

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2019-117375

(P2019-117375A)

(43) 公開日 令和1年7月18日 (2019.7.18)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G03B 17/38 (2006.01)</b>	G03B 17/38 B	2H020
<b>G03B 15/00 (2006.01)</b>	G03B 15/00 R	2H105
<b>G03B 17/56 (2006.01)</b>	G03B 15/00 P	5C122
<b>H04N 5/232 (2006.01)</b>	G03B 17/56 A	5D018
<b>G10L 15/00 (2013.01)</b>	H04N 5/232 220	5D220
審査請求 未請求 請求項の数 20 O L (全 38 頁) 最終頁に続く		

(21) 出願番号	特願2018-207634 (P2018-207634)	(71) 出願人	000001007
(22) 出願日	平成30年11月2日 (2018.11.2)		キヤノン株式会社
(31) 優先権主張番号	特願2017-250108 (P2017-250108)		東京都大田区下丸子3丁目30番2号
(32) 優先日	平成29年12月26日 (2017.12.26)	(74) 代理人	100076428
(33) 優先権主張国	日本国 (JP)		弁理士 大塚 康德
		(74) 代理人	100115071
			弁理士 大塚 康弘
		(74) 代理人	100112508
			弁理士 高柳 司郎
		(74) 代理人	100116894
			弁理士 木村 秀二
		(74) 代理人	100130409
			弁理士 下山 治
		(74) 代理人	100134175
			弁理士 永川 行光

最終頁に続く

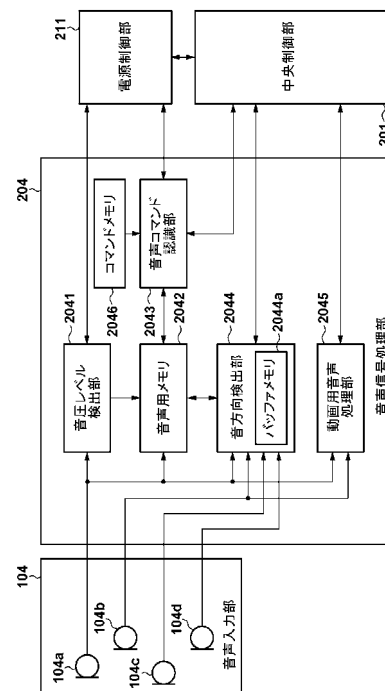
(54) 【発明の名称】 撮像装置及びその制御方法及びプログラム

## (57) 【要約】

【課題】 特別な操作を行わずとも、ユーザの意図したタイミングで意図した構図の画像を撮像する。

【解決手段】 撮像部を有する撮像装置であって、撮像部を駆動する駆動部と、ユーザのいる方向を検出する第一の検出部と、撮像装置の動きを検出する第二の検出部と、音声を集音するための複数の集音部と、複数の集音部を用いて音声の音源の方向を検出する第三の検出部と、制御部と、を有する。ここで制御部は、第一の検出部によって検出したユーザのいる方向と、および、第二の検出部によって検出した撮像装置の動きとに基づいて、2つ以上の集音部を複数の集音部から決定する。また、第三の検出部は、決定された2つ以上の集音部を用いて音声の音源の方向を検出する。そして、第三の検出部が決定された2つ以上の集音部を用いて音声の音源の方向を検出した場合、制御部は第三の検出部が検出した音源の方向に、撮像部の撮像方向を向けるよう駆動部を制御する。

【選択図】 図2



**【特許請求の範囲】****【請求項 1】**

撮像手段を有する撮像装置であって、  
前記撮像手段を駆動する駆動手段と、  
ユーザのいる方向を検出する第一の検出手段と、  
前記撮像装置の動きを検出する第二の検出手段と、  
音声を集音するための複数の集音手段と、  
前記複数の集音手段を用いて前記音声の音源の方向を検出する第三の検出手段と、  
制御手段と、を有し、  
前記制御手段は、前記第一の検出手段によって検出したユーザのいる方向と、および、  
前記第二の検出手段によって検出した前記撮像装置の動きとに基づいて、2つ以上の集音手段を前記複数の集音手段から決定し、  
前記第三の検出手段は、前記決定された2つ以上の集音手段を用いて音声の音源の方向を検出し、  
前記第三の検出手段が前記決定された2つ以上の集音手段を用いて音声の音源の方向を検出した場合、前記制御手段は前記第三の検出手段が検出した音源の方向に、前記撮像手段の撮像方向を向けるよう前記駆動手段を制御することを特徴とする撮像装置。

10

**【請求項 2】**

前記第三の検出手段によって検出した音声の音源の方向に複数の候補があった場合、前記第一の検出手段によって検出したユーザのいる方向ではない方向に前記撮像手段の撮像方向を向けるよう前記制御手段は前記駆動手段を制御することを特徴とする請求項 1 に記載の撮像装置。

20

**【請求項 3】**

前記第二の検出手段は前記撮像装置の動きを、前記撮像装置の加速度および角速度に基づいて検出することを特徴とする請求項 1 または 2 に記載の撮像装置。

**【請求項 4】**

前記複数の集音手段は、すべての前記集音手段が直線上に並ばないように配置されることを特徴とする請求項 1 から 3 のいずれか 1 項に記載の撮像装置。

**【請求項 5】**

撮像手段と、  
前記撮像手段を駆動する駆動手段と、  
音声を集音するための複数の集音手段と、  
前記複数の集音手段を用いて集音した音声の音源の方向を検出する検出手段と、  
制御手段と、を有し、  
前記制御手段は前記撮像手段によって撮像した画像データに基づいて、撮影の必要のない領域を設定し、  
前記検出手段によって検出した音声の音源の方向が、前記撮影の必要のない領域にない場合、前記撮像手段の撮像方向が前記検出手段によって検出した音声の音源の方向に向くよう前記制御手段は前記駆動手段を制御することを特徴とする撮像装置。

30

40

**【請求項 6】**

前記制御手段は、前記撮像手段によって撮像した画像データの輝度が所定の閾値よりも暗い場合、または画像データに映る被写体と前記撮像装置との距離が所定の閾値より小さい場合、その撮像方向を撮影の必要のない領域として設定することを特徴とする請求項 5 に記載の撮像装置。

**【請求項 7】**

前記撮像装置が持ち運ばれていると判断した場合、前記制御手段は前記撮影の必要のない領域を再度設定することを特徴とする請求項 5 または 6 に記載の撮像装置。

50

**【請求項 8】**

前記駆動手段を所定の時間駆動するよう制御した後、前記制御手段はさらに前記撮像手段によって撮像した画像データに基づいて、現在の撮像手段の撮像方向が撮影の必要な領域かどうかを判断し、その判断にしたがって撮影の必要のない領域を再度設定する

ことを特徴とする請求項 5 または 6 に記載の撮像装置。

**【請求項 9】**

撮像手段と、

前記撮像手段をパン動作およびチルト動作によって駆動する駆動手段と、

音声を集音するための複数の集音手段と、

前記複数の集音手段を用いて前記音声の音源の方向のパン角を検出する検出手段と、

制御手段と、を有し、

前記制御手段は前記撮像手段によって被写体を撮像した場合、前記被写体の方向を向いた前記撮像手段の撮像方向のパン角およびチルト角を被写体情報として記録し、

前記検出手段によって検出したパン角と前記被写体情報に含まれるパン角と比較しその 2 つの角度の差が閾値以下だった場合、前記制御手段は前記検出手段によって検出したパン角および前記被写体情報に含まれるチルト角に撮像手段の撮像方向を向けるよう前記駆動手段を制御し、

前記検出手段によって検出したパン角と前記被写体情報に含まれるパン角と比較しその 2 つの角度の差が閾値を超える場合、前記制御手段は、前記検出手段によって検出したパン角にいる被写体に前記撮像手段の撮像方向を向けるよう前記駆動手段を制御する

ことを特徴とする撮像装置。

**【請求項 10】**

前記制御手段は前記検出手段によって検出したパン角および前記被写体情報に含まれるチルト角に撮像手段の撮像方向を向けるよう前記駆動手段を制御し、

前記制御手段は前記検出手段によって検出したパン角および前記被写体情報に含まれるチルト角の方向に被写体を検出した場合、前記制御手段は前記被写体情報を現在の撮像手段の撮像方向のパン角およびチルト角になるよう更新し、

ことを特徴とする請求項 9 に記載の撮像装置。

**【請求項 11】**

前記制御手段は前記検出手段によって検出したパン角および前記被写体情報に含まれるチルト角に撮像手段の撮像方向を向けるよう前記駆動手段を制御し、

前記検出手段によって検出したパン角および前記被写体情報に含まれるチルト角の方向に被写体を検出しない場合、前記制御手段は前記被写体情報を削除する

ことを特徴とする請求項 9 または 10 に記載の撮像装置。

**【請求項 12】**

前記検出手段によって検出したパン角との差が閾値以下になる被写体情報が複数ある場合、

前記制御手段は前記複数の被写体情報のそれぞれのチルト角が所定の範囲内に入るようチルト角を決定し、

前記制御手段は、前記検出手段によって検出したパン角および前記決定したチルト角の方向に前記撮像手段の撮像方向を向けるよう前記駆動手段を制御する

ことを特徴とする請求項 9 から 11 のいずれか 1 項に記載の撮像装置。

**【請求項 13】**

撮像手段と、

前記撮像手段を駆動する駆動手段と、

音声を集音するための複数の集音手段と、

前記複数の集音手段を用いて前記音声の音源の方向を所定の角度の分解能で検出する検出手段と、

制御手段と、

前記制御手段は、前記所定の角度を前記撮像手段の画角より小さくなるように設定し、

前記集音手段によって音声を集音した場合、前記検出手段によって前記所定の角度の分解能で検出した音声の音源の方向に前記撮像手段の撮像方向を向けるよう前記駆動手段を前記制御手段は制御する

ことを特徴とする撮像装置。

【請求項 14】

前記制御手段は、前記撮像手段のズーム倍率を増加させた場合、前記撮像手段の画角より小さくなるように前記所定の角度を大きくするよう設定し、

前記制御手段は、前記撮像手段のズーム倍率を減少させた場合、前記撮像手段の画角より小さくなるように前記所定の角度を小さくするよう設定する

ことを特徴とする請求項 13 に記載の撮像装置。

10

【請求項 15】

さらに音声による指示を認識する認識手段を有し、

前記認識手段が前記撮像手段のズーム倍率を変更する指示を認識した場合、前記制御手段は、前記撮像手段のズーム倍率を前記指示に従って変更する

ことを特徴とする請求項 13 または 14 に記載の撮像装置。

【請求項 16】

撮像手段と、

前記撮像手段を駆動する駆動手段と、

音声を集音するための複数の集音手段と、

前記複数の集音手段を用いて前記音声の音源の方向を検出する検出手段と、を有する撮像装置の制御方法であって、

20

ユーザのいる方向を検出する第一の検出ステップと、

前記撮像装置の動きを検出する第二の検出ステップと、

前記第一の検出ステップで検出したユーザのいる方向と、および、前記第二の検出ステップで検出した前記撮像装置の動きとに基づいて、2 つ以上の集音手段を前記複数の集音手段から決定する決定ステップと、

前記決定ステップで決定された 2 つ以上の集音手段を用いて音声の音源の方向を前記検出手段によって検出する検出ステップと、

前記検出ステップで音声の音源の方向を検出した場合、前記検出ステップで検出した音源の方向に、前記撮像手段の撮像方向を向けるよう前記駆動手段を制御するステップと、を有する

30

ことを特徴とする制御方法。

【請求項 17】

撮像手段と、

音声を集音するための複数の集音手段と、を有する撮像装置の制御方法であって、

前記複数の集音手段を用いて集音した音声の音源の方向を検出する検出ステップと、

前記撮像手段によって撮像した画像データに基づいて、撮影の必要のない領域を設定する設定ステップと、

前記検出ステップによって検出した音声の音源の方向が、前記設定ステップで設定した前記撮影の必要のない領域にない場合、前記撮像手段の撮像方向が前記検出ステップによって検出した音声の音源の方向に向くよう駆動する駆動ステップと、を有する

40

ことを特徴とする制御方法。

【請求項 18】

撮像手段と、

前記撮像手段をパン動作およびチルト動作によって駆動する駆動手段と、

音声を集音するための複数の集音手段と、を有する撮像装置の制御方法であって、

前記撮像手段によって被写体を撮像した場合、前記被写体の方向を向いた前記撮像手段の撮像方向のパン角およびチルト角を被写体情報として記録する記録ステップと、

前記複数の集音手段を用いて前記音声の音源の方向のパン角を検出する検出ステップと

50

前記検出ステップで検出したパン角と前記記録ステップで記録した前記被写体情報に含まれるパン角と比較しその２つの角度の差が閾値以下だった場合、前記検出ステップで検出したパン角および前記記録ステップで記録した前記被写体情報に含まれるチルト角に撮像手段の撮像方向を向けるよう前記駆動手段を制御するステップと、

前記検出ステップで検出したパン角と前記記録ステップで記録した前記被写体情報に含まれるパン角と比較しその２つの角度の差が閾値を超える場合、前記検出ステップで検出したパン角にいる被写体に前記撮像手段の撮像方向を向けるよう前記駆動手段を制御するステップと、を有する

ことを特徴とする制御方法。

【請求項 １ ９】

10

撮像手段と、

前記撮像手段を駆動する駆動手段と、

音声を集音するための複数の集音手段と、を有する撮像装置の制御方法であって、

前記複数の集音手段を用いて前記音声の音源の方向を所定の角度の分解能で検出する検出ステップと、

前記所定の角度を前記撮像手段の画角より小さくなるように設定する設定ステップと、

前記集音手段によって音声を集音した場合、前記検出ステップで前記所定の角度の分解能で検出した音声の音源の方向に前記撮像手段の撮像方向を向けるよう前記駆動手段を制御するステップと、を有する

ことを特徴とする制御方法。

20

【請求項 ２ ０】

コンピュータを請求項 １ から １ ５ の何れか １ 項に記載の撮像装置の各手段として機能させるための、コンピュータが読み取り可能なプログラム。

【発明の詳細な説明】

【技術分野】

【０ ０ ０ １】

本発明は、撮像装置及びその制御方法及びプログラムに関するものである。

【背景技術】

【０ ０ ０ ２】

カメラ等の撮像装置による静止画・動画撮影においては、ユーザがファインダー等を通して撮影対象を決定し、撮影状況を自ら確認して撮影画像のフレーミングを調整することによって、画像を撮影するのが通常である。このような撮像装置では、ユーザの操作ミスを検知してユーザに通知したり、外部環境の検知を行い、撮影に適していない場合にユーザに通知したりする機能が備えられている。また、撮影に適した状態になるようにカメラを制御する仕組みが従来から存在している。

30

【０ ０ ０ ３】

このようなユーザの操作により撮影を実行する撮像装置に対し、ユーザが撮影指示を与えることなく断続的および継続的に撮影を行うライフログカメラが存在する（特許文献 １）。

【先行技術文献】

40

【特許文献】

【０ ０ ０ ４】

【特許文献 １】特表 ２ ０ １ ６ - ５ ３ ６ ８ ６ ８ 号公報

【発明の概要】

【発明が解決しようとする課題】

【０ ０ ０ ５】

しかしながら、ユーザの身に着けるタイプの、これまでのライフログカメラでは、定期的に自動撮影を行うものであるので、得られる画像はユーザの意図とは無関係なものとなる場合があった。

【０ ０ ０ ６】

50

本発明は上記問題に鑑みなされたものであり、特別な操作を行わずとも、ユーザの意図したタイミングでユーザの意図した構図の画像を撮像する技術を提供しようとするものである。

【課題を解決するための手段】

【0007】

この課題を解決するため、例えば本発明の撮像装置は以下の構成を備える。すなわち、撮像手段を有する撮像装置であって、

前記撮像手段を駆動する駆動手段と、

ユーザのいる方向を検出する第一の検出手段と、

前記撮像装置の動きを検出する第二の検出手段と、

音声を集音するための複数の集音手段と、

前記複数の集音手段を用いて前記音声の音源の方向を検出する第三の検出手段と、

制御手段と、を有し、

前記制御手段は、前記第一の検出手段によって検出したユーザのいる方向と、および、前記第二の検出手段によって検出した前記撮像装置の動きとに基づいて、2つ以上の集音手段を前記複数の集音手段から決定し、

前記第三の検出手段は、前記決定された2つ以上の集音手段を用いて音声の音源の方向を検出し、

前記第三の検出手段が前記決定された2つ以上の集音手段を用いて音声の音源の方向を検出した場合、前記制御手段は前記第三の検出手段が検出した音源の方向に、前記撮像手段の撮像方向を向けるよう前記駆動手段を制御することを特徴とする。

【発明の効果】

【0008】

本発明によると、第1には、特別な操作を行わずとも、ユーザの意図したタイミングで意図した構図の画像を撮像する技術を提供しようとするものである。

【0009】

また、他の発明によれば、上記第1の効果に加えて、利用形態に合わせて方向検出に利用するマイクの数を変動させることで、省電力化を図り、且つ、ユーザの身に装着した場合等での衣類とのこすれ音による音方向のご検出を防ぐことが可能になる。

【0010】

また、他の発明によれば、上記第1の効果に加えて、無意味となる撮像方向への変更が無くすることができる。

【0011】

また、他の発明によれば、上記第1の効果に加えて、利用開始からの経過時間によって撮像手段の撮像方向の被写体に向かう動作が効率良くなっていく。

【0012】

また、他の発明によれば、上記第1の効果に加えて、音源の方向の精度が、撮像手段の倍率に依存することになり、音源の方向を検出する精度を常に高くする必要がなくなり、消費電力を削減できる。

