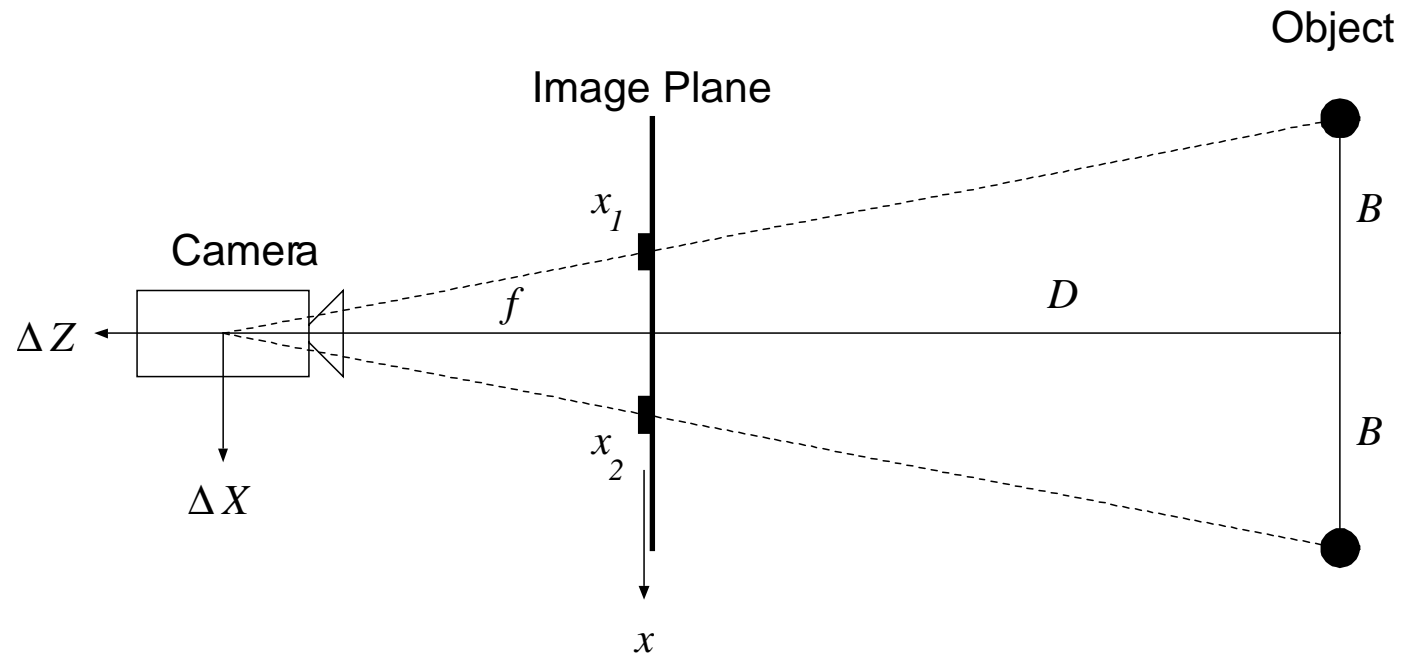
感度

$$Jv_k = d_k u_k$$

- 最大特異値に対応した特異ベクトル方向のカメラの動きに対しては画像は大きく変化するが、最小特異ベクトル方向の動きに対しては画像はあまり変化しない。
- もし、最小特異値がゼロであれば、最小特異ベクトル方向の動きに対しては画像は変化せず、ロボットは動きつづける。
- 最小特異値を感度と定義し、それを大きくするように特徴量を選択する。



2自由度の例



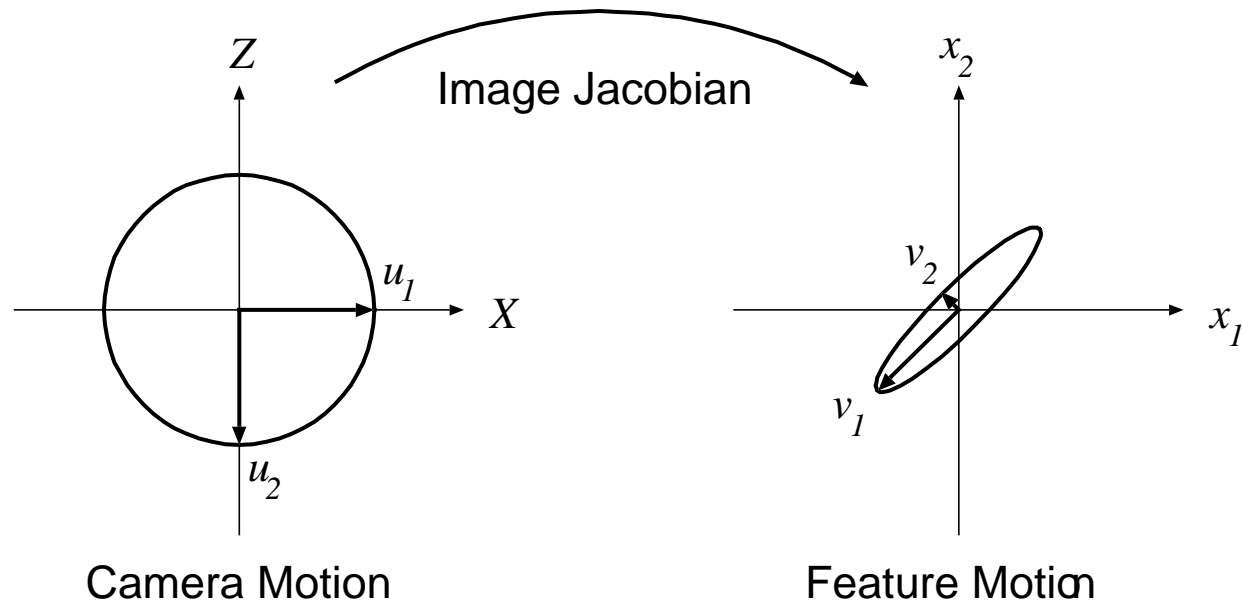
$$\xi = (x_1, x_2) = \left(-f \frac{B}{D}, f \frac{B}{D}\right), \quad J = \begin{bmatrix} -\frac{f}{D} & \frac{fB}{D^2} \\ -\frac{f}{D} & -\frac{fB}{D^2} \end{bmatrix}$$



楕円表現

- 数值例: $D = 1000$, $B = 100$, $f = 100$
- 特異値分解

$$U = \begin{bmatrix} -0.707 & -0.707 \\ -0.707 & 0.707 \end{bmatrix}, D = \begin{bmatrix} 0.141 & 0 \\ 0 & 0.014 \end{bmatrix}, V = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$



大域的安定化

制御則: $u = \lambda J^\dagger(\xi_d - \xi)$

- 初期画像と目標画像がある関係を見たとき、この制御則は機能しない(不安定現象)。
- また、初期画像と目標画像が大きく異なるとき、目標に到達する前にカメラが静止することがある(安定領域の局所性)。

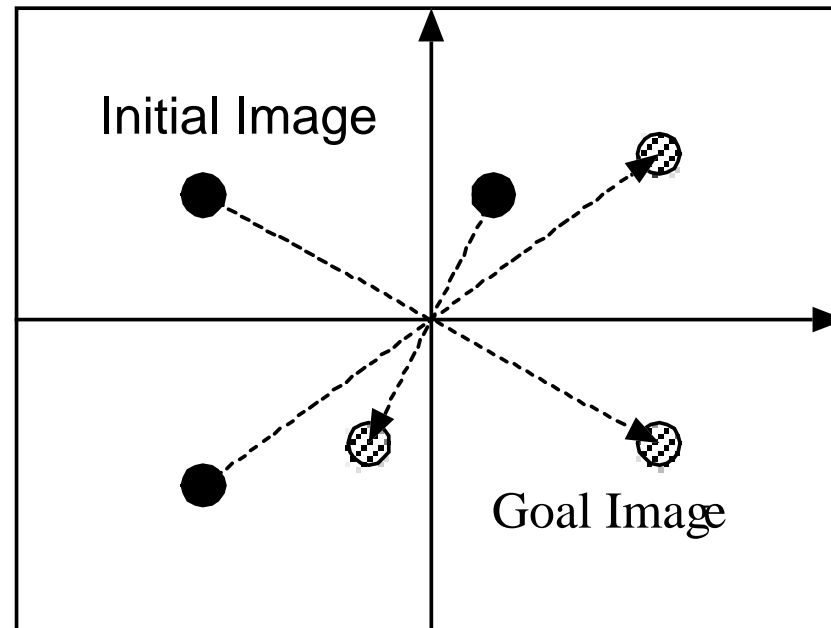
上記の制御則に限定した問題ではない。



不安定な例

画像面の原点に関して点対象の場合(180度回転)

- すべての特徴点は原点へと向かう
- カメラの動きは、オブジェクトから遠ざかる方向



安定領域の局所性

制御則: $u = \lambda J^\dagger(\xi_d - \xi)$

- 特徴量を冗長にする場合 (制御できる変数の数より特徴量を多く選ぶ場合)
- 画像ヤコビアンは縦長
- 一般化逆行列 $J^\dagger = (J^T J)^{-1} J^T$
- 次式をみたす特徴量が存在する

$$\xi_d - \xi \in \text{Ker} J^T$$

- このとき、カメラは停止する

$$u = \lambda J^\dagger(\xi_d - \xi) = 0$$



ポテンシャル

ポテンシャル $V = (\xi_d - \xi)^T (\xi_d - \xi)$

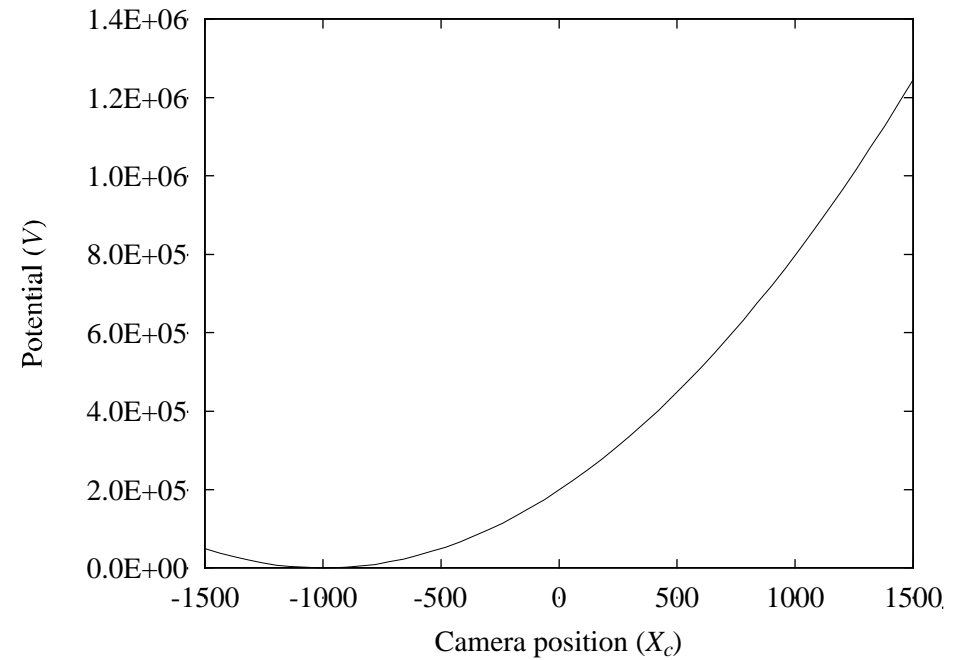
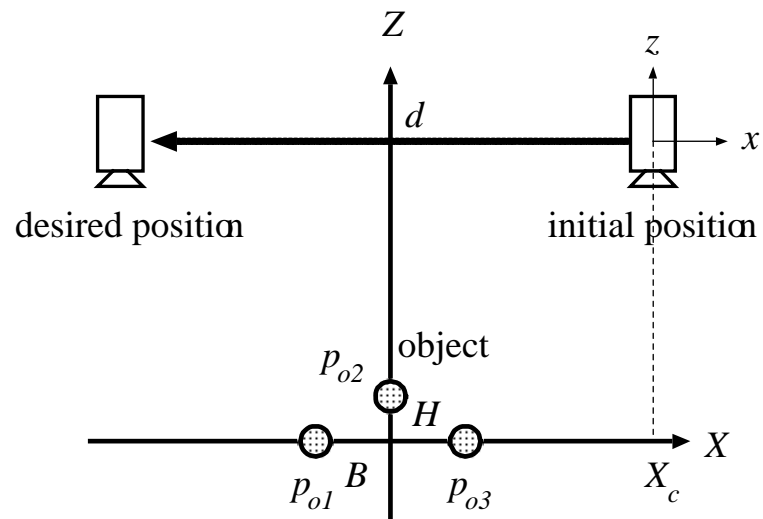
- 制御則 $u = \lambda J^\dagger (\xi_d - \xi)$
- ポテンシャルの最急降下法
- ポテンシャルの極値において更新が停止
- 冗長な特徴量に対するビジュアルサーボは、特徴量空間における経路計画問題

