音源位置同定によるカメラの首振りに関する研究

國田, 政志
九州大学システム情報科学研究院知能システム学部門

有田, 大作
九州大学システム情報科学研究院知能システム学部門

谷口, 倫一郎
九州大学システム情報科学研究院知能システム学部門

出版情報：情報処理学会研究報告. CVIM, [コンピュータビジョンとイメージメディア]. 2002 (34), pp.41-48, 2002-05-09. 情報処理学会
バージョン：

コンピュータビジョンと
イメージメディア 133－6
（2002．5．9）

音源位置同定によるカメラの首振りに関する研究

國田 政志† 有田 大作‡ 谷口 倫一郎‡

† 九州大学 大学院システム情報科学府
‡ 九州大学 大学院システム情報科学研究院

概要 画像処理を行う際に対象を画像内に確実に大きく捉えるために、状況に応じて撮影時にカメラを能動的に操作する手法がよく用いられる。この場合、カメラで撮影する方位を決定するための情報が重要となり、対象物体の存在する大まかな位置情報の獲得が望まれる。本研究では対象物体が発する音を用いて位置情報を得ようとする試みを行った。複数のマイクロフォンを空間中に配置し、観測される信号をMUSIC法と呼ばれる手法で解析する。その結果、マイクロフォンを2つ用いた場合実空間で誤差約 ±2[°] 以内の精度で音源の方位を推定することが可能であることが分かった。また、音源の方位へカメラを向けるシステムを構築し、ほぼ実時間でカメラを制御できることを示す。

Pan-Tilt Camera Control using Sound Source Localization

Masashi KUNITA Daisaku ARITA Rin-ichiro TANIGUCHI
Graduate School of Information Science and Electrical Engineering, Kyushu University

Abstract

Active vision scheme is frequently used to catch targets in the image stably and widely for the following image processing. In this case, acquiring target location information in real-time is required to control cameras. To acquire target location information, we use acoustic information generated by targets. Acoustic signals acquired by microphone array are processed by MUSIC method which estimates sound source locations. The experimental results shows that the direction of sound source is estimated with high accuracy. And we construct a prototype system with microphone array to show a pan-tilt camera is well controlled to capture the target image in real-time.

1 はじめに

コンピュータや映像機器の発達により外界の多量の情報を獲得し処理することが可能となるに伴って、画像処理の分野では情報の高度な利用法の研究が行われている。しかし、画像として得られる情報には、観測範囲、解像度、情報量の問題が付き纏い、十分な解像度の獲得には情報の範囲が犠牲になり、また十分に範囲を確保すれば解像度に不足をきたすというジレンマは解消さないままである。そのため、カメラを

必要に応じて動かして対象を観測するというアクティブビジョンのアプローチ [3] がよく用いられる。この場合、カメラをどのような方向に向けるか、あるいはどのようにズームするかという制御情報を得ることが重要となってくる。

本研究ではマイクロフォンアレーを用い、対象物体が発する音から対象物体の存在する大まかな位置を推定し、そこにカメラを向けるというカメラ制御を試みる。大まかな位置の特定に音を用いることで映像のみでは難しかった全方位の情報が連続的に得られ、またその情報量も比較的少なくて済む利点がある。

2 音源位置同定法

音は空間を伝わる波である。複数のマイクロフォンを用いれば、波の到達する時刻に差が生じ、その情報を解析すれば音源の位置を同定することができる。しかし、実際にマイクロフォンアレーで得られる情報から解析を行うにはいくつかの問題がある。ここでは、音源の位置を同定する手法として MUSIC 法 [1] [2] について述べると共に、その他の音源位置同定法と比較を行う。

2.1 MUSIC 法の原理

マイクロフォンアレーとは、複数のマイクロフォンを空間に配置したものであり、ここではそれぞれのマイクロフォンで観測された音から、信号処理によって音元位置を推定することを考える。マイクロフォンアレーを用いた手法はいくつも提唱されており、MUSIC 法もその中の一つである。

MUSIC 法は、マイクロフォンより少ない数であれば複数の音源に対して有効であり、近傍音場であればマイクロフォンからの距離も特定できるよう拡張することができる。ここで、近傍音場とは波面が球面と見なせる場合のことである。信号源とセンサの距離が離れ、波面が平面に近い状態では、信号の到来方位しか特定することはできない。

