

# マイクロホンアレーと遺伝的アルゴリズムを用いた移動音源の実時間追跡

宮城工業高等専門学校 情報デザイン学科

本郷 哲

〒981-1239 宮城県名取市愛島塩手字野田山 48

PHONE 022-381-0203

FAX 022-381-0340

E-mail hongo@miyagi-ct.ac.jp

## 1. はじめに

### 1.1 研究の背景

遠隔会議システムの発話者追尾や警備システムなどでは音を利用し人間の追尾が要求されることがある。また、ロボットと人間のコミュニケーションにおいては、ロボットが発話者のいる方を向き対話することが必要である。画像処理による顔認識手法も検討されているが、複数の人が視野領域にいる場合には音響信号によって話者の方向を特定する必要がある。

これまで、このような音源位置の追跡のために、マイクロホンアレーを用いる手法が検討され、逆フィルタ法や適応型フィルタ処理などの研究が行われてきている。

一方、近年、多峰性関数の探索や膨大な空間の探索のために、生物の進化をシミュレートした遺伝的アルゴリズム[Generic Algorithm](以下 GA)が研究されている。GA は生物の進化をシミュレートする性質上、環境変化に追従でき、適応型のシステムにも応用できる。音源の移動を音環境の変化と見做すことによって、GA を用いて音源位置の探索を行いながら、同時に移動追跡もできるのではないかと考え、本研究の着想に至った。

### 1.2 研究の目的

本研究では、遺伝的アルゴリズムとマイクロホンアレーを用いて、移動する音源の位置を探索・追跡する有効な方法を理論的に追究するとともに、その高速探索性を生かしてリアルタイムの移

動音源追跡方法を開発することを目的とする。

本研究で提案・検討を行う音源追跡方法の原理の概要は、次の通りである。

- 1)空間内に M 個のマイクロホン配置する。
- 2)未知音源がある音場下に、遺伝子で表現した座標位置に N 個の仮想音源 (固体) を配置する。
- 3)各仮想音源 (個体) からマイクロホンアレーまでの理論上の伝達関数を算出する。マイクロホン出力波形にその逆伝達特性を乗じる。
- 4)マイクロホン出力に逆伝達特性を乗じた波形は、実音源と仮想音源が一致した場合にのみ、全マイクロホンとも同様な波形が得られる。このような場合、補正波形の相互相関関数を計算し、そのピークまでのラグ (時間ずれ) は、すべてのマイクロホン組み合わせにおいて 0 になる。そこで、このラグの総和の逆数を適応度として利用し、GA を適用することにより音源位置を推定する。
- 5)収束後に音源が移動した場合には、遺伝子表現が少しずつの変化をすることから生物が環境に従って進化すると同様に音源を追跡できる。

### 1.3 本研究の意義と役割

マイクロホンアレーを用いた音源位置の探索方法は様々提案されている。また GA を用いた音源位置探索法についても 1996 年浜田らが定式化を行っている<sup>[1]</sup>。しかし、遺伝的アルゴリズムの特徴である環境の適応性を利用し、能動的で逐次的な音源の追跡については、詳細な検討がなされていない。また、マイクロホンアレーを用いた音源位置探索技術は、実験室レベルでは多くの検討

がなされているが、リアルタイムで探索・追跡できるシステムの検討は不十分である。

これまでの研究によって、次のことが確認されている<sup>[2][3]</sup>。

- ① 単一音源の追跡は十分できる。
- ② 探索空間の計算を大幅に減少できる (100 万分の 1 以下の探索時間)

この GA を移動音源位置追跡に適用することは、マイクロホンアレーを用いた音源位置定位の研究の中で独創的であり移動音源追跡を行う試みにおいては、国内初の提案あり、本研究はこの点で独自の位置づけになっている。

## 2. マイクロホンアレーと遺伝的アルゴリズムを用いた移動音源追跡の原理

本章では、マイクロホンアレーを用い、その出力信号の処理に遺伝的アルゴリズムを適用した移動音源追跡法の原理を述べる。

まず、追跡探索空間内に  $M$  個のマイクロホンを配置する。各マイクロホンの座標を  $[x_m, y_m, z_m]$  ( $1 \leq m \leq M$ ) とする。

次に、この探索空間内に遺伝子により表現した座標位置に  $I$  個の仮想音源 (個体) を配置する。一般に、GA においては遺伝子は固体の特性を表すことになり、本研究では仮想音源 (個体) のある  $(x, y, z)$  座標を遺伝子として表現する。

