

研究テーマ 全方向ステレオシステムの開発

研究者	棚橋英樹	財団法人ソフトピアジャパン	雇用研究員
	桑島茂純	株式会社ビュープラス	共同研究員
	桑島裕之	株式会社ビュープラス	共同研究員
	佐藤 淳	名古屋工業大学	兼業研究員
	池田尚志	岐阜大学	共同研究員
	速水 悟	岐阜大学	共同研究員
	伊藤 昭	岐阜大学	共同研究員
	斉藤文彦	岐阜大学	共同研究員
	加藤邦彦	岐阜大学	共同研究員

フェーズ I

1 研究の概要

マンマシン・インターフェースやロボットの遠隔操作を検討した場合、動的環境を理解することが重要であり、広範囲で高解像度な 3 次元情報を獲得することが求められる。そのためのセンサとして、本研究では、完全に死角のない全方向のカラー画像と 3 次元情報をリアルタイムに取得可能な全方向ステレオシステム(Stereo Omni-directional System (SOS))の開発を行った。SOS は、正 20 面体の各面に 3 眼ステレオユニットを配置した計 60 個のカメラから構成することにより、完全に死角のない全方向のカラー画像と 3 次元情報をリアルタイムに取得することが可能となった。

2 研究の目標

従来のカメラシステムには、レンズと撮像素子によって決まる「限られた画角」が当然の制約条件として存在していた。コンピュータビジョンで用いられるカメラシステムについても例外ではなく、画角によって決まる限られた範囲しか注目できない、対象あるいはその一部がフレームアウトしてしまう、などの問題を引き起こし、本来解くべき課題をより複雑にしていた。このような問題の解決を検討する場合、特にコンピュータビジョン分野においては旧来のカメラシステムの概念に捕らわれる必要はなく、カメラシステムそのものから見直すというアプローチを検討することが出来る。本研究では 60 個のカメラを用いて正 20 面体の各面上に 3 眼のステレオビジョンユニットを配置することによって、観測点を中心とした全ての方向の 3 次元カラー情報を均一な解像度で同時に得ることができる全方向ステレオシステム(SOS)の開発を行う。

3 実施内容

我々は家庭内の限られたメンバーが生活する環境で、家電品を人の自然な動作によって制御する知的家電システム及び人間とその周辺の状況を正しく理解することで人間活動を支援する生活支援システム等の実現を目的としている。そのための 3 次元シーン情報をセンシングするシステムとしては、以下の要求を満たす必要があると考えられる。

- (1) センサを中心とした全方向の情報を同時に取得できること。
- (2) リアルタイムで取得可能であること。
- (3) ステレオ画像が取得できること。
- (4) 方向による解像度の差がない高分解能の情報が得られること。
- (5) 高精度の全方向カラー情報が得られること。

そこで、以上の 5 つの要求を満たすセンサとして、正 20 面体上に 3 眼のステレオビジョンシステムを配置することにより、全方向の 3 次元シーン情報を取得可能なシステムを設計、試作した。本システムは、各カメラの相対位置関係を確定しているため、計測したい場所にカメラを移動させることが容易であり、より適用範囲の広い利用が可能である。本システムの概観を図 1 に示す。



図1 試作システムの概観

3.1 試作システムの構成



図2 測地ドーム型カメラ配置



図3 ステレオユニット

3次元空間を等間隔に分割する方法として、正多面体を用いて、さらにその各面を規則的に3角形に分割する方式として測地ドームがある。まず最初は測地ドーム方式として複数のカメラを配置し、3次元空間の全ての方向を等解像度で等間隔に画像を得る構成を考えた(図2)。このようなカメラ配置を行った場合、各カメラは3次元空間の異なる方向を向いており高解像度の画像を得ることができるが、画像から3次元情報を得る場合、エピソード拘束が全ての画像対で異なり、ステレオ視における対応点探索を効率的に行うことが困難である。

ステレオ画像対に対して水平または垂直のエピソード拘束を用いることが可能なカメラ配置を行うことができれば、対応点探索を高速かつ効率的に行うことができる。

ところで画像から3次元距離情報を計算する場合、最低2個のカメラがあればステレオ視が可能であるが、もう一つカメラを追加することにより、正確な対応付けが可能となる。こうした考え方をを用いて、市販されているステレオビジョンシステムに、TRICLOPSがある。

そこで同様な考え方をを用いて、本システムではカメラを図3のような3個のカメラ対からなるステレオユニットに分割し、このユニットを正20面体の各面に配置した。ユニット上の各カメラは、一つを基準カメラとして上下と左右の2つのステレオペアを構成する。また、各カメラの画像面は同一平面上にあり、光軸が互いに平行になるようにカメラを配置した。

ステレオユニット毎に各カメラのキャリブレーションを行うことにより、レンズ歪み、焦点距離、画像中心等のカメラ固有パラメータ及び3個のカメラ間の位置関係を求め、これらのキャリブレーション

ションデータを用いて、ソフトウェアで標準ステレオ画像を生成する。これらの画像は、正方ピクセルを持ったピンホールカメラモデルの性質を有し、上下、左右のステレオ画像対に対して水平、垂直のエピポーラ拘束を満たしている。そのため、ステレオペアにおける対応点探索を高速に行うことが可能である。また、この2つのステレオペアを用いることにより、それぞれのステレオペアから得られる結果をその信頼性の評価値を用いて統合することにより、計測の信頼性を上げることが可能となる。濃度パターンが一方のエピポーラ線に平行な場合でも、もう一方のステレオペアにより対応付けが可能となり、計測不可能な箇所を軽減することができる。

3.2 システムのダウンサイジング

ステレオ処理においては、カメラから遠く離れた3次元情報を精度良く求めるためには、ステレオペアを構成する2個のカメラ間隔（ベースライン長）を長くする必要がある。

しかし、ベースライン長を長くするとステレオユニットが大きくなり、そのユニットをそのまま正20面体の面内に配置すると、面の一边の長さは大きくなり、システムのサイズを大きくせざるを得ない。システムのサイズが大きくなると、隣り合ったカメラ面のどちらかからも見えない距離 d （システム中心 O からの距離）が大きくなり、システムの中心近くで近接死角を生じる。このため全ての方向が見え始めるシステム中心からの距離 d を小さくするためには、多面体の大きさを小さくする必要がある。