Y. Mezouar and F. Chaumette: Path planning in image space for robust visual servoing, ICRA 2000, ICRA 2001

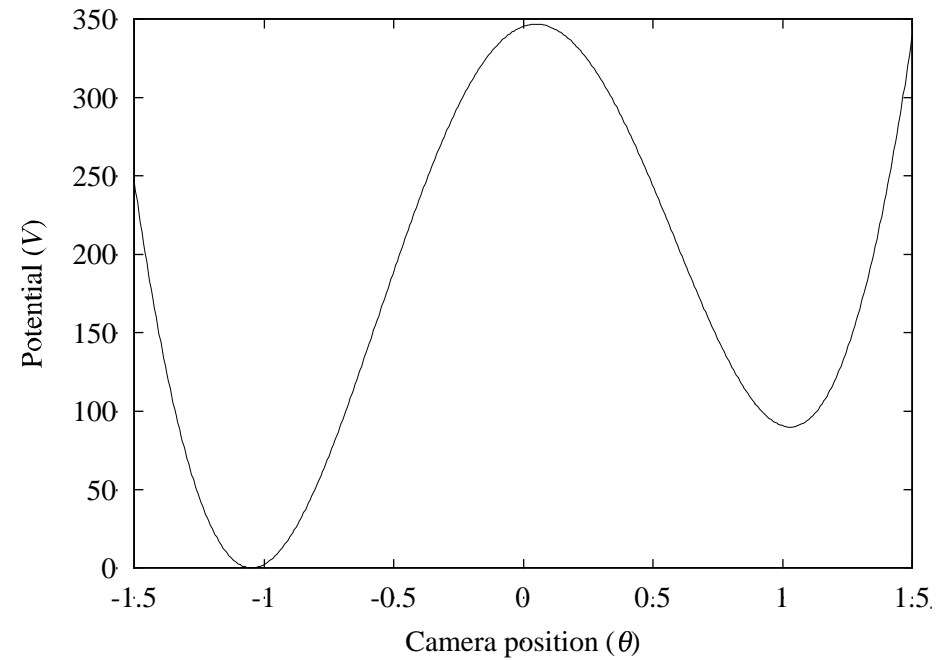
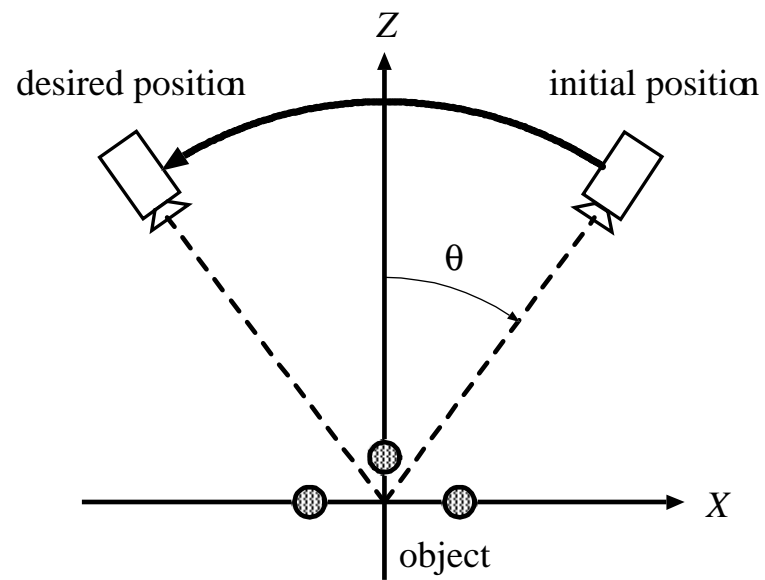
N. J. Cowan and D. E. Koditschek: Planar image based visual servoing as a navigation problem, ICRA 1999



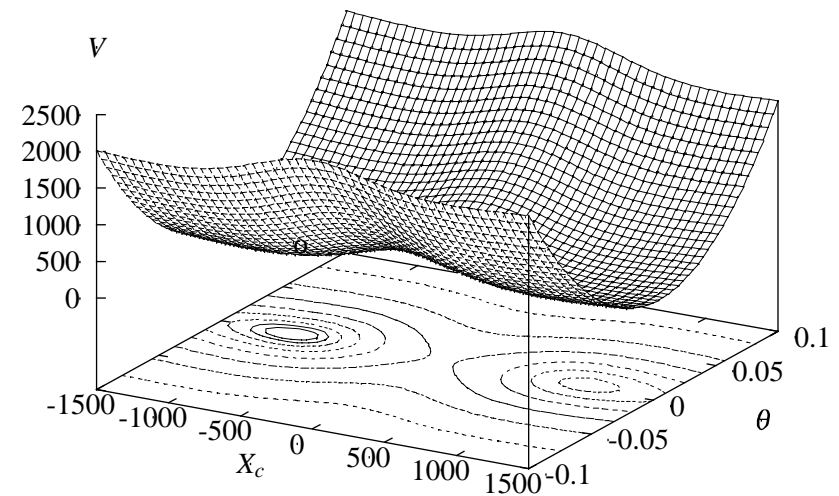
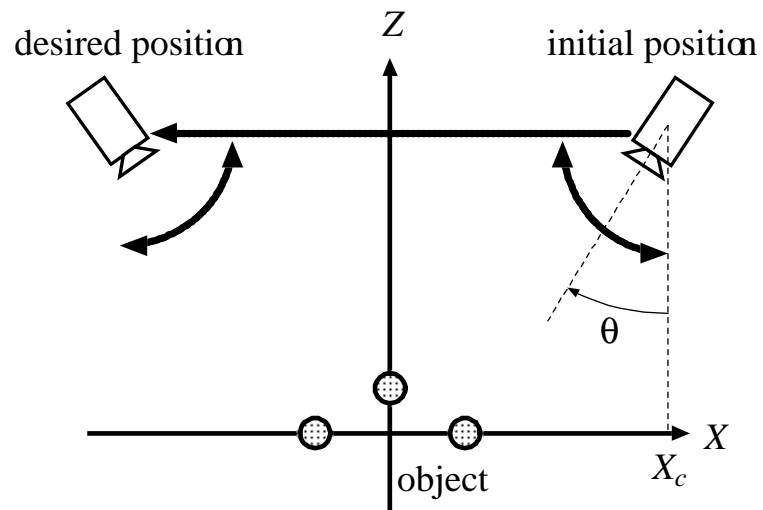
ポテンシャルの例(1自由度並進)



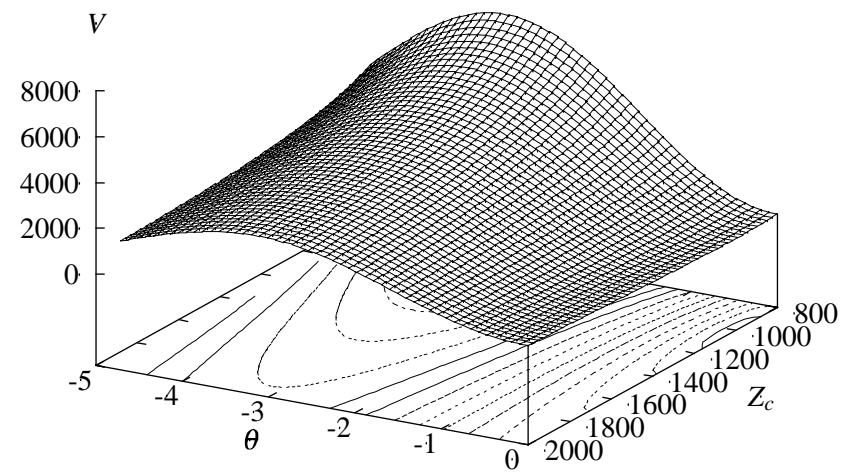
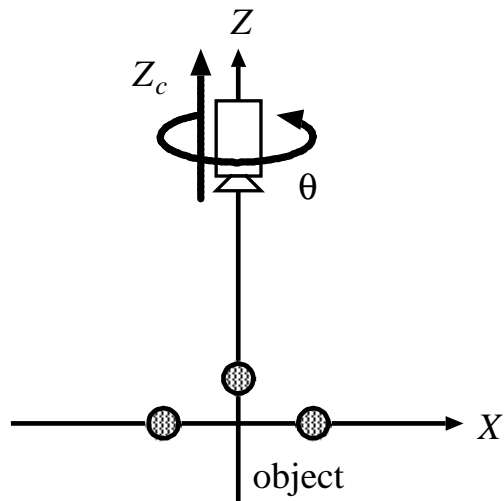
1自由度回転



2自由度直線



2自由度回転



大域的安定化のための制御則

ステレオ

- E. Malis, F. Chaumette and S. Boudet: 2-1/2-D visual servoing, IEEE Trans. RA, 1999.

軌道計画

- N. J. Cowan and D. E. Koditschek: Planar image based visual servoing as a navigation problem, ICRA, 1999
- 橋本浩一、田中浩平、則次俊郎: 視覚サーボにおけるポテンシャル切り替え制御、計測自動制御学会論文誌、2000
- Y. Mezouar and F. Chaumette: Path planning in image space for robust visual servoing, ICRA, 2000

回転と並進の分解

- 出口光一郎、石山墨: 画像ベースド視覚サーボにおける並進と回転の非干渉化によるロボットの最適軌道制御、日本ロボット学会誌、1999
- P. I. Corke and S. Hutchinson: A new hybrid image-based visual servo control scheme, CDC, 2000



Essential Model (Soatto, 1996)

- Dynamical model

$$X_i(t + 1) = R(t)X_i(t) + T(t); X_i(0) = X_{i0}$$

$$Q(t + 1) = Q(t) + n_Q(t), Q \in E$$

$$y_i(t) = x_i(t) + n_i(t), x_i(t) = \pi(X_i(t))$$

$$x_i(t + 1)^T Q(t)x_i(t) = 0 \quad \forall i = 1, \dots, N$$

- Essential Filter

$$x_i(t + 1)^T Q(t)x_i(t) = 0 \Leftrightarrow \chi(t + 1)\hat{Q}(t) = 0$$

$$\oplus : R^{3 \times 3} \times R^{3 \times 3} \rightarrow E$$

$$M_1, M_2 \mapsto Q = pr_{\langle E \rangle}(M_1 + M_2)$$



Essential Model (Soatto, 1996)

- Extended Kalman Filter

$$\hat{Q}(t+1|t) = \hat{Q}(t|t); \quad \hat{Q}(0|0) = Q_0$$

$$P(t+1|t) = P(t|t) + R_Q; \quad P(0|0) = P_0$$

$$\hat{Q}(t+1|t+1) = \hat{Q}(t+1|t) \oplus L(t+1)\chi(t+1)\hat{Q}(t+1|t)$$

$$P(t+1|t+1) = \Gamma(t+1)P(t+1|t)\Gamma^T(t+1) \\ + L(t+1)R_n(t+1)L^T(t+1)$$

$$L(t+1) = -P(t+1|t)\chi^T(t+1)\Lambda^{-1}(t+1)$$

$$\Lambda(t+1) = \chi(t+1)P(t+1|t)\chi(t+1) + R_n(t+1)$$

$$\Gamma(t+1) = I - L(t+1)\chi(t+1)$$



Snake

- Parametric Snake Model

$$E = \int_0^1 \frac{1}{2} [\alpha |x'(s)| + \beta |x''(s)|^2] + E_{ext}(x(s)) ds$$