【図面の簡単な説明】

【0013】

【図1】実施形態に係る撮像装置のブロック図。

【図2】実施形態に係る音声入力部と音声信号処理部の詳細ブロック図。

【図3】実施形態に係る撮像装置の外観図と使用例を示す図。

【図4】実施形態に係る撮像装置のパン動作とチルト動作を示す図。

【図5A】実施形態における中央制御部の処理手順を示すフローチャート。

【図5B】実施形態における中央制御部の処理手順を示すフローチャート。

【図6】図5Bにおける音声コマンド処理の詳細を示すフローチャート。

【図7】実施形態における音声コマンドの意味と音声コマンドとの関係を示す図。

【図8】実施形態における起動時から動作撮影開始コマンドに至るまでのタイミングチャ

10

20

30

40

50

ート。

【図 9】実施形態に係る音方向検知法を説明するための図。

【図 10】撮像装置の真上に音源が存在する場合の検出法を説明するための図。

【図 11】第 1 の実施形態における設置位置の検出処理を示すフローチャート。

【図 12】第 1 の実施形態における設置位置別の音源方向の検出原理を示す図。

【図 13】第 1 の実施形態における設置位置別の音源の検出範囲を示す図。

【図 14】第 2 の実施形態における撮像装置 1 の利用形態と、対応のマスク領域との関係を示す図。

【図 15】第 2 の実施形態における中央制御部の処理手順を示すフローチャート。

【図 16】第 3 の実施形態における問題点を説明するための図。

10

【図 17】第 3 の実施形態における中央制御部の処理手順を示すフローチャート。

【図 18】第 3 の実施形態における改善された動作を説明するための図。

【図 19】第 3 の実施形態の変形例における中央制御部の処理手順を示すフローチャート。

。

【図 20】第 3 の実施形態の変形例における改善された動作を説明するための図。

【図 21】第 4 の実施形態における音方向の感度と画角との関係を示す図。

【図 22】第 4 の実施形態におけるズーム倍率を上げた場合お音方向の感度と画角との関係を示す図。

【図 23】音方向の検出の分解能と処理の負担との関係を示す図。

【図 24】第 4 の実施形態における水平方向の撮影画角と音方向検出時の水平方向の検出分解能の関係を示す図である。

20

【図 25】第 4 の実施形態におけるズーム倍率の音声コマンド受信時の中央制御部の処理手順を示すフローチャート。

【図 26】第 4 の実施形態における撮像装置の動作内容を説明するための図。

【発明を実施するための形態】

【0014】

以下図面に従って本発明に係る実施形態を詳細に説明する。

【0015】

[第 1 の実施形態]

図 1 は、第 1 の実施形態に係る撮像装置 1 のブロック構成図である。撮像装置 1 は、光学レンズユニットを含み、撮像する撮像方向（光軸方向）が可変の可動撮像部 100、及び、可動撮像部 100 の駆動制御および、撮像装置全体を制御する中央制御部（CPU）を含む支持部 200 で構成される。

30

【0016】

なお、支持部 200 は、圧電素子を含む複数の振動体 11～13 が可動撮像部 100 の面に対し接触するように設けられている。これらの振動体 11～13 の振動の制御により、可動撮像部 100 がパン、チルト動作を行う。なお、パン、チルト動作はサーボモータ等で実現しても構わない。

【0017】

可動撮像部 100 は、レンズ部 101、撮像部 102、レンズアクチュエータ制御部 103、及び、音声入力部 104 を有する。

40

【0018】

レンズ部 101 は、ズームレンズ、絞り・シャッタ、および、フォーカスレンズなどの撮影光学系で構成される。撮像部 102 は、CMOS センサや CCD センサなどの撮像素子を含み、レンズ部 101 により結像された光学像を光電変換して電気信号を出力する。レンズアクチュエータ制御部 103 は、モータドライバ IC を含み、レンズ部 101 のズームレンズ、絞り・シャッタ、および、フォーカスレンズ等の各種アクチュエータを駆動する。各種アクチュエータは、後述する支持部 200 内の中央制御部 201 より受信した、アクチュエータ駆動指示データに基づいて駆動される。音声入力部 104 はマイクロフォン（以降マイク）を含む音声入力部であり複数のマイク（本実施形態では 4 つ）で構成さ

50

れており、音声を電気信号に変換し、さらに電気信号をデジタル信号（音声データ）に変換して出力する。

#### 【0019】

一方、支持部200は、撮像装置1の全体の制御を行うための中央制御部201を有する。この中央制御部201は、CPUと、CPUが実行するプログラムを格納したROM、及び、CPUのワークエリアとして使用されるRAMで構成される。また、支持部200は、撮像信号処理部202、映像信号処理部203、音声信号処理部204、操作部205、記憶部206、表示部207を有する。更に、支持部200は、入出力端子部208、音声再生部209、電源部210、電源制御部211、位置検出部212、回動制御部213、無線通信部214、並びに、先に説明した振動体11～13を有する。

10

#### 【0020】

撮像信号処理部202は、可動撮像部100の撮像部102から出力された電気信号を映像信号へ変換する。映像信号処理部203は、撮像信号処理部202から出力された映像信号を用途に応じて加工する。映像信号の加工は画像切り出し、及び、回転加工による電子防振動作や、被写体（顔）を検出する被写体検出処理も含まれる。

#### 【0021】

音声信号処理部204は、音声入力部104から出力されたデジタル信号に対して音声処理を行う。音声入力部104がアナログ信号を出力するマイクであれば、音声信号処理部204において、アナログ信号からデジタル信号に変換する構成が含まれても構わない。なお、音声入力部104を含めた音声信号処理部204の詳細については図2を用いて後述する。

20

#### 【0022】

操作部205は、撮像装置1とユーザとの間のユーザインターフェースとして機能するものであり、各種スイッチ、ボタン等で構成される。記憶部206は、撮影により得られた映像情報などの種々のデータを記憶する。表示部207は、LCDなどのディスプレイを備え、映像信号処理部203から出力された信号に基づいて、必要に応じて画像表示を行う。また、この表示部207は、各種メニュー等を表示することで、ユーザインターフェースの一部として機能する。外部入出力端子部208は、外部装置との間で通信信号および映像信号を入出力する。音声再生部209はスピーカを含み、音声データを電気信号に変換し、音声を再生する。電源部210は、撮像装置の全体（各要素）の駆動に必要な電力供給源であり、本実施形態では充電可能なバッテリーであるものとする。

30

#### 【0023】

電源制御部211は、撮像装置1の状態に応じて、上記の各構成要素への電源部210からの電力の供給／遮断を制御するものである。撮像装置1の状態によっては、不使用の要素が存在する。電源制御部211は、中央制御部201の制御下で、撮像装置1の状態によって不使用な要素への電力を遮断して、電力消費量を抑制する機能を果たす。なお、電力供給／遮断については、後述する説明から明らかにする。

#### 【0024】

位置検出部212はジャイロ、加速度センサ、GPS等で構成され、撮像装置1の動きを検出する。この位置検出部212は、撮像装置1がユーザに身に着ける場合にも対処するためである。回動制御部213は、中央制御部201からの指示に従って振動体11～13を駆動する信号を生成し、出力する。振動体11～13は圧電素子で構成され、回動制御部213から印加される駆動信号に応じて振動する。振動体11～13は、回動駆動部（パン・チルト駆動部）を構成する。この結果、可動撮像部100は、中央制御部201が指示した方向に、パン動作、チルト動作する。

40

#### 【0025】

無線部214はWi-Fi（登録商標）やBLE（Bluetooth（登録商標） Low Energy）などの無線規格に準拠して画僧データ等のデータ送信を行う。

#### 【0026】

次に、本実施形態における音声入力部104および音声信号処理部204の構成と、音

50



方向検出処理を図2を参照して説明する。同図は、音声入力部104および音声信号処理部204の構成と、音声信号処理204、中央制御部201及び電源制御部211の接続関係を示している。

【0027】

音声入力部104は、4つの無指向性のマイク（マイク104a、マイク104b、マイク104c、マイク104d）で構成される。各マイクはA/Dコンバータを内蔵しており、予め設定されたサンプリングレート（コマンド検出、方向検出処理：16kHz、動画録音：48kHz）で音声を集音し、内蔵のA/Dコンバータにより集音した音声信号をデジタルの音声データとして出力する。なお、本実施形態では音声入力部104は4つのデジタルマイクで構成されるものとしているが、アナログ出力のマイクで構成されても構わない。アナログマイクの場合、音声信号処理部204内に、対応するA/Dコンバータを設ければよい。また、本実施形態におけるマイクの数4つとするが、3つ以上であればよい。

10

【0028】

マイク104aは、撮像装置1の電源がONの場合には無条件に電力が供給され、集音可能状態となる。一方、他のマイク104b、マイク104c、マイク104dは、中央制御部201の制御下での電源制御部211による電力供給/遮断の対象となっており、撮像装置1の電源がONとなった初期状態では、電力は遮断されている。

【0029】

音声信号処理部204は、音圧レベル検出部2041、音声用メモリ2042、音声コマンド認識部2043、音方向検出部2044、動画用音声処理部2045、及び、コマンドメモリ2046で構成される。

20

【0030】

音圧レベル検出部2041は、マイク104aから出力された音声データの音圧レベルが予め設定された閾値を超えると、音声検出を表す信号を電源制御部211及び音声用メモリ2042に供給する。

【0031】

電源制御部211は、音圧レベル検出部2041から音声検出を表す信号を受信した場合、音声コマンド認識部2043への電力供給を行う。

【0032】

音声用メモリ2042は、中央制御部201の制御下での電源制御部211による電力供給/遮断の対象の1つである。また、この音声用メモリ2042は、マイク104aから出力された音声データを一時的に記憶するバッファメモリである。マイク104aによるサンプリングレートが16kHzであり、1サンプリングにつき2バイト（16ビット）の音声データを出力し、最長の音声コマンドが仮に5秒であった場合、音声用メモリ2042は、約160キロバイト（ $5 \times 16 \times 1000 \times 2$ ）の容量を有する。また、音声用メモリ2042は、マイク104aからの音声データで満たされた場合、古い音声データが新たな音声データで上書きされる。この結果、音声用メモリ2042は、直近の所定期間（上記例では約5秒）の音声データが保持される。また、音声用メモリ2042は、音圧レベル検出部2041から音声検出を示す信号を受信したことをトリガにして、マイク104aからの音声データをサンプリングデータ領域に格納していく。

30

40

【0033】

コマンドメモリ2046は不揮発性のメモリで構成され、本撮像装置が認識する音声コマンドに係る情報を予め記憶（登録）している。詳細は後述するが、コマンドメモリ2046に格納される音声コマンドの種類は例えば図8に示す通りであり、「起動コマンド」をはじめとして、複数種類のコマンドの情報がコマンドメモリ2046に格納されている。

【0034】

音声コマンド認識部2043は、中央制御部201の制御下での電源制御部211による電力供給/遮断の対象の1つである。なお、音声認識そのものは周知技術であるので、

50

ここでの説明は省略する。この音声コマンド認識部 2043 は、コマンドメモリ 2046 を参照し、音声用メモリ 2042 に格納された音声データの認識処理を行う。そして、音声コマンド認識部 2043 は、マイク 104a により集音した音声データが、音声コマンドであるか否か、並びに、コマンドメモリ 2046 に記憶されている登録音声コマンドに一致するのかの判定を行う。そして、音声コマンド認識部 2043 は、コマンドメモリ 2046 に記憶されたいずれかの音声コマンドに一致する音声データを検出したとき、いずれのコマンドであるかを示す情報、並びに、音声用メモリ 2042 内の、その音声コマンドを決定づけた最初と最後の音声データのアドレス（或いは音声コマンドを受け付けたタイミング）を中央制御部 201 に供給する。

#### 【0035】

音方向検出部 2044 は、中央制御部 201 の制御下での電源制御部 211 による電力供給 / 遮断の対象の 1 つである。また、音方向検出部 2044 は、4 つのマイク 104a 乃至 104d からの音声データに基づき、周期的に音源の存在する方向の検出処理を行う。音方向検出部 2044 は、内部にバッファメモリ 2044a を有し、検出した音源方向を表す情報をバッファメモリ 2044a に格納する。なお、音方向検出部 2044 による音方向検出処理を行う周期（例えば 16 kHz）は、マイク 104a のサンプリング周期に対して十分に長く構わない。ただし、このバッファメモリ 2044a は、音声用メモリ 2042 に格納可能な音声データの期間と同じ期間分の音方向情報を記憶するための容量を有するものとする。

#### 【0036】

動画用音声処理部 2045 は、中央制御部 201 の制御下での電源制御部 211 による電力供給 / 遮断の対象の 1 つである。動画用音声処理部 2045 は、4 つのマイクのうち、マイク 103a とマイク 104b の 2 つの音声データをステレオ音声データとして入力し、各種フィルタ処理、ウィンドカット、ステレオ感強調、駆動音除去、ALC (Auto Level Control)、圧縮処理といった動画音声用の音声処理を行う。詳細は後述する説明から明らかになるが、本実施形態ではマイク 104a はステレオマイクの L チャンネル用マイク、マイク 104b は R チャンネル用マイクとして機能する。

#### 【0037】

なお、図 2 では、消費電力や回路構成を考慮し、音声入力部 104 の各マイクと音声信号処理部 204 に含まれる各ブロックとの接続は、4 つのマイクにおける必要最低限の接続を示す。しかし、電力および回路構成の許す限り、複数のマイクを音声信号処理部 204 に含まれる各ブロックで共有して使用しても構わない。また、本実施形態ではマイク 104a を基準のマイクとして接続しているが、どのマイクを基準としても構わない。

#### 【0038】

図 3 (a) 乃至 (e) を参照して、撮像装置 1 の外観図および使用例を説明する。同図 (a) は、本実施形態に係る撮像装置 1 の外観の上面及び正面図を示している。撮像装置 1 の可動撮像部 100 は、略半球体形であり、底面と平行な面を水平面とし、この面を 0 度したとき、- 20 度から垂直方向を示す 90 度の範囲の切欠き窓を有し、図示矢印 A が示す水平面にて 360 度に亘って回動可能な第 1 の筐体 150 を有する。また、可動撮像部 100 は、この切欠き窓に沿って図示の矢印 B が示す水平から垂直の範囲内で、レンズ部 101 及び撮像部 102 と一緒に回動可能な第 2 の筐体 151 を有する。ここで、第 1 の筐体 150 の矢印 A の回動動作はパン動作、第 2 の筐体 151 の矢印 B の回動動作はチルト動作に対応し、これらは振動体 11 ~ 13 の駆動によって実現している。なお、本実施形態における撮像装置のチルト可能な範囲は、上記の通り、- 20 度から + 90 度の範囲であるものとする。

#### 【0039】

マイク 104a、104b は、第 1 の筐体 150 の切欠き窓を挟む前面側の位置に配置されている。また、マイク 104c、104d は、第 1 の筐体 150 の後方側に設けられている。図 3 (a) からわかるように、第 2 筐体 151 を固定にした状態で、第 1 筐体 150 を矢印 A に沿ってどの方向にパン動作させたとしても、レンズ部 101 及び撮像部

10

20

30

40

50

102に対する、マイク104a、104bの相対的な位置は変わらない。つまり、撮像部102の撮像方向に対して左側にマイク104aが常に位置し、右側にマイク104bが常に位置する。また、マイク104aおよびマイク104bは撮像部102の撮像方向に対して対称に配置されるので、マイク104aはステレオマイクのLチャンネルへの入力を担い、マイク104bはステレオマイクのRチャンネルへの入力を担う。それ故、撮像部102による撮像して得た画像が表す空間と、マイク104a、104bによる取得した音場は一定の関係を維持できる。

#### 【0040】

なお、本実施形態における4つのマイク104a、104b、104c、103dは、撮像装置1の上面から見て、図3(a)に示すように長方形の各頂点の位置に配置されている。また、これら4つのマイクは、図3(a)における1つの水平面上に位置するものとするが、多少のずれがあっても構わない。

#### 【0041】

マイク104aとマイク104bとの距離は、マイク104aとマイク104cとの距離よりも大きい。なお、隣りあうマイク間の距離は、10mm～30mm程度が望ましい。また、本実施形態ではマイクの数4つとしているが、直線上に並ばないという条件を満たせば、マイク数は3つ以上であれば構わない。また、図3(a)のマイク104a乃至104dの配置位置は一例であって、これらの配置方法は、メカ的制約やデザイン制約等によって適宜変更しても構わない。

#### 【0042】

図3(b)乃至(e)は、本実施形態における撮像装置1の利用形態を示している。図3(b)は、机などに撮像装置1が載置される場合で、撮影者自身やその周囲の被写体の撮影を目的とした利用形態を説明するための図である。図3(c)は、撮像装置1を撮影者の首にぶら下げる例であり、主に、撮影者の行動の前方の撮影を目的とした利用形態を説明するための図である。図3(d)は、撮像装置1を撮影者の肩に固定した使用例であり、撮影者の周囲の前後、及び、右側の撮影を目的とした利用形態を説明するための図である。そして、図3(e)は、撮像装置1をユーザが持つ棒の端に固定する使用例であり、ユーザが望む所望の撮影位置(高所や手が届かない位置)に撮像装置1を移動させることで、撮影を行うことを目的とした利用形態を説明するための図である。

#### 【0043】

図4(a)乃至(c)を参照して、本実施形態の撮像装置1のパン動作、およびチルト動作を更に詳しく説明する。ここでは図3(b)のように据え置いた使用例で前提として記載するが、そのほかの使用例においても同様である。

#### 【0044】

図4(a)はレンズ部101が水平を向いている状態を示している。図4(a)を初期状態とし、第1筐体150を、上方向から見て反時計回りに90度パン動作させると、図4(b)のようになる。一方、図4(a)の初期状態から、第2筐体151の90度チルト動作をおこなうと、図4(c)のようになる。第1筐体150、第2筐体151の回動は、先に説明したように、回動制御部213により駆動される振動体11～13による振動にて実現している。

#### 【0045】

次に、本実施形態における撮像装置1の中央制御部201の処理手順を図5A、5Bのフローチャートに従って説明する。同図に係る処理は、撮像装置1のメイン電源がONされた場合の中央制御部201の処理を示している。