マイクロホンアレーのマイクの数を M 個とする。マイクで観測した信号を短区間 FFT したものを

$$\mathbf{x}(\omega, t) = [X_1, ..., X_M]^T$$

で表すことにする。$\mathbf{x}$ の変数 ω は周波数を表し、t は、FFT を行った時間区間を示す指標であり、共に整数値である。

ここで、入力ベクトルが以下のモデルに従うと仮定する。

$$\mathbf{x}(\omega, t) = \mathbf{A}(\omega)\,\mathbf{s}(\omega, t) + \mathbf{n}(\omega, t)$$

$\mathbf{A}(\omega)$ は各音源から各マイクまでの伝達関数の行列、$\mathbf{s}(\omega, t)$ は複数ある音源 (ここでは D 個とする) の信号をそれぞれ短区間 FFT したものを表す列ベクトル、$\mathbf{n}(\omega, t)$ は各マイクで観測される雑音成分を表している。

MUSIC 法では、このモデルに従って観測値 $\mathbf{x}(\omega, t)$ の共分散行列の特性を利用して方位推定を行う。まず雑音が信号と無相関であると仮定して観測値の共分散行列 $\mathbf{R}(\omega)$ を求めると、

$$\begin{aligned}\mathbf{R}(\omega) &= E\left[\mathbf{x}(\omega, t)\,\mathbf{x}^H(\omega, t)\right] \\ &= \mathbf{A}(\omega)\,\mathbf{P}(\omega)\,\mathbf{A}^H(\omega) + \mathbf{K}(\omega)\end{aligned}$$

となる。ここで

$$\begin{aligned}\mathbf{P}(\omega) &= E\left[\mathbf{s}(\omega)\,\mathbf{s}^H(\omega)\right] \\ \mathbf{K}(\omega) &= E\left[\mathbf{n}(\omega)\,\mathbf{n}^H(\omega)\right]\end{aligned}$$

である。

さらに簡単にするために $\mathbf{K}(\omega) = 0$ の場合を考える。マイクの数 M が音源の数 D よりも多い場合を考えると、共分散行列のランクは D となり、$\mathbf{R}(\omega)$ が D 個の非零固有値と M−D 個の零固有値を持つことになる。したがって固有値が大きい順にソートされているとすると、$\mathbf{R}(\omega)$ の対角化は

$$\begin{aligned}\mathbf{E}^H\mathbf{R}(\omega)\,\mathbf{E} &= \left[\mathbf{E}^H\mathbf{A}\right]\mathbf{P}\left[\mathbf{E}^H\mathbf{A}\right]^H \\ &= diag(\lambda_1, ..., \lambda_D, 0, ..., 0)\end{aligned}$$

で表すことができる。

さらに零固有値と非零固有値に分けると、

$$\begin{aligned}\mathbf{E}_s &= [e_1, ..., e_D] \\ \mathbf{E}_n &= [e_{D+1}, ..., e_M]\end{aligned}$$

$$\begin{aligned}\left[\mathbf{E}_n^H \mathbf{A}\right] P \left[\mathbf{E}_n^H \mathbf{A}\right]^H &= diag\,(0, ..., 0) \\ \left[\mathbf{E}_s^H \mathbf{A}\right] P \left[\mathbf{E}_s^H \mathbf{A}\right]^H &= diag\,(\lambda_1, ..., \lambda_D)\end{aligned}$$

したがって

$$\begin{aligned}\mathbf{E}_s^H \mathbf{A} &\neq \mathbf{0} \\ \mathbf{E}_n^H \mathbf{A} &= \mathbf{0}\end{aligned}$$

という関係が導かれる。これらの $\mathbf{E}_s^H$ および $\mathbf{E}_n^H$ はそれぞれ信号部分空間の基底ベクトル、直交補空間の基底ベクトルと呼ばれる。

MUSIC 法では後者の $\mathbf{E}_n^H \mathbf{A} = \mathbf{0}$ の式に注目する。この式は、音源から各マイクまでの伝達関数 $\mathbf{a}$ が与えられたとき、この式の値を計算すれば $\mathbf{0}$ になるということを示している。したがって任意の位置からマイクまでの伝達関数を求めれば、この式からその位置に音源が存在するかどうかが判断できる。

