

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7292853号

(P7292853)

(45)発行日 令和5年6月19日(2023.6.19)

(24)登録日 令和5年6月9日(2023.6.9)

(51)国際特許分類

F I

G 0 3 B 17/38 (2021.01)

G 0 3 B 17/38

B

G 0 3 B 15/00 (2021.01)

G 0 3 B 15/00

R

G 0 3 B 17/56 (2021.01)

G 0 3 B 15/00

P

H 0 4 N 23/60 (2023.01)

G 0 3 B 15/00

D

H 0 4 N 23/65 (2023.01)

G 0 3 B 15/00

Q

請求項の数 12 (全25頁) 最終頁に続く

(21)出願番号 特願2018-207633(P2018-207633)

(22)出願日 平成30年11月2日(2018.11.2)

(65)公開番号 特開2019-117374(P2019-117374

A)

(43)公開日 令和1年7月18日(2019.7.18)

審査請求日 令和3年10月25日(2021.10.25)

(31)優先権主張番号 特願2017-250106(P2017-250106)

(32)優先日 平成29年12月26日(2017.12.26)

(33)優先権主張国・地域又は機関

日本国(JP)

(73)特許権者 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(74)代理人 110003281

弁理士法人大塚国際特許事務所

(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

(72)発明者 飯田 吉信

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

審査官 登丸 久寿

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)【特許請求の範囲】

【請求項1】

撮像手段と、

音声を集音する集音手段と、

前記集音手段によって集音した音声が所定の音圧より大きいか否かを検出する検出手段と、

前記集音手段によって集音した音声、前記撮像手段による撮影を指示する音声であることを認識する認識手段と、

前記集音手段によって集音した音声、前記検出手段によって前記所定の音圧より大きいと検出され、かつ前記認識手段によって音声コマンドが認識された場合、当該認識された音声コマンドに応じて前記撮像手段を制御する制御手段とを有し、

前記認識手段は、前記集音手段によって集音した音声、前記所定の音圧より大きいと前記検出手段により検出されたことに応じて起動し、

前記制御手段は、

前記認識手段が起動に関する音声コマンドを認識した場合、前記撮像手段を起動することで撮像可能な状態に移行させ、

前記撮像手段が前記撮像可能な状態で、前記認識手段が撮像に関する音声コマンドを認識した場合、当該音声コマンドに応じて前記撮像手段による撮像に係る制御を実行する

ことを特徴とする撮像装置。

【請求項2】

10

20

前記撮像手段をパン動作およびチルト動作させる駆動手段をさらに有し、
前記集音手段は複数のマイクで構成され、
前記複数のマイクは前記駆動手段によるパン動作の回転軸に対して垂直な平面に配置される

ことを特徴とする請求項 1 に記載の撮像装置。

【請求項 3】

前記複数のマイクは第一のマイクおよび第二のマイクを含み、
前記撮像手段の光軸を含む、前記撮像装置の底面に対して垂直な平面で前記撮像装置を二つの領域に分けた場合、一方に前記第一のマイクが配置され、他方に前記第二のマイクが配置される

10

ことを特徴とする請求項 2 に記載の撮像装置。

【請求項 4】

前記第一のマイクの配置および第二のマイクの配置は前記撮像手段の光軸に関して対称である

ことを特徴とする請求項 3 に記載の撮像装置。

【請求項 5】

前記複数のマイクは、前記第一のマイクおよび前記第二のマイクを結ぶ直線上ではない位置に配置される第三のマイクをさらに含み、

前記第一のマイクおよび前記第二のマイクを用いて検出した前記音声の音源の方向と、
前記第三のマイクおよび前記第三のマイク以外のマイクを用いて検出した前記音源の方向と、に基づいて、前記音声の音源の方向を検出する方向検出手段を有する

20

ことを特徴とする請求項 3 または 4 に記載の撮像装置。

【請求項 6】

前記方向検出手段は、前記第一のマイクおよび前記第二のマイクで構成される第一のペアを用いて検出した前記音声の音源の方向が前記第一のマイクの位置および前記第二のマイクの位置を結ぶ直線に対して垂直であると判断し、かつ前記第三のマイクおよび前記第三のマイク以外のマイクで構成される第二のペアを用いて検出した前記音声の音源の方向が前記第三のマイクの位置および前記第三のマイク以外の前記マイクの位置を結ぶ直線に対して垂直であると判断した場合、前記音源は前記回転軸の軸方向にあると判断する

ことを特徴とする請求項 5 に記載の撮像装置。

30

【請求項 7】

前記方向検出手段は、前記第一のペアの 2 つのマイクの前記音声を集音した時間差から求めた前記音声の音源の方向と、前記第二のペアの 2 つのマイクの前記音声を集音した時間差から求めた前記音声の音源の方向と、から前記回転軸に垂直な面における前記音声の音源の方向を検出する

ことを特徴とする請求項 6 に記載の撮像装置。

【請求項 8】

前記検出手段は、前記複数のマイクのうち 1 つのマイクを用いて、前記音声が所定の音圧を超えたことを検出し、

前記方向検出手段は、前記複数のマイクのうち、すべてのマイクを用いて前記音声の音源の方向を検出することを特徴とする

40

請求項 5 から 7 のいずれか 1 項に記載の撮像装置。

【請求項 9】

前記駆動手段によってパン動作する距離を二つの区間に分けた場合、その前半の区間では第一の速度で前記撮像手段をパン動作させ、その後半の区間では前記第一の速度から減速された第二の速度で前記撮像手段をパン動作させるよう、前記制御手段は前記駆動手段を制御し、

前記撮像手段は、前記前半の期間ではフォーカシングを行わず、前記後半の期間ではフォーカシングを行う

ことを特徴とする請求項 2 から 8 のいずれか 1 項に記載の撮像装置。

50

【請求項 1 0】

前記撮像手段による撮影を指示する音声に応じて、前記制御手段は前記撮像手段の撮像方向を被写体に追尾するように前記駆動手段を制御し、

前記撮像手段は、静止画の撮影及び記録、または音声付の動画像の撮影及び記録することを特徴とする請求項 2 から 9 のいずれか 1 項に記載の撮像装置。

【請求項 1 1】

撮像手段を有する撮像装置の制御方法であって、

音声を集音する集音ステップと、

前記集音ステップで集音した音声が所定の音圧より大きいと否かを検出する検出ステップと、

前記集音ステップで集音した音声、前記撮像手段による撮影を指示する音声であることを認識する認識ステップと、

前記集音ステップで集音した音声、前記検出ステップで前記所定の音圧より大きいと検出され、かつ前記認識ステップで音声コマンドが認識された場合、当該認識された音声コマンドに応じて前記撮像手段を制御する制御ステップとを有し、

