解 説

# ロボットの能動視覚

Active Vision for Robots

出 口 光一郎* *東京大学工学部計数工学科
Koichiro Deguchi* *University of Tokyo

## 1. はじめに—ロボットの視覚

ロボットのための視覚，あるいはロボットが視覚を持つ…，と考えたとき，単なる画像処理の技術との際立った違いが思い起こされる．ロボットという語感からは，「動き回る」というニュアンスは切り離せない．つまり，ロボットの視覚は，カメラ自身がロボットに搭載され共に運動する（Eye IN Hand）ことを言う（図 1）．その結果，ロボット視覚では，カメラ自身が動くことによってできる動画像を扱うことになる．

動いている対象を静止カメラで捉えた動画像と，自分が動いて捉えた動画像とでは，実は，その研究の醍醐味がずいぶんと違う．しかも，能動的にカメラの動きをコントロールできるとなると，例えば，対象により認識を容易にしたり効率的にする観測者自身の運動といったものが有り得るのではないか，また，すでに得ている画像からもっと詳細を知るにはどう動くべきか，などといったずっと高級な興味が沸いてくる．単に対象を眺めるコンピュータビジョンの枠を一歩出た認識のための戦略が立てられるかも知れない．

さて，ここで扱う画像は，三次元の空間が投影され二次元に縮退したものである．しかし，実際，単一の画像から様々な三次元情報が読み取れる [1]．さらに，我々の視覚は常に動画像をここで言う能動的に得ている．二次元の画像から三次元の空間を認識することのできる仕掛けの多くは，インテリジェンスを伴った能動的な動作にある．

ここでは，カメラを搭載したロボット自身の動きによって，しかも，能動的に動きをコントロールできるとすると，そのとき得られる動画像系列から何が読み取れるようになるのかを論じていこう．

## 2. 能 動 視 覚

「対象の三次元形状」と「カメラの動き（ここではカメラを搭載しているロボットの動き）」と「動画像（あるいは画像系列）」との 3 者は，互いに相補的な関係にある（図 2）．つまり，このうち二つが与えられると後の一つが決定する．対象の形状を与え，カメラの動きをあらかじめ決めれば，アニメーションとしての動画像を作ることができるし，これから得られるであろう画像を予測することができる．カメラの動きが分かっていると，動画像から対象の三次元形状を復元できる．対象の形状が分かっていて動画像が与えられれば，そのときのカメラの動きを導ける．

ロボットが視覚を用いて自律的に行動するためには，一つは周囲の環境を認識し，そしてもう一つは，それに応じて自分の運動を決定する．当初は，これら二つははっきりと分けて考えることが自然であった．まず，周囲の環境の三次元形状，自分，対象物体，目的の移動先での位置姿勢を画像から求める．そして，その環境・位置情報に基づいてロボットの位置制御を行う．すなわち，「見て，認識し，それから行動する」ということであった．

図 2 の関係でいえば，どの二つが既知で何を求めるかを，

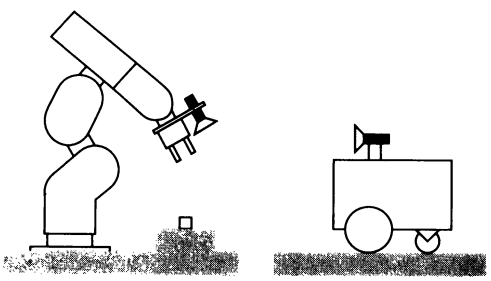


図 1 Eye IN Hand システム．カメラ自身もロボットと共に運動する

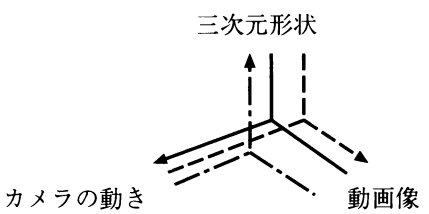



図 2 動画像とカメラの運動と対象形状の関係

原稿受付 1998 年 5 月 6 日
キーワード：Active Vision, Robot Vision, Computer Vision, Visual Servoing, Visual Perception
*〒 113-8656 東京都文京区本郷 7-3-1
*Bunkyo-ku, Tokyo