#### 【0046】

中央制御部201は、ステップS101にて、撮像装置1の初期化処理を行う。この初期化処理にて、中央制御部201は、現在の可動撮像部100の撮像部102の撮像方向における、水平面内の方向成分をパン動作の基準角度(0度)として決定する。

#### 【0047】

これ以降、可動撮像部100のパン動作を行った後の撮像方向のうち水平面の成分は、

10

20

30

40

50

この基準角度からの相対的な角度で表されるものとする。また、音方向検出部 2044 が検出する音源方向のうちの水平面の成分も、上記基準角度に対する相対的な角度で表されるものとする。また、詳細は後述するが、音方向検出部 2044 は、撮像装置 1 の真上の方向（パン動作の回転軸の軸方向）に音源があるか否かの判定も行う。

【0048】

なお、この段階で、音声用メモリ 2042、音方向検出部 2044、動画用音声処理部 2045、並び、マイク 104b 乃至 104d への電力は遮断されている。

【0049】

初期化処理を終えると中央制御部 201 は、ステップ S102 にて、電源制御部 211 を制御して、音圧レベル検出部 2041、マイク 104a への電力の供給を開始する。この結果、音圧レベル検出部 2041 は、マイク 104a から出力された音声データに基づいて、この音声データに変換される前の音声の音圧レベルの検出処理を実行し、この音声

10

が予め設定された閾値を超える音圧レベルであると判定した場合にその旨を中央制御部 201 に通知する。なお、この閾値は、例えば 60dB SPL (Sound Pressure Level) とするが、撮像装置 1 が環境等に応じて変更してもよいし、必要な周波数帯域だけに絞るようにしてもよい。

【0050】

中央制御部 201 は、ステップ S103 にて、音圧レベル検出部 2041 による閾値を超える音圧レベルである音声を検出されるのを待つ。閾値を超える音圧レベルである音声

20

が検出されると、ステップ S104 にて、音声メモリ 2042 はマイク 104a からの音声データの受信、格納処理を開始する。

【0051】

また、ステップ S105 にて、中央制御部 201 は、電源制御部 211 を制御し、音声コマンド認識部 2043 への電力供給を開始する。この結果、音声コマンド認識部 2043 は、コマンドメモリ 2046 を参照した音声用メモリ 2042 に格納されていく音声データの認識処理を開始する。そして、音声コマンド認識部 2043 は、音声用メモリ 2042 に格納された音声データの認識処理を行い、コマンドメモリ 2046 内のいずれかの音声コマンドと一致すると音声コマンドを認識した場合、その認識された音声コマンドを

30

特定する情報と、音声用メモリ 2042 内の、認識した音声コマンドを決定づけた最初と最後の音声データのアドレス（或いは音声コマンドを受け付けたタイミング）情報とを含む情報を中央制御部 201 に通知する。

【0052】

ステップ S106 にて、中央制御部 201 は、音声コマンド認識部 2043 から、音声コマンドが認識されたことを示す情報を受信したか否かを判定する。否の場合、中央制御部 201 は、処理をステップ S108 に進め、音声コマンド認識部 2043 を起動させてからの経過時間が、予め設定された閾値を超えたか否かを判定する。そして、経過時間が閾値以内である限り、中央制御部 201 は、音声コマンド認識部 2043 による音声コマンドが認識されるのを待つ。そして、閾値が示す時間が経過しても、音声コマンド認識部 2043 が音声コマンドを認識しなかった場合、中央制御部 201 は処理をステップ S109 に進める。このステップ S109 にて、中央制御部 201 は、電源制御部 211 を制

40

御して音声コマンド認識部 2043 への電力を遮断する。そして、中央制御部 201 は、処理をステップ S103 に戻す。

【0053】

一方、中央制御部 201 が、音声コマンド認識部 2043 から、音声コマンドが認識されたことを示す情報を受信した場合、処理をステップ S107 に進める。このステップ S107 にて、中央制御部 201 は、認識された音声コマンドが、図 8 に示される起動コマンドに対応するか否かを判定する。そして、認識された音声コマンドが起動コマンド以外のコマンドであると判定した場合、中央制御部 201 は処理をステップ S108 に進める。また、認識された音声コマンドが起動コマンドであった場合、中央制御部 201 は処理をステップ S107 からステップ S110 に進める。

50

## 【 0 0 5 4 】

ステップ S 1 1 0 にて、中央制御部 2 0 1 は、電源制御部 2 1 1 を制御し、音方向検出部 2 0 4 4、マイク 1 0 4 b 乃至 1 0 4 d への電力供給を開始する。この結果、音方向検出部 2 0 4 4 は、4 つのマイク 1 0 4 a 乃至 1 0 4 d からの同時刻の音声データに基づく、音源方向の検出処理を開始する。音源の方向の検出処理は、所定周期で行われる。そして、音方向検出部 2 0 4 4 は、検出した音方向を示す音方向情報を、内部のバッファメモリ 2 0 4 4 a に格納していく。このとき、音方向検出部 2 0 4 4 は、音方向情報を決定に利用した音声データのタイミングが、音声メモリ 2 0 4 2 に格納された音声データのどのタイミングであったのかを対応付くように、バッファメモリ 2 0 4 4 a に格納する。典型的には、バッファメモリ 2 0 4 4 a に格納するのは、音方向と、音声メモリ 2 0 4 2 内の音声データのアドレスとすればよい。なお、音方向情報には、水平面における、先に説明した基準角度に対する音源の方向との差を表す角度とする。また、詳細は後述するが、音源が撮像装置 1 の真上に位置する場合には、真上方向にあることを示す情報が音方向情報にセットされるものとする。

10

## 【 0 0 5 5 】

ステップ S 1 1 1 にて、中央制御部 2 0 1 は、電源制御部 2 1 1 を制御し、撮像部 1 0 2、及び、レンズアクチュエータ制御部 1 0 3 への電力供給を開始する。この結果、可動撮像部 1 0 0 は、撮像装置として機能し始める。

## 【 0 0 5 6 】

次に、ステップ S 1 5 1 にて、中央制御部 2 0 1 は、音声コマンド認識部 2 0 4 3 から、音声コマンドが認識されたことを示す情報を受信したか否かを判定する。否の場合、中央制御部 2 0 1 は、処理をステップ S 1 5 2 に進め、現在、ユーザからの指示に従った実行中のジョブがあるか否かを判定する。詳細は図 6 のフローチャートの説明から明らかになるが、動画撮影記録や追尾処理等がジョブに相当する。ここでは、そのような実行中のジョブは存在しないものとして説明を続ける。

20

## 【 0 0 5 7 】

ステップ S 1 5 3 にて、前回の音声コマンドを認識してからの経過時間が、予め設定された閾値を超えるか否かを判定する。否の場合、中央制御部 2 0 1 は処理をステップ S 1 5 1 に戻し、音声コマンドの認識を待つ。そして、実行中のジョブが無く、且つ、前回の音声コマンドを認識してから閾値を超える時間が経過しても、更なる音声コマンドが認識されない場合、中央制御部 2 0 1 は処理をステップ S 1 5 4 に進める。このステップ S 1 5 4 にて、中央制御部 2 0 1 は、電源制御部 2 1 1 を制御し、撮像部 1 0 2、レンズアクチュエータ 1 0 3 への電力を遮断する。そして、中央制御部 2 0 1 は、ステップ S 1 5 5 にて、電源制御部 2 1 1 を制御し、音方向検出部 2 0 4 4 への電力も遮断し、処理をステップ S 1 0 6 に戻す。

30

## 【 0 0 5 8 】

さて、中央制御部 2 0 1 が音声コマンド認識部 2 0 4 3 から音声コマンドが認識されたことを示す情報を受信したとする。この場合、音声コマンド認識部 2 0 4 3 は、処理をステップ S 1 5 1 からステップ S 1 5 6 に進める。

## 【 0 0 5 9 】

本実施形態における中央制御部 2 0 1 は、認識した音声コマンドに応じたジョブを実行するに先立って、音声コマンドの発生した人物を、可動撮像部 1 0 0 の撮像部 1 0 2 の視野内に入れる処理を行う。そして、撮像部 1 0 2 の視野内に人物が入っている状態で、認識した音声コマンドに基づくジョブを実行する。

40

## 【 0 0 6 0 】

上記を実現するため、中央制御部 2 0 1 は、ステップ S 1 5 6 にて、音声コマンド認識部 2 0 4 3 で認識された音声コマンドに同期する音方向情報を、音方向検出部 2 0 4 4 のバッファメモリ 2 0 4 4 a から取得する。音声コマンド認識部 2 0 4 3 は、音声コマンドを認識したとき、音声用メモリ 2 0 4 2 内の音声コマンドを表す先頭と終端を表す 2 つのアドレスを中央制御部 2 0 1 に通知する。そこで、中央制御部 2 0 1 は、この 2 つのアド

50

レスが示す期間内で検出した音方向情報をバッファメモリ2044aから取得する。2つのアドレスが示す期間内に複数の音方向情報が存在することもある。その場合、中央制御部201はその中の時間的に最も後の音方向情報をバッファメモリ2044aから取得する。時間的に後の音方向情報の方が、その音声コマンドを発した人物の現在の位置を表している蓋然性が高いからである。

#### 【0061】

ステップS157にて、中央制御部201は、取得した音情報が表す音源の方向が、撮像装置の真上の方向であるか否かを判定する。なお、音方向が撮像装置の真上であるか否かの判定についての詳細は項後述する。

#### 【0062】

音源が撮像装置1の真上の方向にある場合、中央制御部201は処理をステップS158に進める。このステップS158にて、中央制御部201は、回動制御部213を制御し、レンズ部101及び撮像部102の撮像方向を図4(c)に示す真上方向になるように、可動撮像部100の第2筐体151を回動させる。撮像部102の撮像方向が真上方向になった場合、中央制御部201は、ステップS159にて、映像信号処理部203から撮像画像を受信し、撮像画像内に音声発生原となるオブジェクト(人物の顔)が存在するか否かを判定する。否の場合、中央制御部201は処理をステップS151に戻す。一方、撮像画像内にオブジェクトが存在する場合、中央制御部201は処理をステップS164に進め、既に認識した音声コマンドに対応するジョブを実行する。なお、このステップS164の詳細は図6を用いて後述する。

#### 【0063】

ステップS157にて、中央制御部201が、音情報が示す方向が真上以外の方向であると判定した場合、処理をステップS160に進める。このステップS160にて、中央制御部201は、回動制御部213を制御して、可動撮像部100のパン動作を行い、現在の撮像部102の水平面の角度を、音情報が示す水平面の角度に一致させる。そして、ステップS161にて、中央制御部201は、映像信号処理部203から撮像画像を受信し、撮像画像内に音声発生原となるオブジェクト(顔)が存在するか否かを判定する。否の場合、中央制御部201は処理をステップS162に進め、回動制御部213を制御して、目標とするオブジェクトに向かって可動撮像部100のチルト動作を行う。そして、ステップS163にて、中央制御部201は、撮像部102の撮像方向のチルトの向きの角度が、チルト動作の上限(本実施形態では水平方向に対して90度)に到達したか否かを判定する。否の場合には、中央制御部201は処理をステップS161に戻す。こうして、中央制御部201は、チルト動作を行いながら、映像信号処理部203からの撮像画像内に音声発生原となるオブジェクト(顔)が存在するか否かを判定していく。そして、撮像部102の撮像方向のチルトの向きの角度がチルト動作の上限に到達してもオブジェクトが検出されない場合、中央制御部201は処理をステップS163からステップS151に戻す。一方、撮像画像内にオブジェクトが存在した場合、中央制御部201は処理をステップS164に進め、既に認識した音声コマンドに対応するジョブを実行する。

#### 【0064】

次に、図6のフローチャート、並びに、図7に示す音声コマンドテーブルに基づいて、ステップS164の処理の詳細を説明する。図7の音声コマンドテーブルに示される“Hi, Camera”等の音声コマンドに対応する音声パターンデータはコマンドメモリ2046に格納されるものである。なお、図7には代表的な音声コマンドを示す。なお音声コマンドはこれに限らない。また、以下の説明における音声コマンドは、図5BのステップS151のタイミングで検出された音声コマンドである点に注意されたい。

#### 【0065】

まず、ステップS201にて、中央制御部201は、音声コマンドが、起動コマンドであるか否かを判定する。

#### 【0066】

この起動コマンドは、撮像装置1に対し、撮像可能な状態に遷移させるための音声コマ

ンドである。この起動コマンドは、図 5 A のステップ S 1 0 7 で判定されるコマンドであり、撮像に係るジョブを実行させるためのコマンドではない。よって、中央制御部 2 0 1 は、認識した音声コマンドが起動コマンドである場合には、そのコマンドについては無視し、処理をステップ S 1 5 1 に戻す。

【 0 0 6 7 】

ステップ S 2 0 2 にて、中央制御部 2 0 1 は、音声コマンドが、停止コマンドであるか否かを判定する。この停止コマンドは、一連の撮像可の状態から、起動コマンドの入力を待つ状態に遷移させるコマンドである。よって、中央制御部 2 0 1 は、認識した音声コマンドが停止コマンドである場合には、処理をステップ S 2 1 1 に進める。ステップ S 2 1 1 にて、中央制御部 2 0 1 は、電源制御部 2 1 1 を制御し、既に起動している撮像部 1 0 2、音方向検出部 2 0 4 4、音声コマンド認識部 2 0 4 3、動画用音声処理部 2 0 4 5、マイク 1 0 4 b 乃至 1 0 4 d 等への電力を遮断し、これらを停止する。そして、中央制御部 2 0 1 は、処理を起動時のステップ S 1 0 3 に戻す。

【 0 0 6 8 】

ステップ S 2 0 3 にて、中央制御部 2 0 1 は、音声コマンドが静止画撮影コマンドであるか否かを判定する。この静止画撮影コマンドは、撮像装置 1 に対して 1 枚の静止画の撮影・記録ジョブの実行の要求を行うコマンドである。よって、中央制御部 2 0 1 は、音声コマンドが静止画撮影コマンドであると判定した場合、処理をステップ S 2 1 2 に進める。ステップ S 2 1 2 にて、中央制御部 2 0 1 は、撮像部 1 0 2 で撮像した 1 枚の静止画像データを例えば J P E G ファイルとして、記憶部 2 0 6 に記録する。なお、この静止画撮影コマンドのジョブが、1 枚の静止画撮影記録により完結するので、先に説明した図 5 B のステップ S 1 5 2 で判定する対象のジョブとはならない。

【 0 0 6 9 】

ステップ S 2 0 4 にて、中央制御部 2 0 1 は、音声コマンドが動画撮影コマンドであるか否かを判定する。動画撮影コマンドは、撮像装置 1 に対して動画画像の撮像と記録を要求するコマンドである。中央制御部 2 0 1 は、音声コマンドが動画撮影コマンドであると判定した場合、処理をステップ S 2 1 3 に進める。このステップ S 2 1 3 にて、中央制御部 2 0 1 は、撮像部 1 0 2 を用いて動画画像の撮影と記録を開始し、処理をステップ S 1 5 1 に戻す。本実施形態では、撮像した動画画像は記憶部 2 0 6 に格納されるものとするが、外部入出力端子部 2 0 8 を介してネットワーク上のファイルサーバに送信しても構わない。動画撮影コマンドは、動画画像の撮像、記録を継続させるコマンドであるので、このコマンドによるジョブは、先に説明したステップ S 1 5 2 で判定する対象のジョブとなる。

【 0 0 7 0 】

ステップ S 2 0 5 にて、中央制御部 2 0 1 は、音声コマンドが動画撮影終了コマンドであるか否かを判定する。中央制御部 2 0 1 は、音声コマンドが動画撮影終了コマンドであり、尚且つ、現に動画画像の撮像・記録中である場合には、その記録（ジョブ）を終了する。そして、中央制御部 2 0 1 は処理をステップ S 1 5 1 に戻す。

【 0 0 7 1 】

ステップ S 2 0 6 にて、中央制御部 2 0 1 は、音声コマンドが追尾コマンドであるか否かを判定する。追尾コマンドは、撮像装置 1 に対して、撮像部 1 0 2 の撮像方向に、ユーザを継続して位置させることを要求するコマンドである。中央制御部 2 0 1 は、音声コマンドが追尾コマンドであると判定した場合、処理をステップ S 2 1 4 に進める。そして、ステップ S 2 1 4 にて、中央制御部 2 0 1 は、映像信号処理部 2 0 3 で得られた映像の中心位置にオブジェクトが位置し続けるように、回動制御部 2 1 3 の制御を開始する。そして、中央制御部 2 0 1 は処理をステップ S 1 5 1 に戻す。この結果、可動撮像部 1 0 0 がパン動作、或いはチルト動作を行い、移動するユーザを追尾する。ただし、ユーザを追尾するものの、撮像した画像の記録は行わない。また、追尾している間は、先に説明した図 5 B のステップ S 1 5 2 で判定する対象のジョブとなる。そして、追尾終了コマンドを受信して初めて、中央制御部 2 0 1 はこの動画画像の撮影記録を終了する。なお、追尾中に、例えば静止画撮影コマンドや動画撮影コマンドのジョブを実行しても構わない。

## 【 0 0 7 2 】

ステップ S 2 0 7 にて、中央制御部 2 0 1 は、音声コマンドが追尾終了コマンドであるか否かを判定する。中央制御部 2 0 1 は、音声コマンドが追尾終了コマンドであり、尚且つ、現に追尾中である場合には、その記録（ジョブ）を終了する。そして、中央制御部 2 0 1 は処理をステップ S 1 5 1 に戻す。