前記認識ステップは、前記集音ステップにおいて集音した音声、前記検出ステップにより前記所定の音圧より大きいと検出されたことに応じて実行され、

前記制御ステップでは、

前記認識ステップにおいて起動に関する音声コマンドが認識された場合、前記撮像手段を起動することで撮像可能な状態に移行させ、

前記撮像手段が前記撮像可能な状態で、前記認識ステップにおいて撮像に関する音声コマンドを認識した場合、当該音声コマンドに応じて前記撮像手段による撮像に係る制御を実行する

ことを特徴とする撮像装置の制御方法。

【請求項 1 2】

コンピュータを請求項 1 から 1 0 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。

【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】**【特許文献】****【0 0 0 4】**

【文献】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 】

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】

【 0 0 0 7 】

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、
撮像手段と、
音声を集音する集音手段と、
前記集音手段によって集音した音声が所定の音圧より大きいか否かを検出する検出手段と、

10

前記集音手段によって集音した音声、前記撮像手段による撮影を指示する音声であることを認識する認識手段と、

前記集音手段によって集音した音声、前記検出手段によって前記所定の音圧より大きいと検出され、かつ前記認識手段によって音声コマンドが認識された場合、当該認識された音声コマンドに応じて前記撮像手段を制御する制御手段とを有し、

前記認識手段は、前記集音手段によって集音した音声、前記所定の音圧より大きいと前記検出手段により検出されたことに応じて起動し、

20

前記制御手段は、
前記認識手段が起動に関する音声コマンドを認識した場合、前記撮像手段を起動することで撮像可能な状態に移行させ、
前記撮像手段が前記撮像可能な状態で、前記認識手段が撮像に関する音声コマンドを認識した場合、当該音声コマンドに応じて前記撮像手段による撮像に係る制御を実行することを特徴とする。

【発明の効果】

【 0 0 0 8 】

本発明によれば、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ことが可能になる。

30

【図面の簡単な説明】

【 0 0 0 9 】

【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 5 A】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 5 B】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 5 B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ート。

40

【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 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 12】第 2 の実施形態における撮像装置の動作を説明するための図。
【発明を実施するための形態】

【 0 0 1 0 】

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。

50

【 0 0 1 1 】

〔 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 1 0 0、及び、可動撮像部 1 0 0 の駆動制御および、撮像装置全体を制御する中央制御部（C P U）を含む支持部 2 0 0 で構成される。

【 0 0 1 2 】

なお、支持部 2 0 0 は、圧電素子を含む複数の振動体 1 1 ~ 1 3 が可動撮像部 1 0 0 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 1 1 ~ 1 3 の振動の制御により、可動撮像部 1 0 0 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。

10

【 0 0 1 3 】

可動撮像部 1 0 0 は、レンズ部 1 0 1、撮像部 1 0 2、レンズアクチュエータ制御部 1 0 3、及び、音声入力部 1 0 4 を有する。

【 0 0 1 4 】

レンズ部 1 0 1 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 1 0 2 は、C M O S センサや C C D センサなどの撮像素子を含み、レンズ部 1 0 1 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 1 0 3 は、モータドライバ I C を含み、レンズ部 1 0 1 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 2 0 0 内の中央制御部 2 0 1 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 1 0 4 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。

20

【 0 0 1 5 】

一方、支持部 2 0 0 は、撮像装置 1 の全体の制御を行うための中央制御部 2 0 1 を有する。この中央制御部 2 0 1 は、C P U と、C P U が実行するプログラムを格納した R O M、及び、C P U のワークエリアとして使用される R A M で構成される。また、支持部 2 0 0 は、撮像信号処理部 2 0 2、映像信号処理部 2 0 3、音声信号処理部 2 0 4、操作部 2 0 5、記憶部 2 0 6、表示部 2 0 7 を有する。更に、支持部 2 0 0 は、入出力端子部 2 0 8、音声再生部 2 0 9、電源部 2 1 0、電源制御部 2 1 1、位置検出部 2 1 2、回動制御部 2 1 3、無線通信部 2 1 4、並びに、先に説明した振動体 1 1 ~ 1 3 を有する。

30

【 0 0 1 6 】

撮像信号処理部 2 0 2 は、可動撮像部 1 0 0 の撮像部 1 0 2 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 2 0 3 は、撮像信号処理部 2 0 2 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。

【 0 0 1 7 】

音声信号処理部 2 0 4 は、音声入力部 1 0 4 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 1 0 4 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 2 0 4 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 1 0 4 を含めた音声信号処理部 2 0 4 の詳細については図 2 を用いて後述する。

40

【 0 0 1 8 】

操作部 2 0 5 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 2 0 6 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 2 0 7 は、L C D などのディスプレイを備え、映像信号処理部 2 0 3 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 2 0 7 は、各種メニュー等を表示することで、ユーザインターフ

50

エ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 はスピーカーを含み、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。

【0019】

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。

10

【0020】

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。

【0021】

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。

20

【0022】

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、音声信号処理部 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。

【0023】

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。

30

【0024】

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。

40

【0025】

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。

【0026】

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。

【0027】

50

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。

【0028】

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 16 kHz 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。

10

【0029】

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。

20

【0030】

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、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。

30

【0031】

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て構わない。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。

40

【0032】

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。

【0033】

50

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。

【0034】

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が示す水平面に於て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。

【0035】

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)からもわかるように、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。

【0036】

なお、本実施形態における4つのマイク104a、104b、104c、104d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。

【0037】

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。

【0038】

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

10

20

30

40

50

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。

【 0 0 3 9 】

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。

【 0 0 4 0 】

図 4 (a) はレンズ部 1 0 1 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 1 5 0 を、上方向から見て反時計回りに 9 0 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 1 5 1 の 9 0 度チルト動作をおこなうと、図 4 (c) の様になる。第 1 筐体 1 5 0、第 2 筐体 1 5 1 の回動は、先に説明したように、回動制御部 2 1 3 により駆動される振動体 1 1 ~ 1 3 による振動にて実現している。

10

【 0 0 4 1 】

次に、本実施形態における撮像装置 1 の中央制御部 2 0 1 の処理手順を図 5 A , 5 B 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 2 0 1 の処理を示している。

【 0 0 4 2 】

中央制御部 2 0 1 は、ステップ S 1 0 1 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 2 0 1 は、現在の可動撮像部 1 0 0 の撮像部 1 0 2 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。