$$E_{ext}(x, y) = -|\nabla I(x, y)|^2$$

$$E_{ext}(x, y) = -|\nabla [G_\sigma(x, y) * I(x, y)]|^2$$

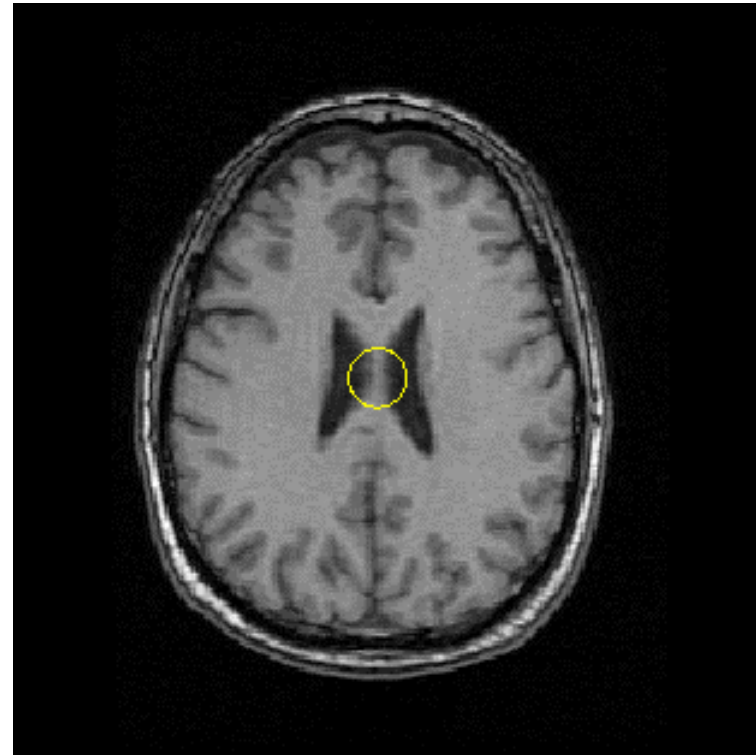
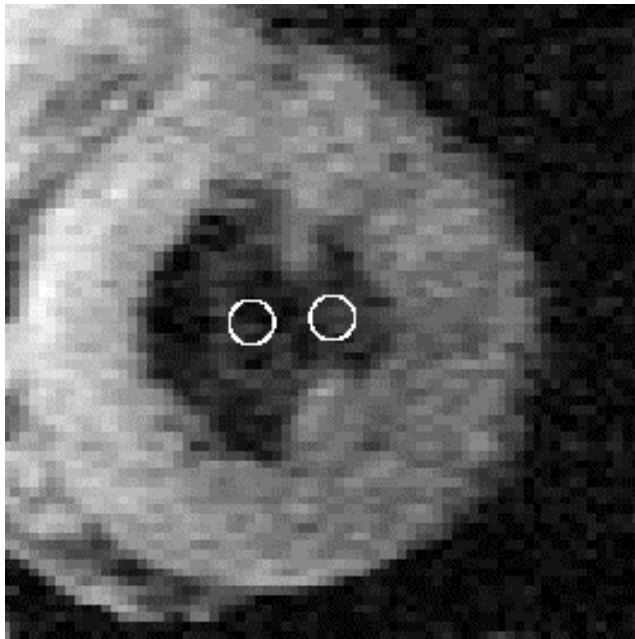
- A snake that minimizes E

$$\alpha x''(s) - \beta x''''(s) - \nabla E_{ext} = 0$$

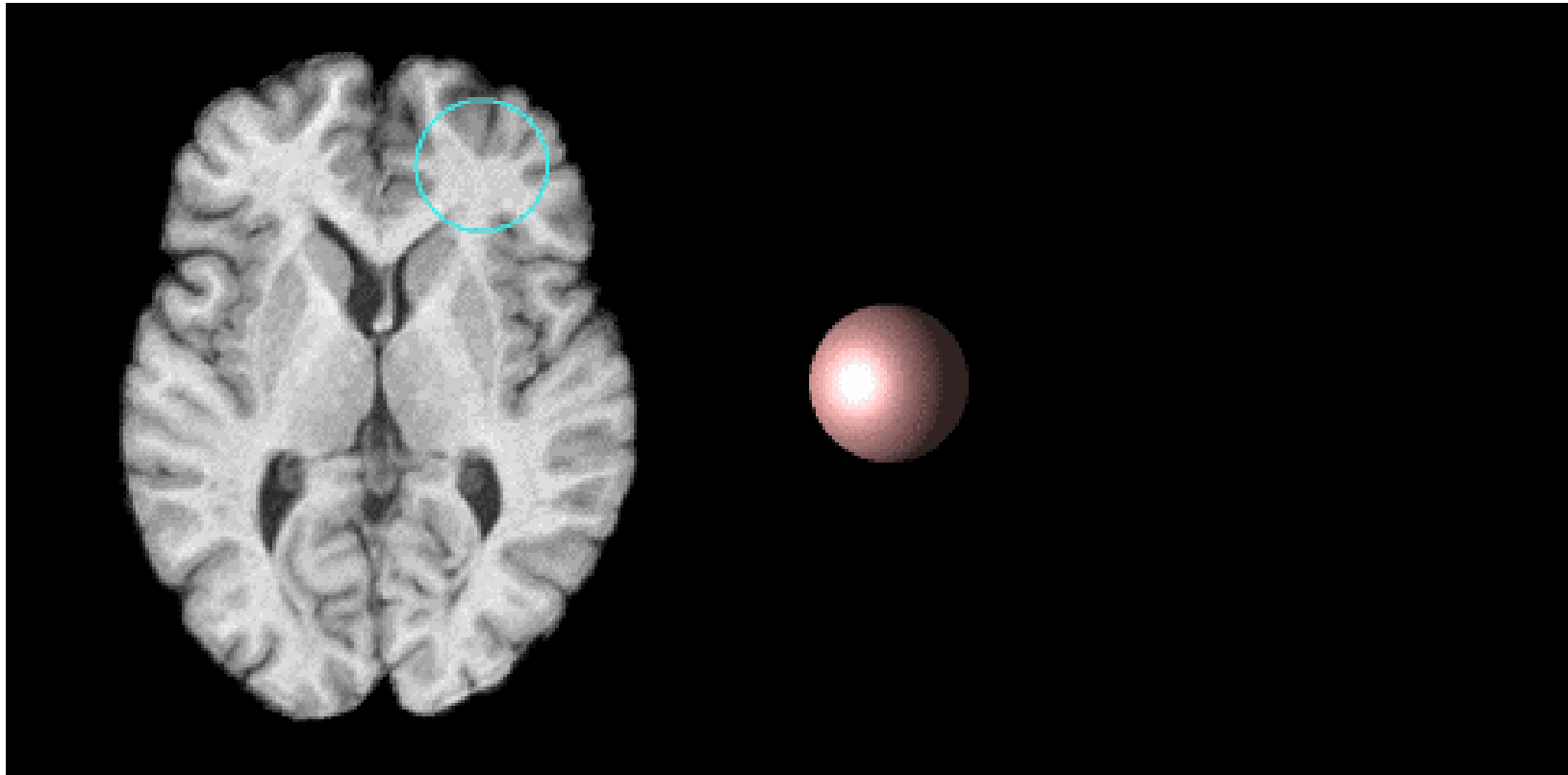
- How to design E_{ext}



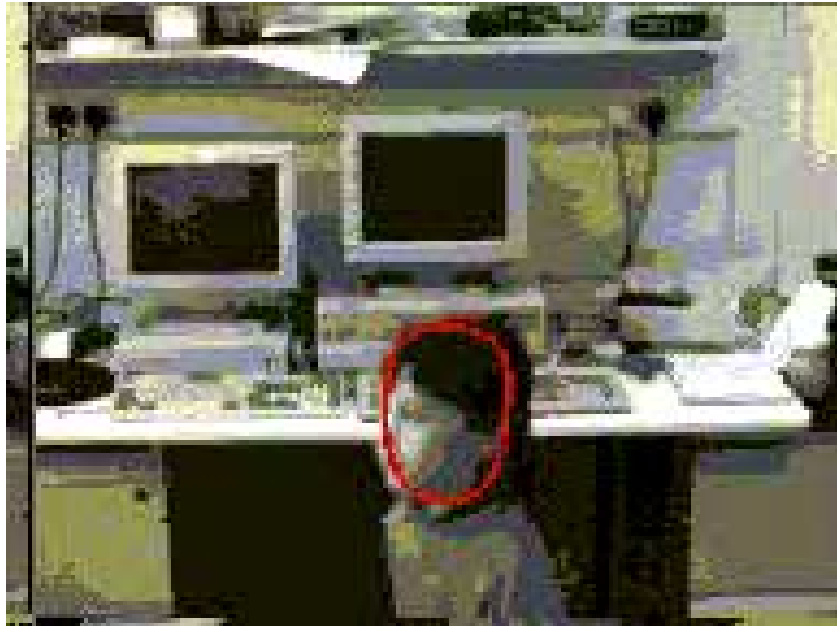
Example of Snake



Example of Snake



Example of Snake



Example of Sanke



Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

研究テーマ

ロボットは人を超えられるか？

人間の限界
その理由
それを超えるためには



Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

知能システムの基本構造

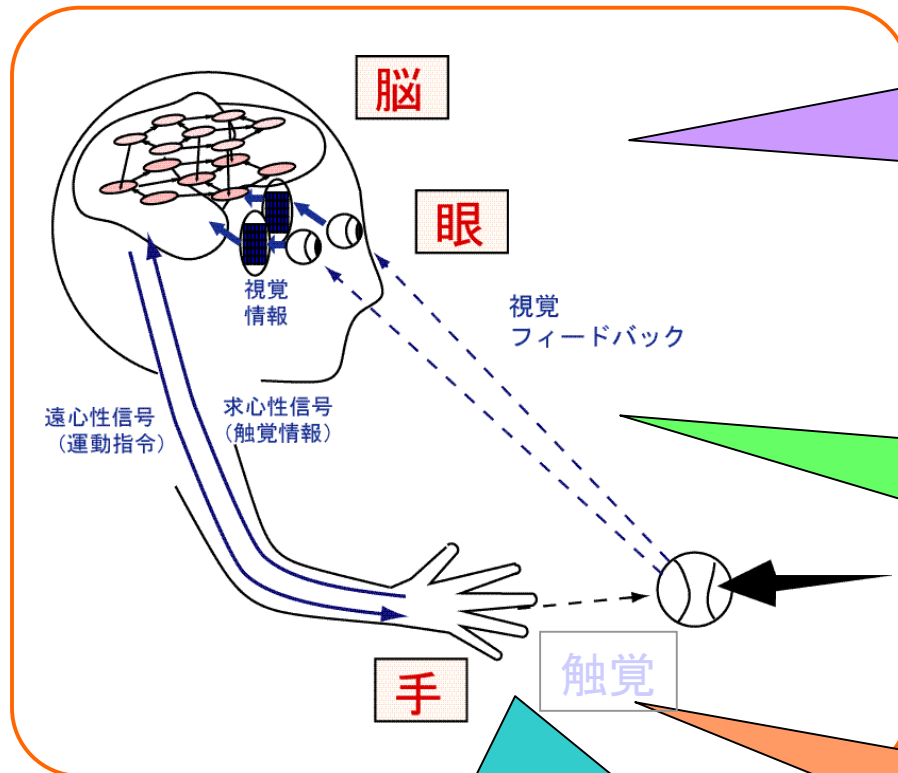


知能システムにおける情報の流れを解析
ロボットと人間の特性の違い

- ◆人間の特性をロボットの目標にしてよいのか？
- ◆人間を超えるロボット＝スーパーロボットは実現できるのか？



人間の認識行動系



脳

- ニューロンによる並列処理
- ・超並列処理
- ・遅い神経伝達速度(20m/s)
- ・高次の学習能力

視覚

- ・低解像度+能動性
- ・遅い処理速度(30fps)
- ・優れたパターン処理能力

触覚

- ・大面積+柔軟性+能動性
- ・低精度、低速処理
- ・優れたパターン処理能力

手

- ・高出力+軽量+多自由度
- ・低速(150km/h)、低精度
- ・器用さ、学習能力



ロボットにおける情報処理

並列処理

LSIプロセッサによる並列処理

- 低い並列性+高い粒度
- 高速な情報伝達速度
- 論理・計算主体の処理

ビジョンチップ

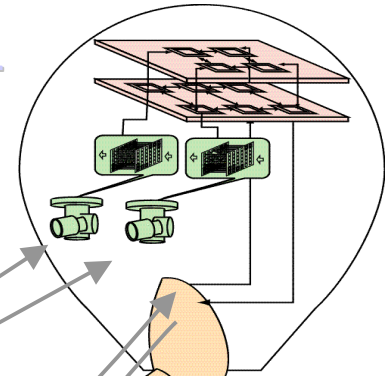
- 高解像度+(能動性)
- 高速な処理能力(1000fps)
- フィードバックセンサ

触覚・力覚センサ

- 小型、非柔軟性、(能動性)
- 高精度、高速な処理
- フィードバックセンサ

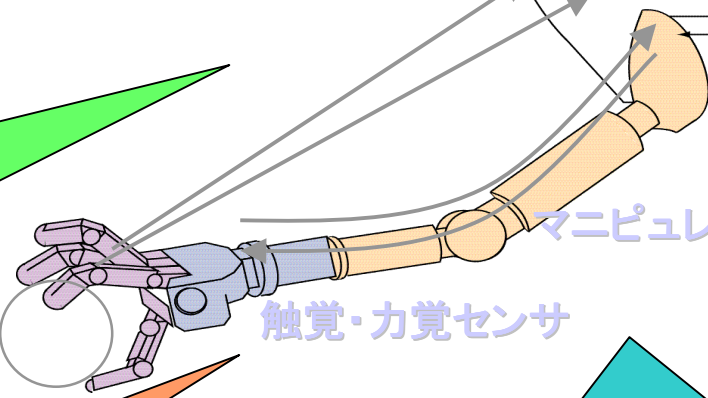
並列プロセッサ

視覚センサ



マニピュレータ

触覚・力覚センサ



マニピュレータ

- 低出力、大重量、少ない自由度
- 高速、高精度
- 剛体性、(学習能力)