## 【 0 0 7 3 】

ステップ S 2 0 8 にて、中央制御部 2 0 1 は、音声コマンドが自動動画撮影コマンドであるか否かを判定する。中央制御部 2 0 1 は、音声コマンドが自動動画撮影コマンドであると判定した場合、処理をステップ S 2 1 7 に進める。このステップ S 2 1 7 にて、中央制御部 2 0 1 は、撮像部 1 0 2 による動画像の撮影と記録を開始し、処理をステップ S 1 5 1 に戻す。この自動動画撮影コマンドにより実行されるジョブと、先に説明した動画撮影コマンドにより実行されるジョブとの違いは、発声がある度に、その発声の音源の方向にレンズ部 1 0 1 の撮像方向を向けつつ動画像を撮影、記録を行う点である。例えば、複数の話者が存在するミーティングの環境下で、発言があるたびにその発言者をレンズ部 1 0 1 の画角内に収めるために、パン動作、およびチルト動作を行いながら、動画像を記録する。なお、この場合、この自動動画撮影コマンドのジョブを実行中はジョブを終了させる音声コマンドを受け付けない。このジョブの終了は、操作部 2 0 5 に設けられた所定のスイッチ操作によって終了するものとする。また、このジョブを実行中、中央制御部 2 0 1 は、音声コマンド認識部 2 0 4 3 を停止させる。そして、中央制御部 2 0 1 は、音圧レベル検出部 2 0 4 1 により、閾値を超える音圧レベルを検出したタイミングでの、音方向検出部 2 0 4 4 が検出した音方向情報を参照して、可動撮像部 1 0 4 のパン動作、チルト動作を行う。

## 【 0 0 7 4 】

なお、図 6 には示していないが、認識した音声コマンドが拡大コマンドである場合、中央制御部 2 0 1 はレンズアクチュエータ制御部 1 0 3 を制御し、予め設定された値だけ、現在のズーム倍率を増加させる。また、認識した音声コマンドが拡大コマンドである場合、中央制御部 2 0 1 はレンズアクチュエータ制御部 1 0 3 を制御し、予め設定された値だけ、現在のズーム倍率を減少させる。なお、レンズ部 1 0 1 が既にテレ端、或いは、ワイド端にあるとき、それを超えた拡大率、縮小率は設定できないので、このような音声コマンドがあった場合、中央制御部 2 0 1 はその音声コマンドを無視する。

## 【 0 0 7 5 】

以上であるが、上記以外の音声コマンドについては、ステップ S 2 0 7 以降で実行されるが、ここでの説明は省略する。

## 【 0 0 7 6 】

ここで、本実施形態における撮像装置 1 におけるメイン電源 ON からの処理のシーケンスの一例を図 8 に示すタイミングチャートに従って説明する。

## 【 0 0 7 7 】

撮像装置 1 のメイン電源が ON になると、音圧レベル検出部 2 0 4 1 はマイク 1 0 1 4 a からの音声データの音圧レベルの検出処理を開始する。タイミング T 6 0 1 にて、ユーザは、起動コマンド “Hi, Camera” の発声を開始したとする。この結果、音圧レベル検出部 2 0 4 1 が閾値を超える音圧を検出する。そして、これがトリガになって、タイミング T 6 0 2 にて、音声用メモリ 2 0 4 2 がマイク 1 0 4 a からの音声データの格納を開始し、音声コマンド認識部 2 0 4 3 が音声コマンドの認識を開始する。ユーザが起動コマンド “Hi, Camera” の発声を終わると、タイミング T 6 0 3 にて、音声コマンド認識部 2 0 4 3 がその音声コマンドを認識し、且つ、認識した音声コマンドが起動コマンドであることを特定する。

## 【 0 0 7 8 】

中央制御部 2 0 1 は、この起動コマンドが認識されたことをトリガにして、タイミング T 6 0 3 にて音方向検出部 2 0 4 4 に電力供給を開始する。また、中央制御部 2 0 1 は、タイミング T 6 0 4 にて撮像部 1 0 2 への電力供給も開始する。



## 【 0 0 7 9 】

ユーザは、タイミング T 6 0 6 にて、例えば “Movie start” の発声を開始したとする。この場合、発生の開始のタイミングの音声データは、タイミング T 6 0 7 から順に音声用メモリ 2 0 4 2 に格納されていく。そして、タイミング T 6 0 8 にて、音声コマンド認識部 2 0 4 3 が、音声データを “Movie start” を表す音声コマンドとして認識する。音声コマンド認識部 2 0 4 3 は、音声用メモリ 2 0 4 2 内の “Movie start” を表す音声データの先頭と終端のアドレスと、認識結果を中央制御部 2 0 1 に通知する。中央制御部 2 0 1 は、受信した先頭と終端のアドレスが表す範囲を有効範囲として決定する。そして、中央制御部 2 0 1 は、音方向検出部 2 0 4 4 のパッファ 2 0 4 4 a 内の、有効範囲内から、最新の音方向情報を抽出し、タイミング T 6 0 9 にて、その抽出した情報に基づいて、駆動制御部 2 1 3 を制御して、可動撮像部 1 0 0 のパン動作、チルト動作を開始する。

10

## 【 0 0 8 0 】

可動撮像部 1 0 0 のパン動作、チルト動作中に、タイミング T 6 1 2 にて、撮像信号処理部 2 0 2 が、撮像部 1 0 2 を用いて生成された画像に被写体（オブジェクト；顔）を検出した場合、中央制御部 2 0 1 はパン動作、チルト動作を停止する（タイミング T 6 1 3）。また、タイミング T 6 1 4 にて、中央制御部 2 0 1 は、動画用音声処理部 2 0 4 5 に電力を供給して、マイク 1 0 4 a、及び、1 0 4 b によるステレオ音声の集音状態にする。そして、中央制御部 2 0 1 は、タイミング T 6 1 5 にて、音声付動画像の撮像と記録を開始する。

## 【 0 0 8 1 】

次に、本実施形態における音方向検出部 2 0 4 4 による音源方向の検出処理を説明する。この処理は、図 5 A のステップ S 1 1 0 以降、周期的に、且つ、継続的に行われるものである。

20

## 【 0 0 8 2 】

まず、図 9 ( a ) を用いて、マイク 1 0 4 a とマイク 1 0 4 b の 2 つのマイクを用いた簡易の音方向検知を説明する。同図は、マイク 1 0 4 a とマイク 1 0 4 b が平面上（パン動作の回転軸に垂直な平面上）に配置されているとする。マイク 1 0 4 a とマイク 1 0 4 b の距離を  $d [ a \quad b ]$  と表す。距離  $d [ a \quad b ]$  に対して、撮像装置 1 と音源間の距離は十分に大きいと仮定する。この場合、マイク 1 0 4 a とマイク 1 0 4 b の音声を比較することによって、両者間の音声の遅延時間を特定することができる。

30

## 【 0 0 8 3 】

到達遅延時間に音速（空気中は約  $340 \text{ m/s}$ ）を乗じることで、距離  $I [ a \quad b ]$  を特定することができる。その結果、次式で音源方向角度  $[ a \quad b ]$  を特定することができる。

$$[ a \quad b ] = \arccos ( I [ a \quad b ] / d [ a \quad b ] )$$

しかしながら、2 つのマイクで求めた音方向は、求めた音源方向と  $[ a - b ]$  と  $[ a \quad b ]'$ （図 9 ( a )）との区別ができない。つまり、2 つの方向のいずれであるのかまでは特定できない。

## 【 0 0 8 4 】

そこで、本実施形態における音源の方向の検出方法を以下、図 9 ( b )、( c ) を用いて説明する。具体的には、2 つのマイクで推定できる音源方向は 2 つあるので、それら 2 つの方向を仮方向として扱う。そして、更なる 2 つのマイクで音源の方向を求め、仮方向を 2 つ求める。そして、これらに共通している方向が、求める音源の方向として決定する。なお、図 9 ( b )、( c ) の上方向を可動撮像部 1 0 0 の撮像方向とする。可動撮像部 1 0 0 の撮像方向は、レンズ部 1 0 1 の光軸方向（主軸方向）とも言い換えられる。

40

## 【 0 0 8 5 】

図 9 ( b ) は 3 つのマイクで行う方式である。マイク 1 0 4 a、マイク 1 0 4 b、マイク 1 0 4 c を用いて説明する。図 3 ( a ) で示したような配置図であると、マイク 1 0 4 a、マイク 1 0 4 b の並ぶ方向に直交する方向がレンズ部 1 0 1 の撮像方向となる。

## 【 0 0 8 6 】

50

図9(a)で説明したように、マイク104a、マイク104bより、距離 $d[a\ b]$ は既知であり、音声データより距離 $I[a\ b]$ を特定することができれば、 $[a\ b]$ を特定できる。さらにマイク104a、マイク104c間の距離 $d[a\ c]$ も既知であるので、音声データより距離 $I[a\ c]$ も特定することができ、 $[a\ c]$ を特定できる。 $[a\ b]$ 及び $[a\ c]$ が算出できれば、マイク104a、104b、104cの配置と同一2次元平面上(パン動作の回転軸に垂直な平面上)における、それらに共通な方角が、正確な音声発生方向として決定できる。

【0087】

図9(c)を用いて、4つのマイクで音源方向を決定する方法を説明する。図3(a)に示すマイク104a、マイク104b、マイク104c、マイク104dの配置により、マイク104a、マイク104bの並ぶ方向に直交する方向がレンズ部101の撮像方向(光軸方向)となる。マイク4つを利用する場合、対角線上に位置するマイク104aと104dのペアと、マイク104bとマイク104cのペアの2つのペアを用いると精度よく音源方向を算出できる。

10

【0088】

マイク104a、マイク104d間の距離 $d[a\ d]$ は既知であるので、音声データから距離 $I[a\ d]$ を特定できるので、 $[a\ d]$ も特定できる。

【0089】

更にマイク104b、マイク104c間の距離 $d[b\ c]$ も既知であるので、音声データより距離 $I[b\ c]$ を特定できるので、 $[b\ c]$ を特定できる。

20

【0090】

よって、 $[a\ d]$ 及び $[b\ c]$ がわかれば、マイクの配置と同一2次元平面上では正確な音声発生方向を検知することが可能である。

【0091】

さらに、 $[a\ b]$ 、 $[c\ d]$ と検知角度を増やしていけば、方向検知の角度の精度を高めることも可能である。

【0092】

以上のような処理を行うため、マイク104aとマイク104bとマイク104cおよびマイク104dは図3のように長方形の4つの頂点に配置した。なお、マイクの数3つであっても、それらが直線状に並ばないのであれば、必ずしも4つである必要はない。

30

【0093】

上記の方法のデメリットとして、同一2次元平面上の音方向しか検知しかできない。そのため、音源が撮像装置1の真上に位置する場合には、その方向を検出できない。そこで、次に、音方向検出部2044における、音源の存在する方向として真上であるか否かの判定原理を図10(a)、(b)を参照して説明する。

【0094】

図10(a)は3つのマイクで行う方式を説明するための図である。マイク104a、マイク104b、マイク104cを用いて説明する。図3(a)で示したような配置図であると、マイク104a、マイク104bの並び方向に直交する方向がレンズ部101の撮像方向(光軸方向)である。マイク104a、マイク104bの並び方向とは、マイク104aの中心点とマイク104bの中心点とを結ぶ直線方向である。

40

【0095】

音声入力部104の配置されている平面に対して、垂直に交わる直線状、すなわち上方からマイク104a、マイク104b、マイク104cに音声が入ってきたときについて記載する。

【0096】

ここで、撮像装置1の真上に音源が位置する場合、その音源からマイク104aとマイク104bは等距離にあると見なせる。つまり、音源からこれら2つのマイク104aと104bに到達する音の時間差は無い。そのため、マイク104aとマイク104bを結ぶ直線に対して、垂直に交わる方向に音源があると認識される。

50

## 【0097】

さらに、マイク104aとマイク104cも同様に音源からは等距離にあると見なせるので、やはり音源からこれら2つのマイク104aと104cに到達する音の時間差は無い。そのため、マイク104aとマイク104cを結ぶ直線に対して、垂直に交わる方向に音源があると認識される。

## 【0098】

つまり、マイク104aとマイク104bで検出した音の時間差の絶対値を  $T1$  とし、マイク104aとマイク104cで検出した音の時間差の絶対値を  $T2$  とし、予め設定された十分に小さい閾値  $\theta$  との関係が次の条件を満たす場合、音源が撮像装置1の真上に位置すると判定できる。

条件： $T1 < \theta$  かつ  $T2 < \theta$

## 【0099】

図10(b)を参照し、4つのマイク104a、マイク104b、マイク104c、マイク104dを用いた、撮像装置1の真上に位置する音源の検出法を説明する。図3(a)に示すように、マイク104a、マイク104dのペアと、マイク104bとマイク104cのペアについて考察する。

## 【0100】

撮像装置1の真上に音源が存在する場合、その音源からマイク104aとマイク104dは等距離になるので、これらマイク104aとマイク104dで検出する音の時間差の絶対値  $T3$  はゼロか、非常に小さい値となる。つまり、マイク104aとマイク104dを結ぶ直線に対して、垂直に交わる方向に音源があると認識となる。

## 【0101】

さらに、マイク104bとマイク104cも、音源からは等距離になるため、これらマイク104bとマイク104cで検出する音の時間差の絶対値  $T4$  もゼロか、非常に小さい値となる。つまり、マイク104bとマイク104cを結ぶ直線に対して、垂直に交わる方向に音源があると認識となる。故に、次の条件を満たす場合、音源が撮像装置1の真上に位置すると判定できる。

条件： $T3 < \theta$  且つ  $T4 < \theta$

## 【0102】

以上のように、3つ以上のマイクのうちの2つのペアについて、音の到達時間差の絶対値を求め、それら2つの絶対値が共に十分に小さい閾値未満となった場合に、音源の存在方向を真上であると決定できる。なお、2つのペアを決めるとき、それら2つのペアの向きが互いに非平行となるように決定すれば、どのような組み合わせでもよい。

## 【0103】

以上、第1の実施形態を説明した。上記実施形態によれば、音声コマンドを発声した人物(の顔)以外を誤って被写体とすることを抑制できる。また、音声コマンドを発した人物の意図したジョブを実行することも可能になる。

## 【0104】

更に、上記実施形態で説明したように、マイク104a乃至104d、音声信号処理部204を構成する各要素は、実際にそれらが利用する段階で初めて中央制御部201の制御の下で電力供給が行われるので、全構成要素が可動状態にある場合と比較して、電力消費量を抑制できる。

## 【0105】

次に、上記実施形態の説明を踏まえ、具体的な利用形態を説明する。図3(b)乃至(e)に示すように、本実施形態における撮像装置1の利用形態は様々である。

## 【0106】

ここで、例えば、図3(c)の様に、ユーザの首にぶら下げる場合について考察する。この場合、レンズ部101の撮像方向(光軸方向)がユーザの体方向に向くと、不要な画像を撮像してしまうのは容易に理解できよう。それ故、レンズ部101の撮像方向(光軸方向)は、ユーザにとって常に前方に向かうことが望ましい。この場合、図3(a)に示

10

20

30

40

50

すように、4つのマイクのうち、マイク104c、194dがユーザの体に触れる可能性が高い。つまり、このマイク104c、1094dが、ユーザの衣類とのこすれ音を集音する可能性が高くなり、4つのマイクでの音方向検出部2044による音方向検出の際の支障を来す。そこで、本実施形態では、撮像装置1をユーザの首にぶら下げる利用形態の場合、中央制御部201は、マイク104c、104dへの電力を遮断すると共に、音方向検出部2044に対し、マイク104a、104bの2つのみを用いた音方向検出を行うよう指示する。この場合、図9(a)を参照して説明した、2つのマイクのみでの音源方向を求めると音源方向が2つ検出されてしまうという問題は無い。なぜなら、音源方向は、少なくともユーザの前方の範囲内にあると見なしてよいからである。つまり、音方向検出部2044は、マイク104a、104bの2つのみを用いるため、計算上は2つの音方向を検出するものの、ユーザの前方にある音源方向を有効な音源方向として検出する。なお、ユーザの体の有る方向の検出は、例えば首にぶら下げられていると判定した後、360度(一周)のパン動作を行い、測距した距離の最短となる方向(図3(c)のユーザの胸部の方向)を中心とする適当な角度(例えば図3(c)では180度)の範囲をユーザの存在する方向として決定すればよい。また、中央制御部201は、決定した方向を記憶部206に基準方向として保存する。

10

#### 【0107】

次に、図3(d)の様に、ユーザの肩に装着する場合についても考察する。この場合、4つのマイクのうち、いずれか1つがユーザの頭に近い位置になり、ユーザの頭、或いは衣類に接する可能性が高くなる。そこで、この場合には、4つのマイクのうち、ユーザの頭に近い1つを音方向検出の際に不使用(電源OFF)とし、残り3つで音源方向を検出する。一旦、ユーザの肩に撮像装置1を装着(固定)した場合、ユーザの動きとは無関係に、撮像装置1としてのユーザの頭の相対方向は変化しない。そこで、中央制御部201は、この方向をユーザの頭の方角として記憶部206に記憶する。そして、中央制御部201は、4つのマイクのうち、記憶した方向とレンズ部101の撮像方向(光軸方向)に基づき、ユーザの頭に近い側の1つのマイクを方向検出に不使用(電力を遮断)とし、且つ、音方向検出部2044には残りの3つマイクを用いて方向検出を行うように設定する。なお、ユーザの頭の有る方向の検出は、例えば肩に装着されていると判定した後、360度のパン動作を行い、測距した距離の最短となる方向を中心とする適当な角度(例えば90度)の範囲をユーザの存在する方向として決定すればよい。また、中央制御部201は、測距した距離の最短となる方向(ユーザの頭の方角)は記憶部206に基準方向として保存する。

20

30

#### 【0108】

そして、図3(b)、(e)の利用形態の場合、音方向検出部2044は4つのマイクを用いた音方向検出を行えばよい。

#### 【0109】

ここで、図3(b)乃至(e)のいずれの利用形態とするかは、ユーザが支持部200の操作部205から設定する。ただし、ユーザが操作部205により、オート検出モードを設定した場合、中央制御部201による利用形態の自動検出を行う。以下、オート検出モードが設定されている場合の中央制御部201の自動検出の処理を説明する。