20

【 0 0 4 3 】

これ以降、可動撮像部 1 0 0 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、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 2 0 4 4 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 2 0 4 4 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。

【 0 0 4 4 】

なお、この段階で、音声用メモリ 2 0 4 2、音方向検出部 2 0 4 4、動画用音声処理部 2 0 4 5、並び、マイク 1 0 4 b 乃至 1 0 4 d への電力は遮断されている。

【 0 0 4 5 】

30

初期化処理を終えると中央制御部 2 0 1 は、ステップ S 1 0 2 にて、電源制御部 2 1 1 を制御して、音圧レベル検出部 2 0 4 1、マイク 1 0 4 a への電力の供給を開始する。この結果、音圧レベル検出部 2 0 4 1 は、マイク 1 0 4 a 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、この音声の音圧レベル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 2 0 1 に通知する。なお、この閾値は、例えば 6 0 dB 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。

【 0 0 4 6 】

中央制御部 2 0 1 は、ステップ S 1 0 3 にて、音圧レベル検出部 2 0 4 1 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を検出されると、ステップ S 1 0 4 にて、音声メモリ 2 0 4 2 はマイク 1 0 4 a からの音声データの受信、格納処理を開始する。

40

【 0 0 4 7 】

また、ステップ S 1 0 5 にて、中央制御部 2 0 1 は、電源制御部 2 1 1 を制御し、音声コマンド認識部 2 0 4 3 への電力供給を開始する。この結果、音声コマンド認識部 2 0 4 3 は、コマンドメモリ 2 0 4 6 を参照した音声用メモリ 2 0 4 2 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 2 0 4 3 は、音声用メモリ 2 0 4 2 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 2 0 4 6 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

50

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。

【0048】

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。

10

【0049】

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。

20

【0050】

ステップ S110 にて、中央制御部 201 は、電源制御部 211 を制御し、音方向検出部 2044、マイク 104b 乃至 104d への電力供給を開始する。この結果、音方向検出部 2044 は、4 つのマイク 104a 乃至 104d 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 2044 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 2044a に格納していく。このとき、音方向検出部 2044 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 2042 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 2044a に格納する。典型的には、バッファメモリ 2044a に格納するのは、音方向と、音声メモリ 2042 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。

30

【0051】

ステップ S111 にて、中央制御部 201 は、電源制御部 211 を制御し、撮像部 102、及び、レンズアクチュエータ制御部 103 への電力供給を開始する。この結果、可動撮像部 100 は、撮像装置として機能し始める。

40

【0052】

次に、ステップ S151 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 S152 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。

【0053】

ステップ S153 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 201 は処理をステップ S1

50

51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201は処理をステップS154に進める。このステップS154にて、中央制御部201は、電源制御部211を制御し、撮像部102、レンズアクチュエータ103への電力を遮断する。そして、中央制御部201は、ステップS155にて、電源制御部211を制御し、音方向検出部2044への電力も遮断し、処理をステップS106に戻す。

【0054】

さて、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とする。この場合、音声コマンド認識部2043は、処理をステップS151からステップS156に進める。

10

【0055】

本実施形態における中央制御部201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100の撮像部102の視野内に入れる処理を行う。そして、撮像部102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。

【0056】

上記を実現するため、中央制御部201は、ステップS156にて、音声コマンド認識部2043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2044のバッファメモリ2044aから取得する。音声コマンド認識部2043は、音声コマンドを認識したとき、音声用メモリ2042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201に通知する。そこで、中央制御部201は、この2つのアド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。

20

【0057】

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。

30

【0058】

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。

40

【0059】

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

50

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 S 163 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 S 161 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 S 163 からステップ S 151 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 S 164 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。

10

【0060】

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 S 164 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 5 B のステップ S 151 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。

【0061】

まず、ステップ S 201 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。

20

【0062】

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 107 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 201 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 151 に戻す。

【0063】

ステップ S 202 にて、中央制御部 201 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 201 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 211 に進める。ステップ S 211 にて、中央制御部 201 は、電源制御部 211 を制御し、既に起動している撮像部 102、音方向検出部 2044、音声コマンド認識部 2043、動画用音声処理部 2045、マイク 104b 乃至 104d 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 201 は、処理を起動時のステップ S 103 に戻す。

30

【0064】

ステップ S 203 にて、中央制御部 201 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 201 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 212 に進める。ステップ S 212 にて、中央制御部 201 は、撮像部 102 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 206 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 152 で判定する対象のジョブとはならない。

40

【0065】

ステップ S 204 にて、中央制御部 201 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 に対して動画の撮像と記録を要求するコマンドである。中央制御部 201 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 213 に進める。このステップ S 213 にて、中央制御部 201 は、撮像部 102 を用いて動画の撮影と記録を開始し、処理をステップ S 151

50

に戻す。本実施形態では、撮像した動画画は記憶部 206 に格納されるものとするが、外部入出力端子部 208 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 152 で判定する対象のジョブとなる。

【0066】

ステップ S 205 にて、中央制御部 201 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 201 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 201 は処理をステップ S 151 に戻す。

【0067】

ステップ S 206 にて、中央制御部 201 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 102 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 201 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 214 に進める。そして、ステップ S 214 にて、中央制御部 201 は、映像信号処理部 203 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 213 の制御を開始する。そして、中央制御部 201 は処理をステップ S 151 に戻す。この結果、可動撮像部 100 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 5B のステップ S 152 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 201 はこの動画画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。

【0068】

ステップ S 207 にて、中央制御部 201 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 201 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 201 は処理をステップ S 151 に戻す。

【0069】

ステップ S 208 にて、中央制御部 201 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 201 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 217 に進める。このステップ S 217 にて、中央制御部 201 は、撮像部 102 による動画画の撮影と記録を開始し、処理をステップ S 151 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 101 の撮像方向を向けつつ動画画を撮影、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 101 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、動画画を記録する。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 205 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 201 は、音声コマンド認識部 2043 を停止させる。そして、中央制御部 201 は、音圧レベル検出部 2041 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 2044 が検出した音方向情報を参照して、可動撮像部 104 のパン動作、チルト動作を行う。

【0070】

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 201 はレンズアクチュエータ制御部 103 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 201 はレンズアクチュエータ制御部 103 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 101 が既にテレ端、或いは、ワイ

10

20

30

40

50

ド端にあるとき、それを越えた拡大率、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 201 はその音声コマンドを無視する。