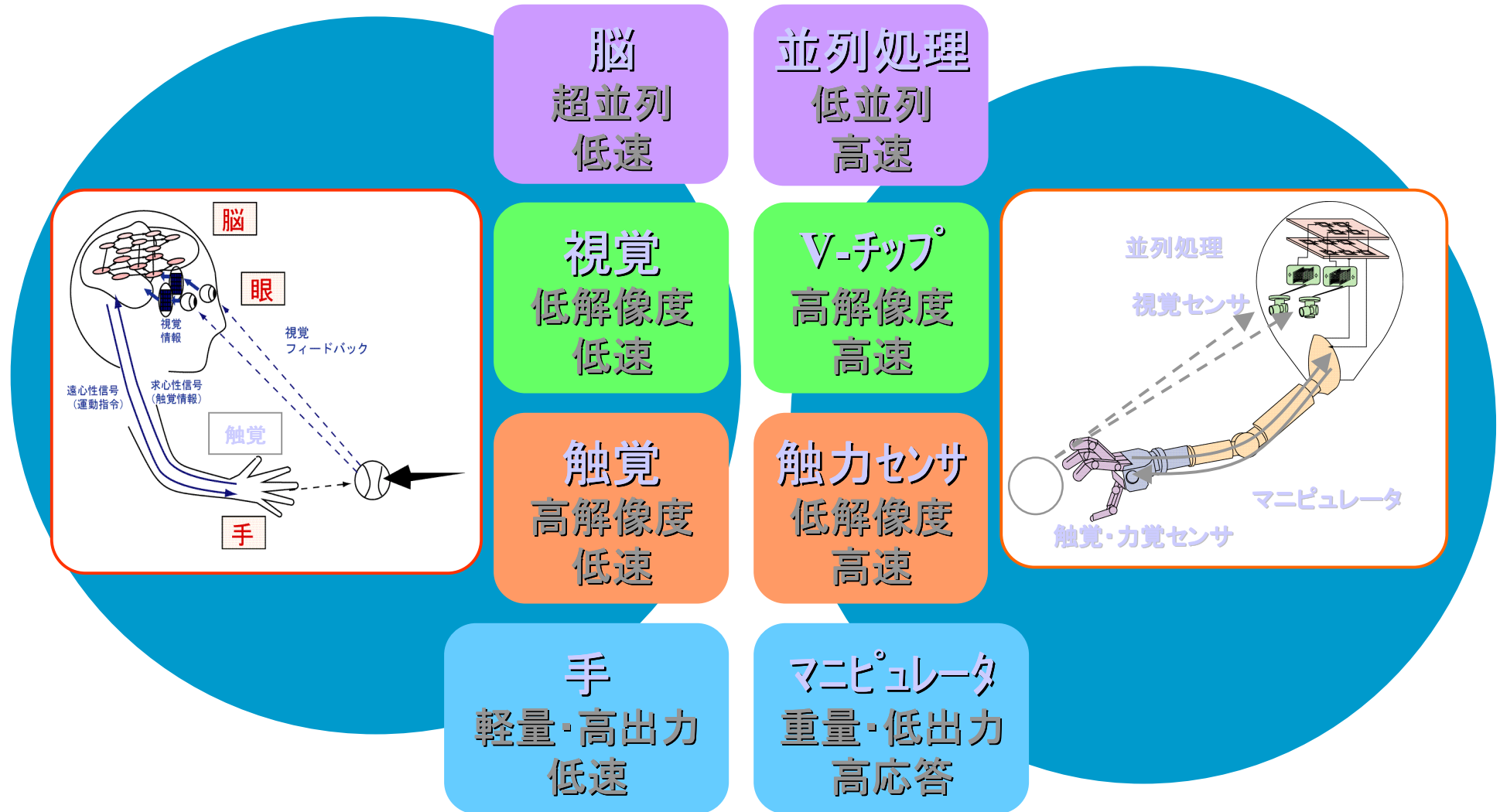


Ishikawa Hashimoto Laboratory

UNIVERSITY OF TOKYO

<http://www.k2.t.u-tokyo.ac.jp/>

人とロボットとの違い



Ishikawa Hashimoto Laboratory

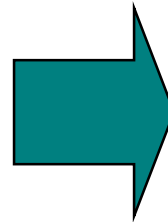
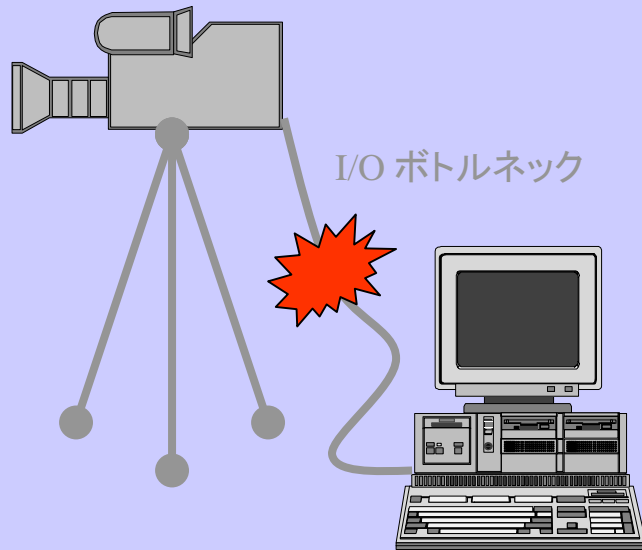
UNIVERSITY OF TOKYO

<http://www.k2.t.u-tokyo.ac.jp/>

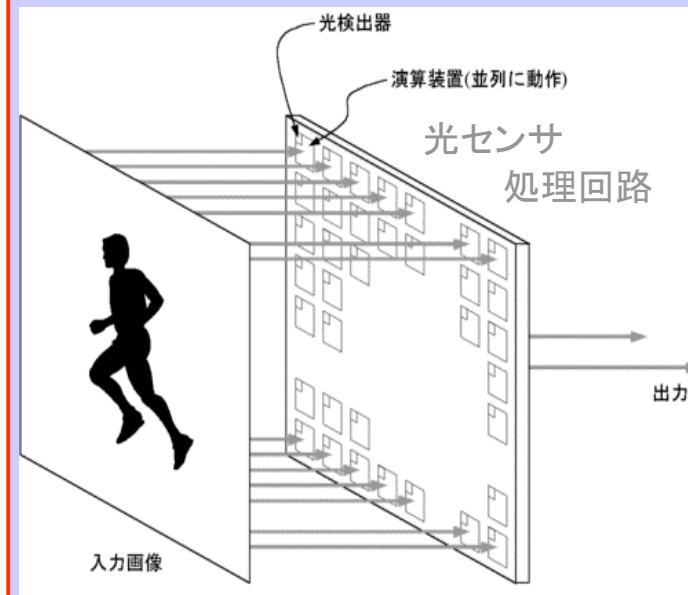
人間の目を超える資格の実現

走査によるフレームレートの制限

ビデオカメラ



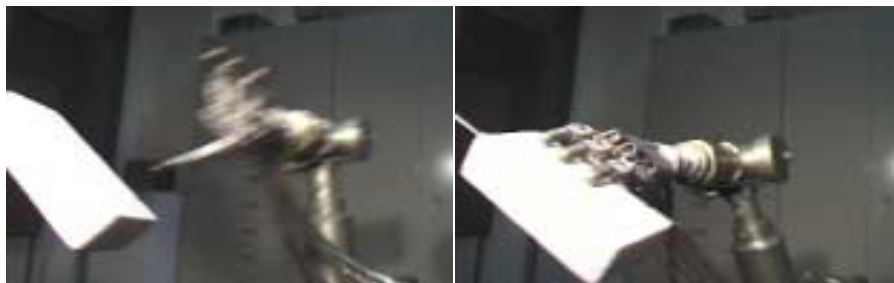
光センサと汎用PEを一体化



汎用デジタルビジョンチップの応用例

汎用デジタルビジョンチップのスケールアップモデル **SPE-256** により実現

1ms 感覚統合システムを用いた
高速移動物体把握の例



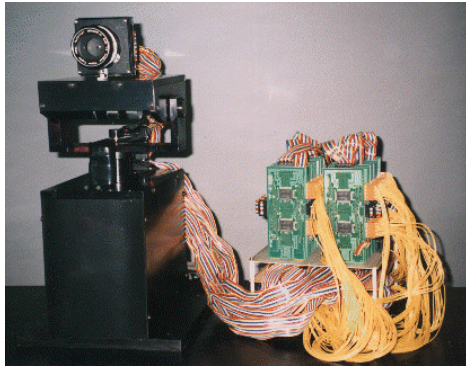
1ms 対象追跡システムを用いた
ジェスチャー認識の例



Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

汎用デジタルビジョンチップのワンチップ集積化

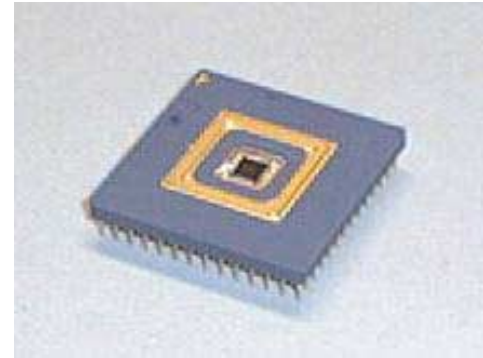
スケールアップモデル



集積化



ワンチップVLSI



目的

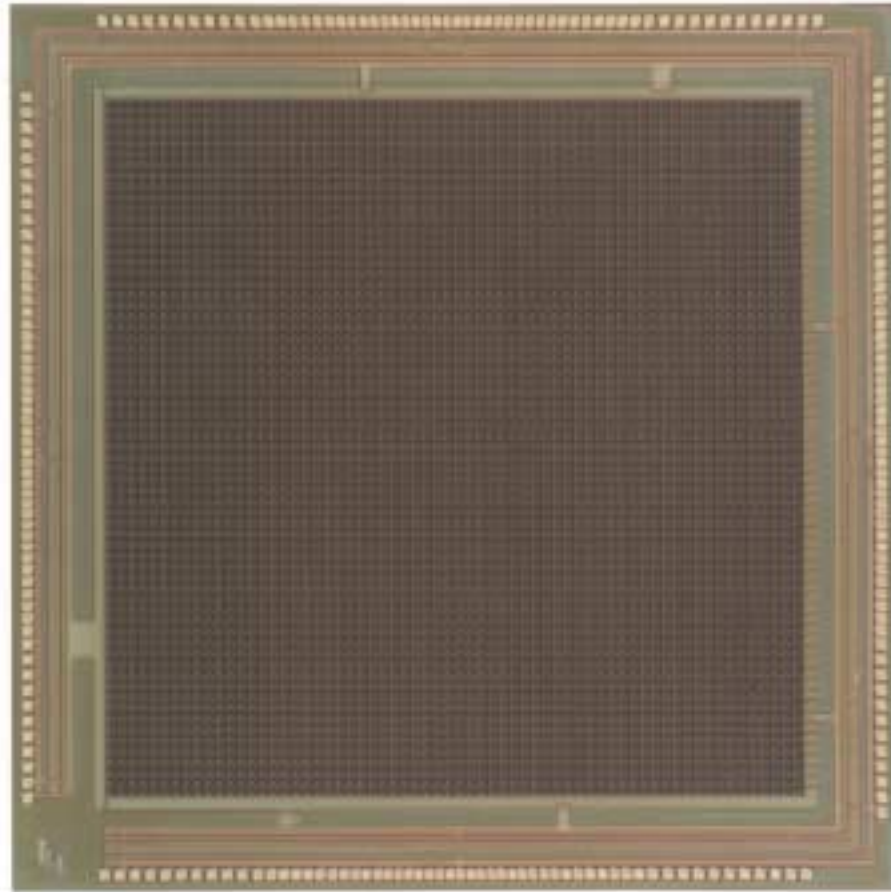
半導体集積化技術を用いて汎用デジタルビジョンチップのワンチップ集積化を図る。

→ コンパクトな設計が必要



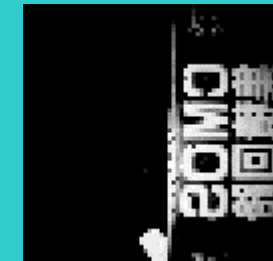
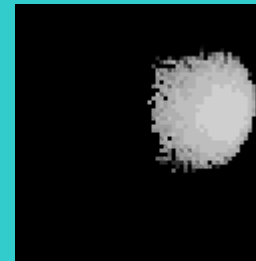
Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