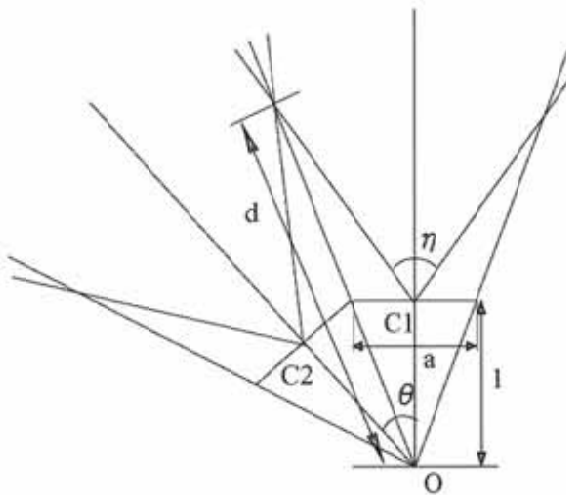


図4 全方位取得可能な距離と視野角の関係

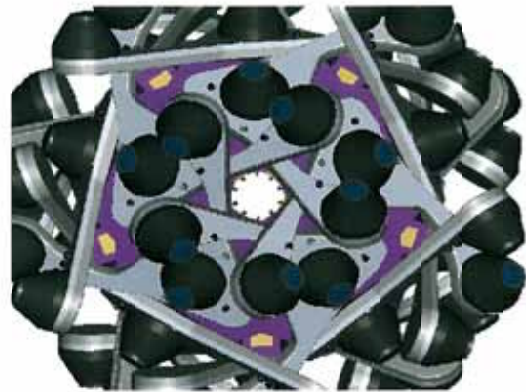


図5 ユニット配置

ちなみにシステム中心から全方向の完全な画像が得られる最初の距離 d は、図4より、次式のようになる。

$$d = l \left(\cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\theta}{2}\right) \tan\left(90 + \frac{\theta - \eta}{2}\right) \right) \quad (1)$$

$$l = a \frac{3 + \sqrt{5}}{4\sqrt{3}} \quad (2)$$

ここで

l : 正20面体の中心から面までの距離

a : 正20面体の面の1辺の長さ

θ : 隣接する面のなす角 (deg)

η : カメラ視野角 (deg)

O : システム中心

C_1, C_2 : カメラ中心

以上のように、一般的にシステムを大きくすると3次元情報の精度は向上するが、逆に近接死角は大きくなりこれら2つを同時に満たすシステムの条件は相反する。そのため、本システムでは十分な3次元情報の精度を持つステレオユニットを、隣り合うステレオユニットがなす基準平面が互いに交差するように配置することによってカメラのベースラインと面のサイズを独立させる構成を考案した。これによってサイズを大幅に小さくすることが可能になる。次なるサイズの限界は各ユニットが互いに隣り合うユニットのカメラの視野を遮らないようにユニットを配置することであり、3次元CADを用いて各カメラ毎に視野角のシミュレーションを行った。これにより図5に示すような配置の3次元形状を得た。こうしたユニットの配置により、システムから遠く離れた3次元情報を精度良く取得可能でありながら、近接死角を小さくすることが可能となった。

3.3 光学パラメータの設定

本システムでは、ステレオユニットを正20面体の各面上に配置し、3次元空間上の全ての方向の3次元情報を取得するために、少なくとも41.81度以上の画角が必要である。本システムでは、人及び人の手が届く範囲の全ての空間の情報を得ることを目的とし、水平画角96.6度、垂直画角71.9度のレンズを使用した。これによって中心から400mm以上の空間は死角なく計測することが可能となった。また、正確な3次元情報を得るために、3つのカメラの幾何学的位置関係のキャリブレーションが必要であるが、これらのキャリブレーションをあらかじめステレオユニット毎に行っておくだけで良いと言うメリットがある。

これらのキャリブレーションには、Tsaiの手法を用いて、レンズ歪み、焦点距離、カメラ中心、水平垂直のスケール比のカメラ固有パラメータ及び3個のカメラ間の位置関係を求めた。また、得られたキャリブレーションパラメータを用いて、カメラから得られた画像からレンズ歪みとカメラ間のアライメントずれの影響をソフトウェアで取り除いた。こうして得られた画像は、正方ピクセルでピンホールカメラモデルの特徴を有し、ステレオユニット上の左右の画像及び上下の画像は、それぞれ水平、垂直のエピポーラ拘束を満たしている。

図6に本システム上のステレオユニットの1個のカメラから得られた画像と補正後の画像を示す。図6(a)は入力画像であり、図6(b)はレンズ歪みを補正した画像である。



(a) 入力画像

(b) 補正後画像

図6 レンズキャリブレーション

3.4 システムの試作

以上の設計に基づいて、図7に示すようなシステムを試作した。また表1に試作システムの諸元を示す。

ベースとなる正20面体は専用のスタンドに取付られている。また各ステレオユニットの基準カメラの中心は面の重心位置に配置した。ユニットにおける上下、左右のステレオベースラインは、約90mmでありながら、3.2節で述べたユニットの配置により全体として直径27cmまでにコンパクトに構成できた。

本試作システムは、センサユニット、画像メモリユニット、ステレオPCから構成される。センサユニットは、2つのステレオユニットを1組とした10組から構成される(図7)。

試作システム上のカメラによる画像の取得は、同期を取って全ての画像を取得する。各ユニットからは、1枚のカラー画像と3枚の白黒画像からなるステレオ画像が取得され、全方向の20枚のカラー画像とステレオ画像を1セットとして、15セット/秒でメモリユニットに転送する。その後、PCIバスによりパソコン上に画像を転送し、ステレオ処理を行う。

表1 試作システムの緒元

基本形状	正 20 面体
撮像素子	1/3' CMOS Color Image Sensor
素子の解像度	640(H) × 480(V)
焦点距離	2.9 mm
画角	96.6 deg(H) × 71.0 deg(V)
ベースライン	90mm
カメラヘッド直径	27cm
重量	4.5kg

4 結果

図8に本システム上の1個のステレオユニットから得られた画像を示す。図8(a)は視差画像であり、図8(b)、(c)、(d)は、それぞれ基準カメラから見て上のカメラからの画像、基準カメラから見て左のカメラからの画像、基準カメラからの画像である。図8(a)は、視差が大きい点ほど輝度値を高く表示している。黒い点は、視差情報が得られなかった点を示す。輝度変化やテクスチャ変化が少ない部分では誤対応が生じているが、人間領域と背景領域の位置関係を把握するのに十分な精度で3次元情報が得られているのがわかる。なおステレオ処理は、ステレオビジョンソフトウェアライブラリを用いて行った。