【0071】

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 S207 以降で実行されるが、ここでの説明は省略する。

【0072】

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。

【0073】

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 2041 はマイク 1014 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 T601 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 2041 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 T602 にて、音声用メモリ 2042 がマイク 104 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 2043 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 T603 にて、音声コマンド認識部 2043 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。

【0074】

中央制御部 201 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 T603 にて音方向検出部 2044 に電力供給を開始する。また、中央制御部 201 は、タイミング T604 にて撮像部 102 への電力供給も開始する。

【0075】

ユーザは、タイミング T606 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 T607 から順に音声用メモリ 2042 に格納されていく。そして、タイミング T608 にて、音声コマンド認識部 2043 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 2043 は、音声用メモリ 2042 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 201 に通知する。中央制御部 201 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 201 は、音方向検出部 2044 のバッファ 2044 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 T609 にて、その抽出した情報に基づいて、駆動制御部 213 を制御して、可動撮像部 100 のパン動作、チルト動作を開始する。

【0076】

可動撮像部 100 のパン動作、チルト動作中に、タイミング T612 にて、撮像信号処理部 202 が、撮像部 102 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 201 はパン動作、チルト動作を停止する（タイミング T613）。また、タイミング T614 にて、中央制御部 201 は、動画用音声処理部 2045 に電力を供給して、マイク 104 a、及び、104 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 201 は、タイミング T615 にて、音声付動画像の撮像と記録を開始する。

【0077】

次に、本実施形態における音方向検出部 2044 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 S110 以降、周期的に、且つ、継続的に行われるものである。

【0078】

まず、図 9 (a) を用いて、マイク 104 a とマイク 104 b の 2 つのマイクを用いた簡易の音方向検出を説明する。同図は、マイク 104 a とマイク 104 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 104 a とマイク 104 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

10

20

30

40

50

は十分に大きいと仮定する。この場合、マイク 104 a とマイク 104 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。

【0079】

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

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

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 $[a, b]$ （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。

【0080】

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 100 の撮像方向とする。可動撮像部 100 の撮像方向は、レンズ部 101 の光軸方向（主軸方向）とも言い換えられる。

【0081】

図9（b）は3つのマイクで行う方式である。マイク 104 a、マイク 104 b、マイク 104 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 104 a、マイク 104 b の並ぶ方向に直交する方向がレンズ部 101 の撮像方向となる。

【0082】

図9（a）で説明したように、マイク 104 a、マイク 104 b 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 104 a、マイク 104 c 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 104 a、104 b、104 c 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、それらに共通な方向が、正確な音声発生方向として決定できる。

【0083】

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 104 a、マイク 104 b、マイク 104 c、マイク 104 d の配置により、マイク 104 a、マイク 104 b 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 104 a と 104 d のペアと、マイク 104 b とマイク 104 c 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。

【0084】

マイク 104 a、マイク 104 d 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。

【0085】

更にマイク 104 b、マイク 104 c 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。

【0086】

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。

【0087】

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。

【0088】

以上のように処理を行うため、マイク 104 a とマイク 104 b とマイク 104 c および

10

20

30

40

50

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 が 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。

【0089】

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。

【0090】

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。

10

【0091】

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、すなわち上方向からマイク 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。

【0092】

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。

20

【0093】

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。

【0094】

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 T_1 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 T_2 とし、予め設定された十分に小さい閾値 θ 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。

30

条件: $T_1 < \theta$ かつ $T_2 < \theta$

【0095】

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。

【0096】

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 T_3 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。

40

【0097】

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 T_4 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: $T_3 < \theta$ 且つ $T_4 < \theta$

50

【 0 0 9 8 】

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。

【 0 0 9 9 】

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。

【 0 1 0 0 】

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。

【 0 1 0 1 】

〔 第2の実施形態 〕

第2の実施形態を説明する。本第2の実施形態における装置構成は第1の実施形態と同じとする。また、音声コマンド認識部2043が音声コマンドとして動画撮影コマンドを既に認識し、且つ、音方向検出部2044がその音声コマンドを発声したユーザの方向も既に検出しているものとして説明する。

【 0 1 0 2 】

上記第1の実施形態では、撮像装置1は音声コマンドとして動画撮影コマンドを認識した場合、その音声コマンドの発生源の存在する方向と、レンズ部101の撮像方向（光軸方向）とを一致させるべく、可動撮像部100のパン動作、チルト動作を行った。そして、撮像装置1はこのパン動作、チルト動作にて、被写体検出処理を行った。そして、最終的に、撮像装置1は被写体を撮像部102による撮像画像内に被写体が存在することを条件に、動画の撮影記録を開始した。

【 0 1 0 3 】

従って、ユーザは、動画撮影コマンドを発声したとしても、そのジョブが実行されるまで多少の時間だけ待つ必要がある。また、パン動作中は、撮像部102から得られる画像は水平方向に流れ、レンズ部101によるフォーカシングが難しいという問題もある。

【 0 1 0 4 】

そこで、本第2の実施形態では、撮像部102による上記待ち時間を短くしつつ、いわゆるピンボケの状態でのジョブを開始することを抑制する例を説明する。具体的には、音方向検出部2044で検出した音方向（以下、目標方向という）に向かうべく、中央制御部201は回動制御部性213による高速な第1のパン動作速度でパン動作を行う。この第1のパン動作速度で駆動を行う期間は、例えば、パン動作を開始する直前のレンズ部101の撮像方向（光軸方向）と目標方向との角度差の半分になるまでとし、撮像装置1はこの期間（第1のパン動作速度で駆動を行う期間）でのフォーカシング制御は行わない。そして、レンズ部101の撮像方向（光軸方向）と目標方向との角度差が、音声を認識した時点の角度の半分の角度になった場合、中央制御部201は、第1のパン動作速度よりも遅く、且つ、フォーカシングが可能な第2のパン動作速度となるよう回動制御部213の制御を行う。そして、中央制御部201は、第2のパン動作速度の期間中に、レンズアクチュエータ103を制御して、レンズ部101のフォーカシング制御を開始する。そして、中央制御部201は次の条件を満たした場合、動画の撮影記録を開始する。

・条件：撮像部102による画像が合焦状態、且つ、レンズ部101の撮像方向（光軸方向）と目標方向との角度差が予め設定された閾値 以下。

ここで閾値 は、撮像部102が撮像可能な画角の左右端とその中心との成す角度とする。閾値 以下になれば、被写体が撮像部102による撮像画像内に入り始めると推定できるからである。ここで、第2のパン動作速度 v_2 は初期速度である第1のパン動作速度 v_1