40

#### 【0110】

実施形態における、位置検出部212はジャイロセンサ、加速度センサ、GPSセンサといった撮像装置1の動きを検出する構成を有することは既に説明した。そこで、撮像装置1のメイン電源がONになり、図5AのステップS101の初期化処理後は、基本的に図3(b)の状態、すなわち、撮像装置1は固定状態にあるとして、音方向検出部2044が音方向検出を行う。

#### 【0111】

一方、図5AのステップS101の初期化処理後に、ユーザが撮像装置1を持ってその利用形態を決める操作を行うと、当然、位置検出部212の加速度センサやジャイロ等のセンサにより閾値より大きい変化を検出することになる。また、ユーザがこの操作を行う

50

タイミングは、撮像装置 1 のメイン電源の投入したタイミングとする。例えば、位置検出部 2 1 2 は、初期化処理後の予め設定された期間内にセンサのうちの 1 つでも閾値より大きい変化を検出した場合、ユーザによる撮像装置 1 の設置操作を行っているとは推定し、中央制御部 2 0 1 に割り込み信号を発する。

【 0 1 1 2 】

図 1 1 に示すフローチャートは、この割り込み処理（撮像装置 1 の設置位置の検出処理）を示している。以下、同図を参照して、中央制御部 2 0 1 の処理を説明する。

【 0 1 1 3 】

まず、ステップ S 1 1 0 1 にて、中央制御部 2 0 1 は、位置検出部 2 1 2 が有する各センサが出力してくるデータを、予め設定された期間分、記憶部 2 0 6 に保存を行う。この保存する期間は、ユーザが利用形態に係る操作を完了するまでに十分な時間が望ましい（例えば 1 分）。

【 0 1 1 4 】

この保存期間を終えると以下に説明するように、中央制御部 2 0 1 は、保存したデータに基づく、撮像装置 1 の設置位置判定を行い、音方向検出部 2 0 4 4 による音方向検出法を決定する。なお、以下の説明にて、x、y 軸が表す平面は、撮像装置 1 のパン動作の回転軸に垂直な面を表し、z 軸は撮像装置 1 のパン動作の回転軸の軸方向を表すものとして説明する。

【 0 1 1 5 】

ユーザは、撮像装置 1 を肩に装着する（図 3（d）の）場合、x、y、z 軸のいずれかの方向への移動量が図 3 に示す（b）、（c）、（e）の場合と比較し、かなり大きくなる傾向がある。それ故、中央制御部 2 0 1 は、ステップ S 1 1 0 2 にて、保存された x、y、z 軸それぞれの加速度のいずれかが予め設定された閾値を超えるか否かを判定する。閾値を超える角速度があった場合、中央制御部 2 0 1 は、撮像装置 1 がユーザの肩に装着されたものと推定し、ステップ S 1 1 0 3 にて、ユーザの頭部に近い 1 つのマイクを除く、残り 3 つのマイクによる音方向検出法（又はルール）に従って音源方向の検出を行うよう、音方向検出部 2 0 4 4 に設定し、本処理を終える。

【 0 1 1 6 】

ステップ S 1 1 0 2 にて、x、y、z 軸のいずれの方向の加速度も閾値以下であった場合、中央制御部 2 0 1 は処理をステップ S 1 1 0 4 に進める。

【 0 1 1 7 】

首にぶら下げる場合、肩載せに比べ、x、y、z、方向への移動量が小さくなる傾向がある。且つ、撮像装置 1 を首にかける為には、図 3（c）に示す通り、上下をひっくり返す動作が必要になる。このため、首にかける動作を行う場合は、ある特定軸の角速度が大きくなる傾向がある。また、z 軸回りの回転は少ない。

【 0 1 1 8 】

そこで、ステップ S 1 1 0 4 にて、中央制御部 2 0 1 は、x、y、z 軸に沿った角速度の検出と閾値比較を行う。具体的には、z 軸に対する角速度（ヨー）が予め設定された閾値以下であり、且つ、x 軸、或いは y 軸に対する角速度（ロール、ピッチ）が予め設定された閾値より大きいか判定を以上であるか判定を行う。

【 0 1 1 9 】

この条件を満たす場合、中央制御部 2 0 1 は、撮像装置 1 はユーザの首にぶら下げられているものと推定する。そして、4 つのマイクのうち、マイク 1 0 4 a、1 0 4 b の 2 つのマイクのみを用い、且つ、マイク 1 0 4 c、1 0 4 d とは逆方向が音源の存在位置であるものと見なす音方向検出法にしたがった音源方向検出を行うよう、中央制御部 2 0 1 は音方向検出部 2 0 4 4 に設定し、本処理を終える。

【 0 1 2 0 】

一方、ステップ S 1 1 0 4 にて、ヨー方向の角速度が閾値より大きく、且つ、ロールまたはピッチの角速度が閾値以下であると判定された場合、中央制御部 2 0 1 がステップ 1 1 0 6 にて、撮像装置 1 が、ユーザの手で適当な位置に固定されたものと見なす。それ故

10

20

30

40

50

、中央制御部 201 は、ステップ S 1106 にて、4つのマイクを用いた音方向検出法に従って音源方向の検出を行うよう、音方向検出部 2044 に設定し、本処理を終える。

【0121】

図 12 (a) は、撮像装置 1 がユーザの首にぶら下げた場合の音方向検出方法を示す図であり、図 12 (b) は、撮像装置 1 がユーザの肩に固定された場合の音方向検出方法を示す図である。そして、図 12 (c) は、撮像装置 1 が固定された場合の音方向検出方法を示す図である。

【0122】

図 13 (a) 乃至 (c) は、図 12 (a) 乃至 (c) それぞれの方法で得られるマイクの指向性を示した図である。なお、図 12 (a) 乃至 (c) による音源方向の決定法が、図 9 (a) 乃至 (c) と同じため詳細な説明を省略し、以下、簡単に説明する。

10

【0123】

図 12 (a) は、図 11 の処理にて、撮像装置 1 がユーザの首にかけられたと判断された場合の音方向の検出法を示している。音源の方向導出原理そのものは、図 9 (a) と同じである。マイク 104a、マイク 104b 間の距離  $d[a-b]$  を一辺とし、 $[a-b]$  を求める。音源の方向は角度  $[a-b]$  および角度  $[a-b]'$  の 2 つの候補があるが、ユーザの体の方向の角度  $[a-b]'$  は、無視すればよい。また、先に説明したように、マイク 104c、104d への電力は遮断してよい。なお、図 13 (a) に示す参照符号 1101 が示す破線範囲が、この検出法による音源方向の検出範囲を示している。図示のように、前方に対する音方向の検出範囲が後方よりも広がるが、後方はユーザの体になるので問題

20

【0124】

図 12 (b) は、図 11 の処理にて、撮像装置 1 がユーザの肩に装着されたと判断された場合の音方向の検出法を示している。ユーザの頭の方向は、図示の左下方向であるものとしている。撮像装置 1 はユーザの肩に装着された場合、マイク 104a、マイク 104b 間の距離  $d[a-b]$  を一辺とし、 $[a-b]$  を求める。その後、マイク 104b、マイク 104c 間の距離  $d[c-b]$  を一辺とした、 $[c-b]$  を求め、 $[a-b]$  との相関をとることで音源位置の角度を求める。電源が OFF となるのは、4つのマイクのうち 1 つであり、残り 3 つは、音方向検出部 2044 が可動中である限り、電源が供給される。図 13 (b) の参照符号 1102 が示す範囲は、この検出法による音源方向の検出範囲を示している。図示のように、音方向の検出範囲が左下方向が狭くなるが、この方向にユーザの頭が存在することになるので、特に問題はない。

30

【0125】

図 12 (c) は、図 11 の処理にて、撮像装置 1 がユーザ等の移動体に装着されるのではなく、固定される場合の音方向の検出法を示している。この場合、4つのマイクの全てへの電源が供給され、これら 4つのマイクを用いた音方向が行われる。図 13 (c) の参照符号 1103 が示す範囲が、この検出法による音源方向の検出範囲を示している。図示のように、音方向の検出範囲が全方向に対して偏りなく音源方向が検出できる。

【0126】

以上、撮像装置の取り付け位置を検出し、該検出情報に応じた音方向検知方法を選択することで、音方向検知時に、取り付け部位に適したマイクの指向性を確保し、検知精度を向上することが可能となる。

40

【0127】

[第2の実施形態]

第2の実施形態を説明する。装置構成は、上記第1の実施形態と同じであるとし、その説明は省略し、異なる点について説明する。

【0128】

撮像装置 1 を室内の人員を撮影するため、部屋の隅に固定する場合を考察する。しかし、何等かの原因により、音方向検出部 2044 が設置位置の近くの壁の方向に音源があると誤検出した場合、上記実施形態によれば、レンズ部 101 の撮像方向（光軸方向）は一

50

旦、その壁の方向への無意味なパン動作をしてしまう。

【0129】

そこで、本第2の実施形態では、中央制御部201は、音方向検出部2044に対して、音方向の有効範囲（或いは無効範囲）を設定する。そして、音方向検出部2044は、音方向検出処理にて、検出した音方向が有効範囲内にある場合のみ、その方向を示す音情報を、内部のバッファに2044aに格納する例を説明する。換言すれば、音方向検出部2044は、音方向検出処理にて、検出した音方向が無効範囲内にある場合には、その検出した音方向を示す情報を内部のバッファに2044aに格納せず、無視（マスク）する例を説明する。

【0130】

図14（A）乃至（F）は、本第2の実施形態にて想定した撮像装置1の利用形態と、対応のマスク領域との関係を示す図である。

【0131】

図14（A）は撮像装置1をユーザの首から下げた例である。図示の矢印Aをユーザの前方と定義したとき、図14（B）は、撮像装置1を底面から見た透視図である。図示の様に、マイク104a及びマイク104b側が撮像装置1の撮影できる領域である。逆にマイク104c及び、マイク104dが配置されている側は撮影する必要のない領域であることがわかる。そこで、中央制御部201は、音方向検出部2044に対し、ユーザの体の方向を中心とする所定の範囲（図示では180度の範囲）を、音方向検出のマスク領域として設定する。この設定を受け、音方向検出部2044は、検出した音方向が、設定されたマスク領域内にある場合には、その音方向を表す音方向情報をバッファメモリ2044aに格納しない。換言すれば、音方向検出部2044は、検出した音方向が、設定されたマスク領域外にある場合のみ、音方向情報をバッファメモリ2044aに格納する。この結果、中央制御部201は、レンズ部101の撮像方向（光軸方向）を、マスク領域に向かうようなパン動作を行わないようになる。

【0132】

図14（C）は撮像装置1を部屋の隅の壁際においた例である。この時、図14（D）が表すように、撮像装置1を上面から見たとき、四隅に向かう方向を中心とする適当な角度範囲（例えば200度）をマスク領域とする。

【0133】

図14（E）は撮像装置1をユーザの肩に装着した例である。図14（F）は、ユーザの上方から見た場合のマスク領域を示している。図示の如く、ユーザの頭部が存在する方向を示す領域をマスク領域とする。

【0134】

次に、第2の実施形態における、中央制御部201による処理を図15（a）のフローチャートを参照して説明する。図15（a）は、中央制御部201によるマスク領域設定を含む、主要な処理を示している点に注意されたい。また、以下では、図6のステップS217の自動動画撮影記録のジョブが実行されているものとして説明する。

【0135】

中央制御部201は、自動動画撮影モードに移行すると、ステップS1502にて、撮像部102、撮像信号処理部202の出力より、現在の画角範囲が撮影の必要な領域かどうかを確認する。判断方法としては、得られた画像が所定値以上の輝度があるか、レンズアクチュエータ制御部103によってフォーカスが合う位置に被写体がいるか、もしくは被写体が近すぎないかを判断する。測距センサや、距離マップなどを用いて被写体までの距離を求めて判断してもよい。

【0136】

現在の画角の一部もしくは全体が、撮影不要と判断された場合、中央制御部201は、ステップS1503にて、その角度を音方向検知マスク領域として、記憶部206に保存する。

【0137】

10

20

30

40

50

中央制御部 201 は、ステップ S 1504 にて、回動制御部 213 を制御し、予め設定された単位角度だけ、可動撮像部 100 をパン動作させる。そして、中央制御部 201 は、ステップ S 1505 にて、このパン動作が 360 度（一周）したと判定するまで、ステップ S 1502 以降の処理を繰り返す。この結果、記憶部 206 には、マスクする複数の角度が格納されているので、中央制御部 201 は、複数の角度のうちの両端の角度で挟まれる複数の角度を含む範囲をマスク領域として決定する。ここまで、初期の音方向検知マスク領域判定動作を完了とする。

#### 【0138】

この後、ステップ S 1506 にて、音方向検出部 2044 が音源の方向を検出したとする。この場合、ステップ S 1507 にて、音方向検出部 2044 は、その音源の方向が先に決定したマスク領域内にあるか否かを判定する。検出した音源方向がマスク領域内にあるとき、音方向検出部 2044 はその音源方向を無視する。つまり、音方向検出部は、音方向情報を、内部のバッファメモリ 2044a に格納せず、処理をステップ S 1506 に戻す。

#### 【0139】

一方、検出した音方向がマスク領域外であった場合、音方向検出部 2044 は、検出した方向を内部バッファ 2044a に格納する。この結果、中央制御部 201 は、音方向検出部 2044 による音方向検出があったことがわかるので、ステップ S 1508 にて、回動制御部 213 を制御し、可動撮像部 100 を音源方向に向かうようにパン動作させる。

#### 【0140】

そして、ステップ S 1509 にて、中央制御部 201 が、映像信号処理部 203 を介して取得した画像内に被写体を検出できなかった場合は、処理をステップ S 1506 に戻し、音方向検知の待機状態を継続する。

#### 【0141】

一方、撮像画像内に被写体が含まれている場合、中央制御部 201 は、ステップ S 1510 にて、顔認識、追尾、静止画撮影や動画撮影といったジョブを実行する。ここで、ステップ S 1511 にて、位置検出部 212 のジャイロや加速度センサを用いて撮像装置 1 の移動の検出が行われる。ここで位置検出部 212 によって撮像装置 1 の移動が検出された場合、中央制御部 201 は撮像装置 1 が持ち運ばれていると判断する。そして、中央制御部 201 は、ステップ S 1502 に戻し、再度の音方向検知マスク領域の設定処理を行う。

#### 【0142】

図 15 (a) は、マスク領域設定処理を、撮像装置 1 の通常利用する事前処理で行うものであったが、随時、音方向検知マスク領域を更新する処理を図 15 (b) のフローチャートを参照して説明する。なお、以下の説明でも、中央制御部 201 によるマスク領域設定を含む、主要な処理を示している点に注意されたい。つまり、図 15 (b) では、第 1 の実施形態で説明した起動コマンド等の電源制御については省略し、マスク領域の設定と、音方向検出から音声コマンドに基づく処理の主要部のみを示している。

#### 【0143】

中央制御部 201 は、ステップ S 1522 にて、音方向検出部 2044 による音方向検出を待つ。音方向検出があると、中央制御部 201 は、ステップ S 1523 にて、検出した音源の方向が音検出のマスク領域内にあるか否かを判定し、音源の方向がマスク領域内であれば、その音方向は無視し、ステップ S 1522 に処理を戻す。なお、初期状態では、音方向検出のマスク領域が設定されていない。それ故、中央制御部 201 は、ステップ S 1524 に処理を進め、回動制御部 213 を制御し、可動撮像部 100 を音源方向に向かうようにパン動作を開始させる。

#### 【0144】

所定の時間のパン動作後、ステップ S 1525 にて、中央制御部 201 は、映像信号処理部 203 の出力より、画角範囲が撮影の必要な領域かどうかを確認する。判断方法としては、得られた画像が所定値以上の輝度があるか、レンズアクチュエータ制御部 103 に

10

20

30

40

50



よってフォーカスが合う位置に被写体がいるか、もしくは被写体が近すぎて合焦不可となるかを判断する。測距センサや、距離マップなどを用いて被写体までの距離を求めて判断しても構わない。

#### 【0145】

現在の画角の一部もしくは全体が、撮影の必要な領域と判断された場合、中央制御部201は、ステップS1526にて、その方向（角度）は音方向検知マスク領域を解除して保存する。逆に、画角の一部もしくは全体が撮影不要と判断された場合、中央制御部201は、ステップS1527にて、その方向（角度）を音方向検知マスク領域として保存する。

#### 【0146】

10

そして、中央制御部201は、ステップS1528にて、先のステップS1522にて検出した音源の方向に到達したか否かを判定する。否の場合、中央制御部201はステップS1529にて、パン動作を行う。そして、中央制御部201は、処理をステップS1525に戻す。

#### 【0147】

ステップS1528にて、中央制御部201が、音方向検出部2044の方向へのパン動作を行ったと判断した場合、処理をステップS1530に進める。このステップS1530にて、中央制御部201は、映像信号処理部203を介して得た画像内に被写体（顔）を検出する。検出できなかった場合、中央制御部201は、処理をステップS1522に戻し、音方向検知の待機状態に戻す。一方、映像信号処理部203で得た画像中に被写体を検出できた場合、中央制御部201は、処理をステップS1531に進め、認識した音声コマンドに応じた、追尾、静止画撮影や動画撮影といった所定動作を行う。

20

#### 【0148】

以上のようにして、音方向検知マスク領域を拡大、縮小していくことによって、最適な方向のみの音方向検出部2044の結果を得ることができる。

#### 【0149】

以上のようにして、音方向検知マスク領域を拡大、縮小の更新処理を行うことによって、最適な方向のみの音方向検出部2044の結果を得ることができる。

#### 【0150】

##### [第3の実施形態]

30

本第3の実施形態は、図6のステップS217の自動動画記録のジョブに適用した例を説明する。図16は、撮像装置1が演壇1605上に固定されており、被写体（の顔）1603、1604が異なる高さとなっている模式図である（一方の人物が起立しており、他方が着席していると考えるとわかりやすい）。