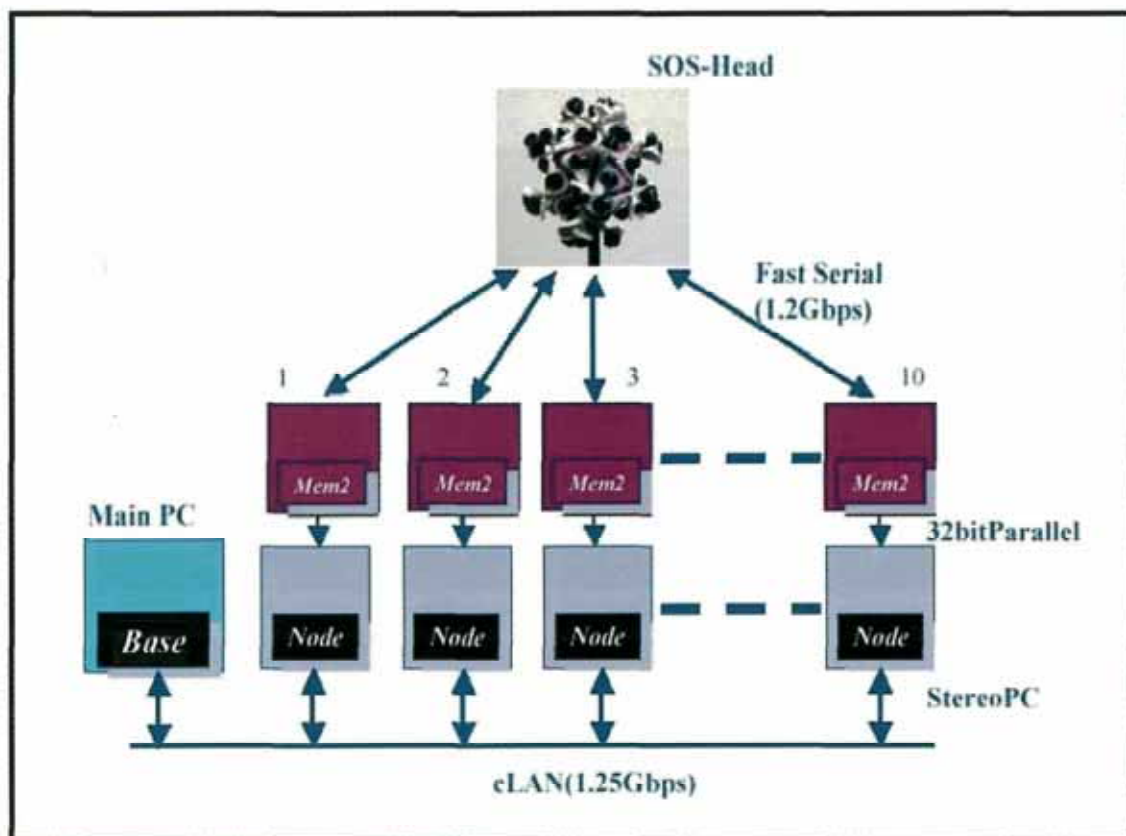


図7 システム構成

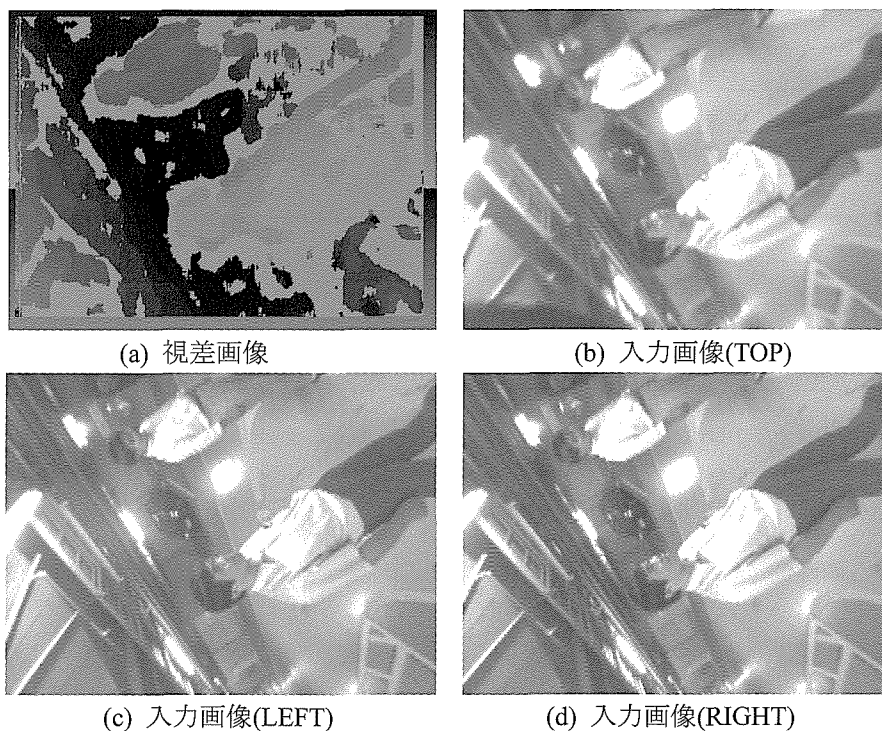


図 8 システムからの画像例

試作したシステムにより取得された 20 枚のカラー画像から全方向パノラマ画像を生成した。図 9 は、取得した 20 枚のカラー画像をカメラ視野角を用いて、システムから約 1m 付近の範囲の画像を切り出し、正 20 面体の展開図面上に張り付けた作成した全方向パノラマ画像である。空間全ての方向を 20 個のカメラで取得しているため高精度なカラー情報が得られている。

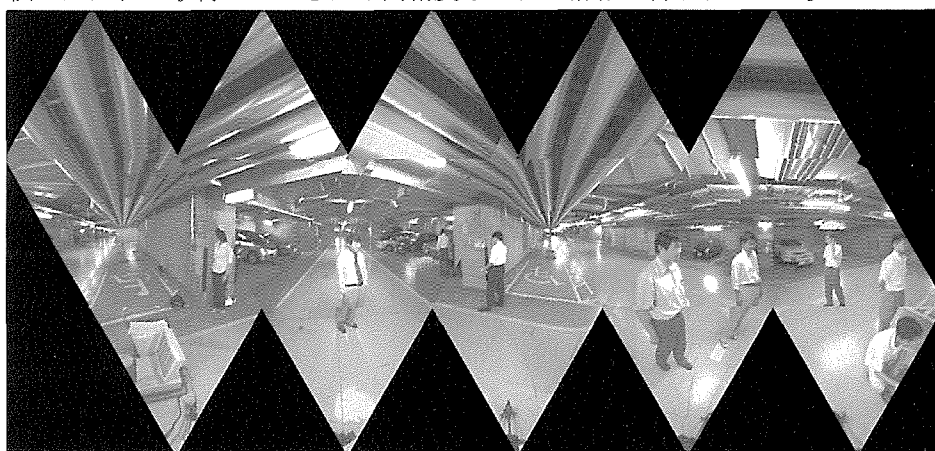


図 9 全方向パノラマ画像

最後に個々のユニットから得られた 3 次元情報を統合して 3 次元空間を構築した。図 10 は、再構成した 3 次元空間を任意方向から見た透視投影画像である。人間とその周りの環境の位置関係等が 3 次元的に再構築されているのが分かる。

本論文では、人間とその周囲の実環境の全ての方向のカラー画像と距離情報を同時に取得可能なシステムを試作し、取得されたカラー画像から全方向パノラマ画像の生成と 3 次元カラー情報を用いた 3 次元環境の構築例を示した。

本システムは、方向による解像度の偏りがなく、高解像度のカラー画像と距離画像をリアルタイムで取得可能である。また、複数のステレオビジョンユニットを用いたことにより、ステレオ処理におけるキャリブレーションは、ユニット毎に行うことが可能である。さらにこれらのユニットを、お互いの視野を遮らないように配置することによってステレオから得られる距離精度を維持しながら、システム全体の小型化が実現できた。

本システムの開発によって 3 次元空間全ての方向の複数の人や物体の動きを同時に追跡する可能性を開いた。



図 10 構築した 3 次元空間透視投影画像

フェーズ II

1 研究の概要

フェーズ I では、センサを中心とする完全に死角の無い全方向のカラー画像と距離情報をリアルタイムに取得する全方向ステレオシステム (SOS) (プロトタイプ Jupiter) を開発した。フェーズ II では、新たに、移動体ビジョンを指向しセンサの直径 11.6cm と小型化を進めたモデル (Venus)、および環境監視システム等の用途を指向しシステムサイズを小型化しながらも 10cm のステレオベースラインを確保したモデル (Saturn) の開発を行った。また、SOS の画像情報と音響情報との統合したマルチモーダルインタラクションの実現を目指し、正 20 面体の頂点にマイクロフォンを立体的に配置したマイクロフォンアレイシステムを開発した。