それぞれの段階ごとに戦略として切り分けていた．ところが，人間は，例えばものを掴むとき，事前に正確な位置があらかじめ分かっていなくても，最終的にそれを掴むことができる．認識と行動を一体として考えている．動いた結果を視覚情報としてフィードバックしながら動かすことが極めて有効に働いている．言い方を変えると，図2の3者の関係を渾然一体として使っている．

「認識する」ことと「行動する」ことを一体の問題として扱っている．ロボットに求められる視覚とは，このような能動視覚である．

自律的に観測計画を立て，適切な運動制御を行って観測を行い，認識を行う能動視覚システムの研究 [2][3] は近年盛んになってきている．ここでは，そのいくつかの形態を具体的に紹介する．

## 3. 注視制御と三次元形状の獲得

まず，能動的な視覚の例を示そう．

運動している物体上のある点を画像上のある一点に常に留めるようにカメラのパン，チルトを制御して首を振る（注視制御と呼ぶ）と，物体の運動がその点を中心とした回転運動とみなせる画像が得られる（**図3**，**4**）．このことを用いると，注視点以外の点のオプティカルフロー（対象点の動画像上での流れ）[4] から，その物体の三次元形状を復元できる．

カメラを注視制御するためには，まず，注視点が次の時刻において画像上のどこに移動したかを求める．そして，この点が常に画像中心にくるようにカメラの運動をフィードバック制御する．

このとき得られる画像は，**図5**に示すようにカメラと対象物体の相対位置を考えると，カメラを固定して物体を注視点まわりに自転させた場合の画像と等価となる．また，この場合の物体の回転角は，カメラの回転角と大きさが同じで符号が逆向きになる．すなわち，対象物体の表面のある注視点を中心に物体自身がパン，チルト方向にそれぞれある角速度で回転している画像を得ることになる．そのときの物体表面の点の画像は，その回転中心よりどれだけ前後にあるかによって，画像上では特有の動きをする．

