

(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2016-10010

(P2016-10010A)

(43) 公開日 平成28年1月18日(2016.1.18)

(51) Int.Cl. F 1 テーマコード (参考)
H04N 7/15 (2006.01) H04N 7/15 630A 5C164

審査請求 未請求 請求項の数 5 O L (全 14 頁)

(21) 出願番号	特願2014-129638 (P2014-129638)	(71) 出願人	000005810
(22) 出願日	平成26年6月24日 (2014. 6. 24)		日立マクセル株式会社
			大阪府茨木市丑寅一丁目1番88号
		(74) 代理人	100104547
			弁理士 栗林 三男
		(72) 発明者	大坪 宏安
			大阪府茨木市丑寅1丁目1番88号 日立マクセル株式会社内
		Fターム(参考)	5C164 FA10 VA04P VA06P VA35S VA54S VA56S

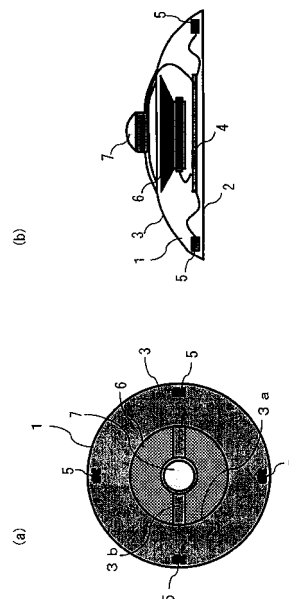
(54) 【発明の名称】 音声入出力機能付き撮像装置およびテレビ会議システム

(57) 【要約】

【課題】低コストに製造可能なテレビ会議用の音声入出力機能付き撮像装置およびテレビ会議システムを提供する。

【解決手段】テレビ会議システムで用いられる音声入出力機能付き撮像装置1は、マイク5、スピーカ6および全方位カメラ7を備え、互いに近傍となる位置にマイク5、スピーカ6および全方位カメラ7が配置されている。音声入出力機能付き撮像装置1は、テレビ会議に参加する参加者が囲んで座るテーブルに置かれて用いられる。会議の発言者は、マイク5に向かって話すとともに、スピーカ6の音を聞くためにスピーカを向く可能性が高い。したがって、全方位カメラ7で、発言者を正面から撮影する状況となる。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

周囲を撮像対象とする全方位カメラと、

前記全方位カメラの近傍に設けられ、外部から入力される音声信号を音声として周囲に出力する音声出力手段と、

前記全方位カメラの近傍に設けられ、周囲の音声を音声信号として入力する音声入力手段とを備え、

前記全方位カメラにより撮像された画像データと、前記音声入力手段により入力された音声信号を出力することを特徴とする音声入出力機能付き撮像装置。

【請求項 2】

前記全方位カメラの近傍で当該全方位カメラによる周囲の撮像を妨げない位置に、周囲の複数方向から視認可能に外部から入力された画像データを表示する複数台のディスプレイが設けられていることを特徴とする請求項 1 に記載の音声入出力機能付き撮像装置。

【請求項 3】

前記音声入力手段は、少なくとも周囲の異なる方向をそれぞれ向いた少なくとも 3 つのマイクを備え、

各マイクに入力する音声の音量から音源の方向を特定する音源方向認識手段と、

前記全方位カメラで撮像された全方位画像データを、前記音源方向認識手段により特定された音源の方向を中心とする画像データに変換する画像処理手段を備えることを特徴とする請求項 1 または請求項 2 に記載の音声入出力機能付き撮像装置。

【請求項 4】

前記全方位カメラで撮像された画像データ中に撮像されている被撮像者の顔を認識するとともに、認識された前記顔の口の動きから前記被撮像者のうちの発言している前記被撮像者を特定する画像認識手段と、

前記全方位カメラで撮像された全方位画像データを、前記画像認識手段により発言していると特定された前記被撮像者を中心とする画像データに変換する画像処理手段とを備えることを特徴とする請求項 1 から請求項 3 のいずれか 1 項に記載の音声入出力機能付き撮像装置。