2 研究の目標

人間共存型の移動マシンを検討する場合、人間の移動生活空間という複雑で、かつ時間的変化の大きい環境への適応性が重要となる。SOS は全方向のカラー画像と 3 次元情報をリアルタイムに取得可能であり、激しい姿勢変化の予想される移動体ビジョンにとって特に有利な条件となる。移動体搭載するために更なる小型モデル (Venus) の開発を行う。環境監視システム等の使用を想定し、Jupiter と比較してカメラヘッドサイズを小型化しながらも、ベースライン長 10cm を確保し、より広範囲の距離情報を安定に取得することを可能とするモデル (Saturn) の開発を行う。また、SOS の画像情報と音響情報を統合したマルチモーダルインタラクションの実現を目指し、音響情報によって 3 次元空間における音源方向の推定をリアルタイムで行うシステムの開発も行う。

3 実施内容

3.1 小型全方向ステレオシステム Venus(miniSOS)の開発

センサを中心とした全方向のカラー画像と距離情報をリアルタイムに取得し、かつセンサ自体を小型軽量とした新しい全方向ステレオシステム (以下 miniSOS) について報告する。

本システムは正十二面体を基本形状とし、各面に 3 眼ステレオユニットを配置する設計によりセンサを中心とする完全に全方向のカラー画像と距離情報を死角無くリアルタイムに取得することを可能とする。

更に本システムは、移動体への搭載を指向し、直径 11.5cm、重量 615g という極めてコンパクトな形状を実現した。センサを中心とした、完全に上下を含め全ての方向が死角無く取得できる性質は、センサ自体の任意の回転に対して常に不変の画像が取得できることを意味しており、特に姿勢変化の大きな移動体等においてそのメリットは大きい。

本稿では本システムの設計とその特徴について述べる。

3.1.1 システムの設計



図 11 miniSOS の外観

表 2 miniSOS の緒元

基本形状	正 12 面体
撮像素子	1/4' カラーCMOS
素子の解像度	640(H)x480(V)
焦点距離	1.9mm
画角	101 deg(H)×76 deg(V)
ベースライン	50mm
フレームレート	15fps
カメラヘッド直径	11.6cm
重量	615g
消費電力	約 9W(15V、約 0.6A)

3.1.1 概要

図 11 にシステムの外観を示す。本システムでは移動体への搭載を強く意識しているため、センサを中心とした全方向のカラー画像と距離情報を死角無くリアルタイムに高解像度で取得する能力に加え、小型かつ軽量であることを重要視した。またセンサに付随する処理系およびその間のインタフェースについてもセンサの移動を考慮し、コンパクトかつシンプルな構成をとる方針とした。

3.1.2 カメラヘッドの設計

本システムでは、正 12 面体を基本形状とし、その各面上に 3 眼ステレオユニットが配置される構造とした(合計 36 カメラ)。距離情報の精度を確保するためにはステレオユニットのカメラ間距離(ベースライン)を適切に確保する必要があるが、一方でベースラインの増加はカメラヘッド全体の大きさを増大させる。本システムでは移動体による周囲の安全確認のための情報収集を最低限可能とするため、50mm のベースラインを確保することとした(理論値で 6cm/1m の精度を想定)。

このようなユニットを正 12 面体の各面内に配置する構造をとると、カメラヘッド全体の大きさが極めて大きくなる。この問題を解決するために従来我々は SOS において、3 眼カメラを L 字型アームによってマウントし、正 20 面体における隣り合うステレオユニットがなす基準平面が互いに交差するように配置することにより、ベースラインと面のサイズを独立させる手法を考案した。

本システムでは、更なる高密度実装を可能とするため、3 眼ユニットを T 字型を基本形とする非対称アームによってマウントする設計とした(図 12)。ここで、ユニット上の各カメラは同一平面上にあり、それぞれの光軸は互いに平行である。また、センタカメラを中心としてベースライン(50mm)が直交するように 2 つのカメラが配置されている。これにより、各ステレオ対がそれぞれ水平、垂直のエピポラ拘束を満たし、対応点探索の処理コストを低減することができる。

図 13 にステレオユニットにより取得された画像の例を示す。画像サイズは通常の 640×480 ピクセルの他、撮像素子レベルで出力を 320×240 ピクセルとする機能を実装している。これにより例えば低解像度でイベント検出を行い、イベントの発生が認められる領域を含むユニットのみ解像度を高めるなどの制御が可能となり、全方向に関する膨大な情報量による、通信系のトラフィック及び処理系の負荷をコントロールすることが可能となる。

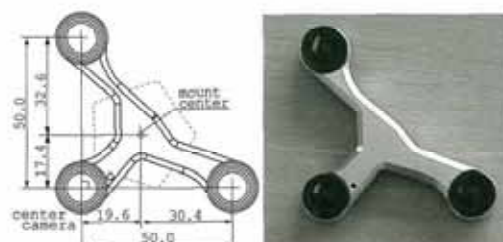


図 12 ステレオユニット



図 13 ステレオユニットからの画像例

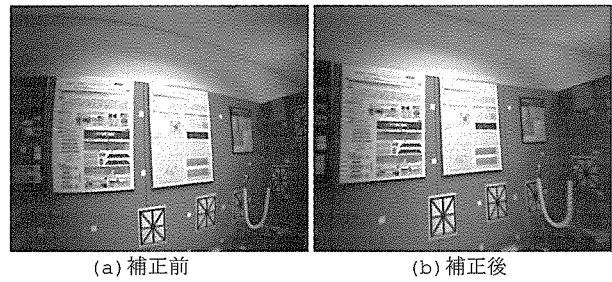


図 14 レンズキャリブレーション

3眼カメラのうちセンタカメラはステレオ対としての役割の他、カラー画像を取得する。他の2眼はモノクロ画像のみを取得する。各ユニットはそれぞれユニット単位でキャリブレーションを行っており、レンズ歪み（図 14）およびカメラ間のアライメントずれの影響をソフトウェアにより除去している。なおステレオ処理に関しては、ステレオビジョンソフトウェアライブラリを使用している。