スーパージョンチップ V4.3

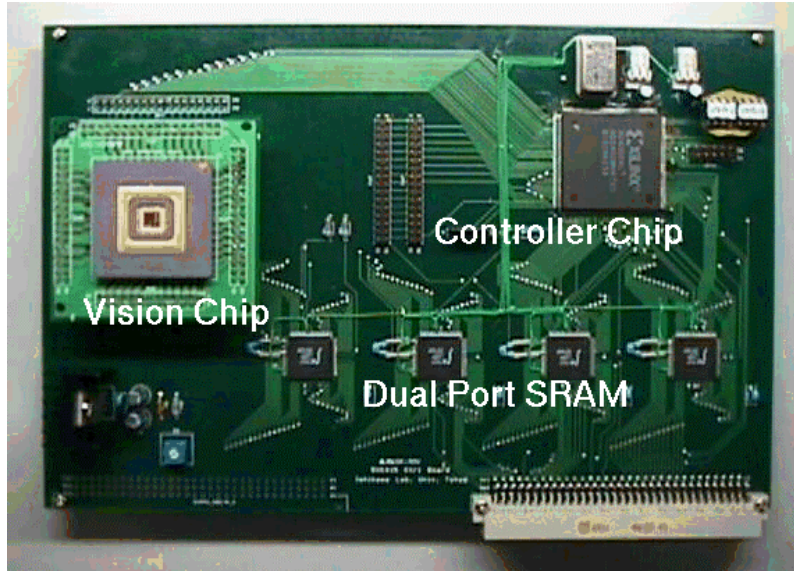


- チップ総面積:
 $8.7 \times 8.7 \text{ mm}^2$
- 画素数: $64 \times 64 = 4096$

撮像実験の結果



ビジョンチップシステムボード



- 以下の要素をワンボードに搭載
 - ビジョンチップ(16×16)
 - コントローラチップ
 - デュアルポートSRAM
 - etc.
- ビジョンチップはドータボードに実装

コントローラ部の仕様

- コントローラ基本部分, 総和ユニット, 1次元メモリ アレイ等を集積
- Verilog-HDL にて設計
- XILINX社のFPGA XC4052XL を使用
- 動作周波数 : 12.827MHz
- 基本構成ブロック(CLB)数: 1616 (83%)
- 換算ゲート数 : 35158

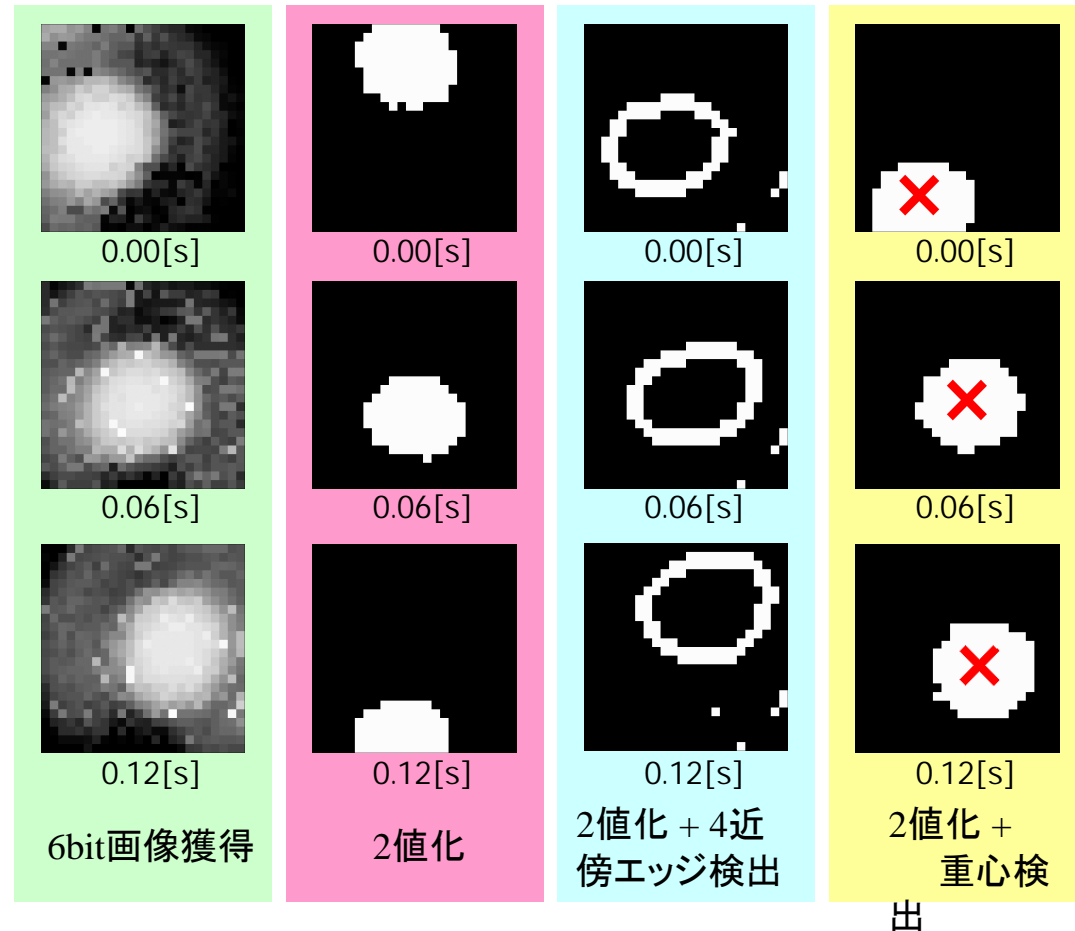


評価結果

実験方法

- システムに光学系を取り付け、レンズ前方 20cm~50cm の位置でクリップライトをランダムに動かし、撮像および処理を行った。
- PDの蓄積時間を 1ms、システムの動作周波数を8MHzとした。

Program	steps	time(μ s)
4-neighbor edge detection	4	2.0
4-neighbor smoothing	8	4.0
4-neighbor erosion, dilation	5	2.5
Image dumping	48	24
Center of gravity	84	42
	462	231



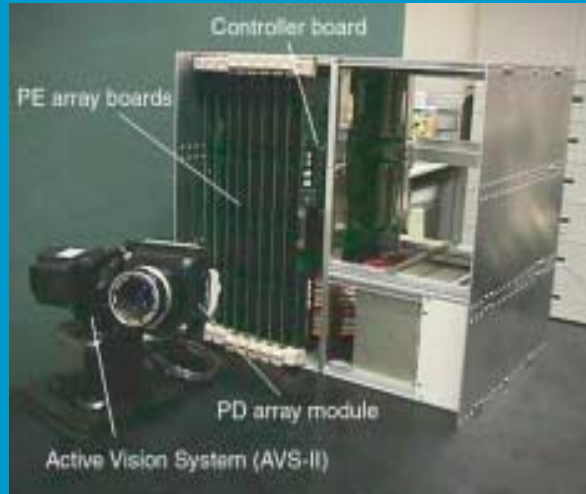
Ishikawa Hashimoto Laboratory

UNIVERSITY OF TOKYO

<http://www.k2.t.u-tokyo.ac.jp/>

ビジョンチップシステムの展開

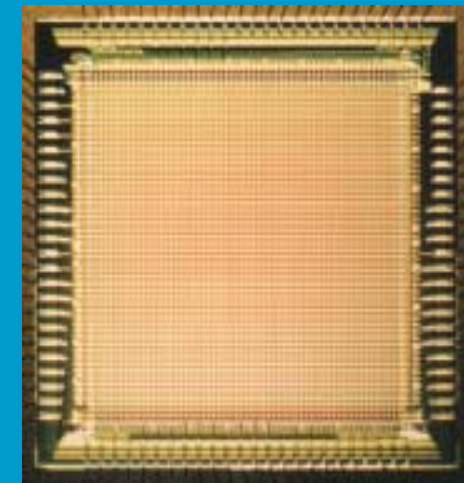
列並列ビジョンシステム Column Parallel Vision System