【請求項 5】

請求項 1 から請求項 4 のいずれか 1 項に記載の前記音声入出力機能付き撮像装置を複数備え、各音声入出力機能付き撮像装置は、他の前記音声入出力機能付き撮像装置に前記画像データと前記音声信号を出力し、かつ、他の前記音声入出力機能付き撮像装置から出力された前記画像データおよび前記音声信号を入力するための通信手段を備えることを特徴とするテレビ会議システム。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、音声入出力機能付き撮像装置およびテレビ会議システムに関する。

【背景技術】**【0002】**

近年、会議室には、テーブル上に配置されたマイクおよびスピーカを備える音声入出力装置と、テーブルの近傍に配置されるディスプレイと、このディスプレイの近傍に配置された動画撮影用のカメラ（例えば録画機能の無いテレビカメラ）とを備え、離れた場所の別の会議室との間で、画像と音声を用いた所謂テレビ会議が可能となるテレビ会議システムが設けられている場合がある。

【0003】

このようなテレビ会議システムでは、テレビカメラの画角を調整して、会議の参加者の全員が撮影範囲に入るようにする場合が多い。この場合に、参加者の着席位置が制限されたり、参加者全員を撮影範囲に収めることが困難であったりする場合がある。また、会議開始前にテレビカメラの画角やズーム等を調整するのに少し時間がかかることがあり、参

10

20

30

40

50

加者が全員揃ってから会議開始までに時間差が生じてしまう。

【0004】

また、会議において、主な発言者が予め決まっている場合には、発言者にテレビカメラの撮影範囲のなるべく中央側に座って貰うなどの対策が可能であるが、参加者の誰が発言するか分からない状態では、発言者が撮影範囲の端の方において、よく見えないなどの問題が生じる。

【0005】

そこで、音声入力用マイクを複数設けるか、複数の広角度カメラを設け、これら複数のマイクの音声信号や複数の広角カメラの画像データから発言者の位置を特定し、発言者の位置に基づいて、発言者の発する音声を主に音声入力するようにマイクを制御し、かつ、発言者を主に撮影するようにカメラを制御する提案がなされている（特許文献1参照）。

10

【0006】

また、近年の会議システムでは、カメラとしてPTZカメラが用いられる。PTZとは、カメラを左右に首振りさせるパン（P）と、上下に首振りさせるチルト（t）、画像を拡大するズーム（Z）が可能なカメラであり、例えば、会議の発言者が中心となるようにカメラの向きとズームを制御することができる。また、上述のように発言者の位置が特定できるシステムの場合に、自動でPTZカメラを発言者に向けることができる。

【先行技術文献】

【特許文献】

【0007】

20

【特許文献1】特開平10-145763号公報

【発明の概要】

【発明が解決しようとする課題】

【0008】

ところで、特許文献1の発明では、複数台のマイクやカメラを用いて発言者の位置を特定し、この特定された発言者の位置に基づいて、発言者が主に撮影されるようにカメラを制御したり、発言者の発言の音声が入力されるようにマイクを制御したりする。したがって、特許文献1では、複数のマイクやカメラが必要で、かつ、マイクやカメラを制御する制御装置が必要であり、会議システムのコストが高くなる。

【0009】

30

例えば、1つの会議室の参加者が数十人を超えるような場合には、発言者の位置を特定し、特定された発言者を撮像するためのカメラの制御や、発言者の発言の音声を抽出するためのマイクの制御が必要となるかもしれないが、1つの会議室の参加者が十数人以下の場合に、コストパフォーマンス的に問題がある。

【0010】

本発明は、前記事情に鑑みてなされたものであり、低コストに製造可能なテレビ会議用の音声入出力機能付き撮像装置およびこの音声入出力機能付き撮像装置を有するテレビ会議システムを提供することを目的とする。