10

20

30

40

50

1 から所定の加速度で減速する速度である。

【0105】

図11は、本第2の実施形態における音声コマンド認識部2043による音声コマンド認識をトリガにして開始される処理を示すフローチャートである。このフローチャートに係るプログラムは、中央制御部201が有するROMに格納されているものである。図12は、本第2の実施形態における音声コマンド認識後の処理のシーケンスを示す図である。以下、これらの図を参照して、第2の実施形態における中央制御部201の処理を説明する。また、ここでは音声コマンドとして動画像撮影コマンドが認識されたものとして説明する。

【0106】

まず、ステップS1001にて、中央制御部201は、音声コマンド認識部2043から、認識した音声コマンドの元になった音声データの先頭と終端のアドレスが示す有効期間内（発声期間内でもある）で検出した音方向情報を、音方向検出部2044のバッファメモリ2044aから取得する。そして、ステップS1002にて、中央制御部201は、取得した音方向情報が示す方向が、現在の可動撮像部100の撮像部102の画角内にあるか否かを判定する。

【0107】

音方向情報が示す方向が現在の撮像部102の画角内にある場合、中央制御部201は、処理をステップS1003に進め、認識した動画像の撮影記録を開始する。

【0108】

一方、音方向情報が示す方向が現在の可動撮像部100の画角外にある場合、中央制御部201は、処理をステップS1004に進める。このステップS1004にて、中央制御部201は、現在のレンズ部101の撮像方向（光軸方向）Aと、S1001で取得した音源方向Cとからパン動作を行う角度を算出し、撮像方向Aからパン動作を行う角度がその半分の角度（前半の区間）では第1のパン速度 v_1 で動作させ、残りの半分の角度（後半の区間）を所定の加速度でパン動作を減速した場合に要する時間（ $= t_4 - t_0$ ）を算出する。

【0109】

レンズ部101は、PI（Photo Interrupter）により、発光部からの光を物体が遮るのを受光部で検出することによって、位置を検出する方法を用いていることを前提に説明する。この場合、現在のレンズ部101の方向をA、目標方向をC、レンズ部101の方向Aと目標方向Cのなす角度の中間の方向をBとする。レンズ部101の方向Aと目標方向Cのなす角度に応じてPIのステップ数PIACが決まる。このステップ数PIACから、第1のパン動作速度で駆動する前半の区間の距離DABおよび第2のパン動作速度で駆動する後半の区間の距離DBCを中央制御部201は算出する。前半の区間を第1のパン動作速度で駆動する期間（ $t_1 - t_0$ ）は、 DAB / v_1 として求めることができる。また、後半の区間を第2のパン動作速度 v_2 で駆動する期間（ $t_4 - t_1$ ）は、 DBC / v_2 （ただし、本実施形態では $DBC = DAB$ とする）となる。ここで前述したように、第2のパン動作速度 v_2 は初期速度である第1のパン動作速度 v_1 から所定の加速度で減速する速度である。

【0110】

ステップS1006にて、中央制御部201は、回動制御部213を制御し、可動撮像部100の第1のパン動作速度 v_1 で目標とする方向に向けてパン動作を開始する。そして、ステップS1007にて、中央制御部201は、パン動作の時間が DAB / v_1 となったと判定されるまで、このパン動作を継続する。

【0111】

そして、パン動作を開始してからの時間が DAB / v_1 になった場合、中央制御部201は、ステップS1008にて、パン動作が終了する時刻 t_4 まで、パン動作速度を第2のパン動作速度 v_2 に設定する。そして、ステップS1009にて、中央制御部201は、現在のレンズ部101の撮像方向（光軸方向）と目標方向との成す角度が予め設定され

10

20

30

40

50

た閾値 以下であるか否かを判定する。閾値 以下になると、被写体が画角内に入っている可能性があるので、ステップ S 1 0 1 0 にて、中央制御部 2 0 1 は、フォーカス状態値 A F を求める。そして、ステップ S 1 0 1 1 にて、中央制御部 2 0 1 は、求めたフォーカス状態値 A F が所定閾値 A F t h 以内であることを判定する。フォーカス状態値 A F が所定閾値 A F t h 以内であるとき、合焦状態である。そのため、中央制御部 2 0 1 はパン動作が完了していなくても、認識した動画撮影コマンドに基づくジョブを開始すべく処理をステップ S 1 0 0 3 に進める。

【 0 1 1 2 】

一方、フォーカス状態値 A F が所定閾値 A F t h を超える場合、中央制御部 2 0 1 は処理をステップ S 1 0 1 2 に進め、当初計画したパン動作時間が経過したか否かを判定する。なお、このステップ S 1 0 1 2 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）が目標方向に到達したか否かを判定しても構わない。そして、ステップ S 1 0 1 2 の判定が否（ N o ）の場合、中央制御部 2 0 1 は処理をステップ S 1 0 0 8 に戻す。

10

【 0 1 1 3 】

一方、ステップ S 1 0 1 2 の判定が是（ Y e s ）の場合、中央制御部 2 0 1 は処理をステップ S 1 0 1 3 に進める。このステップ S 1 0 1 3 にて、中央制御部 2 0 1 は、回動制御部 2 1 3 を制御し、回動駆動部の駆動をパン動作からチルト動作に切換え、チルト動作を開始する。

【 0 1 1 4 】

そして、中央制御部 2 0 1 は、ステップ S 1 0 1 4 にて被写体の検出、もしくは、チルト上限への到達のいずれかが判定されるまで、ステップ S 1 0 1 3 のチルト動作を継続する。被写体が検出された場合（タイミング t 5 ）、中央制御部 2 0 1 は、チルト動作を止め、処理をステップ S 1 0 0 3 に進める。一方、被写体未検出のまま、チルトの上限まで到達した場合（タイミング t 6 ）、中央制御部 2 0 1 はエラー処理を行う。このエラーでは、例えば認識した静止画撮影開始コマンドを実行しない、等の処理となる。

20

【 0 1 1 5 】

以上であるが、本第 2 の実施形態に係る処理を、第 1 の実施形態に適用するのであれば、第 1 の実施形態におけるステップ S 1 6 0 以降の処理を、第 2 の実施形態におけるステップ S 1 0 0 2 以降の処理で置き換えればよい。この場合、第 2 の実施形態におけるステップ S 1 0 1 5 はステップ S 1 6 3 に対応させればよい。