例えば、本研究では  $2[m] \times 2[m] \times 1[m]$  の探索空間を 22bit で表現している。22bit のうち  $x$  軸方向に 8bit ( $2[m]$  を 256 点に分割)、 $y$  軸方向に 8bit ( $2[m]$  を 256 点に分割)、 $z$  軸方向に 6bit ( $1[m]$  を 64 点に分割) に割り振ることにより、探索空間を 420 万点に分割している。即ち、22bit 遺伝子の場所情報をもつ  $I$  個の個体が探索空間に配置される。

次に各仮想音源 (個体) からマイクロホンアレーまでの理論上の伝達関数を算出する。3次元自由音場空間における点音源の波動方程式に従うと仮定すれば、個体  $i$  (座標  $[x_i, y_i, z_i]$ ) から各マ

イクロホン (座標  $[x_m, y_m, z_m]$ ) までの伝達関数は次のように表現できる。

$$H_{im} = \frac{e^{-j\omega r_{im}/c}}{r_{im}}$$

ここで、

$$r_{im} = \sqrt{(x_i - x_m)^2 + (y_i - y_m)^2 + (z_i - z_m)^2}$$

であり、 $c$  は空中音速である。

次にマイクロホン出力に対し、逆伝達特性の位相成分のみを乗じて補正する。即ち、マイクロホン出力スペクトルを  $A_{im}$  とすると、補正スペクトル  $A'_{im}$  は、次のように与えられる。

$$A'_{im} = A_{im} \cdot e^{j\omega t_{im}}$$

マイクロホン出力に逆伝達特性を乗じた波形は、実音源と仮想音源が一致した場合にのみ、図 1 に示すように全波形の位相が揃うことになる。

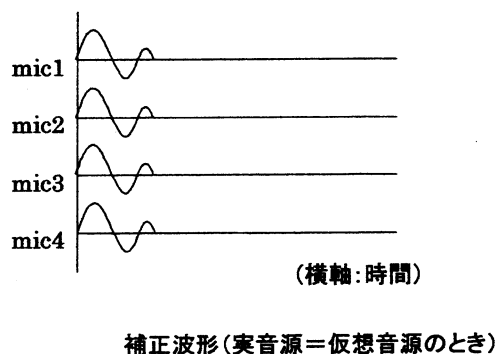
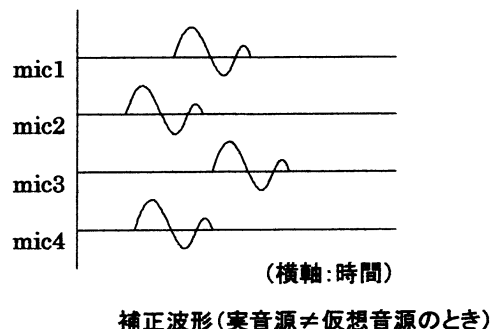


図 1 位相補正波形の概念

次に、図 1 に示す補正波形の揃い方を定量的に求めるために、補正波形の相互相関関数を求める。

2つの系列を  $x_p, y_p$  とすると、それらの相互相

関関数  $\gamma_{xy}(k)$  は次式で表される。

$$\gamma_{xy}(k) = \frac{1}{P\sigma_x\sigma_y} \sum_{p=0}^{P-1} (x_{p-k} - \bar{x})(y_p - \bar{y}),$$

$$\sigma_x = \sqrt{\frac{1}{P} \sum_{p=0}^P (x_{p-k} - \bar{x})^2}, \sigma_y = \sqrt{\frac{1}{P} \sum_{p=0}^P (y_p - \bar{y})^2}$$

$x_p, y_p$  に各マイクロホンの補正波形を入れて、

相互相関関数を求める。

仮想音源（個体）の位置と実音源の位置の距離が一致していれば、ラグ（時間ずれ）0 のところの相互相関関数  $\gamma_{xy}(0)$  はピークとなる。逆に、仮想音源の個体と実音源の位置の距離が違っている場合にはピークの位置はラグ（時間ずれ）0 からずれたところに出現する。