したがって，このような注視点制御画像から，対象の奥行き情報を得ることができる．

**図6～8**に，このような注視制御による対象の動きの追跡と実時間での三次元形状復元システム [5] での，形状認識の結果を示す．

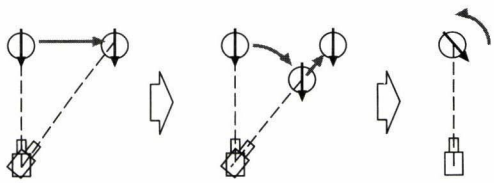


**図5** 注視点制御による画像は，カメラと対象物体の相対位置を考えると，対象が静止したカメラの前で自転しているときの画像と等価となる

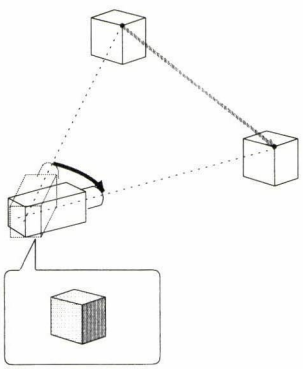


**図3** 動物体のカメラ追跡

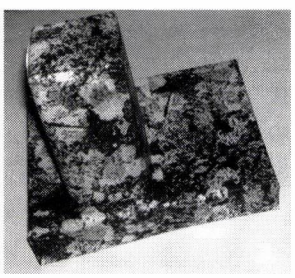


**図6** 三次元形状の復元実験に用いた対象

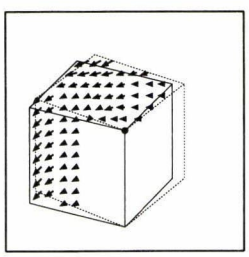


**図4** 注視点制御画像

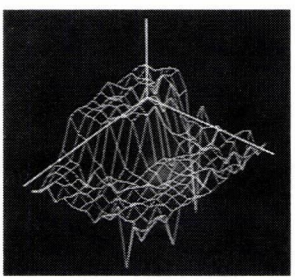


**図7** 観測された三次元形状

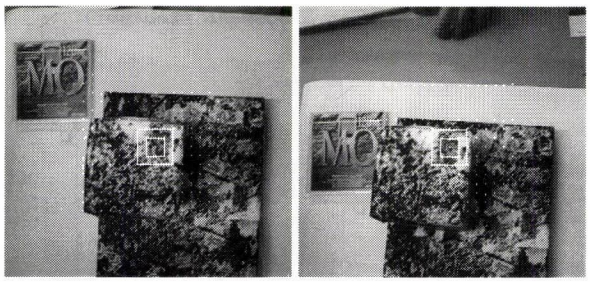


**図 8** 注視制御による対象の動きの追跡画像と検出されたオプティカルフローの例

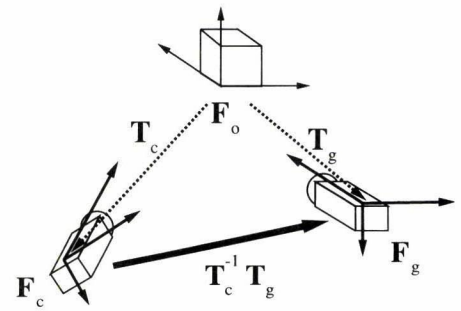



**図 9** 位置ベーストの視覚サーボ・画像から対象物体とカメラの位置との間の相対的な位置関係 $T_c, T_g$ を推定し，その差分 $T_c^{-1}T_g$ を移動することでゴール位置へ向かう

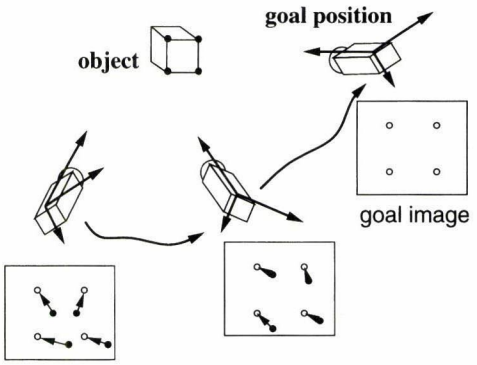



**図 10** 画像ベーストの視覚サーボ．現在の画像とゴール画像を比較し画像が近くなるように移動する

## 4. 視 覚 サ ー ボ

目標位置・姿勢にロボットがあるときどのような画像が得られるかを与えることで，視覚によるフィードバックを用いて，ロボットを与えられた目標位置へ導いたり，対象の動きに追従させる制御は，視覚サーボと呼ばれる [6][7].

まず直感的には，対象の三次元形状が分かっているときは，画像からその対象が与えられたように見える相対的な位置が求められる．したがって，画像から三次元情報を復元してからそれに基づいて行動をする．これを位置ベーストな方法と呼ぶ.

まず現在とゴールのそれぞれの位置姿勢を推定する.

対象物体の三次元幾何モデルとカメラの焦点距離や画素のサイズ，光軸の位置などのいわゆるカメラパラメタが正確に得られていれば，それらのパラメタと画像を用いて，対象物体に対するカメラの現在とゴールのそれぞれの位置姿勢を求めることができる（これをカメラキャリブレーションという）.

**図 9** に示すように，与えられたゴールでの画像から，対象との相対的なゴール位置・姿勢 $T_g$（相対的な平行移動 $(x,y,z)$ と回転角 $(\omega_x,\omega_y,\omega_z)$ をひとまとめにして，このように表す）を求める．また，現在得ている画像から，相対的な位置 $T_c$ を求めれば，$T_c^{-1}T_g$ に対応する運動を行うことで容易にゴール位置へ移動することができる.

これに対して，画像ベーストな手法 [8] では，**図 10** のようにカメラの動きと画像の変化を直接結び付けて考え，各時点での画像のゴール画像からの偏差を最小にする向きに運動を制御する．そのようにフィードバックをかけながら移動していくことで，最終的に現在の画像とゴール画像とを一致させる．この手法では，シーン中の対象の三次元情報を明示的に用いないため，対象の幾何モデルは不要であり，一般に様々な外乱に対して頑強であると言われている.

## 5. カメラ運動と画像変化のモデル

この視覚サーボの手法も含めて，能動視覚では，カメラの運動によって画像の変化がどのように引き起こされるかの解析 [9] が重要である.