#### 【0151】

図16において、撮像装置1が被写体1603を撮影している最中に（参照符号1601がその時の画角）、その後、被写体1604が声を発したとする。この場合、撮像装置1は被写体160の水平方向の角度（パン角）を検知することができるが、被写体1604の垂直方向の角度（チルト角）は検知することができない（図示の参照符号1602はチルト角未定でパン動作が完了した際の画角）。このため、パン動作後に、チルト動作を徐々に行っては被写体を検出する必要がある。

40

#### 【0152】

しかしながら、被写体1603と被写体1604の撮影が交互に繰り返される場合、パン動作するたびに、画角のチルト動作を行って被写体を探索する必要が出てきてしまい、被写体の検出までに、より多くの時間がかかる。また、動画を記録している場合には、違和感のある画角移動をする動画が記録されてしまうといった問題が発生してしまう。

#### 【0153】

そこで、本第3の実施形態では、一度、被写体を認識した場合、その際のレンズ部101の撮像方向（光軸方向）を表すパン角、チルト角を学習（記憶）する。そして、音方向検出部2044が検出した音方向が、学習済みの方向に対して、予め設定された閾値以下

50

の許容範囲内にある場合（２つの方向がほぼ一致する）は、その学習済みの方向に向けてパン動作とチルト動作を同時に実行し、レンズ部１０１の撮像方向（光軸方向）を学習済みの方向に一致させ、パン動作、チルト動作の時間の短縮を図る。なお、パン角、チルト角の学習では、第１の実施形態で説明した基準角、すなわち、撮像装置１の起動時のレンズ部１０１の水平面方向（パン０度）、及びチルト範囲の水平方向（チルト０度）として、その差分を記憶部２０６に記憶するものとする。

#### 【０１５４】

図１７は、本第３の実施形態における、中央制御部２０１の自動動画記録ジョブ（図６のステップＳ２１７）の処理手順を示すフローチャートである。なお、本処理を開始する際に、既に、音声付きの動画像撮影、記録が開始されているものとする。

10

#### 【０１５５】

まず、ステップＳ１７０１において、中央制御部２０１は、音方向検出部２０４４による音源の方向の検出を待つ。音源の方向の検出が行われた場合、中央制御部２０１は、処理をステップＳ１７０２に進め、現在のレンズ部１０１の撮像方向（光軸方向）と検出した音源の方向から、パン動作の方向と角度を決定する。そして、ステップＳ１７０３にて、中央制御部２０１は、今回検出した音源方向に一致する、被写体情報が記憶部２０６に既に登録されているか否かを判断する。本実施形態の撮像装置１においては、過去の被写体情報を記憶部２０６に保存しておくことができる。過去の被写体情報として、被写体検出を行った時間、水平方向の角度（パン角）、垂直方向の角度（チルト角）の情報を蓄積しておくことで、新たな撮影を行う時の被写体検索の有効な手がかりとすることができる。

20

#### 【０１５６】

ステップＳ１７０３において、中央制御部２０１が、今回検出した音源方向に一致する過去の被写体情報があると判断した場合、処理をステップＳ１７０４に移行する。また、ステップＳ１７０３において、中央制御部２０１が、今回検出した音源方向に一致する、被写体情報が無いと判断した場合、処理をステップＳ１７０６に進める。

#### 【０１５７】

ステップＳ１７０４において、中央制御部２０１は、今回検出した音源方向に一致すると判定された被写体情報が表すチルト角と、現在のチルト角とから、チルト動作の方向と角度を決定する。そして、ステップＳ１７０５にて、中央制御部２０１は、先のステップＳ１７０２において決定されたチルト動作の方向と角度の情報を元に、レンズ部１０１の撮像方向（光軸方向）が最短距離で目標方向に向かうべく、パン動作とチルト動作を並列して実行する。このようにすることで、過去の被写体情報を検出した時点から撮像装置１と被写体の位置関係が変化していない場合、一度の画角移動で被写体の検出を行うことができ、被写体の検出までの時間を最小化できる。そのため、撮像装置１を用いて動画を記録している場合にも、ユーザにとって違和感の無い画角移動をする動画を記録することができる。

30

#### 【０１５８】

ステップＳ１７０６にて、中央制御部２０１は、パン動作を行わせ、レンズ部１０１の撮像方向（光軸方向）を、検出した音源に撮像方向（光軸方向）を向ける。そして、中央制御部２０１は、処理をステップＳ１７０７に進める。

40

#### 【０１５９】

ステップＳ１７０７にて、中央制御部２０１が、映像信号処理部２０３より得た現在の撮像画像から、被写体検出を行う。被写体が検出された場合、ステップＳ１７０８に移行し、その被写体の撮影を実施する。このとき、中央制御部２０１は、記憶部２０６内に、現在のパン角に対して許容範囲内の差の被写体情報が存在した場合、その被写体情報のパン角、チルト角を現在のレンズ部１０１の視線に合わせて更新する。また、中央制御部２０１は、記憶部２０６内に、現在のパン角に対して許容範囲内の差の被写体情報が存在しない場合、現在のレンズ部１０１の撮像方向（光軸方向）を示すパン角、チルト角を、新規な被写体情報として、記憶部２０６に登録する。

50

## 【 0 1 6 0 】

一方、ステップ S 1 7 0 7 にて、画角移動後に被写体が発見されなかった場合、中央制御部 2 0 1 は、ステップ S 1 7 0 9 に処理を進める。このステップ S 1 7 0 9 にて、中央制御部 2 0 1 は、レンズ部 1 0 1 の撮像方向（光軸方向）を垂直方向に移動（チルト動作）させ、被写体の探索を行なう。そして、ステップ S 1 7 1 0 にて、中央制御部 2 0 1 は、被写体が発見されたか否かを判定する。被写体が発見された場合には、処理をステップ S 1 7 0 8 に進める。ステップ S 1 7 0 8 に処理が進んだ場合、新規な被写体情報が記憶部 2 0 6 に登録される。

## 【 0 1 6 1 】

また、ステップ S 1 7 1 0 において、被写体が発見されなかった場合、中央制御部 2 0 1 は、処理をステップ S 1 7 1 1 に進め、エラー処理を行う。このエラー処理は、例えば、その位置のまま撮影、記録を継続する処理でもよいが、例えばステップ S 1 7 0 1 にて音源方向を発見したと判定された時点でのレンズ部 1 0 1 の撮像方向（光軸方向）に戻す処理でも良い。また、被写体が移動している可能性があるため、現在のレンズ部 1 0 1 の水平面のパン角に対し許容範囲内のパン角となっている、被写体情報を記憶部 2 0 6 から削除する処理である。

## 【 0 1 6 2 】

図 1 8 は本第 3 の実施形態の撮像装置の制御を模式的に示す図である。被写体 1 6 0 4 が発声したことに起因して、撮像装置 1 がパン動作、チルト動作を行って被写体 1 6 0 4 を検出できたとする。この場合、本実施形態の撮像装置 1 は、次回、被写体 1 6 0 4 が発声した場合、直ちに、レンズ部 1 0 1 の画角を 1 8 0 1 となるように、パン動作とチルト動作を最短距離となるように制御することができるようになる。

## 【 0 1 6 3 】

次に、第 3 の実施形態の変形例を説明する。以下でも、図 6 のステップ S 2 1 7 の自動動画記録のジョブに適用した例を説明する。

## 【 0 1 6 4 】

図 1 9 は、本変形例における中央制御部 2 0 1 による、自動動画記録のジョブ中の処理手順を示すフローチャートである。なお、本処理を開始する際に、既に、音声付きの動画撮影、記録が開始されているものとする。

## 【 0 1 6 5 】

図 1 7 との違いは、ステップ S 1 9 0 1、ステップ S 1 9 0 2 が追加された点である。

## 【 0 1 6 6 】

まず、中央制御部 2 0 1 は、ステップ S 1 7 0 1 において、音方向検出部 2 0 4 4 により音源の方向の検出を待つ。音源方向の検出が行われると、ステップ S 1 7 0 2 にて、中央制御部 2 0 1 は、現在のレンズ部 1 0 1 の撮像方向（光軸方向）と検出した音源の方向に基づき、パン動作の方向と角度を決定する。

## 【 0 1 6 7 】

次に、ステップ S 1 9 0 1 にて、中央制御部 2 0 1 は、記憶部 2 0 6 に目標方向を中心とする、予め設定された範囲内に複数の被写体情報があるかどうかの判断を行う。今回検出した音源方向に複数の被写体情報があると判断した場合、中央制御部 2 0 1 は処理をステップ S 1 9 0 2 に移行する。また、被写体検出情報が 1 つのみ、或いは存在しない場合、中央制御部 2 0 1 は処理をステップ S 1 7 0 3 に進める。

## 【 0 1 6 8 】

ステップ S 1 9 0 2 にて、中央制御部 2 0 1 は、複数の被写体が発見されたレンズ部 1 0 1 の画角内に入る目標チルト角を決定する。そして、中央制御部 2 0 1 は処理をステップ S 1 7 0 5 に進める。

## 【 0 1 6 9 】

ステップ S 1 7 0 3 以降は、図 1 7 と同じであるため、その説明は省略する。

## 【 0 1 7 0 】

上記の処理の結果、複数の被写体が発見された場所、その中の一人が発声した場

10

20

30

40

50

合には、実際に発声した被写体を含む複数の被写体を画角にして撮影することになり、違和感のない動画像を記録することができる。

#### 【0171】

例えば、図20に示すように、被写体1604、1610が近い位置にあり、いずれの被写体情報も記憶部206に登録されている状況下で、仮に被写体1604が発声した場合には、図示の画角2001となるように中央制御部201が可動撮像部100のパン動作及びチルト動作を最短距離となるように動作することになり、自然な動画撮影記録が行えるようになる。

#### 【0172】

以上説明したように、本第3の実施形態並びにその変形例によれば、一旦発声した被写体をレンズ部101の画角内に入れて認識されると、その被写体の方向である、基準方向に対するパン角、チルト角を被写体情報として記憶（学習）する。そして、2回目以降、音方向検出部2044が検出した音方向のパン角が、記憶した被写体情報におけるパン角と実質的に一致した場合には、記憶された被写体情報が示すパン角、チルト角になるよう、パン動作とチルト動作を同時に実行して、可動撮像部100を移動させる。この結果、自然な被写体の切り替わりが行われ、違和感の少ない動画像記録を行うことが可能になる。

#### 【0173】

##### [第4の実施形態]

第4の実施形態を説明する。本第4の実施形態は、音方向検出部2044が検出する音方向の検出精度を可変にする例を説明する。音方向検出部2044による音方向の検出原理については既に説明したが、音方向の検出精度を高くするためには、単位時間当たりの検出回数を上げて、その平均値を求めることである。しかし、単位時間値の検出回数を増やすことは、音方向検出部2044の負担の増加、すなわち、可動率を上げることになり、撮像装置1の消費電力の増大になる。

#### 【0174】

そこで、本第4の実施形態では、音方向検出部2044の音方向検出精度を可変にし、必要に応じて精度を低くしたり、高める例を説明する。

#### 【0175】

図21(a)、(b)、及び、図22(a)乃至(c)は、ある撮影例における撮像装置1の水平方向の撮影画角と、音方向検出時の水平方向の検出分解能の関係を示す図である。図21(a)、(b)、及び、図22(a)乃至(c)において、座標右方向を基準方向の0°とし、反時計回りの回転方向を正方向とする。また、一点鎖線で示される角度が、撮像装置1のレンズ部101の撮影画角とする。図21(a)、(b)では  $\theta = 110^\circ$ 、図22(a)乃至(c)では  $\theta = 40^\circ$  の例である。なお、撮影画角  $\theta$  が小さいほどズーム倍率が高いことを示し、逆に撮影画角  $\theta$  が大きいということはズーム倍率が低いことを示す。ここで、音方向検出部2044の水平方向の角度の分解能を音方向検出分解能  $\phi$  とする。そして、音方向検出部2044が検出した音源の位置を示すのが図示の黒丸部分とする。

#### 【0176】

図21(a)、(b)は、撮影画角  $\theta >$  音方向検出分解能  $\phi$  の場合の撮影例を示している。上記の通り、撮影画角  $\theta$  は  $110^\circ$ 、音方向検出分解能  $\phi$  は  $90^\circ$  である。音方向検出分解能  $\phi$  が  $90^\circ$  とは、音方向検出範囲を4分割しているのと同義である。この場合、音方向検出部2044が出力する音方向検出結果は、 $0 \sim 90^\circ$ 、 $90 \sim 180^\circ$ 、 $180 \sim 270^\circ$ 、 $270^\circ \sim 360^\circ$  ( $0^\circ$ ) の4方向のうちいずれかとなる。

#### 【0177】

図22(a)は撮像装置1の初期状態を示しており、撮影方向は  $90^\circ$  である。また、音声を発する被写体は座標  $270^\circ \sim 360^\circ$  ( $0^\circ$ ) の範囲の点で示した位置に存在する。図21(a)の撮影例において、音方向検出を行った後、図21(b)に示すように、音方向検出された範囲を撮影画角  $\theta$  がカバーするようにパン駆動し撮影方向を変更する

ことで、撮影画角 内に被写体を収めることができる。

【0178】

図22(a)乃至(c)は、撮影画角 < 音方向検出分解能 の場合の撮影例を示している。同図では、撮影画角 は40°、音方向検出分解能 は90°である。図22(a)は撮像装置1の初期状態を示しており、撮影方向は90°である。また、音声を発する被写体は座標270°~360°(0°)の範囲の点で示した位置に存在する。図22(a)で示す撮影例において、音方向検出を行った後、図22(b)、もしくは図22(c)に示すように音方向検出された範囲に撮影画角 が収まるようにパン駆動し撮影方向を変更することになる。撮影方向を図22(c)に示すように変更した場合、撮影画角 内に被写体を収めることができるが、図22(b)に示すように変更後の撮影画角 内に被写体を収めることができない。この場合、被写体を撮影画角 内に収めるために繰り返しパン駆動を行い、図22(c)に示すような撮影方向に変更する必要がある。

10

【0179】

以上、図21(a)、(b)、及び、図22(a)乃至(c)を用いて説明したように、撮影画角 > 音方向検出分解能 の場合は、音声検出方向を一度のパン駆動で撮影画角内に収めることができ、被写体検出を行える。しかし、撮影画角 < 音方向検出分解能 の場合は、音声検出方向を一度のパン駆動で撮影画角内に収めることができない可能性があり、結果、繰り返しのパン駆動により被写体検出に余計な動作時間と消費電力がかかってしまう問題があるのは理解できよう。

【0180】

20

図23は、音方向検出分解能 と音声信号処理部2045の処理量の関係を示した図である。音方向検出分解能 が小さくなるほど音声信号処理部2045の単位時間当たりの処理量は増加し、音方向検出分解能 が大きくなるほど音声信号処理部2045の単位時間当たりの処理量は減少する関係がある。つまり、音方向検出分解能 を必要以上に小さくしてしまうと、音声信号処理部2045の処理量が増大し、他の処理へ影響を与えてしまう問題がある。

【0181】

以上から、撮影画角 と音方向検出分解能 の関係は、撮影画角 > 音方向検出分解能 の条件を満たしつつ、音方向検出分解能 は可能な限り大きくすることが望ましい。

【0182】

30

図24(a)乃至(c)は、本第4の実施形態における撮像装置1の水平方向の撮影画角と、音方向検出時の水平方向の検出分解能の関係を示す図である。図25は、音声コマンド認識部2043が拡大コマンド又は縮小コマンドを認識した際の中央制御部201の処理を表すフローチャートである。図25のフローチャートは、第1の実施形態における図5BのステップS164の処理の一部である。すなわち、図6において省略したステップS208以降に音声コマンドが拡大、もしくは縮小コマンドであると判定された場合の処理である。

【0183】

ステップS2501にて、中央制御部201は、認識された音声コマンドが拡大コマンド、縮小コマンドのいずれであるかを判定する。拡大コマンドである場合、中央制御部201は処理をステップS2502に進める。このステップS2502にて、中央制御部201は、レンズアクチュエータ制御部103から、現在のズームレンズの位置を取得し、その位置がテレ端になっているか否かを判定する。現在のズームレンズの位置がテレ端の位置になっている場合、これ以上の拡大はできない。そこで、中央制御部201は、認識された拡大コマンドは無視し、処理を図5BのステップS151に処理を戻す。

40

【0184】

また、現在のズームレンズの位置がテレ端に至っていないと判断した場合、中央制御部201は、処理をステップS2503に進める。このステップS2503にて、中央制御部201は、レンズアクチュエータ制御部103を制御し、所定倍率だけズーム倍率を増加させる。そして、中央制御部201は、処理を図5BのステップS151に処理を戻す

50

。

## 【 0 1 8 5 】

一方、ステップ S 2 5 0 1 にて、認識された音声コマンドが縮小コマンドであると判定した場合、中央制御部 2 0 1 は処理をステップ S 2 5 0 4 に進める。このステップ S 2 5 0 4 にて、中央制御部 2 0 1 は、レンズアクチュエータ制御部 1 0 3 から、現在のズームレンズの位置を取得し、その位置がワイド端になっているか否かを判定する。現在のズームレンズの位置がワイド端の位置になっている場合、これ以上の縮小はできない。そこで、中央制御部 2 0 1 は、認識された縮小コマンドは無視し、処理を図 5 B のステップ S 1 5 1 に処理を戻す。

## 【 0 1 8 6 】

また、現在のズームレンズの位置がワイド端に至っていないと判断した場合、中央制御部 2 0 1 は、処理をステップ S 2 5 0 5 に進める。このステップ S 2 5 0 5 にて、中央制御部 2 0 1 は、レンズアクチュエータ制御部 1 0 3 を制御し、所定倍率だけズーム倍率を減少させる。そして、中央制御部 2 0 1 は、処理を図 5 B のステップ S 1 5 1 に処理を戻す。