そこで、ある座標における空間スペクトルとして、

$$P = \frac{1}{\|\mathbf{E}_n^H \mathbf{a}\|^2}$$

を定義し空間内を走査することで、音源のある位置にピークが現れ、位置を同定することが可能となる。

図 1 の環境を想定しこの手法を用いてシュミレーションを行った結果が図 2 である。音源の存在する方位に反応があることがうかがえる。

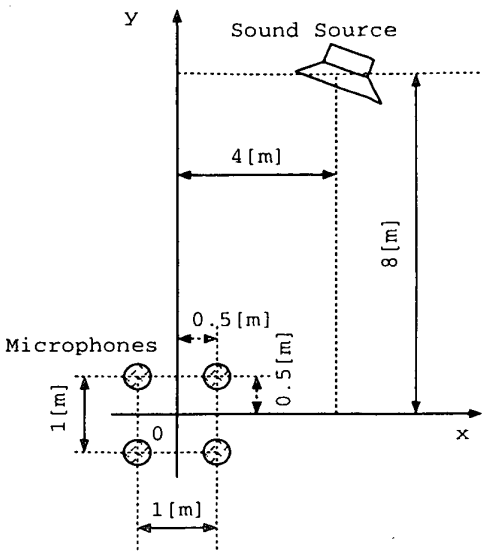

図 1: シミュレーションの想定配置図

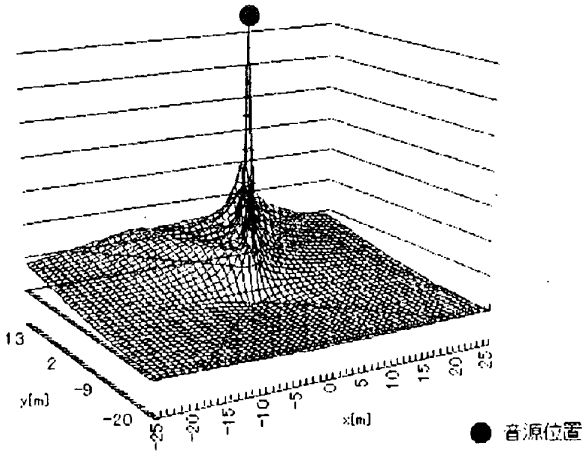

図 2: MUSIC 法による空間スペクトル

2.2 MUSIC 法の利点と課題

MUSIC 法は音源の位置推定において、以下の点で強力な手法と考えられる。

- 複数の観測音に対する包括的な処理が可能である
- 複数の音源に対処可能である
- 波形 (形状) を直視しない解析方法である

一方、MUSIC 法には解決すべき性質がいくつかある。その性質とは、

- 測定環境における音の伝達関数が既知でなければならない
- 音源から発せられる音と相関のある雑音に弱い
- 音源の位置を推定するためには、空間を走査しなくてはならない

の 3 点である。

最初の項目はこの手法の鍵となる部分である。ある注目する座標が音源位置と一致するか否かは、注目位置からマイクロフォンアレーまでの伝達関数と、実際の音源位置からのそれが一致するか否かの判定によって行われる。よって、注目位置からの伝達関数が正確さに欠けている場合は、特定された座標も正確さに欠けたものとなる。これは具体的に言い換えると、気温等の変化によって音速が変化した場合や、反響音、残響音がある環境では、それを十分に考慮しなければならないということである。

次の項目は、先に挙げた伝達関数が既知であるかどうかと密接に関係がある。反響音がある環境で観測を行った場合を例にとると、解析に用いた伝達関数に、反響音を考慮していないものを用いた場合がこれにあたる。また、観測環境の変化によって、それまで存在しなかった反射音が出現した場合も含まれる。

最後の項目は、音源の位置を詳細に調べるほど計算量が増大することを表した計算量の問題である。MUSIC 法は注目する位置に対して音源の存在を判定する手法であり、いわば消極的な手法である。したがって空間のどこに存在するかを知るためには、空間内を繰り返し探索しなければならない。しかし、これは空間に現れるスペクトルにおいて最大値をもつパラメータ(座標) を探し当てる問題であり、種々の探索法を利用して解決することができる。