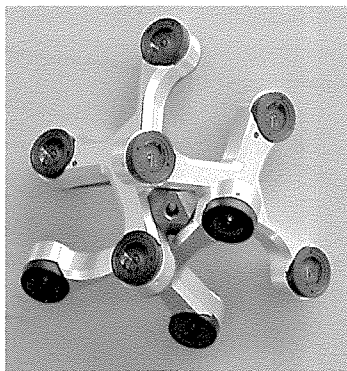


図 15 ステレオユニットセット

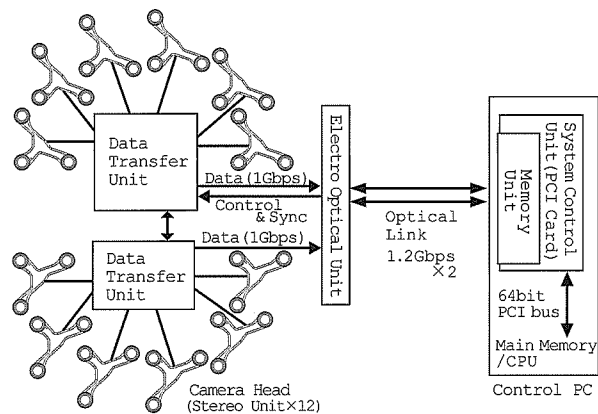


図 16 システム構成

次にステレオユニットを正 12 面体形状へマウントする方法について述べる。本システムでは先に述べたように、移動体搭載のために小型・軽量であることが設計上最もプライオリティの高い項目となる。そこで我々はベースライン 50mm を保ちつつ、かつ各カメラの物理的干渉及び互いの写り込みが発生しない条件において、カメラヘッドサイズが最も小さくなるようにステレオユニットの配置を決定した。また、図 15 に示すように、正 12 面体のベース形状を実体として持たず、各ステレオユニットのアームがカメラヘッド全体の強度を支えるフレーム構造の一部としての役割を兼ねることで高密度化および軽量化を実現している。ステレオユニットは 3 ユニットずつ固定金具によりステレオユニットセットとして固定され（図 15）、これを 4 セット組み合わせることで正 12 面体を構成する。

ここで、各カメラの画角は水平 101° 、垂直 76° であり、システム中心から半径 200mm 以上の空間を死角無くカバーする。

3.1.3 処理系のインタフェース

図 16 に本システムの構成図を示す。カメラヘッドによって取得された画像群は本体支柱下部に設置された光変換ユニットにより 1.2Gbps x 2 の光信号として出力される。メモリユニット及びコントロールユニットは 1 枚の PCI ボード上に実装されており、1 台の PC のみで全てのコントロールおよび画像の取得が可能である。取得された画像群は 1 フレーム分ごとに、同期の取られた全ユニット分の画像を直列に結合した形になるため、それぞれの画像に対するアクセスが容易でソフトウェア開発にかかる負担を軽減することができる。また、カメラヘッドと PC 間の通信手段としては、

理想的には無線接続が望ましいが、現状では12ステレオユニットによって取得される画像群を15fpsで転送するための十分な帯域を確保することが困難であるため、移動体が発するノイズ等に対して頑健であることも考慮して光インタフェースを採用した。これにより、取り回しの容易な細線ケーブル2本のみで十分な帯域をもって画像群を転送することが可能になり、かつカメラヘッドと処理系との距離を十分に確保することを可能とした(現在最長100mの光ケーブルを使用)。

3.1.4 結果



図 17 取得されたカラー画像



図 18 正 12 面体展開画像

本システムから取得されるカラー情報は、図17のように各カメラの取り付け角に応じた回転を含んだ12枚の画像として取得されるため、このままではそれぞれの位置関係を直感的に把握することが難しい。

本システムはコンテンツ製作を指向したものではなく、しかも実際の3次元座標がステレオ処理によって得られることから、2次元画像のみを用いたモザイクキングに重点を置くものではないが、簡易的にカメラ間の位置関係を把握するために用いるモザイクキング処理について検討を行ったので紹介する。

本システムでは、各センタカメラの光学中心が1点に集中していないため、2次元画像のみを用いて中心射影画像を完全に表現することはできないが、対象までの距離を仮定することにより、近似的な合成画像を作成することができる。図18はカメラの視野角を利用してシステムの中心から約1.5m付近の画像を切り出し、正12面体の展開図上に示した全方向画像である。図17と比較してカメラ間の位置関係が容易に判別できることがわかる。



図19 パノラマ展開画像

更に図19は、図18の正12面体を、それを取り囲む半径1.5mのシリンダー上に投影し、展開した全方向パノラマ画像で（上端はセンサの完全に真上、下端はセンサの真下を示す）各カメラの映像を更に連続的な画像として観測することが可能である。

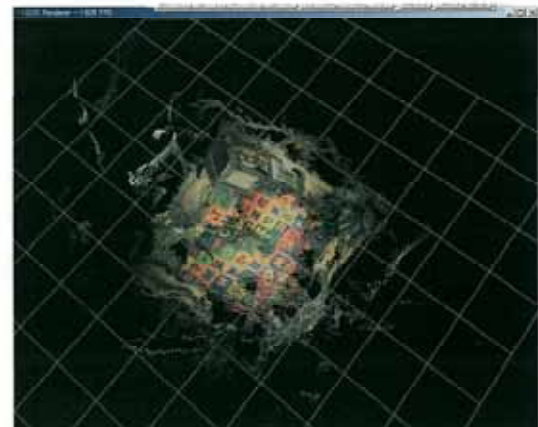


図20 3次元ポイントクラウド画像

図20は、各ユニットから取得された3次元情報を、カメラヘッドの設計時に用いたカメラ配置パラメータ（以下デザインパラメータ）を用いて1つの座標系に統合し、可視化したポイントクラウドデータである。miniSOSでは、先に述べたように各ステレオユニットはそれぞれ単体で十分なキャリブレーションを行っており、ユニット同士の組み付けに関しても、予め高い精度を意識した組み付け加工を行っている。

図20を見ると、デザインパラメータを用いた簡易的な統合ではあるが、各ユニットから得られた距離情報が良好に統合できていることがわかる。なお我々は既に、先に開発したSOSシステムにおいて、並進パラメータをデザインパラメータに固定した上で、その他の運動パラメータを同時に最適化する方法を提案しており、今後同方法に従って更に高精度のユニット間キャリブレーションを行う予定である。

移動体においては障害物回避、行動計画等のために3次元情報が利用できるメリットは大きい。特に本システムではこのような情報を死角無く全方向にわたってリアルタイムに取得できるため、人混みを移動することを想定した移動体などへの応用が検討できる。

センサを中心とした完全に全方向のカラー画像と距離情報をリアルタイムかつ高解像度で取得することが可能で、かつセンサ自体を小型軽量とした新しい全方向ステレオシステムについて報告した。開発したハードウェアの設計、および画像統合等のソフトウェアについて述べた。カメラヘッドを小型軽量とし、かつ処理系及び通信系を扱いやすいシンプルな構成とすることで移動体へのインプリメントを容易にした。

3.2 高精度全方向ステレオシステム Saturn の開発

3.2.1 概要



図 21 Saturn の外観

表 3 Saturn の緒元

基本形状	正 12 面体
撮像素子	1/3' カラーCMOS
素子の解像度	640(H)×480(V)
焦点距離	2.5mm
画角	115.5deg(H)×83.8 deg(V)
ベースライン	100mm
フレームレート	15fps
カメラヘッド直径	21.0cm
重量	1.85kg