## 【 0 1 8 7 】

以上の結果、例えば、今、図 2 6 ( a ) に示すように、撮影画角が 1 1 0 で、レンズ部 1 0 1 が基準方向に対して 9 0 度を向いており、且つ、音方向検出分解能 が 9 0 度であるものとする。そして、このとき、座標 2 7 0 度乃至 3 6 0 度に位置する黒丸が示す人物が拡大コマンドを発声したとする。この場合、音方向検出分解能 は 9 0 度であるので、パン動作の結果のレンズ部 1 0 1 の画角は図 2 6 ( b ) のようになる。つまり、発声した被写体をレンズ部 1 0 1 の画角内に入れることは可能になる。しかし、このコマンドを実行することになるので、レンズ部 1 0 1 の画角が狭くなる。結果、図 2 6 ( c ) に示すように、被写体 ( 黒丸 ) が更新後のレンズ部 1 0 1 の画角外になることが起こり得る。しかし、同一人物が拡大コマンドを発声すると、前回よりも高い分解能である音方向検出分解能 が設定された状態 ( 音方向検出分解能 が 3 0 度 ) で、パン動作することになるので、図 2 6 ( d ) に示すように、レンズ部 1 0 1 の画角内に被写体を入れることができる。つまり、被写体の人物が拡大コマンドを連呼すれば、レンズ部 1 0 1 の撮像方向 ( 光軸方向 ) はより高い精度で被写体に向けうようになり、且つ、どんどん拡大していくことになる。

## 【 0 1 8 8 】

以上、説明したように本第 4 の実施形態によれば、ズーム駆動により撮影画角が変更となった場合にも音声検出分解能 を変更する。この結果、変更後の音声検出分解能 に従って音方向検出を行うことで、それに処理にかかる時間や消費電力を抑えつつ画角外に存在する被写体を効率よく画角内に収めることができる。また、被写体となる人物が拡大コマンドを発声した後、例えば、動画像撮影コマンドを発声した場合、その人物の拡大した状態での動画像撮影と記憶が行われることになる。

## 【 0 1 8 9 】

上記例では、ユーザによるズームに関する音声コマンドに応じて、音方向の分解能を変更するものであった。しかし、音声コマンドに応じてパン動作した際、撮像画像中に複数の被写体が存在する場合、ズーム倍率とは無関係に、発声者を特定するために音方向の分解能を高くするようにしてもよい。

## 【 0 1 9 0 】

( その他の実施例 )

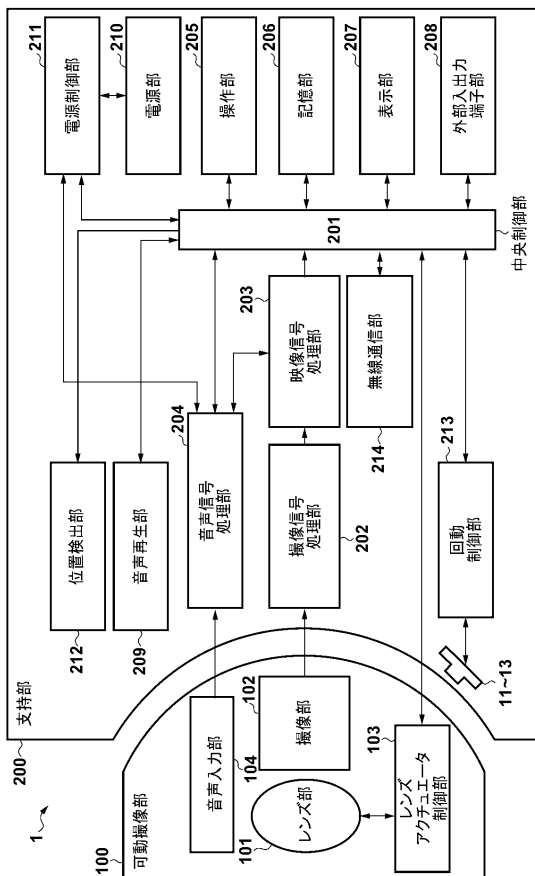
本発明は、上述の実施形態の 1 以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける 1 つ以上のプロセッサがプログラムを読み出し実行する処理でも実現可能である。また、1 以上の機能を実現する回路 ( 例えば、A S I C ) によっても実現可能である。

## 【 符号の説明 】

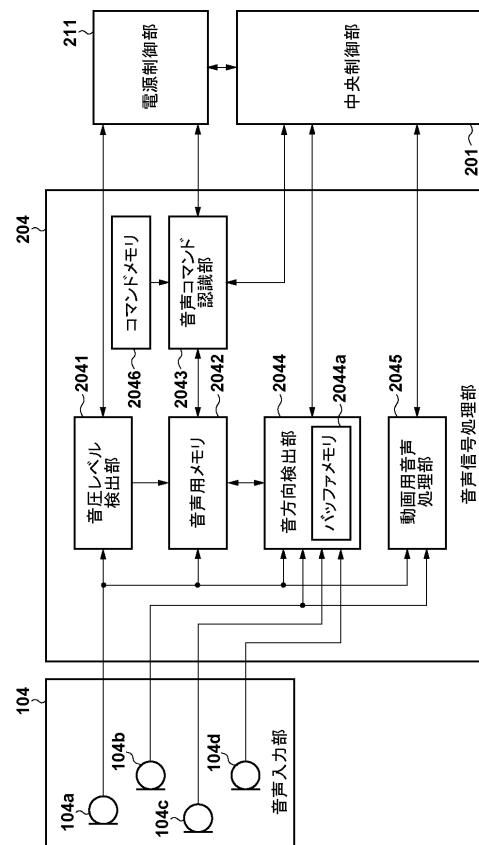
## 【 0 1 9 1 】

1 ... 撮像装置、100 ... 可動撮像部、101 ... レンズ部、102 ... 撮像部、103 ... レンズアクチュエータ制御部、104 ... 音声入力部、104a乃至104d ... マイク、201 ... 中央制御部、204 ... 音声信号処理部、211 ... 電源制御部、2041 ... 音圧レベル検出部、2042 ... 音声用メモリ、2043 ... 音声コマンド認識部、2044 ... 音方向検出部、2045 ... 動画音声処理部、2046 ... コマンドメモリ

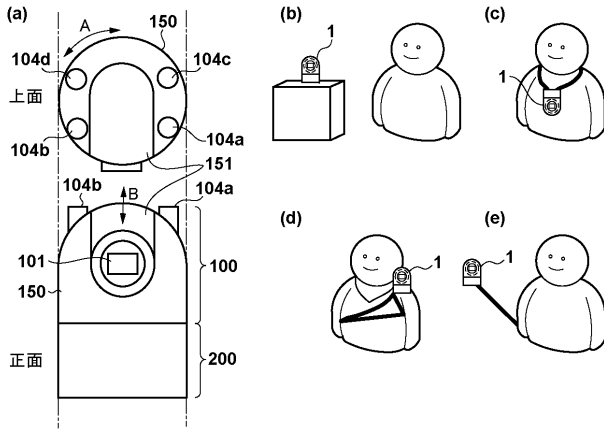
【図1】



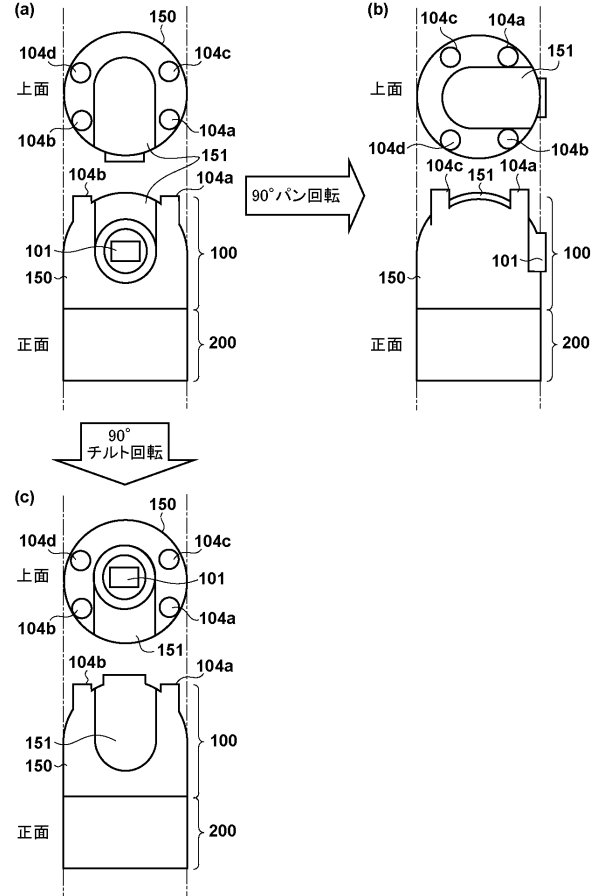
【図2】



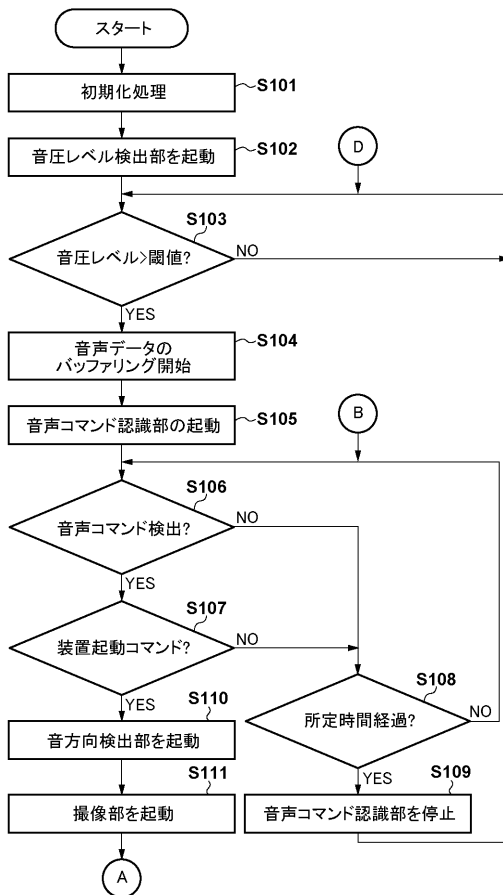
【図 3】



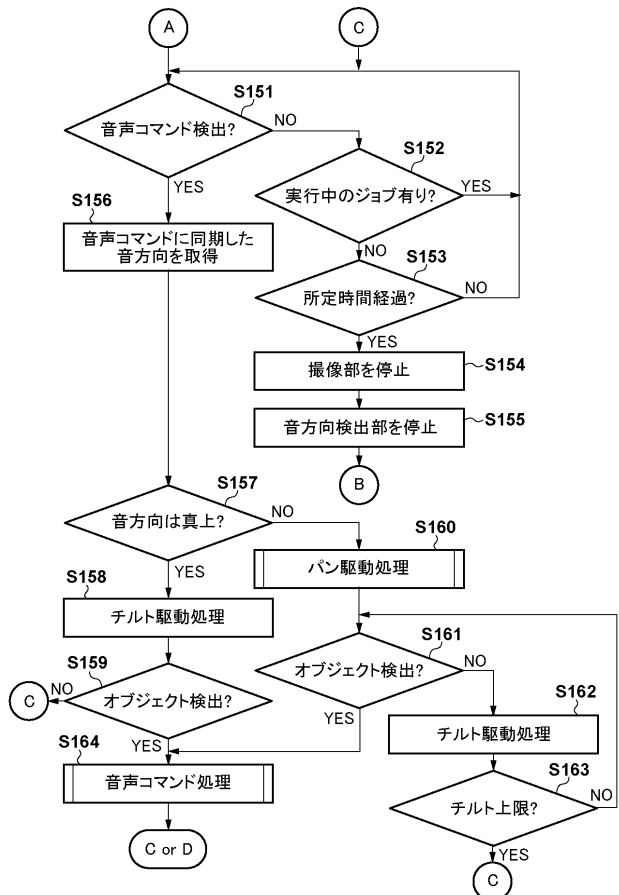
【図 4】



【図 5 A】

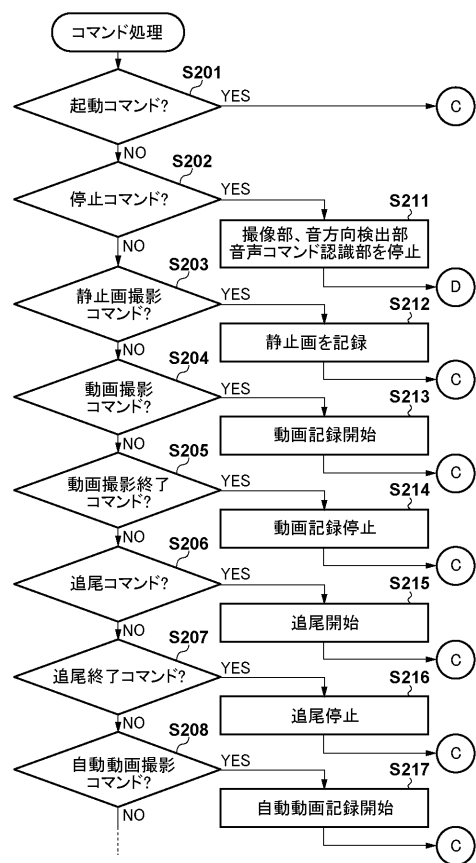


【図 5 B】





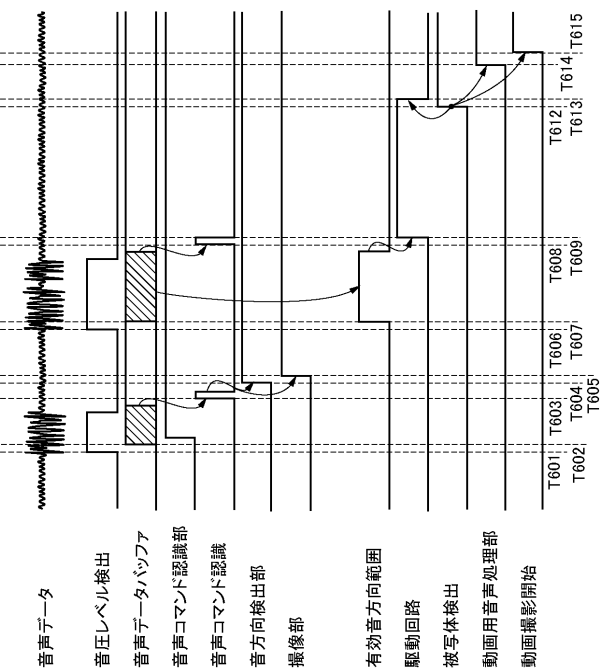
【 図 6 】



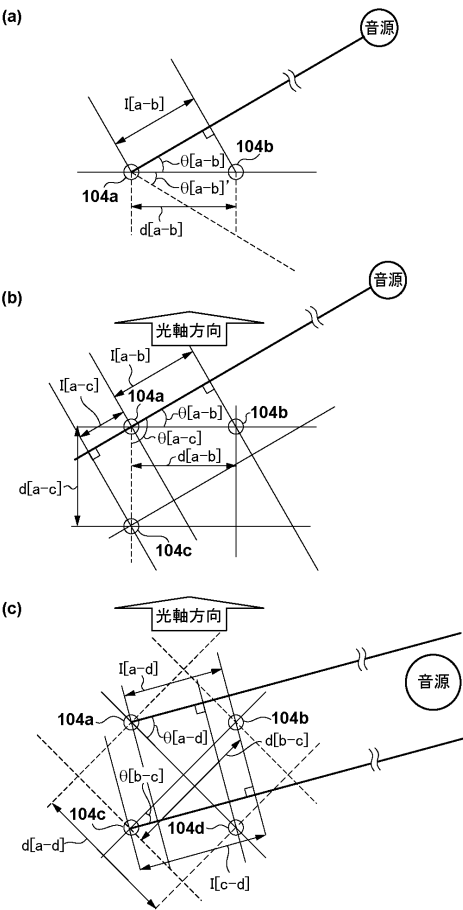
【 図 7 】

コマンド	
起動	"Hi, Camera", "Wake-up, Camera"
停止	"Good bye Camera", "ByeBye Camera"
静止画撮影	"Snap me", "Take a picture"
動画撮影開始	"Start movie", "Take a video"
動画撮影終了	"Stop movie"
追尾開始	"Follow me", "Track me"
追尾終了	"Stop Following", "Stop tracking"
⋮	
自動動画撮影開始	"Start Auto movie"
⋮	
拡大コマンド	"Zoom In"
縮小コマンド	"Zoom Out"