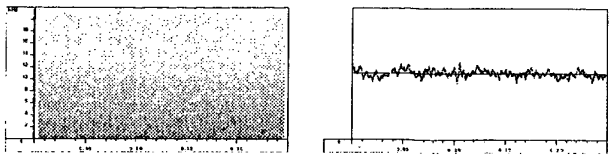

図 3: 部屋の雑音の周波数成分と波形

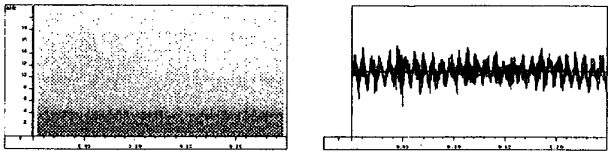

図 4: 観測された音の周波数成分と波形

音源には小型ラジオを用い、信号として FM ラジオが局と同調していない状態の雑音を使用する。 これをマイクロフォンで観測した波形を、部屋の雑音と共に図 3・4 に示す。なお、雑音のほとんどは PC および測定機器のファンから発せられるものである。

この状態で次の図 5 のように実験を行う。マイクロフォンアレーを部屋中央に配置し、そこから半径 $r[m]$、正面からの角度 $\theta[°]$ 離れた位置に音源であるラジオを置く。そして、r と θ を徐々に変化させながら、MUSIC 法によって推定された角度と実際の角度とを比較する。

3 音源位置同定実験

3.1 実験方法

まず、実環境で MUSIC 法を用いた場合の特性を評価するための実験を以下の要領で行う。実験は基本的な特性を測る目的から音源数、マイクロフォン数共に少数で行うものとする。

反響のある室内で、音源を一つのみ用いて音源の位置を推定する。マイクロフォンの数はアレーとして最小の 2 とし、MUSIC 法で解析に用いる伝達関数は距離による遅延のみを考慮した簡易的なものを使用する。実験環境の一覧を表 1 に示す。

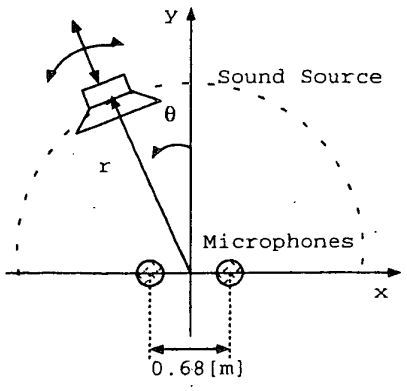

図 5: 実験の配置図

表 1: 実験環境

音源数	1
マイクロフォン数	2
マイクロフォンの配置	間隔 0.68[m]
サンプリング周波数	44100[Hz]
観測時間	0.25[sec]
使用する周波数帯域	100 - 3900 [Hz]
位置推定用伝達関数 (音速)	距離による遅延のみ考慮 (338.65[m/sec])
S/N 比	10.43 [dB]
検索空間	2 次元空間 (マイクロフォンと音源を含む平面)
部屋寸法 (幅 × 奥行き × 高さ)	3.63 × 4.28 × 2.58 [m]
残響時間 (-60dB になるまでの所要時間)	0.6 [sec]

3.2 実験結果と考察

まず、実環境で MUSIC 法の空間スペクトルがどのような分布になるかを確かめるため、音源を異なる二つの位置に置いた場合の空間スペクトルの様子を出力した結果が図 6 である。今回の実験で用いたマイクロフォン数は 2 である。解析に用いた伝達関数は距離による遅延のみしか考慮していないため、これは二つのマイクに入ってくる信号の時間差を見て音源位置を推定していることと等しい。信号の届く時間差が一定であることは、音速が変化しなければ音源から各マイクロフォンまでの距離の差が一定であることを示している。つまり、この場合は音源の位置として考えられる座標はマイクロフォンの位置を焦点とする双曲線状に現れるはずであり、このことは図 6 でも確認できる。

そしてこれらの分布を調べ、音源の存在する角度を推定し、本来の角度との誤差をグラフにしたものが図 7 である。半径 $r = 1.0[m]$ では、おおよそ全体に平均して誤差 2[°] 程度に収まっている。半径 $r = 1.5[m]$ についてはほぼ全域において誤差 1[°] 以内で角度を推定することに成功している。一方半径 $r = 0.5[m]$ の場合においては、正面に近くなるほど精度が上がり、$\theta = \pm 90[°]$ に近づくにつれ精度が落ちる傾向が見受けられる。$\theta = \pm 90[°]$ の場合の観測波形を図 8 示す。