Saturn の外観を図 21 に、緒元を表 3 に示す。初代として開発した SOS(Jupiter)は、正 20 面体を基本形状とし、その各面上に 3 眼ステレオユニットを配置する構造を基本に設計した。ここで、ステレオユニット上の各カメラは同一平面上にあり、それぞれの光軸は互いに平行である。また、センタカメラから対になる 2 つのカメラまでの基線（ベースライン）が直交するように配置されている。これにより各ステレオ対がそれぞれ、水平・垂直のエピポーラ拘束を満たし、対応点探索の処理コストを低減することができる。距離情報の精度確保のためにはベースライン長を適切に確保する必要があるが、一方でベースライン長の増加はカメラヘッド全体の大きさを増大させる。この問題を解決するために我々は 3 眼カメラを L 字型アーム上にマウントし、隣り合うステレオユニットがなす基準平面が互い交差するように配置することにより、ベースライン長と面のサイズを独立させる手法を考案した。

Venus は、移動体ビジョンでの使用を想定したモデルである。Jupiter の開発で得られた知見を活用し、徹底した小型軽量化を進めることで、体積比で 1/12 以下、重量比で 1/7 以下となる直径 11.6cm、重量 615g を実現した。カメラヘッドサイズの小型化のためには、レンズおよび撮像素子の小型化、またベースライン長の縮小が必要であり、距離情報の安定取得が可能となる限界範囲とはトレードオフの関係にある。しかしながら、Venus はより高い精度が必要な場合にはセンサ自体が測定対象に自ら接近することが可能な移動体ビジョンでの使用を想定しているため、障害物回避などのタスクに影響が出ない範囲で小型化を優先させた。また、正 12 面体を基本形状とし、3 眼ユニットを T 字型を基本形とする非対称アームによってマウントすることにより、大幅な小型軽量化を実現しつつ、ベースライン長 5cm (Jupiter 比 1/1.8)を確保した。

Saturn は今回新たに開発した環境監視システム等での使用を想定したモデルである。Venus の開発で得られた小型化に関する知見を活用し、Jupiter と比較してカメラヘッドサイズを小型化しながらも、ベースライン長 10cm (図 22)を確保し、より広範囲の距離情報を安定に取得することを可能にした。

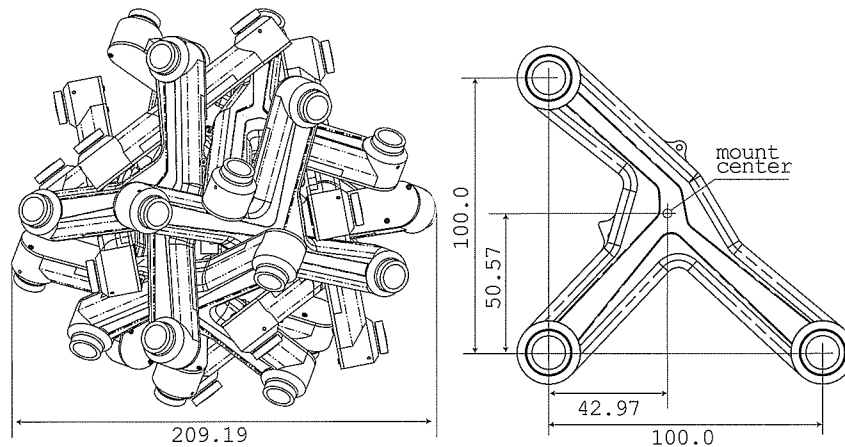


図 22 Saturn の外形(左)およびステレオユニット(右)

3.2.2 処理系とインタフェース

図 16 に Saturn のシステム構成図を示す(Venus と共通)。カメラヘッドによって取得された画像群は本体支柱下部に設置された光変換ユニットにより 1.2Gbps×2 の光信号として出力される。メモリユニット及びコントロールユニットは 1 枚の PCI ボード上に実装されており、1 台の PC のみで全てのコントロールおよび画像の取得(15fps)が可能である。取得された画像群は 1 フレーム分ごとに、同期の取られた全ユニット分の画像が直列に結合された形となるため、目的の画像へのアクセスが容易である。なおステレオ処理に関しては、ステレオビジョンソフトウェアライブラリを使用している。

3.2.2 結果

本システムから取得されるカラー情報は、各カメラの取り付け角に応じた回転を含んだ 12 枚の画像として取得されるため、そのままではそれぞれの位置関係を直感的に把握することが難しい。本システムにおいては、各センタカメラの光学中心が 1 点に集中していないため、2 次元画像のみを用いて中心射影画像を完全に表現することはできないが、対象までの距離を仮定することにより、近似的な合成画像を作成することができる。図 23 はカメラの視野角を利用してシステムの中心から約 1.5m 付近の画像を切り出し、更にそれを取り囲む半径 1.5m のシリンダー上に投影し、展開した全方向パノラマ画像である。簡易的な統合ではあるが、これにより各カメラの映像を連続的な画像として観測することができる。

図 24 は、各ユニットから取得された 3 次元情報を、前節でキャリブレーションを行ったカメラ配置パラメータを用いて 1 つの座標系に統合し、テクスチャ情報をパッチとして貼付した画像である。このように、完全に死角の無い全方向のカラー画像情報と距離情報をリアルタイムに取得可能とするセンサの能力は、次章で述べるサーベイランスシステム等の用途においても人物の動線検出を極めて容易にするなど、ソフトウェア開発にかかる負担を大幅に軽減することができる。



図 23 パノラマ展開画像



図 24 任意仮想視点画像

3.3 マイクロフォンアレイを用いた 3 次元空間における音源方向の推定

近年、人間を含む実環境において、コンピュータによる実環境認識や人間とのインタラクションの実現を目指した研究が行われるようになってきた。音響情報については、音源の位置情報を用いた音源分離の他、テレビ会議システム等での話者方向の特定、移動ロボット等の環境認識システム、監視システム、騒音の分析に代表されるように離れた地点の発話者や音源を認識するシステムなど幅広い分野で発話者・音源の方向推定が求められており、それに対する研究が行なわれている。画像分野においても、全方向ステレオシステムを用いた画像認識など、3次元空間についての研究もさまざまな行われており、音響情報と画像情報を3次元空間で統合させる必要性が出てきた。

音源を推定するシステムとして、時間差に基づく方向推定法、反響音を考慮せず直接入力装置に入ってくる音に着目したビームフォーム法や、遅延和アレー法、MUSIC 法、ESPRIT 法などがある。しかし、時間差による音源方向を検出する方法には、音源から発している音以外にも雑音や壁などによる反射音なども存在し、音源から直接来る音が検出しにくいという欠点があるので難しい。また、遅延和アレー法や ESPRIT 法はマイクの位置を等間隔に置かなくてはならないので物理的制約がある。マイクの位置に制約されず、雑音環境下でも音の到来方向に対して一番正確に検出する方法が、MUSIC 法(MULTIPLE SIGNAL CLASSIFICATION)である。MUSIC 法は音波における音源方向の推定で使われているだけでなく、レーダなどで用いられているように電磁波の到来方向推定でも使用されている。従来、MUSIC 法を用いた 2 次元空間においてリアルタイムで音源を認識するシステムや、リアルタイム性はないが、3次元空間において音源を推定するシステムの研究が行われてきた。