30

【 0 1 1 6 】

なお、上記第 2 の実施形態では、レンズ部 1 0 1 の撮像方向（光軸方向）を目標方向に一致させるためのパン動作期間を二分し、前期間では第 1 のパン動作速度にし、後期間では、それより低速の第 2 のパン動作速度に切り換える例を説明した。しかし、第 1 のパン動作速度での駆動期間に到達したとき、目標方向でパン動作速度がゼロとなるように、一定加速度で減速するようにしてもよい。

【 0 1 1 7 】

以上より、本第 2 の実施形態によれば、特定の音声の発生位置が、現在の画角 A の外にあり、パン・チルト駆動で画角を変更して被写体を捉える必要がある場合にも、不要なパン・チルト駆動やフォーカシング後から、動画記録を開始することができる。さらに、パン動作の速度を切り替えることで、フォーカシングを行うことができ、動画記録の開始タイミングを第 1 の実施形態よりも早めることもできる。

40

【 0 1 1 8 】

上記実施形態では、音圧レベル検出部 2 0 4 1、音声コマンド認識部 2 0 4 3、音方向検出部 2 0 4 4、動画用音声処理部 2 0 4 5 等が、中央制御部 2 0 1 とは独立した処理部とする例を説明した。しかしながら、中央制御部 2 0 1 が、これらの全部或いは一部を、プログラムを実行することで代替しても構わない。