【課題を解決するための手段】

【0011】

40

前記課題を解決するために、本発明の音声入出力機能付き撮像装置は、周囲を撮像対象とする全方位カメラと、

前記全方位カメラの近傍に設けられ、外部から入力される音声信号を音声として周囲に出力する音声出力手段と、

前記全方位カメラの近傍に設けられ、周囲の音声を音声信号として入力する音声入力手段とを備え、

前記全方位カメラにより撮像された画像データと、前記音声入力手段により入力された音声信号を出力することを特徴とする。

【0012】

このような構成によれば、音声入出力機能付き撮像装置を会議システムの撮像装置、音

50

声出力手段としてのスピーカ、音声入力手段としてのマイクとして使用する場合に、当該音声入出力機能付き撮像装置をテーブルに配置し、このテーブルを囲んで会議の複数の参加者に座って貰うことにより、全方位カメラにより参加者の全員を撮影することが可能となる。この場合、テーブルを囲む参加者は、それぞれ、テーブル上の音声入出力機能付き撮像装置を見るか、テレビ会議の他の会場が映し出されたディスプレイを見ることになる。

【0013】

但し、発言者の多くは、基本的に音声入力手段としてのマイクに向かって発言する場合が多く、また、他の参加者の音声が入力されるスピーカの方角を向く可能性も高い。一般に音源が顔の正面方向にある方が、音が聞き取り易く、音が良く聞こえるように音源の方角を見る人が多い。すなわち、テーブル上の全方位カメラの周囲を囲んで会議の参加者が座った場合に少なくとも発言者がマイクやスピーカの方角を向くことで、マイクやスピーカの近傍にある全方位カメラの方角を向くことになり、全方位カメラで正面から参加者を撮影する状態となり、発言者を撮像した画像データ上では、発言者が画像データを見ているテレビ会議の他の会場の参加者を向いて発言しているように見える。

【0014】

すなわち、マイクとスピーカとカメラを略同じ位置に配置することにより、少なくとも会議の参加者が発言する場合に、カメラを向いて発言させるように促すことができ、発言者の画像を明確にすることができる。

【0015】

また、全方位カメラは、テーブル状に配置されて、テーブル周囲に座る参加者を撮影するので参加者と距離が短く、かつ、各参加者による距離の差が少ない。したがって、高い解像度を有する全方位カメラでなくても、参加者の撮影が十分可能であり、高い解像度の全方位カメラを用いる場合よりもコストの低減を図ることができる。

【0016】

なお、全方位カメラで撮像された全方位画像データをそのまま平面に投影した状態に出力すると歪んだ画像となるので、例えば、パノラマ画像に変換したり、各被写体となる会議の参加者毎の画像に変換したりするとともに、歪みをとる画像処理を行う必要がある。また、全方位カメラは、例えば、魚眼レンズを用いた魚眼カメラや、円錐状に近い形状のミラーを用いたカメラや、全天球カメラが含まれる。また、音声出力手段は、例えば、スピーカである。また、音声入力手段は、例えば、マイクである。

【0017】

本発明の前記構成において、前記全方位カメラの近傍で当該全方位カメラによる周囲の撮像を妨げない位置に、周囲の複数方向から視認可能に外部から入力された画像データを表示する複数台のディスプレイが設けられていることが好ましい。

【0018】

このような構成によれば、基本的に会議の参加者は、他の会場の参加者が映し出されたディスプレイ、他の会場の参加者の発言が音声として出力されるスピーカ、他の会場の参加者へ話しかけるためのマイクの方角を向く可能性が高いが、これらディスプレイ、マイク、スピーカが互いに近傍にまとまって存在するので、基本的に会議の参加者は、その多くが自然に全方位カメラの方角を向くことになり、他の会場のディスプレイでは、他の会場の参加者の方向を向いている参加者が映し出されることになる。

【0019】

また、全方位カメラをテーブル上に置いた場合に、各参加者とディスプレイとの距離が短くなり、比較的小さなサイズのディスプレイでも別会場の参加者の識別が可能になるので、ディスプレイを複数用いるものとしても大きなディスプレイを1つ用いる場合よりもコストの低減を図ることができる。なお、四角いテーブルに、参加者が2列で向かい合って座るような場合に、ディスプレイを2つとすることができる。円卓の周囲に参加者が円状に並んで座る場合や、四角いテーブルの4辺のうちの3辺以上に、分かれて参加者が座る場合には、ディスプレイが3つ以上あることが好ましい。