マイクロホン数4個の場合についての具体例について説明する。本研究においてもマイクロホンは4個用いている。この場合2個のペアの組み合わせは6通りになる。この6通りのペアの相互相関関数を求め、 $\gamma_1 \cdots \gamma_6$  とし、それぞれのピークを示す点のラグの絶対値  $|k_i|$ （時間ずれの絶対値）を求める。そして、 $|k_i|$  の総和量  $K$  を求める。

$$K = \sum_{i=1}^6 k_i$$

$K$  が小さくなるほど、最適解に近づくことになるので、本研究では、適応度  $E$  としてその逆数を採用した。

$$E = \frac{1}{K} = \frac{1}{\sum_{i=1}^6 k_i}$$

適応度の定義については検討の余地があるが、後述のような良好な結果が得られている。また、実際のシミュレーションでは、数値の無限大発散を防ぐために極めて小さな値である  $\varepsilon$  を用いて次式のような適応度を採用した。

$$E = \frac{1}{K + \varepsilon} = \frac{1}{\left(\sum_{i=1}^6 k_i\right) + \varepsilon}$$

本報告に記載したシミュレーション結果では  $\varepsilon = 0.001$  とした。つまり評価関数の最大値は  $K = 0$  のときの  $E = 1000$  となっている。

### 3. シミュレーションによる検討

#### 3.1 固定音源探索シミュレーション

本手法の有効性を検証するためにシミュレーションを行った。図2に示すように、(0.2, 0.4, 0.1) に音源を、(-1, -1, 1) (-1, 1, 0) (1, -1, 0) (1, 1, 1) に4個のマイクロホンを配置した。

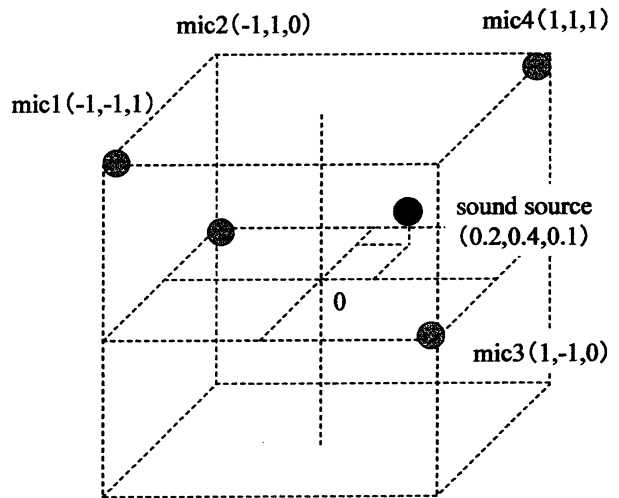


図2 音源とマイクロホンの配置

サンプリング周波数 8000[Hz]、TAP 数 256 とする。音速を 340[m/s] とした。これにより、理論的には誤差 42.5[mm] までの探索が期待できる。

シミュレーションは適合度の最も高い親、まりエリートを無条件にそのまま次の世代へ残すというエリート戦略を導入し、個体数 30、交叉率 50%、突然変異率 4% で行った。探索空間は 2[m] × 2[m] × 1[m] を約 420 万点に分割している。世代毎の仮想音源（個体）の位置、最大適応度、個体数全ての適応度平均、実音源と推定された音源（適応度最大の個体の座標情報）との距離誤差を算出した。乱数の偶然的な影響を無くすために異なる乱数による 10 回平均した。結果を図3に示

す。

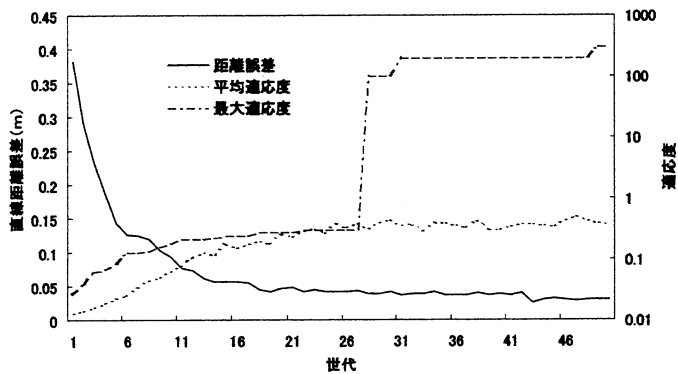


図3 固定音源探索シミュレーションの結果

世代が進むにしたがって収束していくことが確認され、30世代では平均4[cm]の誤差、50世代では平均3[cm]の誤差となった。計900点程度の計算で、420万点の空間の探索が可能であり、GAの高速探索性能が発揮されたことが分かる。