【 0 1 1 9 】

（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

50

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、ＡＳＩＣ）によっても実現可能である。

【符号の説明】

【０１２０】

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…動画像音声処理部、２０４６…コマンドメモリ

10

20

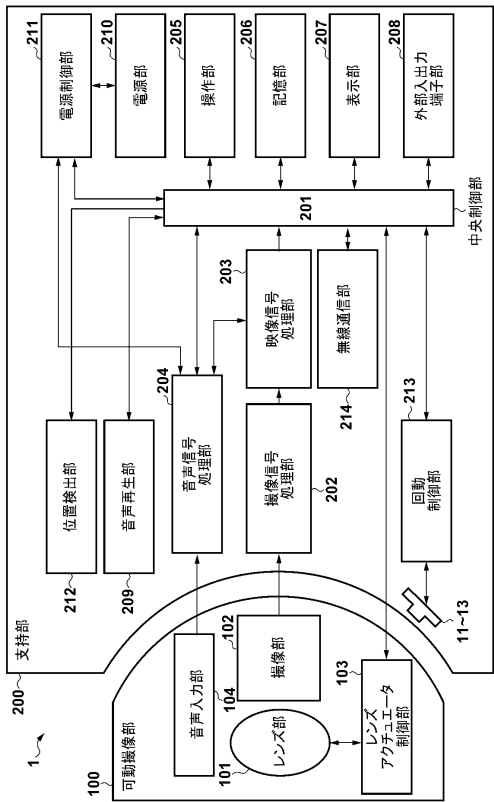
30

40

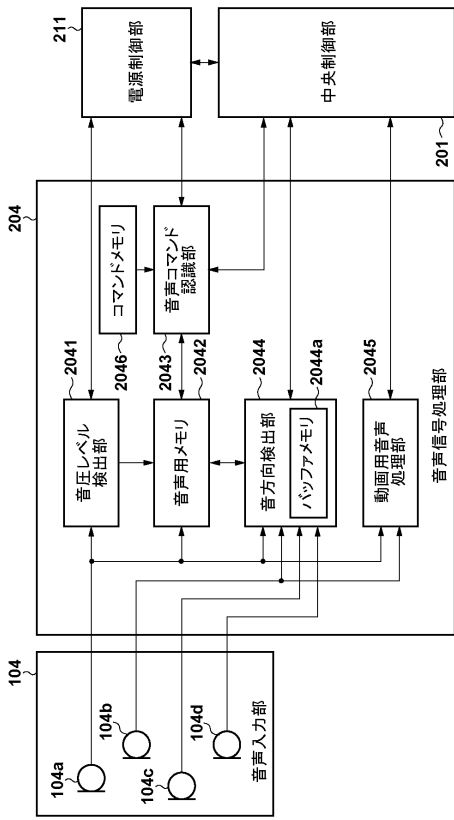
50

【図面】

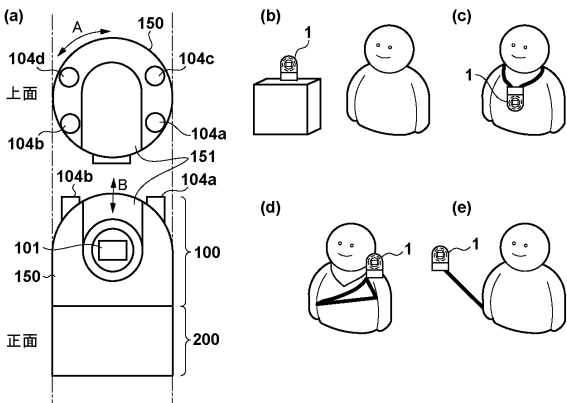
【図 1】



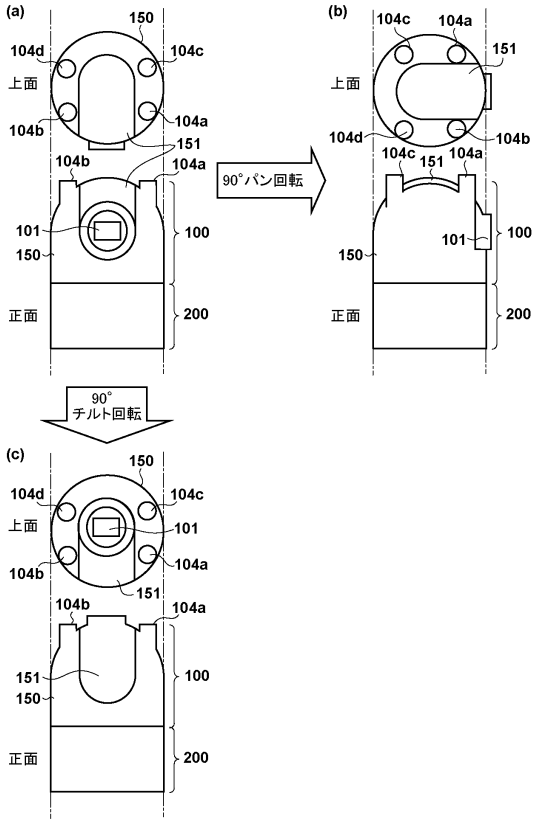
【図 2】



【図 3】



【図 4】



10

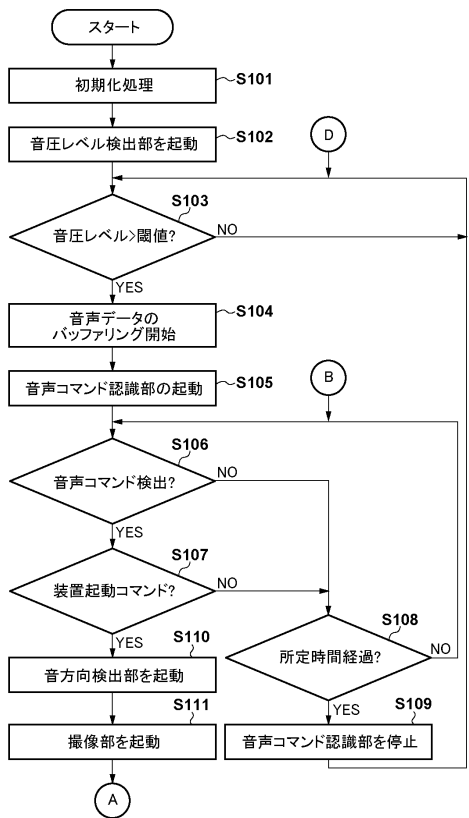
20

30

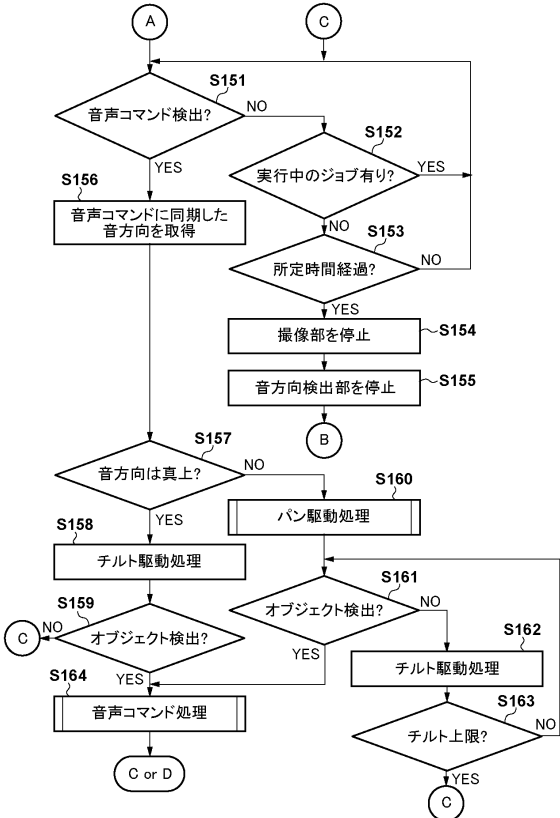
40

50

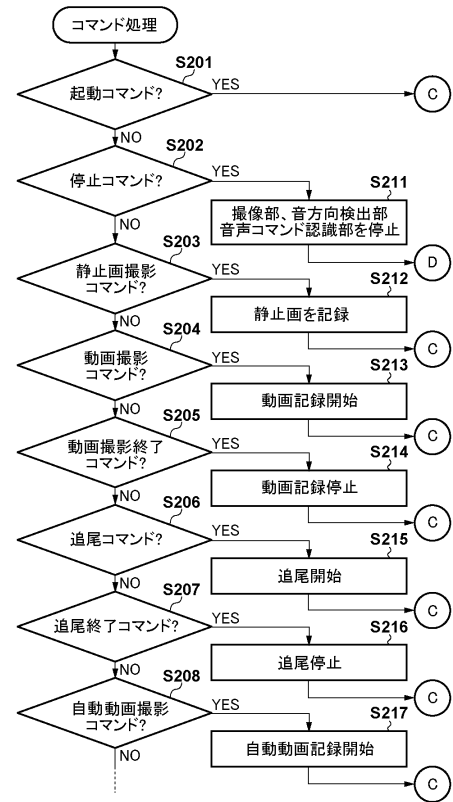
【図 5 A】



【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"