【0020】

本発明の前記構成において、前記音声入力手段は、少なくとも周囲の異なる方向をそれぞれ向いた少なくとも3つのマイクを備え、

各マイクに入力する音声の音量から音源の方向を特定する音源方向認識手段と、

前記全方位カメラで撮像された全方位画像データを、前記音源方向認識手段により特定された音源の方向を中心とする画像データに変換する画像処理手段を備えることが好ましい。

【0021】

このような構成によれば、参加者のうち発言している発言者を特定して、発言者を左右の略中央とするパノラマ画像を他の会議会場のディスプレイに表示したり、発言者を抜き出した状態の画像を他の会場のディスプレイに表示したりすることが可能になる。本発明においては、全方位カメラおよびその近傍のマイクの周囲に参加者がいるので、指向性の高いマイクでなくとも、各マイクの音量を比較することで、比較的容易に音源としての発言者の方向を特定可能であるとともに、発言者の全方位画像上の位置を特定するのに音源の方向さえ特定できれば、音源の位置まで特定する必要がなく、音源の位置を特定するのにマイクアレーや指向性の高いマイク等を用いなくてもよいので、コストの低減を図ることができる。また、マイクにより特定された発言者を中心（主体）とする画像データを作成する際には、全方位画像データ上で方向を指定することにより、容易に発言者を主体とする画像データを作成することができる。

【0022】

本発明の前記構成において、前記全方位カメラで撮像された画像データ中に撮像されている被撮像者の顔を認識するとともに、認識された前記顔の口の動きから前記被撮像者のうちの発言している前記被撮像者を特定する画像認識手段と、

前記全方位カメラで撮像された全方位画像データを、前記画像認識手段により発言していると特定された前記被撮像者を中心とする画像データに変換する画像処理手段とを備えることが好ましい。

【0023】

このような構成におれば、音声の場合と同様に発言者の方向を特定すれば、発言者を主体とする画像データを作成可能であり、位置を特定する必要がないので、複数のカメラを用いる必要がなく、コストの低減を図ることができる。また、音声で発言者の方向を特定した場合と同様に、特定された発言者を主体とする画像データを作成する際には、全方位画像データ上で方向を指定することにより、容易に発言者を主体とする画像データを作成することができる。

【0024】

本発明のテレビ会議システムは、本発明の音声入出力機能付き撮像装置を複数備え、各音声入出力機能付き撮像装置は、他の前記音声入出力機能付き撮像装置に前記画像データと前記音声信号を出力し、かつ、他の前記音声入出力機能付き撮像装置から出力された前記画像データおよび前記音声信号を入力するための通信手段を備えることを特徴とする。

【0025】

このような構成によれば、本発明のテレビ会議システムは、各音声入出力機能付き撮像装置の上述の作用効果を奏することができる。なお、音声入出力機能付き撮像装置には、ディスプレイが無い構成の場合もあるが、他の音声入出力機能付き撮像装置で撮像された画像データが入力されることにより、音声入出力機能付き撮像装置において、外部のディスプレイに画像データを出力することが可能になる。

【発明の効果】

【0026】

本発明の音声入出力機能付き撮像装置およびテレビ会議システムによれば、低コストに製造可能であり、かつ、発言者がディスプレイに表示された場合に、ディスプレイを見る人を向いた状態となる可能性が高くなる。

【図面の簡単な説明】

【0027】

【図1】第1の実施の形態の音声入出力機能付き撮像装置を示すカバーを半透明化した図であって、(a)が平面図であり、(b)が側面図である。

【図2】同、音声入出力機能付き撮像装置の使用状況を説明するための図である。

【図3】同、音声入出力機能付き撮像装置の全方位カメラに撮影された画像を説明するための図である。

【図4】同、音声入出力機能付き撮像装置から出力される画像を説明するための図であって、(a)は全方位画像から変換されたパノラマ画像の概略を示す図であり、(b)は全方位画像から変換されたパノラマ画像を分割して2列にしたものであり、(c)は、発言者の画像を加えたものであり、(d)異なる3か所で撮影された全方位画像をそれぞれパノラマ画像としたものである。