焦点距離を 1 とした透視変換を考え，カメラ座標系で $z=1$ に画像面を考える（**図 11**）．三次元空間の対象点 $\boldsymbol{P}=[x,y,z]^\top$ の画像を $\boldsymbol{p}=[u,v,1]^\top$ とすると，

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{1}$$

カメラの運動と対象点の画像の動き（先のオプティカルフローにあたる）の関係は，カメラの並進速度を $\boldsymbol{v}$，回転速度を $\boldsymbol{\omega}$ とすると，

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = -\boldsymbol{v} - \boldsymbol{\omega} \times \begin{bmatrix} x \\ y \\ z \end{bmatrix} \text{ より，}$$

$$\frac{d}{dt}\begin{bmatrix} u \\ v \end{bmatrix} = \left[\begin{array}{ccc|ccc} \frac{1}{z} & 0 & -\frac{x}{z^2} & \frac{xy}{z^2} & -1-\frac{x^2}{z^2} & \frac{y}{z} \\ 0 & \frac{1}{z} & -\frac{y}{z^2} & 1+\frac{y^2}{z^2} & -\frac{xy}{z^2} & -\frac{x}{z} \end{array}\right] \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{\omega} \end{bmatrix}$$

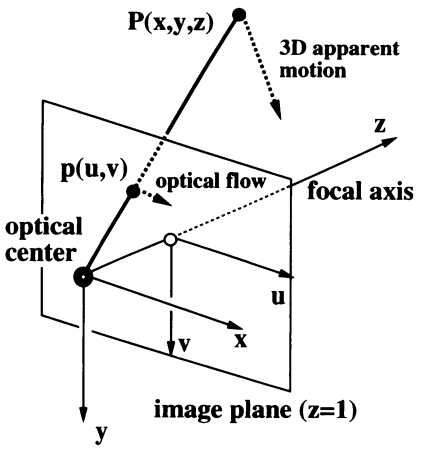



図 11 カメラ運動と画像変化のモデル

$$= \left[\begin{array}{ccc|ccc} \frac{1}{z} & 0 & -\frac{u}{z^2} & uv & -1-u^2 & v \\ 0 & \frac{1}{z} & -\frac{v}{z} & 1+v^2 & -uv & -u \end{array}\right] \left[\begin{array}{c} \boldsymbol{v} \\ \boldsymbol{\omega} \end{array}\right] \tag{2}$$

この係数行列は相互行列（Interaction matrix）と呼ばれる．

現在の画像特徴量を例えば画像上の特徴点の位置 $(u_1,v_1),(u_2,v_2),\cdots$ を並べた $\boldsymbol{f}=[u_1,v_1,u_2,v_2,\ldots]^T$ で表し，同様に，ゴール位置で得られるであろう対応する画像特徴量を $\boldsymbol{f}_g$ とすれば，ここでの目的は，$\boldsymbol{e}=\boldsymbol{f}-\boldsymbol{f}_g$（画像偏差）がゼロになるようにロボットの運動を制御することである．カメラの運動 $\boldsymbol{T}=[\boldsymbol{v}^T\ \boldsymbol{\omega}^T]^T$ と画像上の特徴点の動きの関係は，各点についての相互行列を縦に並べた行列を $L$ として式（2）を縦に並べ，次のように書ける．

$$\frac{d\boldsymbol{f}}{dt}=L\boldsymbol{T} \tag{3}$$

したがって，カメラ運動の制御則は，画像偏差を最小にするような運動，すなわち，$\left|\boldsymbol{e}-\frac{d\boldsymbol{f}}{dt}\right|=\left|(\boldsymbol{f}-\boldsymbol{f}_g)-L\boldsymbol{T}\right|$ を最小化する $\boldsymbol{T}$ として，

$$\boldsymbol{T}=-\lambda L^{+}(\boldsymbol{f}-\boldsymbol{f}_g)$$
$$(\lambda>0:\text{ゲイン},\ L^{+}:L\text{ の疑似逆行列})$$

で与えられる．この制御則を刻々得られる画像に対して繰り返し用いてフィードバック制御をすることで，画像偏差 $\boldsymbol{e}$ がゼロになることが示される．

点以外の画像特徴量を用いるときの相互行列 $L$ の導き方，実際にこの制御を行う際には $L^{+}$ の値を計算しなければならないが，それを簡略に計算する手法，などが研究されている [9][10].