10

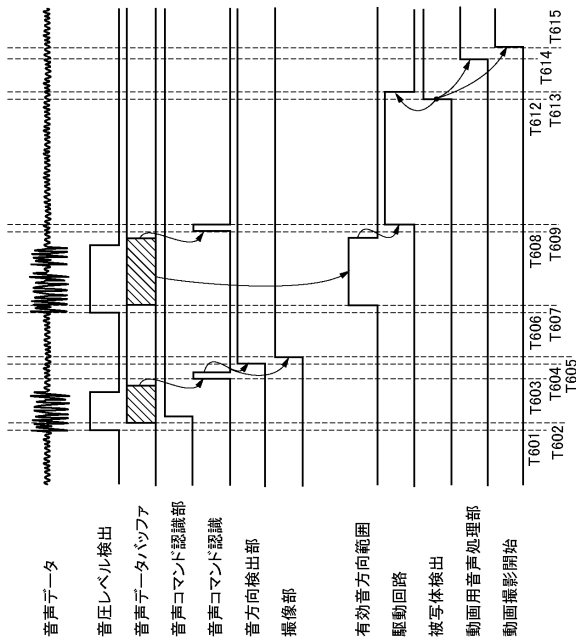
20

30

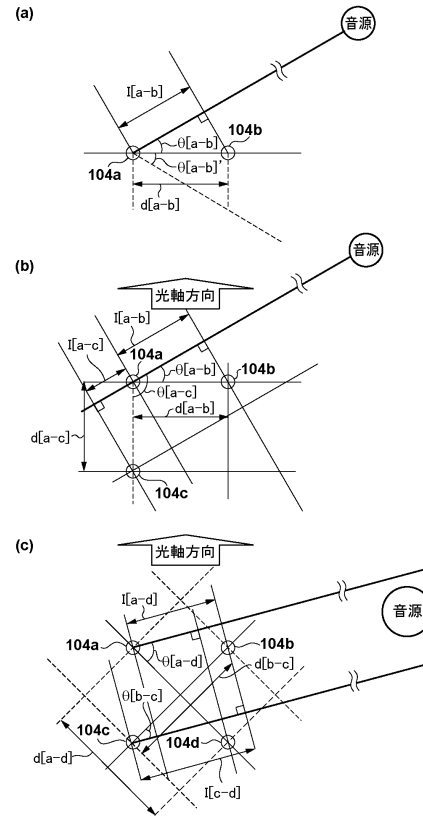
40

50

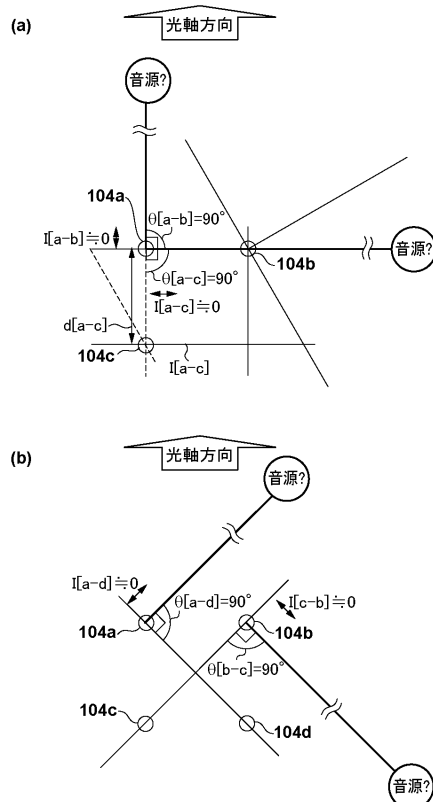
【圖 8】



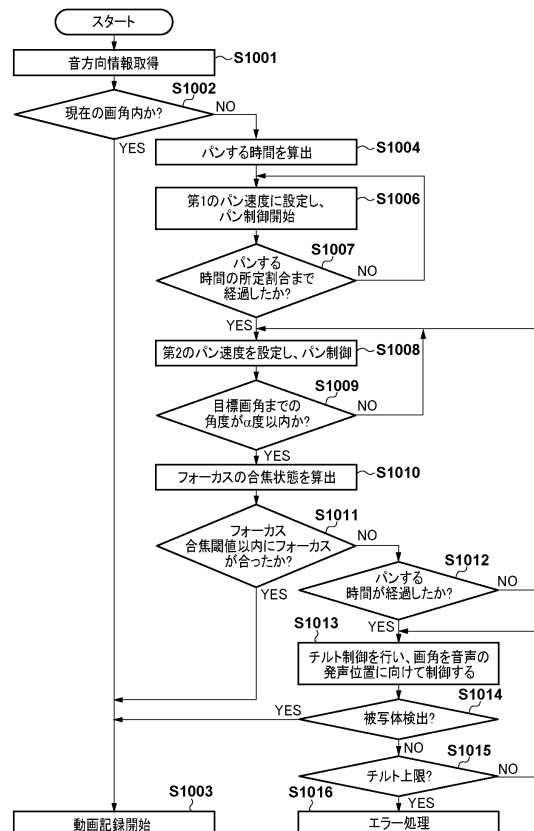
【 図 9 】



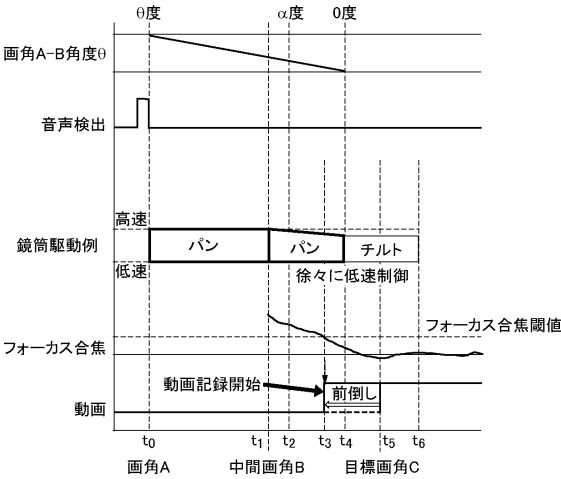
【 図 1 0 】



【 図 1 1 】



【図 12】



10

20

30

40

50

フロントページの続き

(51)国際特許分類

F I

<i>H 0 4 N</i>	<i>23/695 (2023.01)</i>	<i>G 0 3 B</i>	<i>17/56</i>	<i>A</i>
<i>G 1 0 L</i>	<i>15/00 (2013.01)</i>	<i>H 0 4 N</i>	<i>23/60</i>	<i>1 0 0</i>
<i>G 1 0 L</i>	<i>15/10 (2006.01)</i>	<i>H 0 4 N</i>	<i>23/65</i>	
<i>G 1 0 L</i>	<i>15/28 (2013.01)</i>	<i>H 0 4 N</i>	<i>23/695</i>	
<i>G 1 0 L</i>	<i>25/51 (2013.01)</i>	<i>G 1 0 L</i>	<i>15/00</i>	<i>2 0 0 G</i>
<i>H 0 4 R</i>	<i>3/00 (2006.01)</i>	<i>G 1 0 L</i>	<i>15/10</i>	<i>2 0 0 W</i>
<i>H 0 4 R</i>	<i>1/40 (2006.01)</i>	<i>G 1 0 L</i>	<i>15/28</i>	<i>2 3 0 K</i>
<i>G 1 0 L</i>	<i>15/04 (2013.01)</i>	<i>G 1 0 L</i>	<i>25/51</i>	<i>4 0 0</i>
		<i>H 0 4 R</i>	<i>3/00</i>	<i>3 2 0</i>
		<i>H 0 4 R</i>	<i>1/40</i>	<i>3 2 0 A</i>
		<i>G 1 0 L</i>	<i>15/04</i>	<i>3 0 0 A</i>

(56)参考文献

特開 2 0 0 9 - 1 7 7 4 8 0 (J P , A)
 特開 2 0 0 5 - 1 8 4 4 8 5 (J P , A)
 特開 2 0 1 6 - 1 1 9 6 1 5 (J P , A)
 特開 2 0 1 1 - 1 8 8 0 5 5 (J P , A)
 特開 2 0 1 3 - 1 2 1 0 7 8 (J P , A)
 特開平 0 9 - 2 0 5 5 7 4 (J P , A)

(58)調査した分野 (Int.Cl., D B 名)

G 0 3 B *1 7 / 3 8*
G 0 3 B *1 5 / 0 0*
G 0 3 B *1 7 / 5 6*
H 0 4 N *2 3 / 6 0*
H 0 4 N *2 3 / 6 9 5*
H 0 4 N *2 3 / 6 5*
G 1 0 L *1 5 / 0 0*
G 1 0 L *1 5 / 1 0*
G 1 0 L *1 5 / 2 8*
G 1 0 L *2 5 / 5 1*
H 0 4 R *3 / 0 0*
H 0 4 R *1 / 4 0*
G 1 0 L *1 5 / 0 4*