〔 浜松ホトニクスとの
共同研究 〕

- ・汎用用途
- ・高解像度・中速
- ・高感度

〔 日本プレシジョン・サーキット
との共同研究 〕



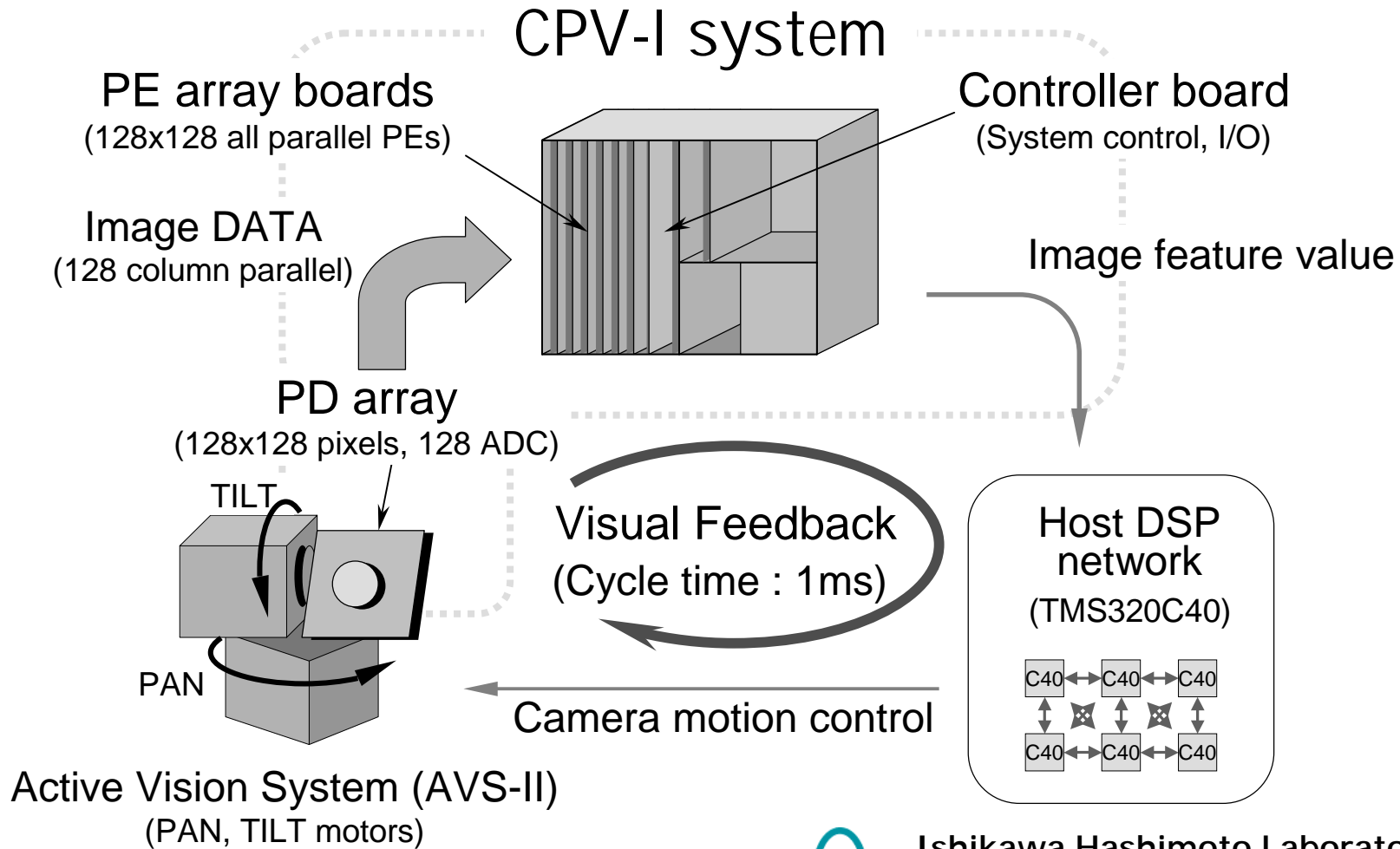
- ・特定用途
（モーメント演算）
- ・高解像度・高速
- ・高集積化

ターゲットトラッキングチップ

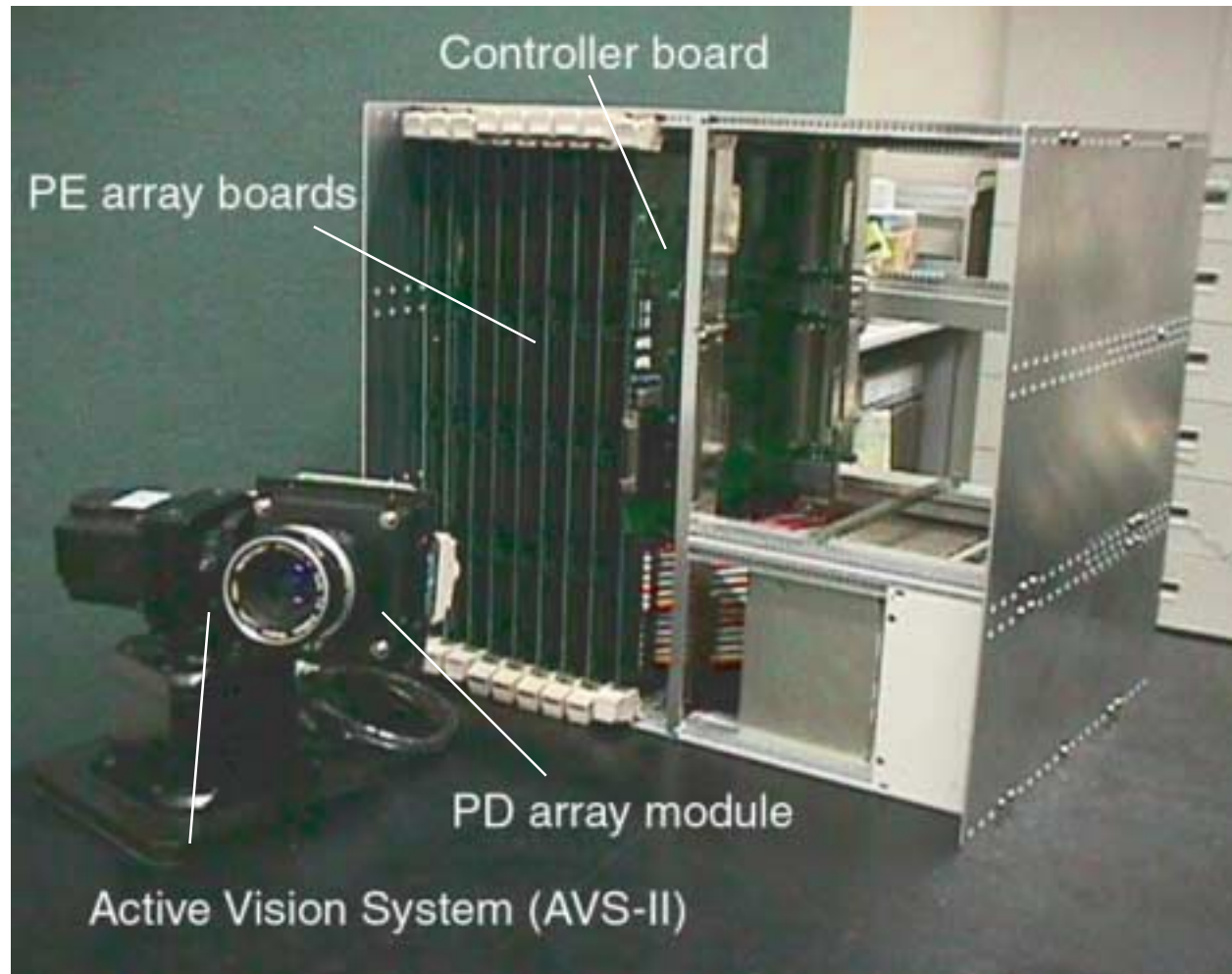


Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

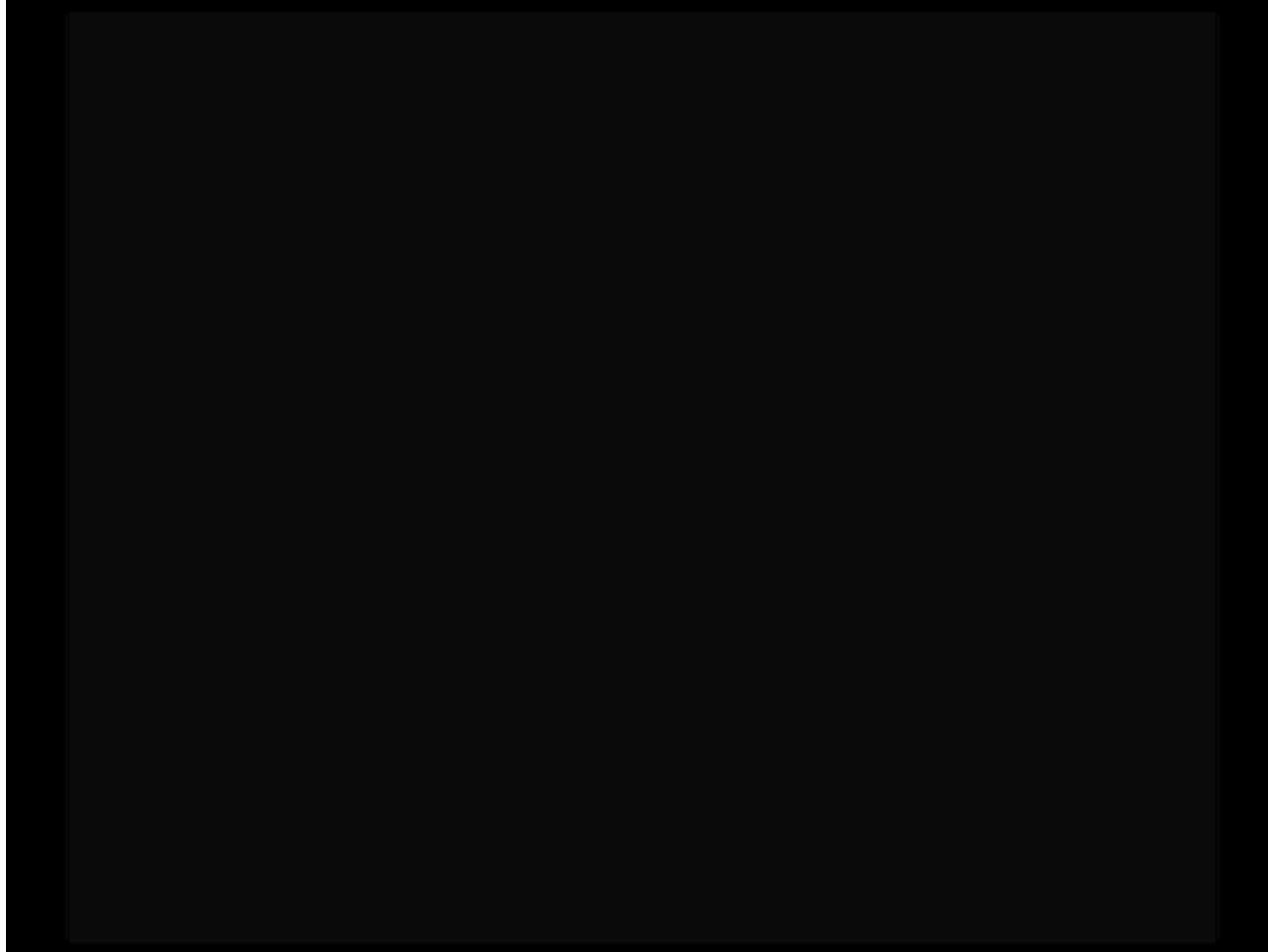
アクティブビジョンシステムの全体構成



超高速・超並列ビジョンシステム (CPV-I)



アクティブビジョンビデオ



Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

センサフュージョンプロジェクト

感覚運動統合システムの構築

人間と同等の機能を持つロボットシステムの構築

- 1ms感覚運動統合システム
- 高速把握システム

高速視覚応用

高速視覚処理システムの開発
高速視覚の特性を利用した動的制御

ヒューマンインタフェース

センサフィードバックを用いた
能動的な人間の作業支援



人間の能力を超えるロボットの実現

—高速センサフィードバックに基づく感覚運動統合システム—

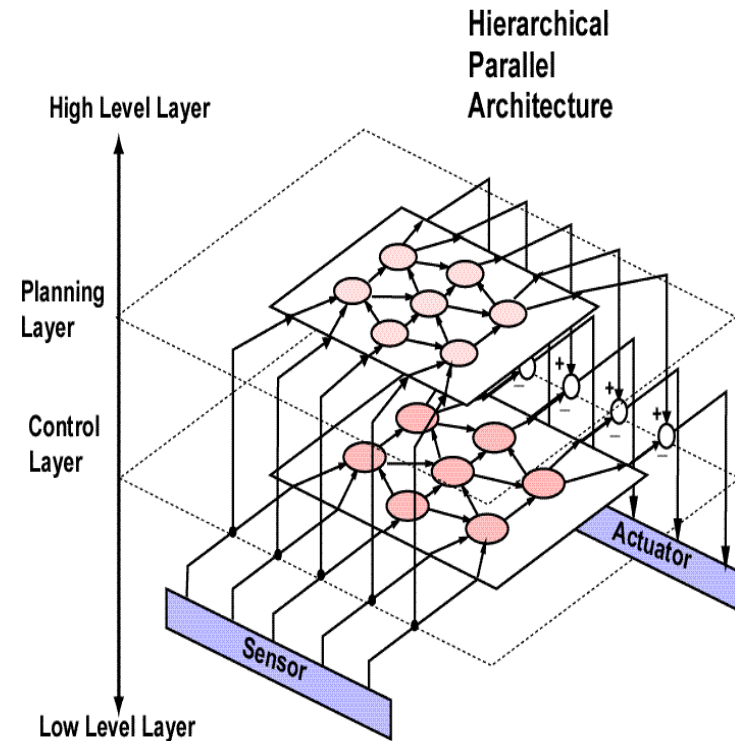
環境の動的な変化に対する**高応答性**
環境の多様な状態に対する**柔軟性**



**高速センサフィードバックに基づく
階層型並列処理アーキテクチャ**

機械の高速な処理能力

➡ 人間を超える処理能力



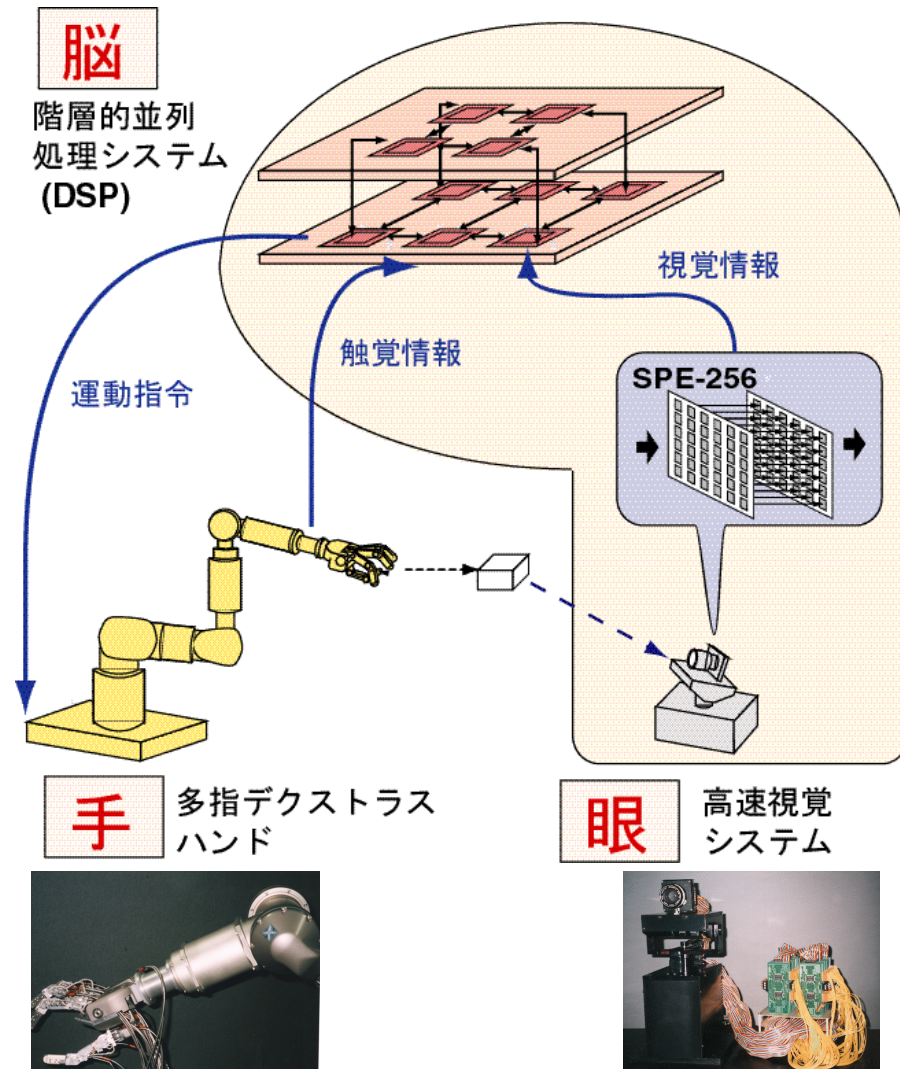
1ms 感覚運動統合システム

DSP階層並列処理システム (複数の並列処理用DSP)

- 多数の並列 I/Oチャンネル
(DAC 32CH, ADC 96CH,
PIO 64bit)
- 高速ビジョンチップシステム
を装備したアクティブビジョン
- カセンサ付4本指多指ハンド
を装備した7軸マニピュレータ