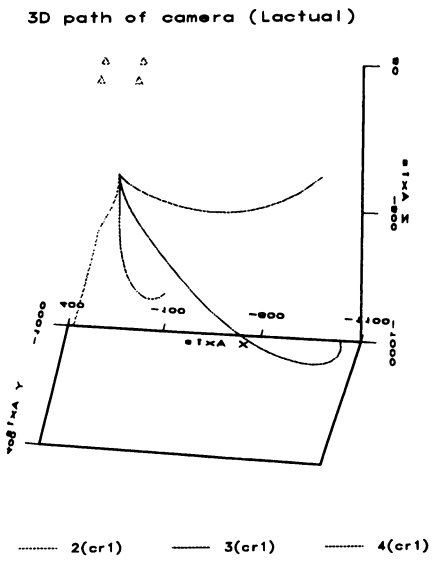



図 12 様々な位置から出発したときの画像ベースト法の視覚サーボによるカメラの三次元空間における軌道

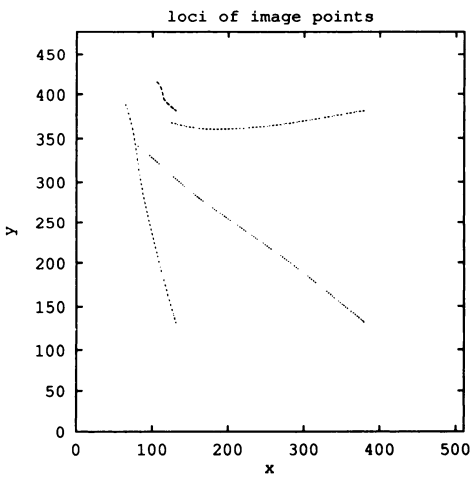



図 13 3 の位置から出発した場合の画像上での各点の動きの軌跡

**図 12** は画像ベーストの制御によりカメラを制御した結果である．対象はゴール位置上方の正方形の角の 4 点で，それを正面から画像いっぱいに見る位置がゴールの位置・姿勢である．**図 13** に示すように，最初左上方にあった 4 点が，正面に見た位置にまで導かれている．

ただし，この図に示すように，画像上では最短距離で各点がゴール位置へ向かっていても，空間の軌跡は回り道をすることがある．これは，画像上の最短軌道とそれを得るカメラの最短軌道とは対応していないからであり，視覚サーボで最適な軌道を得るための手法が提案されている [11].

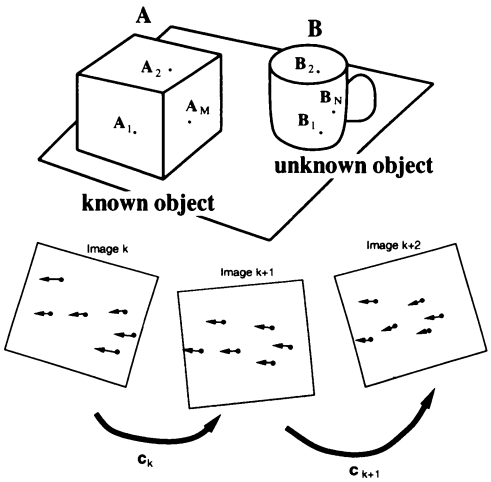



図 14　既知の対象と未知の対象を撮影しながら三次元形状を認識する

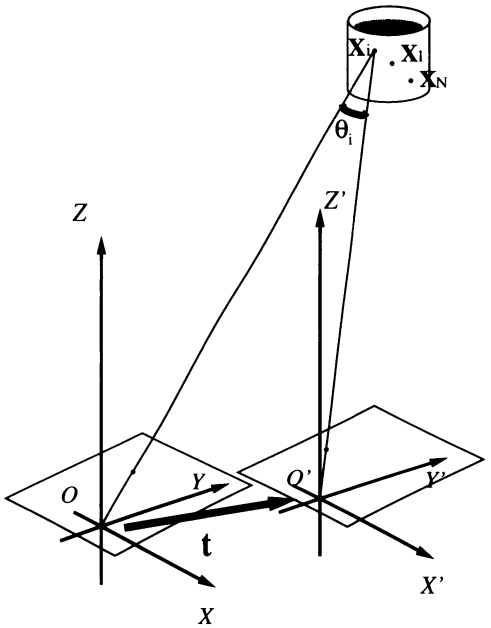



図 15　ある点 $\boldsymbol{X}_i$ の三次元位置を決定するための視差角 $\theta_i$

## 6.　画像系列の利用とカメラの最適運動

もう一つの能動視覚の主要な要素は，対象を長く見続けることで，より完全な三次元形状情報を得ることである．どのように視点を変えれば，よりよく対象を認識することができるか，という問題にも通じる．