10

【図5】第2の実施の形態の音声入出力機能付き撮像装置を示すカバーを半透明化した図であって、(a)が平面図であり、(b)が側面図である。

【図6】第3の実施の形態の音声入出力機能付き撮像装置を示すカバーを半透明化した図であって、(a)が平面図であり、(b)が側面図である。

【図7】第4の実施の形態の音声入出力機能付き撮像装置を示す図であって、(a)が正面図であり、(b)が背面図である。

【発明を実施するための形態】

【0028】

以下、図面を参照しながら本発明の第1の実施の形態について説明する。

20

本実施の形態のテレビ会議システムは、図1(a)、(b)に示す音声入出力機能付き撮像装置1を複数用いるものであり、離れた複数箇所の会議室に音声入出力機能付き撮像装置1を配置することにより、テレビ会議システムが構築される。

【0029】

図1に示す音声入出力機能付き撮像装置1は、略円板状のベース板2と、ベース板2上を覆う略ドーム状のカバー3と、ベース板2の外周部に周方向に沿って等間隔に配置されるとともに後述の制御基板4に接続されたマイク(音声入力手段)5と、ベース板2とカバー3との間に、カバー3で覆われた状態に配置されたスピーカ(音声出力手段)6と、カバー3上に固定された全方位カメラ7とを備える。マイク5と、スピーカ6と、全方位カメラ7は、互いに近接して設けられている。すなわち、マイク5と、スピーカ6と、全方位カメラ7とは互いに近傍となる配置となっている。また、スピーカ6と、全方位カメラ7とは、それらの中心軸が略一致するように配置され、マイク5は、上述の中心軸から略等距離となる位置に配置されている。

30

【0030】

ベース板2は、その上面に、マイク5、スピーカ6、制御基板4を取り付けるための取付構造が設けられている。また、円板状のベース板2の外周部には、ベース板2と略同径のカバー3の円形の下側縁部(外周縁部)を取り付けるための取付構造が設けられている。

【0031】

カバー3は、マイク5に対応する位置に図示しない1つまたは複数の孔が設けられ、マイク5への音声入力を妨げないようになっている。まあ、ドーム状のカバー3の上部(中央部)には、スピーカ6からの音声出力用の開口部3aが設けられている。まあ、カバー3の開口部3aには、全方位カメラ7をカバー3の上部の中央部に固定するための橋状のカメラ固定部3bが設けられている。

40

【0032】

マイク5は、例えば、指向性を有するものであり、最も感度の高い方向を、全方位カメラ7の例えば撮影範囲となる半球面や円筒面の中心軸に直交する半径方向に合わせている。また、マイク5の配置位置は、撮影範囲の中心軸に対して半径方向に等距離で、それぞれ90度ずれた位置(周方向に等間隔)に配置されている。なお、マイク5として無指向性のマイク5を用いてもよい。各マイク5は、制御基板4に接続されており音声音声信

50

号に変換して制御基板4に入力している。なお、音声信号はアナログであってもデジタルであってもよい。

【0033】

スピーカ6は、全方位型のものであり、1つのスピーカ6により音声が入力される全方位に略同等に出力する。なお、全方位型でないスピーカを3つまたは4つ等のように複数用いてもよい。スピーカ6は、制御基板4に接続されており、制御基板4から出力される音声信号を音声に変換して周囲に出力する。

【0034】

全方位カメラ7は、例えば、半球状の撮像範囲を有する魚眼カメラであり、周囲を撮像対象としているが、例えば、複数のカメラで撮影された画像から全方位画像データF（図2に図示）を得るようなものであっても良いし、略円錐状のミラーを介して周囲を撮影するカメラであってもよいし、全天球カメラであってもよい。全方位カメラ7では、テーブルTに置かれた音声入出力機能付き撮像装置1からテーブルTの周囲に座る被写体としての参加者を撮像できればよく、例えば、上方向の画像データは必要としない。

【0035】

また、全方位カメラ7の配置位置が高い場合、例えば、座った参加者の頭部以上の高さを有する場合など、半球状の撮影範囲では、参加者の胸像を撮影することができなくなるので、全方位カメラ7の配置位置が高くなる場合には、全天球カメラを好適に用いることができる。