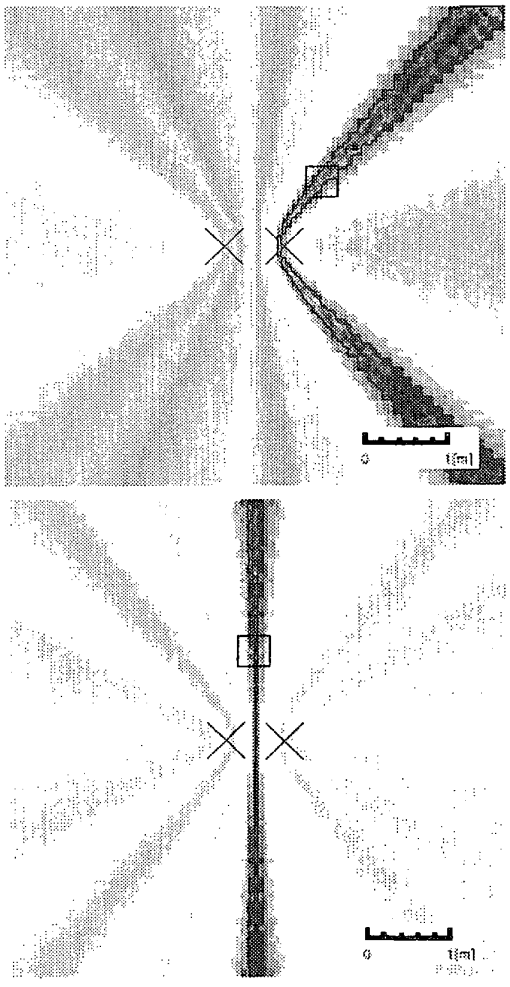

図 6: MUSIC 法によって得られた空間スペクトル (× :マイクロフォン, □:音源)

この現象は、解析に用いた伝達関数が不正確であることで説明ができる。マイクロフォンア

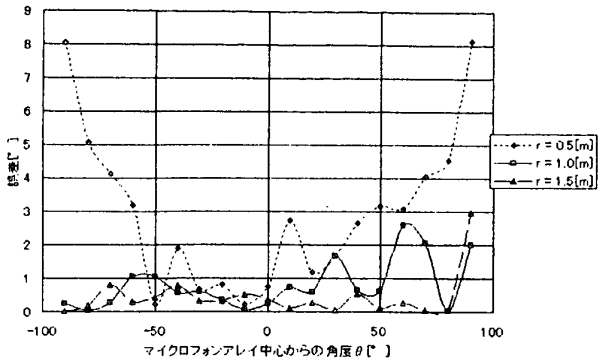

図 7: 距離と角度による推定角度の精度

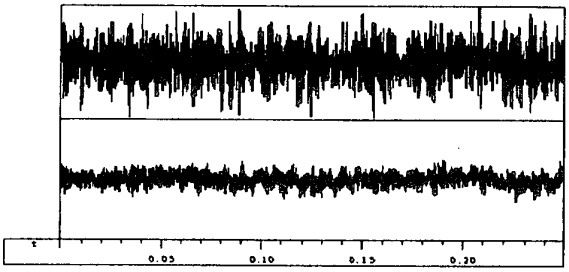

図 8: $r = 0.5[m], \theta = 90[°]$ の時の観測波形

レーから音源までの距離が離れているならば、音が届くまでに空中を伝搬した距離は、各マイクロフォン間であまり差がない。一方、マイクロフォンアレーから音源までの距離が近い場合は、伝播した距離が各マイクロフォン間で相対的に大きく異なってくる。実際、今回の実験では $\theta = -90[°]$ のとき、二つのマイクロフォンまでの伝播距離は、

$$r = 1.5[m] \text{ では} \quad 1.16[m]:1.84[m] \simeq 1:1.586$$
$$r = 0.5[m] \text{ では} \quad 0.16[m]:0.84[m] \simeq 1:5.25$$

と、大きな開きがある。音は空間を同心球状に広がりながら伝播するので、その波のエネルギーは距離の 2 乗に比例して低下する。したがって、 大きく距離に違いがあれば、各マイクロフォンで観測される信号のエネルギーにも大きな差が生じる。図 8 の波形では約 9.4[dB] の差が見られた。一方この実験で想定していた伝達関数は、距離による信号の遅延しか考慮しておらず、ここに大きなずれが生じたと考えられる。そのため、双方のマイクロフォンまでの