シーンの対象の中には，形状が未知のものに，既知のものが混ざっていることがある．ただし，既知といっても不確定さを持つ場合が多くあり，また，画像にも量子化誤差を含む不確定さがある．

例えば，カメラは既知の対象 $\boldsymbol{A}$ と未知の対象 $\boldsymbol{B}$ を撮影する（**図 14**）．この状態からカメラを動かすと，画像上には $\boldsymbol{A}$，$\boldsymbol{B}$ 上の点のオプティカルフローが生じる．$\boldsymbol{A}$ のオプティカルフローと既知である $\boldsymbol{A}$ の三次元形状からこの時点でのカメラの運動量を求めることができる．すると，この計算された運動量と，$\boldsymbol{B}$ のオプティカルフローから $\boldsymbol{B}$ の形状を決定することができる．しかし，1 回の計測だけでは正確な形状を認識することはできない．なぜなら，画像には量子化誤差があり，この影響は避けることができないからである．そこで，この一連の操作を繰り返し，複数枚の時系列の画像を使って，量子化誤差の影響を小さくしていく．

このために，複数枚の画像を有効に利用する手法として拡張カルマンフィルタの利用が提案されている [12]．状態量として対象のカメラ座標系における三次元位置を，観測量としてその点の画像の位置を採用し，先のカメラの運動と画像点の動きの関係式（2）に基づいてカルマンフィルタを各点に対して構成する．カメラが移動するごとに，観測量である画像の情報を用いてカルマンフィルタを更新する．ある程度，観測を繰り返すと，状態量は対象の三次元位置に収束していく．このとき，この推定量の共分散行列

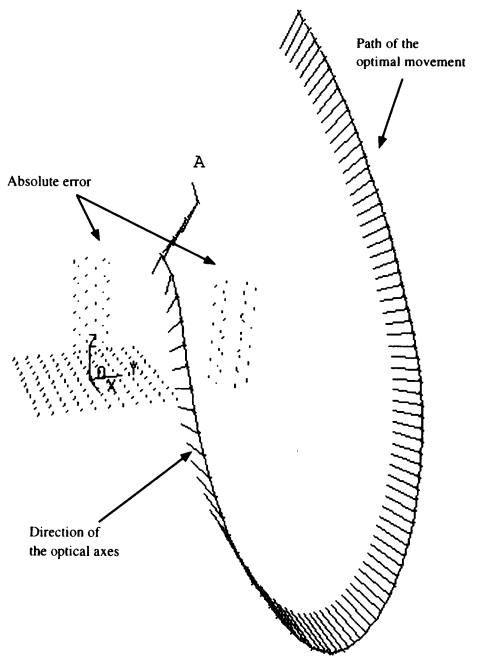



図 16　カメラの最適軌道生成の実験に用いた対象と，100 画像めの時点での三次元位置の絶対誤差とそれまでのカメラの運動

$\boldsymbol{P}$ も同時に得ることができ，これは各点の誤差の見積りと解釈することができる．認識が十分でなく，誤差が大きい点に対しては $|\boldsymbol{P}|$ が大きくなり，十分な精度で対象を認識している場合は $|\boldsymbol{P}|$ の値は小さくなる．

カメラの移動ベクトルを $\boldsymbol{t}$ として，$|\boldsymbol{t}| \leq T_0$ のもとである点 $\boldsymbol{X}_i$ の三次元位置を決定しようとしたら，**図 15** に示

すその点に対する視差角 $\theta_i$ を，なるべく大きくするように移動すべきである．

各点の認識の度合は，カルマンフィルタの状態更新量の共分散行列 $\boldsymbol{P}$ で評価することができた．そこで，各点について重み $w_i = |\boldsymbol{P}|$ を定義し，認識が終わっていない点を重点的に認識するような運動を決定することができる．すなわち，各時点で各点の視差角の認識優先重みを考慮した和を求め，これを最大にするような $\boldsymbol{t}$ を求める．そして，そこで新しい画像を得て，カルマンフィルタの状態量を更新し，さらに次の運動を決定していく．

このようにして決定したカメラ運動の軌道の例を示す[13].