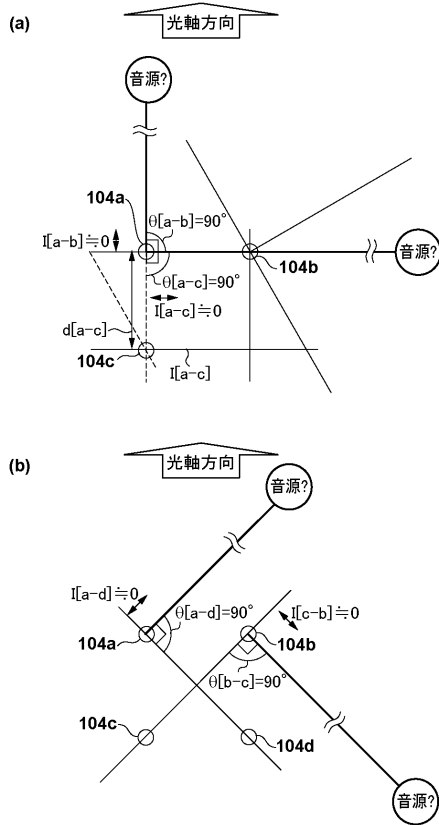
【 図 8 】



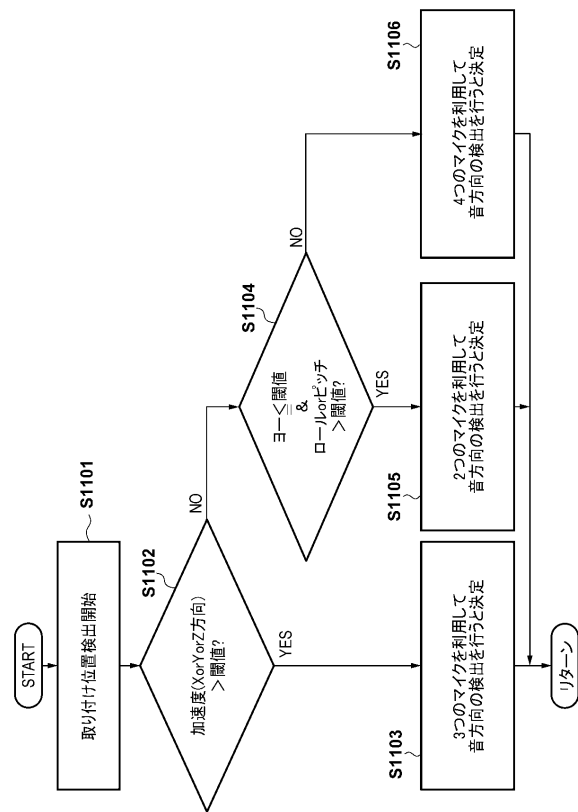
【 図 9 】



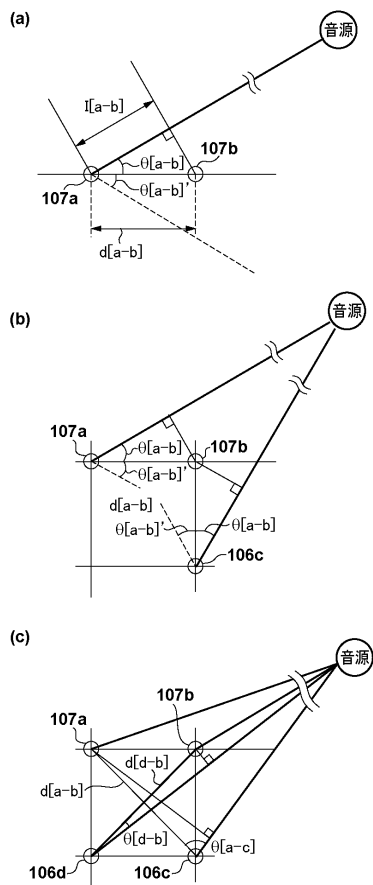
【図 10】



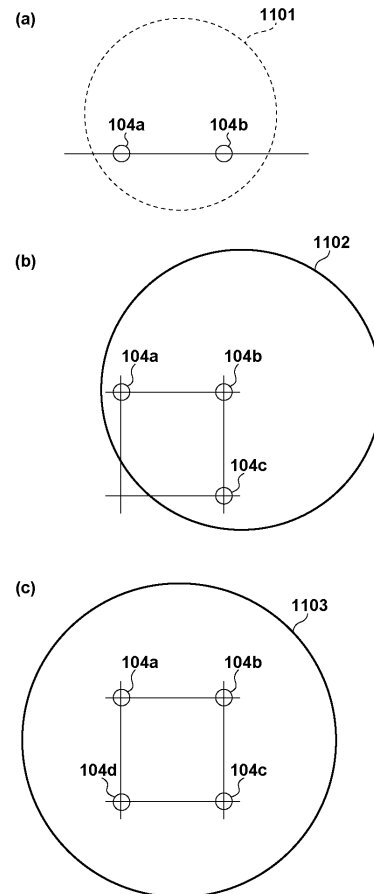
【図 11】



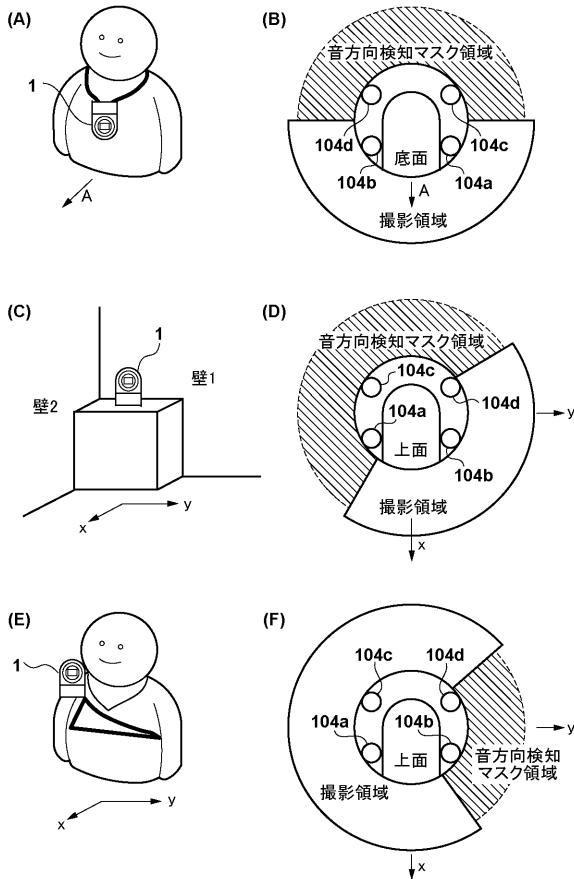
【図 12】



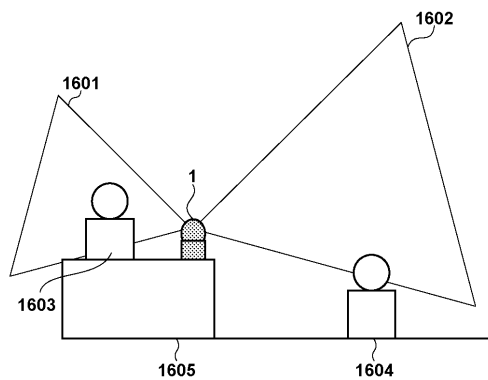
【図 13】



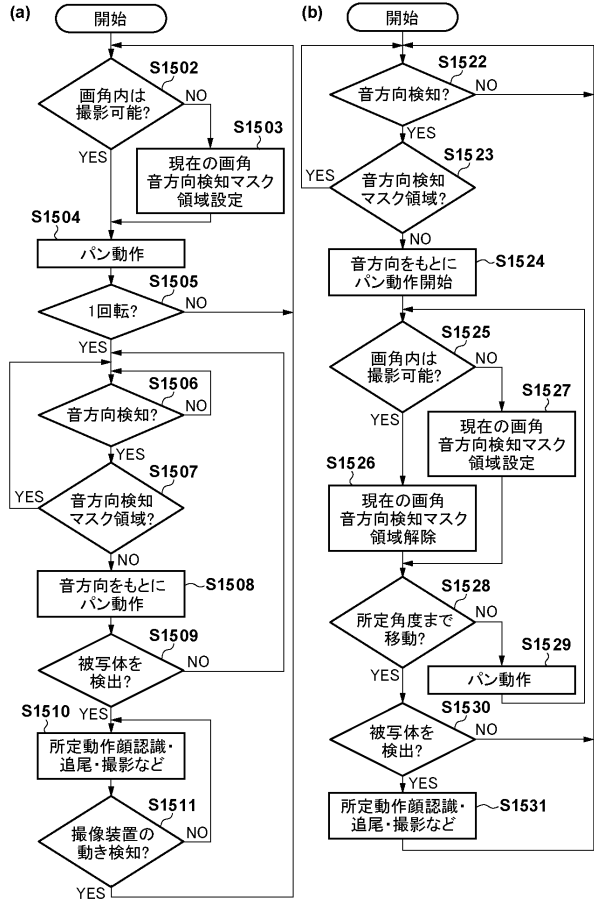
【図 14】



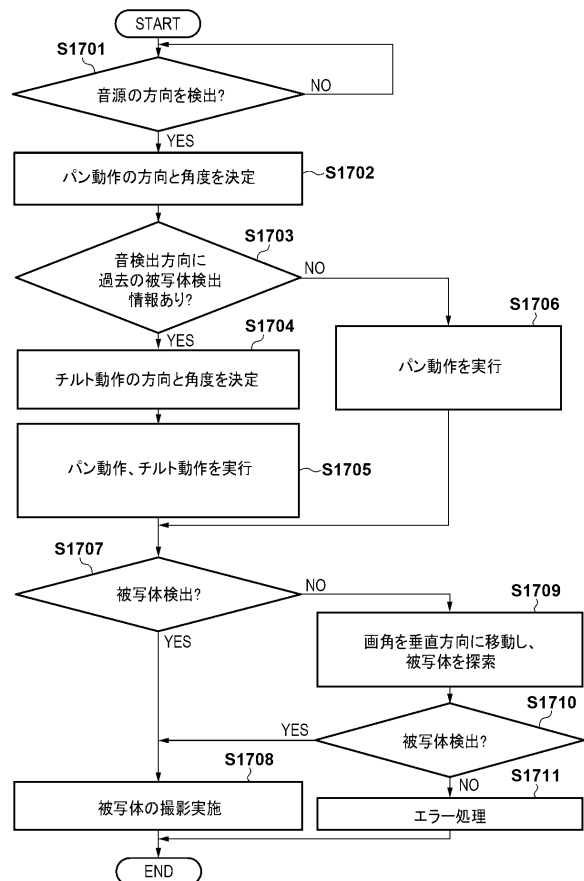
【図 16】



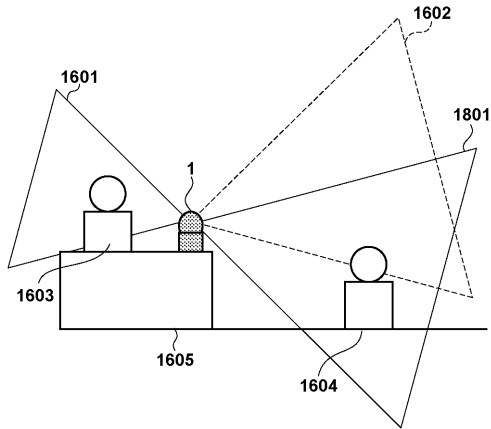
【図 15】



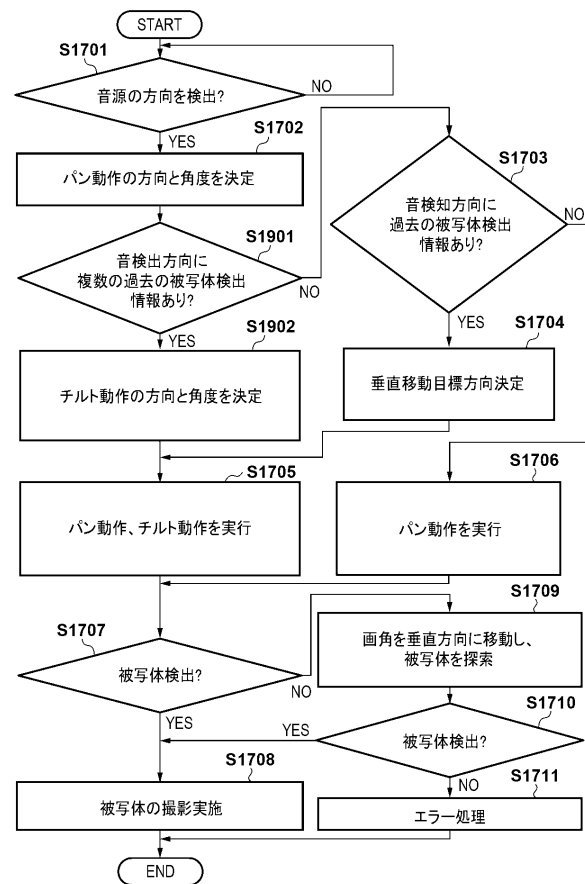
【図 17】



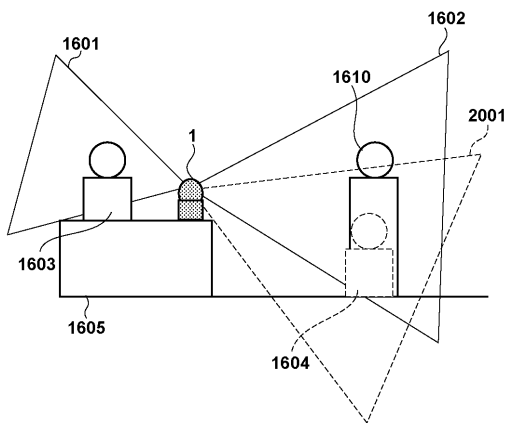
【図 18】



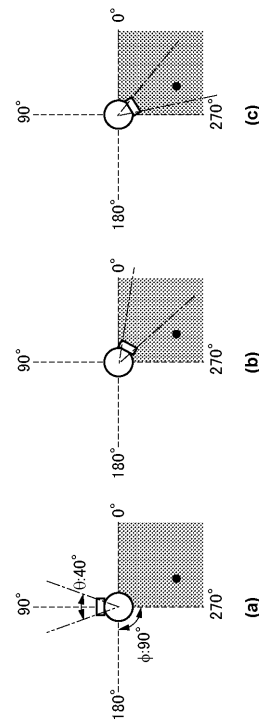
【図 19】



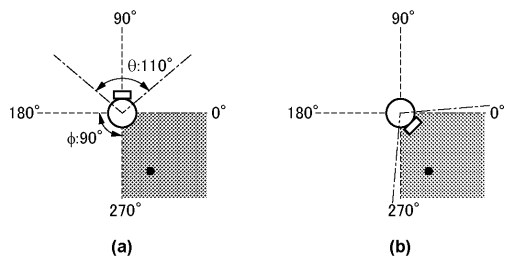
【図 20】



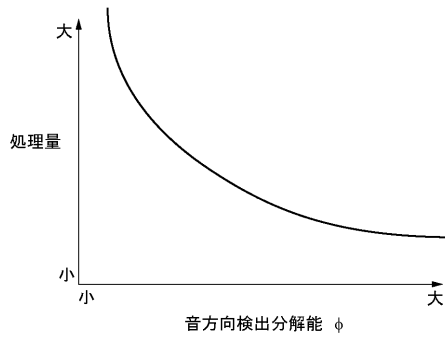
【図 22】



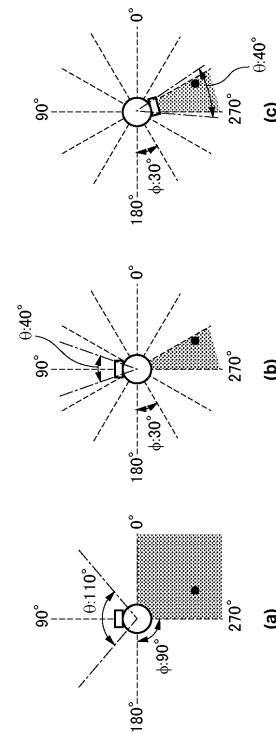
【図 21】



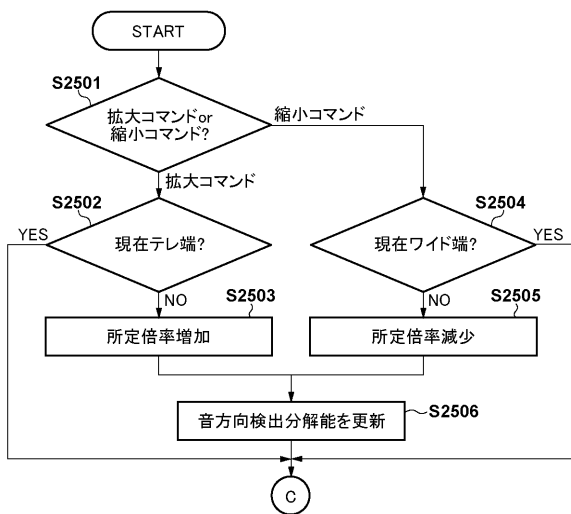
【図 2 3】



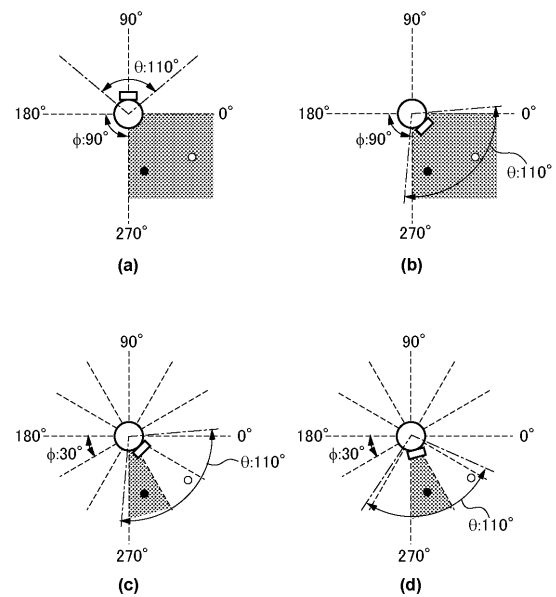
【図 2 4】



【図 2 5】



【図 2 6】



## フロントページの続き

(51)Int.Cl.	F I			テーマコード (参考)		
<b>G 1 0 L 25/51 (2013.01)</b>	H 0 4 N	5/232	9 9 0			
<b>G 1 0 L 15/10 (2006.01)</b>	G 1 0 L	15/00	2 0 0 G			
<b>H 0 4 R 3/00 (2006.01)</b>	G 1 0 L	25/51	4 0 0			
<b>H 0 4 R 1/40 (2006.01)</b>	G 1 0 L	15/10	2 0 0 W			
	H 0 4 R	3/00	3 2 0			
	H 0 4 R	1/40	3 2 0 A			

(72)発明者 鳥海 祐介  
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(72)発明者 風間 規久夫  
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(72)発明者 佐藤 龍介  
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(72)発明者 辻本 悠貴  
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

Fターム(参考) 2H020 FB05 MD15 MD17  
2H105 AA14 EE34  
5C122 EA42 EA69 FH09 FH11 FH14 FJ01 FJ09 FJ14 HA78 HA82  
HA90  
5D018 BB22  
5D220 BA06 BC05