### 3.2 移動音源探索シミュレーション

移動音源に対するGAの有効性を検証するために、音源座標を移動させシミュレーションを行った。初期座標を(0.2, -0.5, 0)とし、世代が進むたびにy座標に0.01[m]を加算し、101世代目では座標が(0.2, 0.5, 0)となるようにした。移動の様子と配置を図4に示す。また、個体数30、交叉率50%、突然変異率4%で行った。結果を図5に示す。

初期の状態から一度収束した以後、うまく追跡が行われていることが分かる。これは、一度収束すると音源付近に個体が集中しているため、連続的な移動にはすぐに収束するためと考えている。

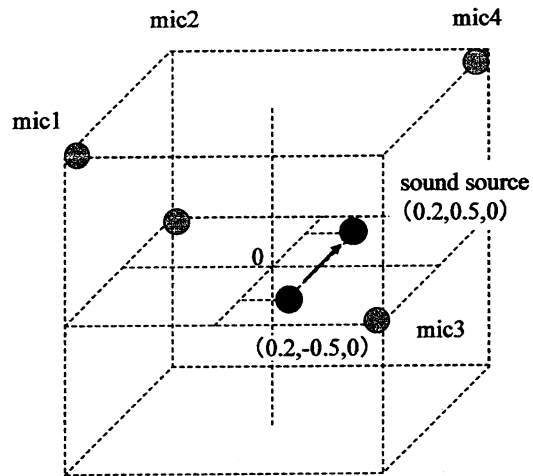


図4 音源の移動とマイクロホンの配置

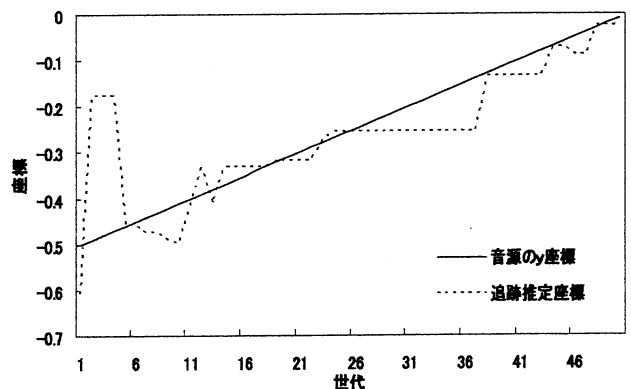


図5 移動音源のy座標と追跡y座標推定値

また、1世代当りの移動速度を変えた結果を図6、図7に示す。移動速度は固定、0.005[m/gene]、0.01[m/gene]、0.02[m/gene]とし、図6はy軸方向、図7はz軸方向に音源を移動させた結果である。ここでも乱数の偶然的な影響を除くために10回平均している。GAパラメータは、個体数30、交叉率50%、突然変異率4%に設定した。

y方向移動、z方向移動どちらも0.02[m/gene]の変化では26世代付近で誤差が増加し、その後再び収束がされていることが見られる。これは、交叉のみでは収束が追いつかず、突然変異により追跡がなされたものと考えている。

移動量により差異が見られるが、20世代以降では平均67[mm]の誤差で追従していることが分かった。

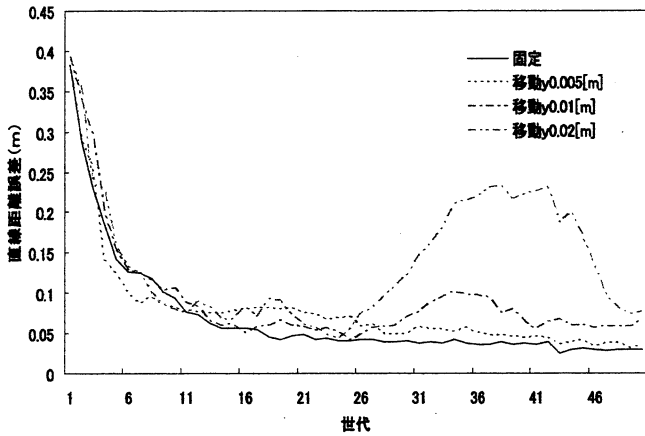


図6 音源と追跡推定位置の距離誤差(y方向移動)