**図 16** で左奥に示す対象は，平面上に格子状に配置された 121 個の点とその上に 6 角柱の形状をなす 36 点が 2 組から成る．6 角柱の 72 点のうち 50 点は最初から位置の分かっている点，その他はこれから認識しようとする点である．画像は，常に量子化誤差を含む位置誤差を持つ．

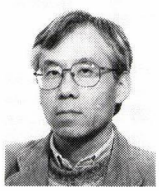
この対象に対して A の位置にあったカメラに最適な運動をさせた場合の 100 画像めの時点での対象の三次元位置検出の絶対誤差とそれまでのカメラの運動の軌跡と，視線の方向を図中に表す．このようにカメラを運動させながら，対象を観察するのが，対象の全体像を把握するのに，一番効率が良いということである．

## 7. おわりに

ロボットの視覚について，能動視覚という考え方とその実現例を見てきた．能動視覚について，いろいろな概念が提案されてきてはいるが，実現はまだまだ難しい．しかし，三次元ビジョンの幾何学的な側面は急速に解明されつつあり，カメラの運動と形状認識の問題はずいぶんと明らかになった．それに基づく戦略の部分がこれからのロボットビジョンの課題であろう．

## 参考文献

[1] 出口：画像と空間—コンピュータビジョンの幾何学—．昭晃堂，1991.
[2] K. Pahlavin, T. Uhlin and J-O. Eklundh: “Dynamic fixation and active perception,” Int. J. Computer Vision, vol.17, no.2, pp.113–135, 1996.
[3] P. Whaite and F.P. Ferrie: “Autonomous exploration: Driven by uncertainty,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.19, no.3, pp.193–205, 1997.
[4] S.S. Beauchemin and J.L. Barron: “The Computation of Optical Flow,” ACM Computing Surveys, vol.27, No.3, pp.433–465, 1995.
[5] 出口ほか：能動カメラによる運動物体追跡と実時間 3 次元形状復元，情報処理学会 CVIM 研究会，vol.111, no.8, 1998.
[6] S. Hutchinson, G.D. Hager and P.I. Corke: “A tutorial on visual servo control,” IEEE Trans. Robotics and Automation, vol.12, no.5, pp.651–670, 1996.
[7] 橋本浩一：“視覚フィードバック制御—静から動へ”，システム/制御/情報，vol.38, no.12, pp.659–665, 1994.
[8] B. Espiau, F. Chaumette and P. Rives: “A new approach to visual servoing in robotics,” IEEE Trans. Robotics and Automation, vol.8, no.3, pp.313–326, 1992.
[9] 出口：“コンピュータビジョンのための幾何学（4）—視覚によるロボットの姿勢制御—”，情報処理，vol.30, no.9, pp.880–887, Sept., 1996.
[10] K. Hosoda and M. Asada: “Versatile visual servoing without knowledge of true jacobian,” In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.186–193, 1994.
[11] 出口，石山：“画像ベースト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”，第 3 回ロボティクスシンポジア，1B13, 1998.
[12] 木下，出口：“能動視覚による 3 次元形状認識”，計測自動制御学会論文集，vol.28, no.1, pp.144–153, 1992.
[13] 木下，出口：“能動視覚のためのカメラの最適運動”，計測自動制御学会論文集，vol.30, no.9, pp.1109–1116, 1994.



**出口光一郎（Koichiro Deguchi）**

1976 年，東京大学大学院修士課程修了（計数工学）．同年より東京大学工学部助手，講師を経て，1984 年，山形大学工学部情報工学科助教授，1988 年，東京大学工学部計数工学科助教授，現在に至る．この間，1991 年～1992 年，米国ワシントン大学客員準教授．コンピュータビジョン，画像計測，並列コンピュータの研究に従事．計測自動制御学会，情報処理学会，電子情報通信学会，形の科学会，IEEE などの会員．（日本ロボット学会正会員）