本研究では、3次元空間において、発話者・音源がどの方向から音が発しているのかをリアルタイムで推定するシステムを構築した。このシステムは、3次元空間での音源方向推定によって、物音などがどの方向から発せられたのか、などの発話者・音源の位置情報を得ることができる。3次元空間において音響情報と画像情報との統合を行うための重要な第一歩と位置づけられる。

3.3.1 音源を推定するシステム

3.3.1.1 音源方向を推定するシステムの流れ

小型ハードウェア rasp を使って、音源方向をリアルタイムで推定する。分解能は水平面に 5 度間隔と垂直面では 3 方向である。

また、MUSIC 法を使うことにより、同時に複数の音源方向の推定もできる。

音源を推定するアルゴリズムは次のようになっている。

- (1) A 行列・FFT のプログラム読み込みなどの準備。
- (2) マイクの音の入力。
- (3) 入力した音データを MUSIC 法で使えるようにするために相関行列にする。
- (4) A 行列のデータとマイクから拾った音データとともに MUSIC 法を用いて計算して、MUSIC スペクトルを求める。
- (5) MUSIC スペクトルの値を大きい順に角度を入れ替える。
- (6) 各方向から出てきた MUSIC スペクトルの最大値を比較して垂直方向の位置を決める。
- (7) rasp に表示できるようにデータを変換する。
- (8) rasp に色別にグラフ表示、ホスト PC に角度、方向を表示する。そして、データを clear して、2 に戻る。

3.3.1.2 3次元空間での音源方向推定

3次元空間で立体的な音源方向を推定するシステムを作るために、2次元における MUSIC 法を複数回用いる。音データと各方向における A 行列を2次元空間と同じように上・水平・下のそれぞれの方角に対して MUSIC 法を用いる。そして、各方向から出てきた MUSIC スペクトルを大きい順にソートを行って、各方向の MUSIC スペクトルの最大値を比較して、その中から一番大きい値を方向として決定する。そして、結果を表示するときは、MUSIC スペクトル値が最大の方向の各角度におけるスペクトル値を使って結果を出す。

図 25 に3次元空間における音源方向の推定を行うアルゴリズムを示す。



図 25 垂直方向を決定するプログラム

3.3.1.3 キャリブレーション

MUSIC 法を用いて、音の到来方向を推定するシステムでは、基礎データとなる A 行列を生成するために、あらかじめデータ収集を行わなければならない。そこで、あらかじめ特性が分かっている音響データを用いて、各角度の音響データを録り、そのデータを基に A 行列を生成する。

計測時のスピーカーの位置は、次の図 26 である。また、キャリブレーションの様子を図 27 に示す。

マイクロフォンアレイとスピーカーの位置の距離は 150cm とした。

垂直面は、

上・・・床から高さ 150cm

水平・・・床から高さ 90cm

下・・・床から高さ 30cm

の位置にスピーカーを置いた。

上・水平・下の各 3 方向に対して、0°から 360°の 5 度間隔で計測するため、測定点は 72 点×3 方向になる。測定点の位置を図 26 に示す。

0°から 355°のそれぞれの点において、tsp(Time-Stretched Pulse)を作り、その tsp を入力させて、インパルス応答を再構成している。この tsp を観測したデータから、A 行列を生成する。

A 行列は、 $72 \times 1 \times 1 \times 257$ のデータ量である。それぞれの配列に入っているデータは次のとおりである。

72・・・ $0^\circ \sim 355^\circ$ の 5° おきの点の数

1・・・距離方向における数

6・・・マイクの数

257・・・512 点フーリエ変換した各周波数点におけるデータ数

この生成された A 行列を使って MUSIC 法によって音源の方向を推定する。

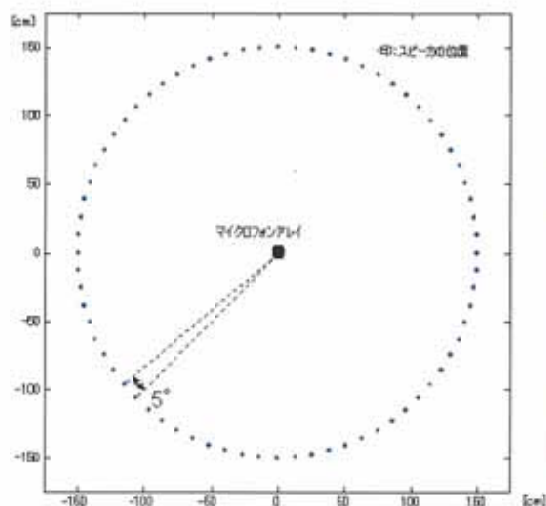


図 26 キャリブレーションの計測位置



図 27 キャリブレーションの様子

3.3.1.4 出力結果

MUSIC 法で音源の方向を推定した結果を出力する。出力は、小型ハードウェア rasp に MUSIC スペクトルのグラフを表示して、ホスト PC のモニターには、音源方向を推定した角度と MUSIC スペクトルの最大値が出力される。

表示方法は次のとおりである。小型ハードウェア rasp はグラフの右側が 0° 、左側が 180° を表示する。また、垂直方向の方向は、グラフの色によって 3 つの方向に分けた。図 28 に rasp の表示例を示す。

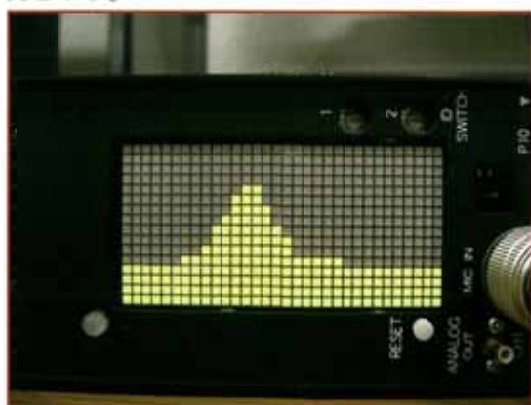


図 28 rasp の表示説明

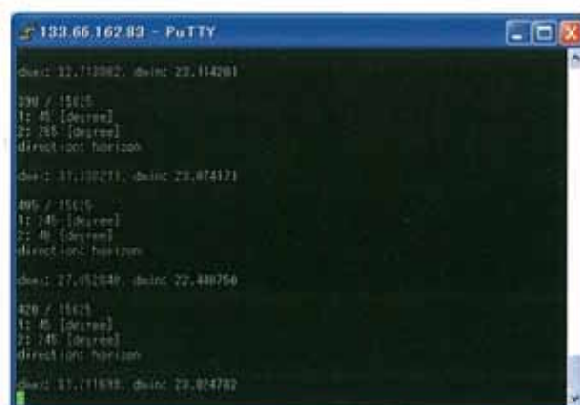


図 29 結果表示画面

このスペクトルは MUSIC スペクトルを表していて、MUSIC スペクトルにおいて山の頂点のところが音源の方向を示す。また、1 目盛は 5° おきに示されている。

音源から音を出力して、システムを動かすと、 $0^\circ \sim 180^\circ$ のとき、音源が垂直方向に対して上方方向にあると rasp の MUSIC スペクトルの色が緑色になり、水平方向に対してはオレンジ色になり、下方方向に対しては赤色になる。

また、ホスト PC にも 0° から 360° の角度の値と垂直方向の決定方向、MUSIC スペクトルの最大値・最小値が表示される。(図 29)