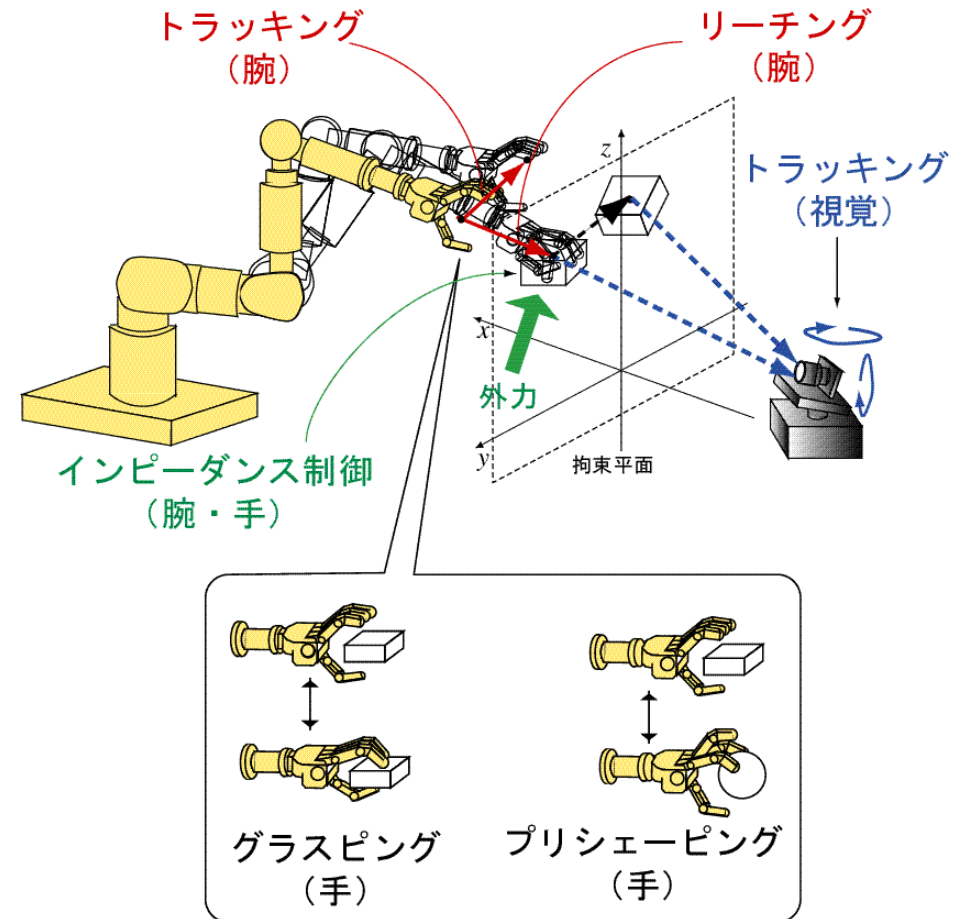
視覚と力覚のフィードバック
を1msで実現



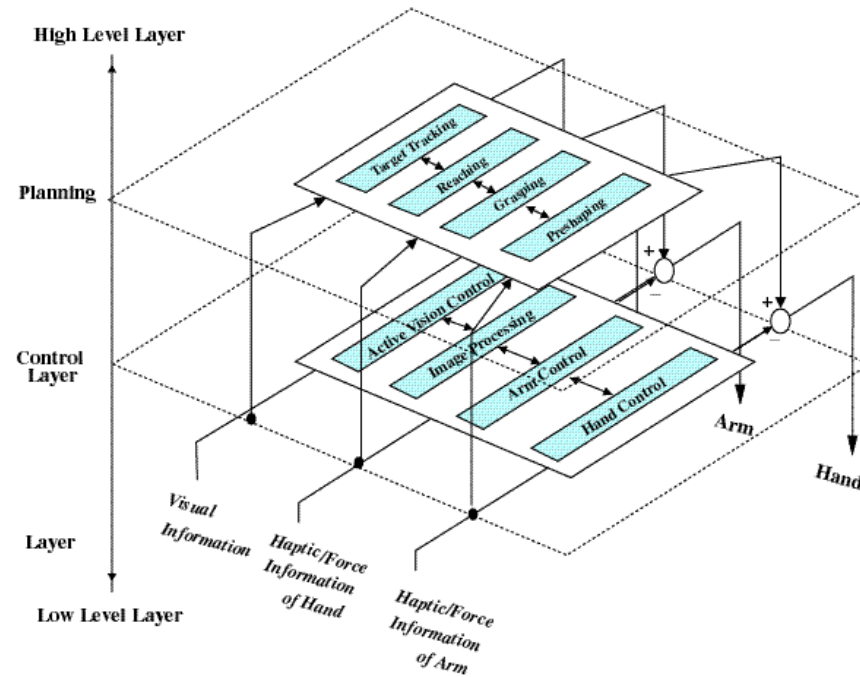
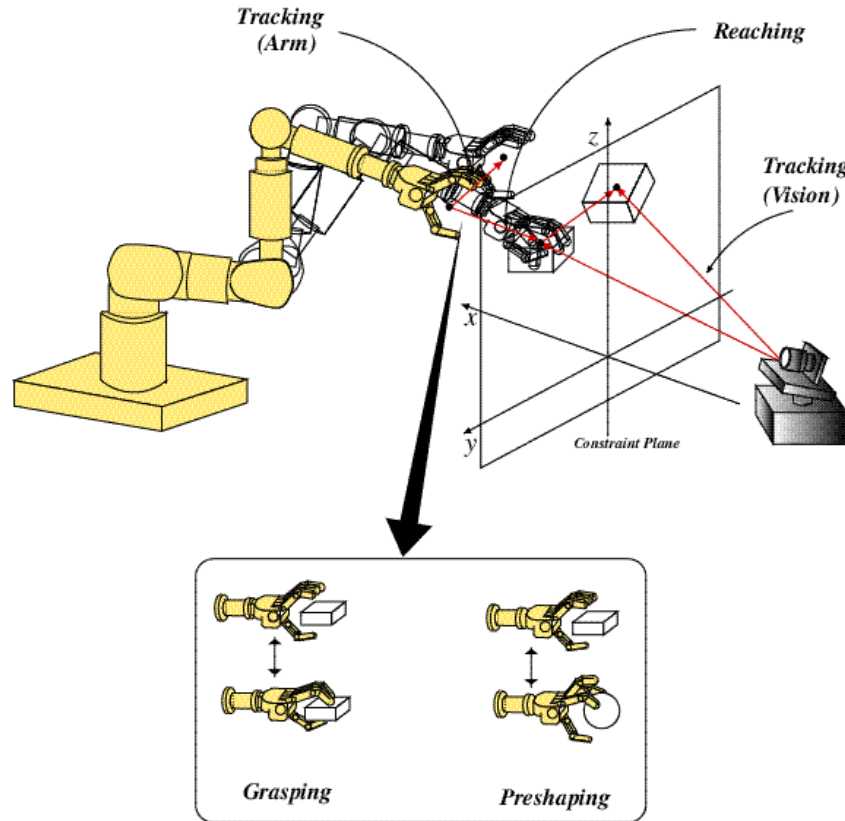
高速把握システム

複数のサブタスクの統合によって把握行動を実現

- **アクティブビジョン**
 - **トラッキング**
対象を視界内に捕える
- **アーム**
 - **トラッキング**
手先を対象に追従させる
 - **リーチング**
手を対象に差し伸ばす
- **ハンド**
 - **グラスピング**
対象を把握する
 - **プリシェーピング**
対象の形状に合わせて手の形状を变化する