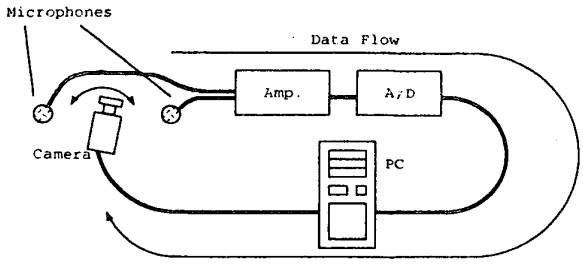

図 9: 自動音源方位撮影システムの構成と処理の流れ

距離が近くなるマイクロフォンアレーの正面に近づくと誤差が小さくなったと思われる。

4 音源位置同定によるカメラ制御実験

MUSIC 法を用いれば、マイクロフォンアレーから音源の方位を推定することが可能であることが示された。次に音源の方位に自動でカメラを向けるシステムを構築する。

4.1 システム構成

システムの構成は図 9 のように、マイクロフォン数 4 のマイクロフォンアレーと信号処理用の PC、さらに音源の方向を撮影する首振りカメラから成る。

マイクロフォンアレーで観測された音は、増幅機、A/D 変換器を経て PC へ入力される。PC では 0.1[scc] を 1 フレームとして短区間 FFT を行い、MUSIC 法を用いて検索空間内のスペクトルを求め、最も反応の強く現れる方位にカメラが向くよう首振りカメラに指令を送る。ただし、スペクトルに大きな反応が見られない場合はカメラは静止させる。

4.2 結果

このシステムによって得られる空間スペクトルは図 10 のように現れる。4 個のマイクロフォンを鉛直平面上に配置したため、図中心に位置

表 2: 自動音源方位撮影システムの詳細

想定音源数	1 以下
マイクロフォン数	4
マイクロフォンの配置	0.30[m] 四方の正方形 (鉛直に配置)
サンプリング周波数	44100[Hz]
観測サイクル時間 (短区間 FFT を行う際の 1 区間)	0.1[sec]
使用する周波数帯域	100 - 3900 [Hz]
位置推定用伝達関数 (音速)	距離による遅延のみ考慮 (338.65[m/sec])
検索空間	2 次元空間 (マイクロフォンを中心とする球面)
検索数	正面から上下左右 ±50[°] の 2000 点
部屋寸法 (幅 × 奥行き × 高さ)	3.63 × 4.28 × 2.58 [m]
残響時間 (-60dB になるまでの所要時間)	0.6 [sec]

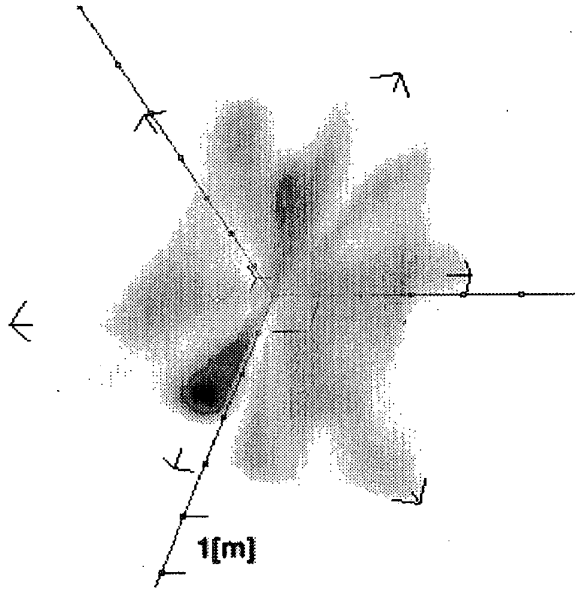

図 10: 空間スペクトル (マイクロフォン数 4)

するマイクロフォンアレーの前方と後方に同様の反応が現れている。音源の存在する位置を照らすスポットライトのような反応が得られることが確認できる。

図 11 は動作状況を示した図である。はじめラジオからは音が出ておらず、その後鳴り始めてからカメラは音源の方位を推定することに成功すると、速やかにその方位へカメラを向ける様子が確認できる。音源は映像の中心に捕らえられており、対象物体の映像を撮影する目的は達成されている。

4.3 処理時間に関する考察

このシステムはカメラの撮影方位のみを決定する目的から探索空間をマイクロフォンアレーを中心とする球面に限ったものとした。そのため、空間全体を探索するのに比べ範囲を大きく削減することができ、音源位置の探索に要する時間を大幅に少なくすることが可能となった。観測された音は、まず 0.1 秒間を 1 区間として短区間フーリエ変換される。すなわち、観測された音は最大でそれだけの時間を経た後に処理されることになる。探索に要する時間は約 0.6 秒であったため、このシステムの遅延は、合わせて約 0.7 秒である。この場合、音源の方位の探索に要する計算時間が遅延の主たる理由であるため、各種研究されている探索法を利用して調べる点の数を削減することによって遅延を短縮できると考えられる。