結果表示画面では、推定した角度の結果、垂直方向の推定結果、MUSIC スペクトル値の最大値・最小値を表示する。

390/15625

1:45 [degree]

2:265 [degree]

direction: horizon

dmax: 31.338211、 dmin: 23.074171

の表示を例に説明する。

“390/15625”は、経過時間/終了時間 のように認識する時間を表している。

この数字の単位は、次式であらわされる。

$$t(\text{時間}) \times \frac{16000(\text{Hz})}{1024(\text{ポイント})}$$

1秒が 15.625 に相当する。

“1:45 [degree]” “2:265 [degree]” は、システムが推定した角度の第1候補が 45°、第2候補が 265° 方向であると推定されている。

“direction: horizon”は、垂直方向における推定方向を示す。

up は、マイクロフォンアレイに対して上方向に音源があると推定し、horizon はマイクロフォンアレイに対して同じぐらいの高さに音源があると推定する。また、down は、マイクロフォンアレイに対して下方向に音源があると推定していることを示す。

また、dmax・dmin はそれぞれ、LED に示される MUSIC スペクトル値の最大値・最小値を示している。

3.3.2 マイクロフォンアレイの形状

マイクロフォンアレイは、正 20 面体の頂点に 12 個のマイクを取り付けてある。そのうちの 6 個のマイクを使って、音源方向の推定を行う。

マイクロフォンアレイの形は正 20 面体である。これは、将来的に全方向ステレオシステム(SOS)と組み合わせて音と映像における方向推定や発音・発話方向のイベントを検出するようなシステムを作るためにこの形にした。

マイクロフォンアレイの外観を図 30 に示す。



図 30 マイクロフォンアレイの外観

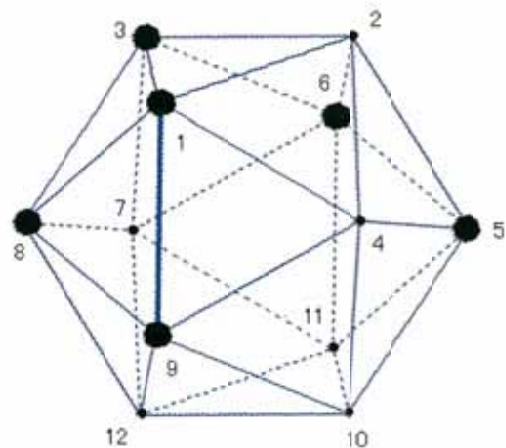


図 31 マイクロフォンアレイのマイクの位置

形状：正 20 面体

高さ：38cm

頂点と頂点との距離：10.5cm

頂点についている 12 個のマイクのうち、このシステムに使用したのは、任意に選んだ 6 個である。(図 31)

図 31 中の太線の方向が 0°方向となるように基準線をとった。

大きい丸点のところにマイクを据え付けた。(図 31 中：1、3、5、6、8、9 番)

3.3.3 結果と考察

このシステムを家庭内環境で使用するために、家庭内環境で使われるような音と、このシステム

が高周波で音源を検出できるかどうかを確かめるために、システムが対象としない高周波成分をもつ楽器音を使用して実験を行った。実験は、0.5 秒間隔でこのシステムが発話・発音区間と非発話・発音区間を正確に分ける検出率を求めた。上・水平・下の垂直方向と、5°きざみの水平方向の両方について推定された音源方向が一致したもののみを正解とした。

検出率は、

$$\text{検出率(\%)} = \frac{\text{発話・発音及び非発話区間を正確に検出できた区間}}{\text{全区間}} \times 100$$

で定義する。表 4 は実験で使った音を示す。

表 4 実験に使用した音

実験名	使用した音
実験 A	発話(with 音楽環境下)
実験 B	ガラスの割れる音
実験 C	携帯電話の着信音
実験 D(ピアノ)	ピアノの音(最大周波数 3 kHz 以上)
実験 D(ピッコロ)	ピッコロの音(最大周波数 3 kHz 以上)
実験 E	足音

実験結果を図 32 に記す。

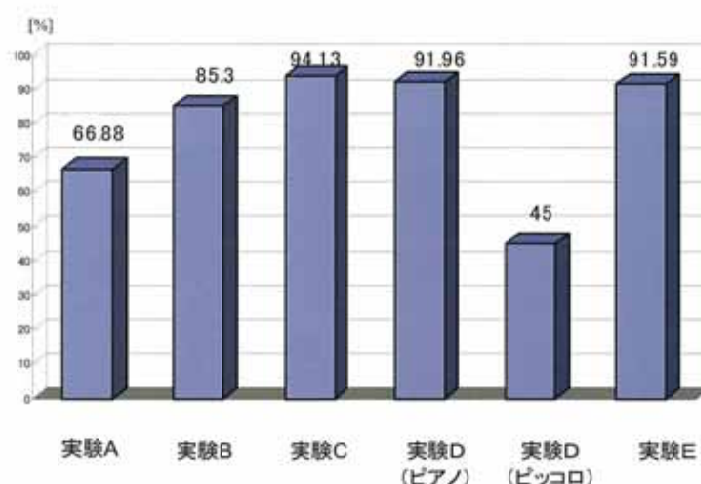


図 32 環境に対する検出率

家庭内での物音については 85%、94%と検出率が高かった。音楽環境における話し声は検出率が低くなった。また、ピッコロとピアノの音は、2 つとも最大の周波数成分がシステムが使用していない 3kHz 以上であるのに、検出率に大きな差があった。これは、ピアノ音において最大の周波数の次に大きい周波数成分をシステムが利用できたからと思われる。

リアルタイムで 3 次元空間における音源方向を推定するために、各垂直方向に対して 2 次元における音源方向推定システムを用い、各垂直方向の MUSIC スペクトル値を比較して、垂直方向の音源を決定する新たなシステムを作成した。また、従来のパーソナルコンピュータとは異なる多チャンネルにおける音響信号処理に特化した信号処理装置を使うことにより複雑な音源推定に関するデータ処理をより手軽に扱えるようになった。

また、評価実験では、家庭内環境における物音や話し声を取り上げて実験を行った。物音は、

80%以上の精度で発音・非発音区間を検出し、方向推定を行うことができた。音楽環境中の話し声は、方向推定も正しい検出率は 67%であったが、発話していることを検出する発話検出率は 100%であった。

以上のように、リアルタイムで 3 次元空間における音源方向の推定をするシステムを作成し、その動作を確認した。

フェーズ III

今後の取り組み

これまでの研究において、完全に死角のない全方向のカラー画像と 3 次元情報を取得可能な全方向ステレオシステム(SOS) (プロトタイプモデル Jupiter) のシステム構成がほぼ確立した。今回開発した移動体を指向した小型モデル Venus、環境監視システム等の用途を指向したモデル Saturn の有効性については 6-2、3、4 にて述べる。今後の取り組みとしては、更に SOS の特性が有効となるアプリケーションを指向した新たな SOS の開発、またフェーズ II にて行った音声認識の研究を踏まえ、3 次元空間における音響情報と SOS にて得られる画像情報の統合を指向したマルチモーダルインタラクションモデルの開発などを検討する。