高速把握システム階層図



アクティブビジョン
アーム
ハンド

トラッキング
トラッキング・リーチング
グラスピング・プリシェーピング

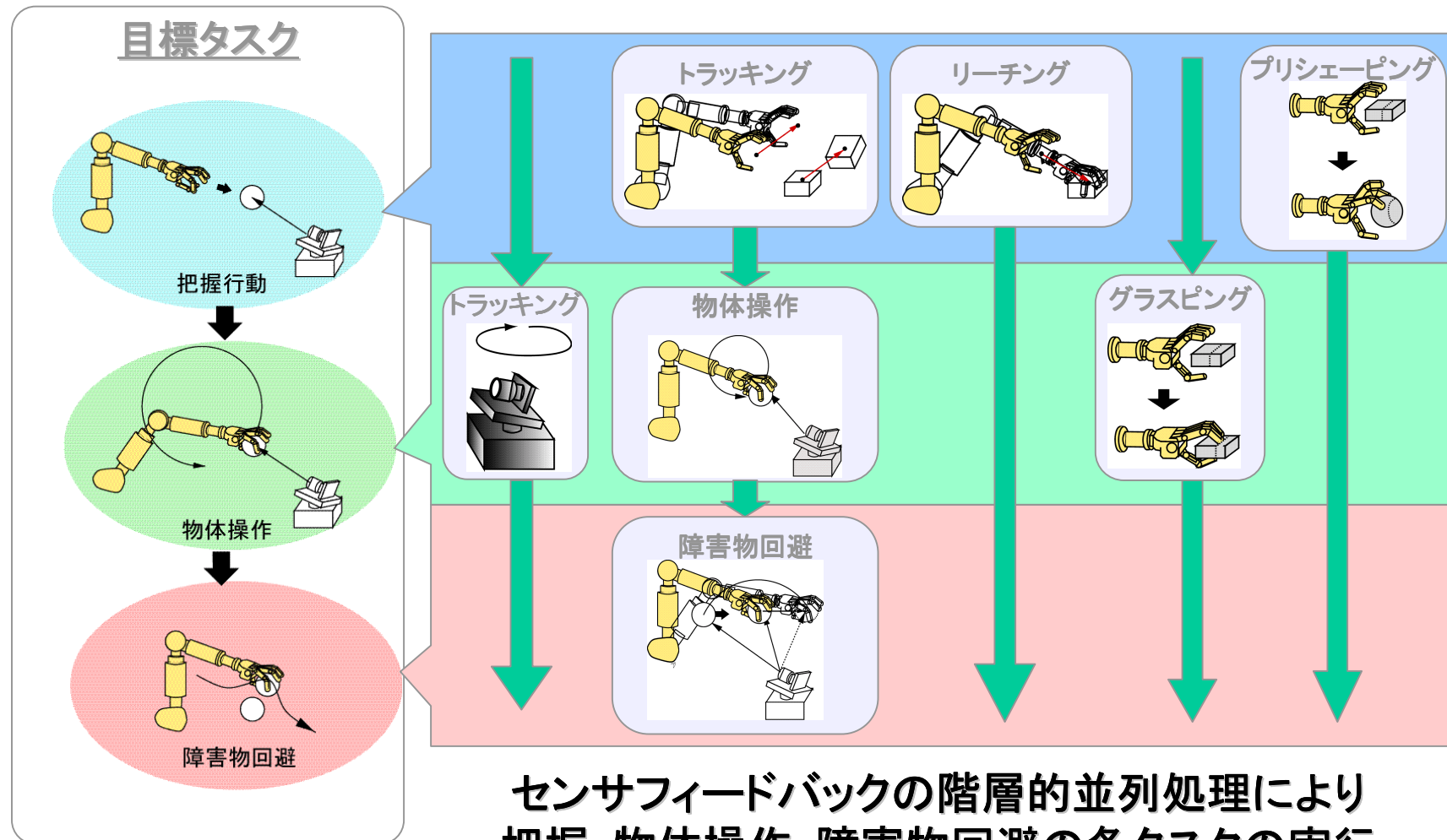


Ishikawa Hashimoto Laboratory

UNIVERSITY OF TOKYO

<http://www.k2.t.u-tokyo.ac.jp/>

階層的並列処理に基づく センサフィードバックの統合

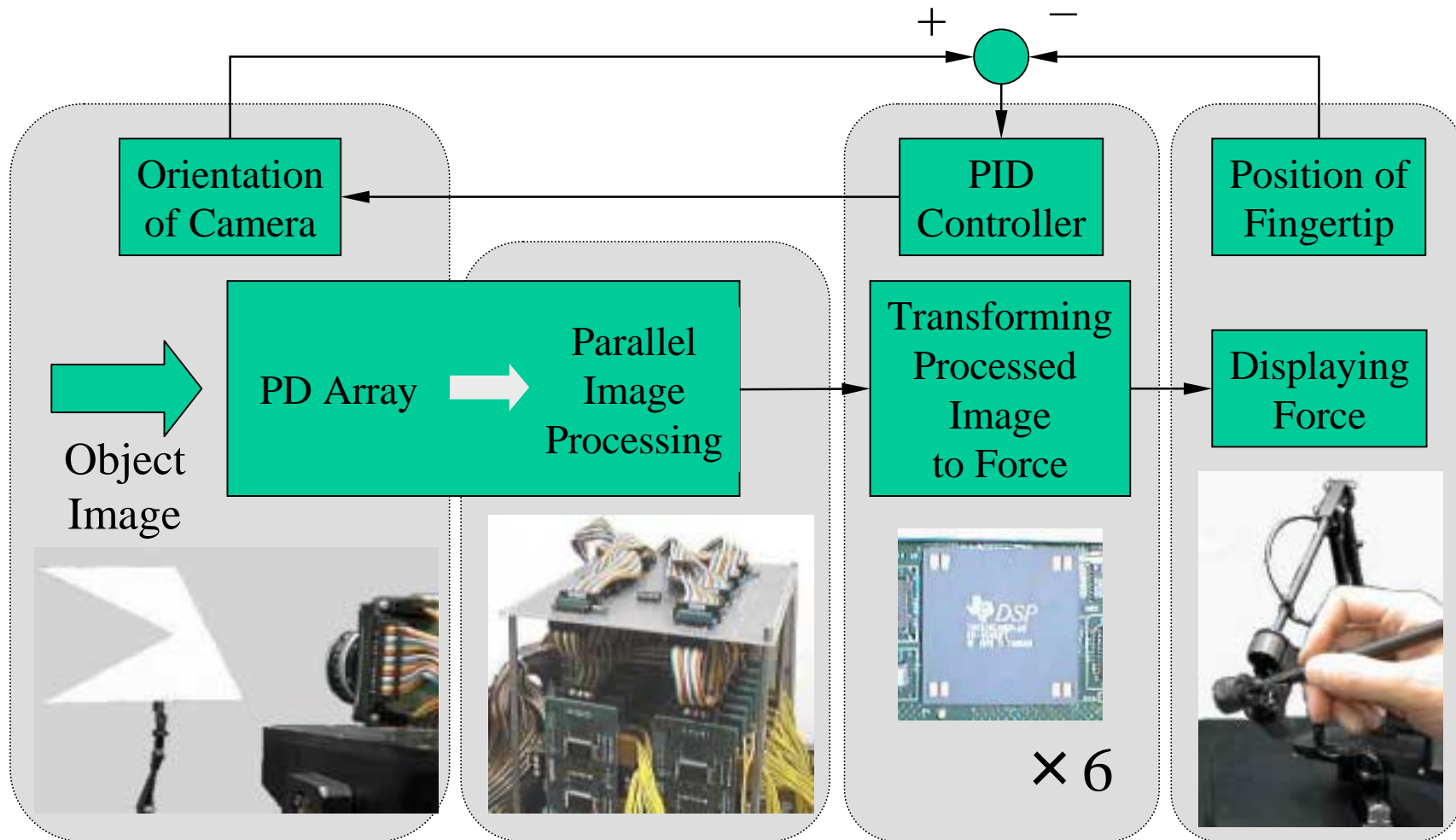


高速把握行動ビデオ



Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

視触覚モダリティ変換システム



Active Vision

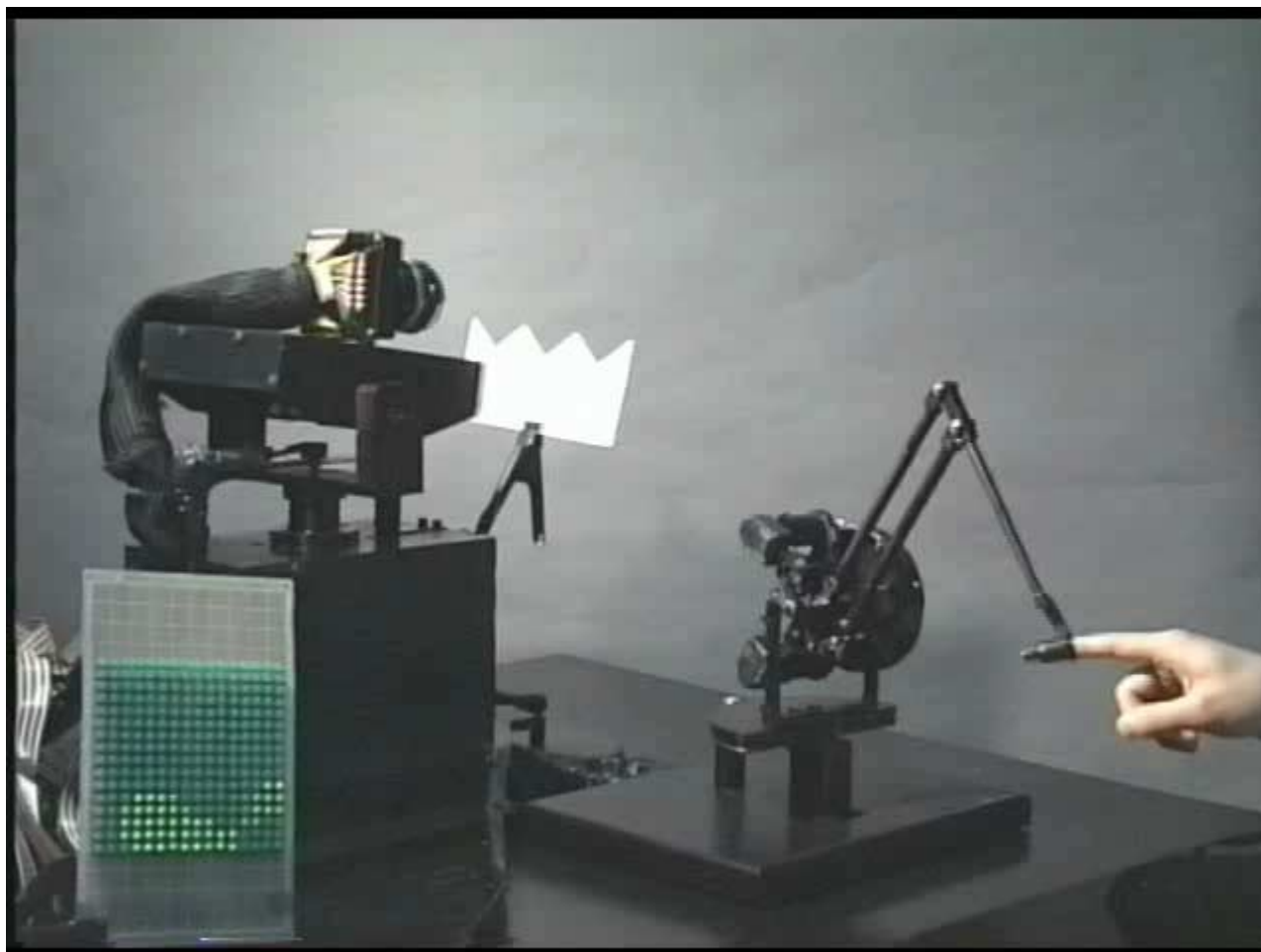
SPE-256

DSP System

PHANToM

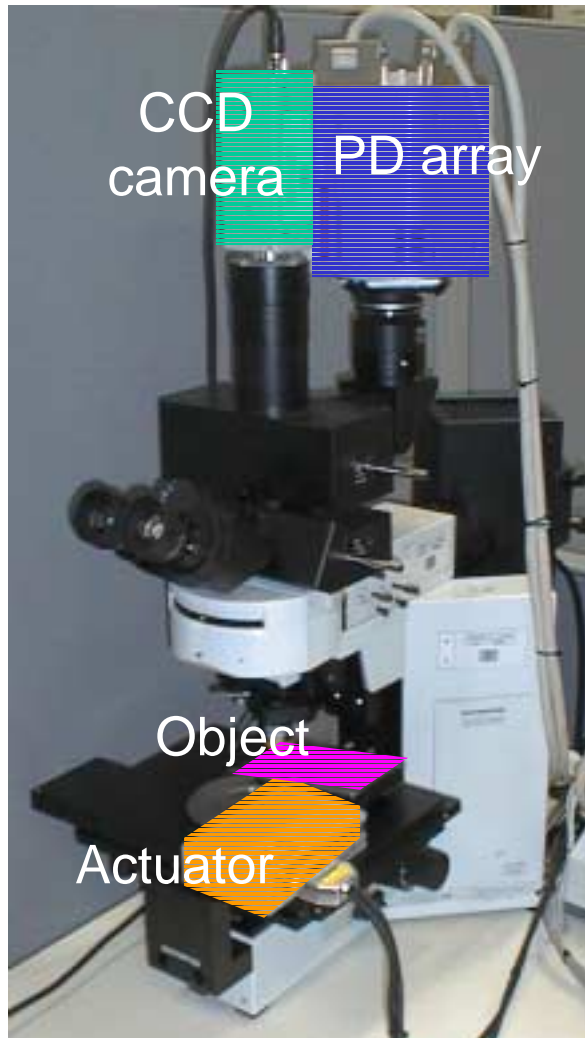


モダリティ変換ビデオ

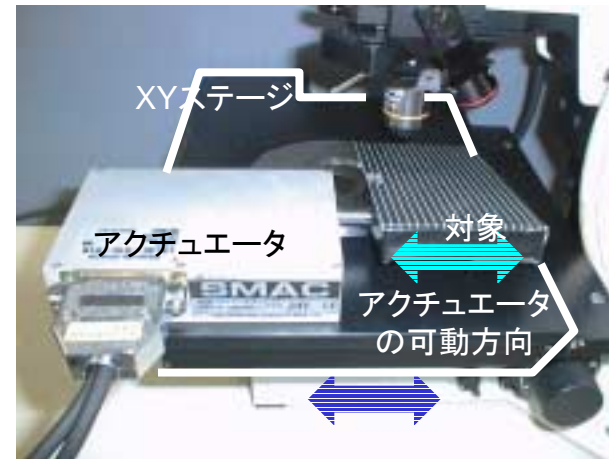


Ishikawa Hashimoto Laboratory
UNIVERSITY OF TOKYO
<http://www.k2.t.u-tokyo.ac.jp/>

新HMVFシステム



- 16 × 16 PD array(浜ホト)+³S PE [高速視覚]
- LAL30015-51(SMC) [リニアコイルアクチュエータ]
- TMS320C40(TI) × 3 [制御プロセッサ]
- BX50WI(OLYMPUS) [顕微鏡]



人による外乱の方向



Ishikawa Hashimoto Laboratory

UNIVERSITY OF TOKYO

<http://www.k2.t.u-tokyo.ac.jp/>

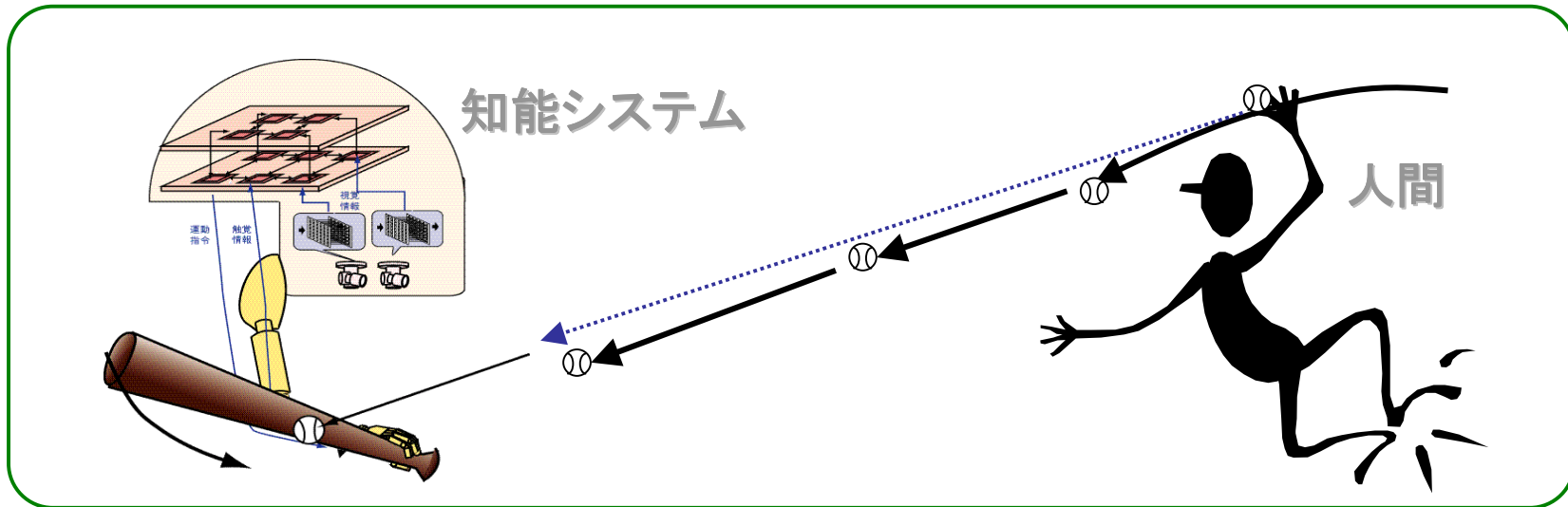
マイクロビジュアルフィードバックビデオ

Without Tracking Control



ロボットシステムの将来

—感覚運動統合システムの目指すもの—



人間を超える能力

- 認識能力 ⇒ 1msレベルでの高速な認識能力
- 運動能力 ⇒ 人間とスポーツで対戦可能な運動能力
- 学習能力 ⇒ ロボットの特性を応用した高次学習能力



知能システムの将来

人間の優れた認識・行動能力をシステムの特성에合わせて実現

センサ系(入力)

人間を超えるセンシング能力
(スーパーセンサ)

運動系(出力)

高速マニピュレータの制御
(スーパーアクチュエータ)

処理系

人間を超える並列処理能力
(スーパーコンピューティング)

- コンピューティング
- 制御
- 信号処理
- システムアーキテクチャ
- サイバネティクス
- 知能化センサ
- 集積化知能システム
- 画像認識
- 音声処理
- 人工現実感
- テレロボティクス