【0036】

制御基板4は、音源方向認識手段として、4つのマイク5から入力される音声信号の音量レベル（音の大きさ）から音源の方向を特定するようになっている。本実施の形態では、音源の方向と音源までの距離を特定することにより音源の位置を特定することはしないので、4つのマイク5の音量レベルから音源の位置を測定する。例えば、音量レベルが高い上位2本の隣り合うマイク5を特定し、これらの2つのマイクの音量の差からこれら2つのマイク5の中間となる方向を決定する。

【0037】

例えば、2つのマイク5で音量に差が無ければ、これらマイク5の略中央となる方向に音源があると特定し、どちらかのマイク5の音量が高ければ、これらマイク5の中央となる方向と、音量が高い方のマイク5の方向との間に音源の方向があることになる。また、音量が2位となるマイク5と、音量が3位となるマイク5とで音量が略同じならば、音量が1位のマイク5が向く方向に音源があることになる。

【0038】

なお、各マイク5における音の位相のずれから音源を特定するものとしてもよい。すなわち、音源からの距離の違いによる各マイク5における音の到達時間の違いに基づいて音源の方向を特定する周知の方法を用いてもよい。

また、画像認識手段としての制御基板4は、全方位カメラ7から入力される全方位画像データFから発言者の方向を特定するようになっている。基本的には、周知の顔認識により全方位画像データFから各参加者（被撮像者）の顔を認識することにより、各参加者の方向を特定する。また、各参加者の口を画像認識し、口（唇）が動いているか否かを判定し、口が動いていると判定された顔の方向を発言者の方向とする。

【0039】

なお、画像処理および画像認識に関しては、インテル（登録商標）オープンCV（Intel Open Source Computer Vision Library）を利用して容易に作成可能である。例えば、顔認識プログラムを作成する場合に、オープンCVに登録されているオブジェクト検出プログラムを用いることができる。画像認識の原理として、学習フェーズと認識フェーズがあり、画像から特徴量を抽出し、学習アルゴリズムによってオブジェクトの特徴を学習することにより、例えば、顔認識等の画像認識が可能となる。オープンCVでは、画像特徴量としてHaar・Like特徴量を用い、学習アルゴリズムとしてAdaBoostと呼ばれるアルゴリズムを使用している。オブ

10

20

30

40

50

ジェクト検出プログラムにおいて、特徴点に基づいて顔の画像か否かを機械学習させることにより、オブジェクト検出プログラムにおいて、顔の画像を顔として認識することが可能となる。なお、画像認識プログラムに必ずしもオープンCVを利用しなくてもよいし、既存のプログラムや、既存の画像認識回路を搭載したチップを利用してもよい。発言者の口の動きの認識も上述のオープンCVのオブジェクト検出プログラムを用いて、機会学習させることにより、例えば、話している口と、黙っている口の違いを認識させることができる。

【0040】

本実施の形態では、顔認識を行って各参加者の方向を認識するとともに、口の動きを検出して発言者の方向を認識する。なお、上述のように制御基板4では、音声によっても発言者としての音源の方向を特定しているので、本実施の形態では、これら音源方向認識と画像認識に基づく発言者の方向が例えば所定角度範囲内（例えば0～10度以内）で一致する場合に、これら音源方向認識と画像認識で求められた2つの方向のうち、例えば、画像認識で得られた方向を、発言者の方向としている。

【0041】

音源方向認識による音源方向と画像認識による発言者の方向とが所定角度範囲以内とまらない場合には、発言者がいないと判定する。これにより、小声で私語を話している参加者や、あくびをしている参加者や、椅子を動かした際に大きな音を出した参加者などが、一時的にでも発言者として認識されて例えば別の会場のディスプレイ8に大きく表示されてしまうような状態を防止している。なお、音源方向認識だけで、発言者の方向を決定してもよいし、画像認識だけで発言者の方向を決定してもよい。