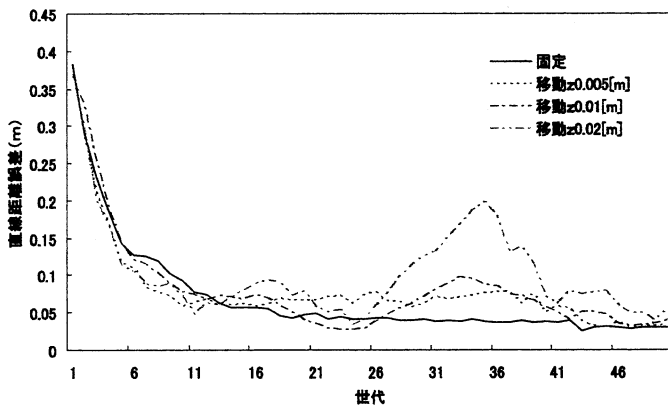


図7 音源と追跡推定位置の距離誤差(z方向移動)

### 3.3 適応度の分布

GAを用いた音源定位において、適応度は音源座標と仮想音源座標との直線距離誤差を評価する尺度である。この適応度の選び方については今後検討の余地が残るが、本手法で選択した適応度の空間分布を確認した。音源座標を  $(0.4, 0.1, 0)$  としたときに、 $x$  座標 $-0.6 \sim 0.6$ [m]、 $y$  座標 $-0.6 \sim 0.6$ [m]、 $z$  座標  $0$ [m] (音源を含む面)の適応度分布表を図8に示す。

適応度は実音源に近づくにつれ滑らかに上がっていることが分かる。適応度は妥当なものであることが確認できた。

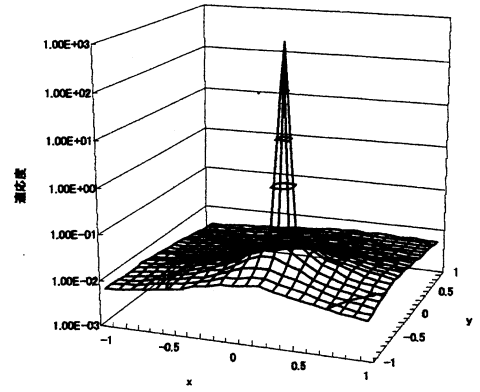


図8 適応度の空間分布

### 3.4 アルゴリズムパラメータの検討

突然変異率により収束に差が見られるか検討した。交叉率50%、突然変異率を0~20%に変化させた際の収束の様子を調べた。初期座標を  $(0.2, -0.5, 0)$  とし、固定音源、 $y$  方向に1世代あたり  $0.005$ [m],  $0.01$ [m],  $0.02$ [m]と移動速度を変えたものについて10回の平均をとった結果を図9, 10, 11, 12に示す。

移動音源に対しては、突然変異率が高い方が収束性が良いことが分かる。特に  $0.02$ [m/gene]では30世代付近で誤差の増加が見られ、その後収束が行われているが、突然変異率が高いものはいち早く収束している。

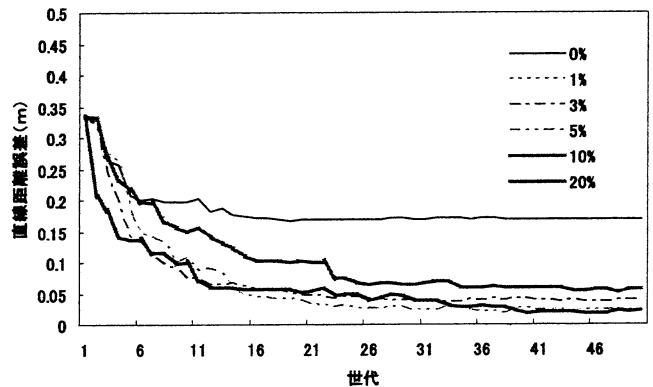


図9 追跡性能に対する突然変異率の影響 (音源固定)