5 おわりに

マイクロフォンアレーで観測された音の情報から音源の位置情報を推定する手法に MUSIC

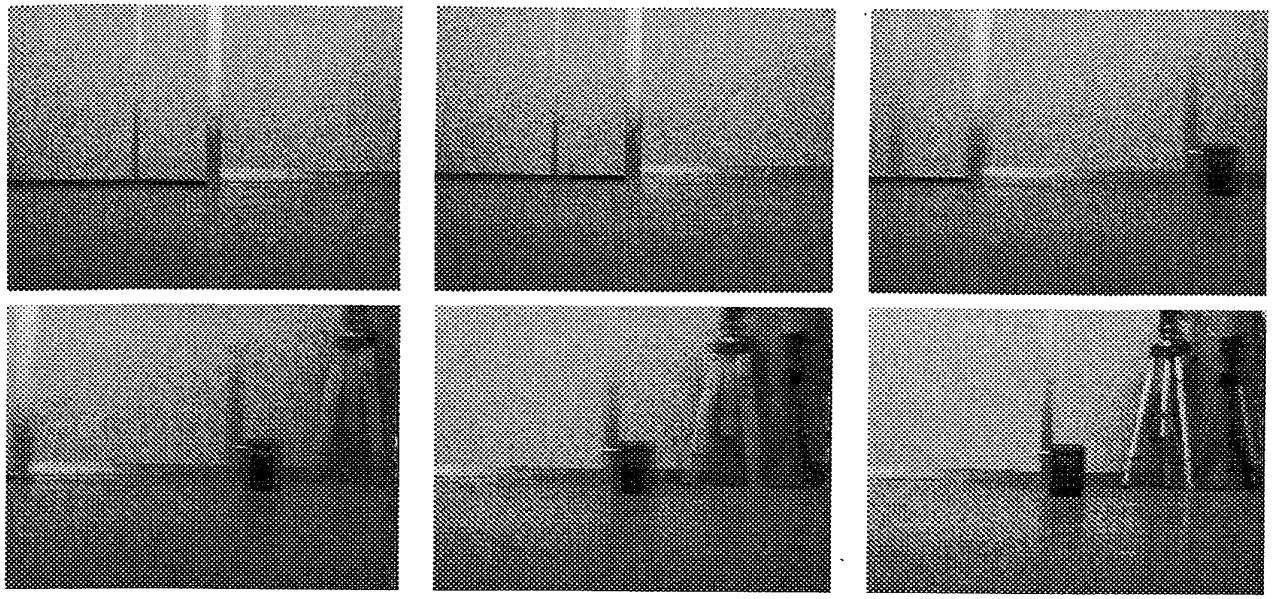

図 11: 自動音源方位撮影システムの動作 (首振りカメラからの視点)

法を使用することで、ある程度反響のある部屋の中でも音源の方位を特定することが可能であることを示した。また、その精度も条件によっては十分実用に耐えるものであった。今後は

- より多くのマイクロフォンアレイを用いたシステムへの展開
- 複数音源への対応
- 方位推定のみではなく距離の推定への展開
- 処理速度の向上およびリアルタイム処理への配慮
- 複数音源での信号の分離

などの展開が考えられる。今回構築されたシステムは映像分野に音声の情報を利用する有効な方法として十分機能するものであり、今後更なる情報の相互利用が期待される。

参考文献

[1] Don H. Johnson - Dan E. Dudgeon. Array Signal Processing: Concepts and Techniques. pp. 373-393. PRENTICE HALL, 1993.

[2] Futoshi Asano, Masataka Goto, Katunobu Itou and Hideki Asoh. Real-time Sound Source Localization and Separation System and Its Application to Automatic Speech Recognition. J. Acoust. Soc. Am., Vol.88, No.1,pp.159-168. 1990

[3] 石黒. 能動視覚とその応用. 松山, 久野, 井宮 (編), コンピュータビジョン 技術評論と将来展望, 第 15 章, pp. 219-229. 新技術コミュニケーションズ, 1998.