【0042】

また、制御基板4は、全方位カメラ7から入力された全方位画像データFを周知の画像処理によりパノラマ画像に変換する画像処理手段として機能する。この際には、全方位画像データFからパノラマ画像の右端および左端となる位置を決定して、全方位画像データFからパノラマ画像データを作成する。上述のように発言者の方向が特定された場合には、発言者の方向から180度、すなわち、発明者の方向の反対となる方向の位置で、全方位画像データFを切り開き、この位置をパノラマ画像の右端および左端の位置とする。また、発言者がいない場合には、例えば、上述のように顔認識された各参加者の間隔を判定し、最も広い間隔の中央をパノラマ画像の左端および右端の位置とする。

【0043】

また、制御基板4は、発言者の方向を特定した場合に、その方向で顔認識された参加者が主に被写体となっている発言者の画像データを作成する。なお、この画像データの作成においては、顔認識された参加者の画像部分を取り出して画像データとしてもよいし、特定された発言者の方向の所定角度範囲の画像部分を発言者の画像データとしてもよい。

【0044】

また、通信手段としての制御基板4は、ローカルエリアネットワーク（LAN）や、インターネットや公衆電話回線網や、携帯電話回線網や専用通信回線等を利用して、離れた場所にある他の音声入出力機能付き撮像装置1とデータ通信を行い、マイク5により入力された音声信号および全方位カメラ7で撮影された全方位画像データFを上述のように画像処理したパノラマ画像データおよび発言者の画像データを他の音声入出力機能付き撮像装置1に送信する。

【0045】

また、他の音声入出力機能付き撮像装置1から送信された音声信号、パノラマ画像データ、発言者の画像データ等を受信する。なお、発言者の画像データは、当該画像データが作成された場合にだけ送受信される。また、本実施の形態では、音声入出力機能付き撮像装置1にはディスプレイ8が無いので、受信された画像データは、ディスプレイ8用の接続端子に出力され、接続端子に接続されたディスプレイ8に画像データを表示する。なお、後述のように音声入出力機能付き撮像装置1にディスプレイ8を含めて受信した画像データを音声入出力機能付き撮像装置1のディスプレイ8に出力するようにしてもよい。

【0046】

また、制御基板4で、音源方向認識、画像認識、画像処理等を行うものとしたが、制御基板4では、主に音声信号、画像データの入出力だけを制御し、制御基板4に有線LANや無線LANやUSB等で接続されたパーソナルコンピュータ（パソコンPC：図2に図示）で音源方向認識、画像認識、画像処理を行うものとしてもよい。また、各種画像処理を、全方位画像を撮影した全方位カメラ7がある音声入出力機能付き撮像装置1で行うものとしたが、画像処理を、画像データを受信する側の音声入出力機能付き撮像装置1またはそれに接続されたパソコンPCで行ってもよい。すなわち、画像データとして全方位カメラ7で撮影された全方位画像データFをそのまま送信して、受信した音声入出力機能付き撮像装置1において、画像処理してディスプレイ8に表示するものとしてもよい。

10

【0047】

このような電話会議システムの音声入出力機能付き撮像装置1は、例えば、図2に示すように、会議室のテーブルTの上に置いて用いられる。会議の参加者Pは、テーブルTを囲んで座ることになる。ここでは、長形状のテーブルTの2つの長辺にそれぞれ参加者Pが2列に座っている。なお、図2では、上述のようにパソコンPCを用いるものとし、ディスプレイ8は、パソコンPCを介して接続されており、パソコンPCで処理された画像データがディスプレイ8に表示される。

【0048】

図2に示す状態で、全方位カメラ7で撮像された全方位画像データFは、図3に示す状態となる。なお、図3では、立体的な全方位画像データFを平面に投影した状態で簡略化して示している。制御基板4では、この全方位画像データFを画像処理して、図4（a）または図4（b）に示すディスプレイ8の表示中に表示されるパノラマ画像G1またはパノラマ画像G1を2つに分割したパノラマ画像G2、G3としている。

20

【0049】

本実施の形態では、図4（b）に示すように、全方位画像データF中の各参加者Pの間隔を判定し、所定間隔（角度）以上の間隔がある場合に、パノラマ画像G1を分離し、分離された部分の間隔をカットすることで、パノラマ画像G2、G3の左右幅を圧縮している。なお、パノラマ画像G1、G2、G3の作成に際し、参加者P同士の間隔を全てカットするようにしてもよい。また、所定幅（所定角度範囲）で各参加者Pの画像データを作成し、これを横に並べることでパノラマ画像を作成してもよい。この場合も、参加者P同士の間隔を表示しないようにできる。なお、図4（b）では、2つに分離した画像データを上下二段に表示することにより、各パノラマ画像G2、G3を大きく表示している。