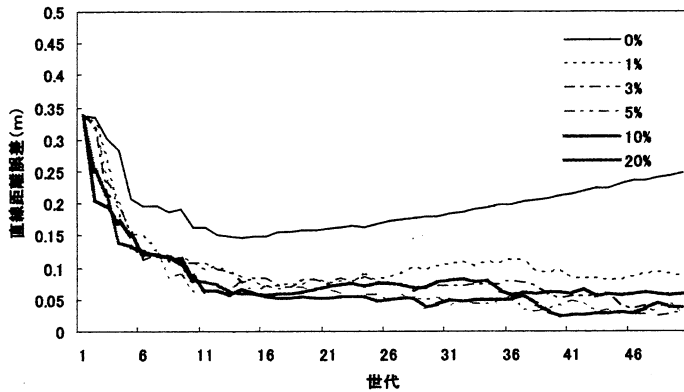


図 10 追跡性能に対する突然変異率の影響  
(y 方向移動速度 0.005 [m/gene])

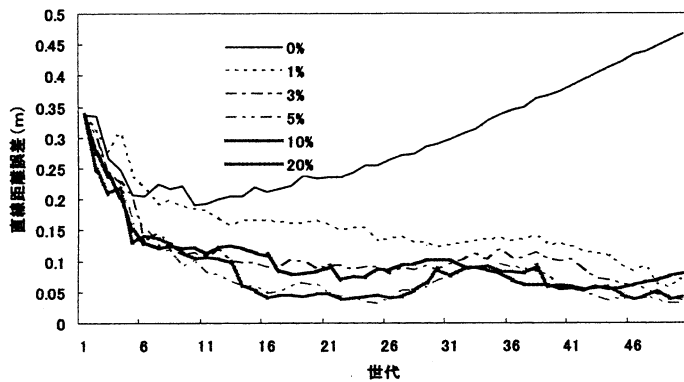


図 11 追跡性能に対する突然変異率の影響  
(y 方向移動速度 0.01 [m/gene])

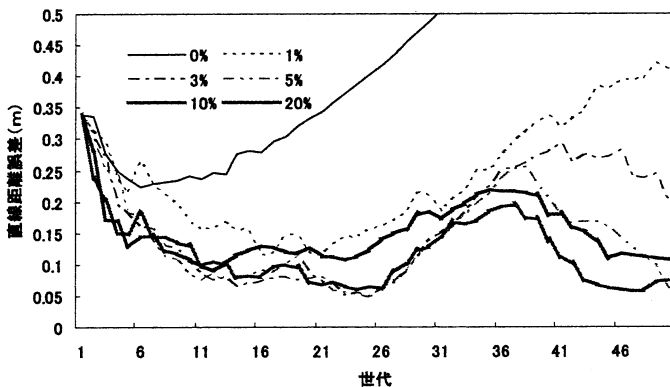


図 12 追跡性能に対する突然変異率の影響  
(y 方向移動速度 0.02 [m/gene])

## 4. 今後の展開と応用分野

### 4.1 シミュレーションによる本手法の精査

現在、本手法の計算機によるシミュレーション方法は確立した。しかし、様々な検討事項が残されており、今後以下の項目を検討する。

#### ① 複数音源の追跡と拡散音場における追跡

単一音源の場合は、空間的サンプリング定理を満たす限り、探索空間内の適応度関数が単峰性であるため、探索が容易である。これに対して、複数の音源が存在する場合や拡散音場（響きが大きい場所）について、GAパラメータや音響的条件を変えて検討を行う。このとき、適応度関数が多峰性になるが、突然変異率を処理中に変えることにより複数の音源を追跡する方法を検討する。

#### ② 実時間追跡のための計算数、メモリ量の検討

本研究は、最終的に実時間追跡を目指す。特にDSPを用いた検討を行う。①、②のシミュレーションを行う際に、演算数やメモリの使用量を考察し、DSPにより演算可能なパラメータを検討する。

### 4.2 実環境における実験

実環境下において本手法の有効性を確認する。これまで確定信号で同期をとって模擬的なアレイを実現して実環境の実験も行ってきたが、多チャンネルAD変換の購入を検討し、実環境追跡実験を行う。

### 4.3 応用分野

#### ① 対話ロボットのヒューマンインタフェース

⇒ロボットに組み込み、話者の方へ顔を向ける。

#### ② スタジオにおける話者の追跡

⇒スタジオ内に取り付け、発話者へのカメラ追跡の自動化。

#### ③ 警備システム

建築物の入り口などに取り付けることで監視カメラの自動追跡へ応用。

### 参考文献

- [1] 浜田他, 信学技法 EA96-1 p.1
- [2] 本郷他, 日本音響学会講演論文集, p. 533-534, (2002-09)
- [3] 本郷他, 電気関係学会東北支部連合大会 p. 280, (2002-08)