30

【0050】

また、発言者を特定した場合には、図4（c）に示すように、図4（a）に示すパノラマ画像G1に加えて発言者を主体とする画像G10を別に表示する。なお、テレビ会議は、2箇所だけで行われるとは限らず、3か所以上で行われる場合があるので、その場合には、例えば、図4（d）に示すように、ディスプレイ8の画面を分割して、各分割箇所にパノラマ画像G1、G4、G5を表示する。図4（d）では、4か所を結んでテレビ会議が行われ、ディスプレイ8がある会議室以外の他の3か所の会議室の画像が表示された状態となっている。

40

【0051】

この音声入出力機能付き撮像装置1を用いたテレビ会議システムでは、各会議室に設置された音声入出力機能付き撮像装置1の上述のように通信手段としての制御基板4において、各会議室で撮影された画像データと入力された音声信号を送受信することにより、上述のようにディスプレイ8に、他の会議室の参加者の画像が表示されるとともに、スピーカ6から他の会議室で入力された音声信号が出力される。

【0052】

このような音声入出力機能付き撮像装置1およびテレビ会議システムにおいては、上述のように全方位カメラ7とマイク5とスピーカ6とが略一体的に構成されており、発言す

50

る参加者（発言者）は、基本的にマイク5に向かって発言しようとする。この場合に、マイク5の近傍に全方位カメラ7があるので、発言者は、全方位カメラ7に向かって発言する状態となり、発言者は正面から撮影される状態となる。この場合に、発言者の画像G10をディスプレイ8に表示した際に、発言者がディスプレイ8を見ている他の会議室の参加者に向かって話しているように見える可能性が高い。

【0053】

また、他の会議室の参加者と話し合っている状態の場合には、他の会議室の発言者の音声、全方位カメラ7の近傍のスピーカ6から聞こえるので、音を聞き取り易くするためにスピーカ6の方を向くことになる。これにより、発言者が全方位カメラ7に向かって話す状態となり安い。したがって、上述のように発言者が他の会議室の参加者の方を向いて話している状態の画像を得易くなる。これらのことから、ディスプレイ8の画面において、発言者が全方位カメラ7以外の方向を向いて話すことによるテレビ会議特有の違和感が生じるのを抑制することができる。言い換えれば、発言者が意識してカメラの方を向くように努力しなくても、自然に全方位カメラ7の方を向くように促すことができる。

【0054】

また、全方位カメラ7により、基本的にテーブルTの周囲に坐っている全ての参加者が略同等の大きさで撮影されているので、特に全方位カメラ7を制御しなくても、上述のように発言している参加者を特定すれば、容易に発言者の画像を得ることができる。

【0055】

次に、本発明の第2の実施の形態を説明する。

図5(a)、(b)に示すように、第2の実施の形態の音声入出力機能付き撮像装置1aは、第1の実施の形態の音声入出力機能付き撮像装置1と同様に、ベース板11、カバー12、図示しない制御基板（図1の制御基板4）、マイク5、スピーカ6、全方位カメラ7を備える。第2の実施の形態の音声入出力機能付き撮像装置1aは、さらにディスプレイ8を備える、すなわち、第1の実施の形態の音声入出力機能付き撮像装置1と第2の実施の形態の音声入出力機能付き撮像装置1aとの違いは、ディスプレイ8が音声入出力機能付き撮像装置1に対して別体になっているか、音声入出力機能付き撮像装置1aに、ディスプレイ8が備えられているかの違いである。

【0056】

本実施の形態において、ベース板11は、矩形板状に形成され、その四隅部のそれぞれにマイク5が備えられている。また、ベース板11の互いに離れた一对の側縁部には、それぞれ表示画面を反対方向（外側）に向けてディスプレイ（例えば、液晶ディスプレイ）8が取り付けられている。また、ベース板11の2つのディスプレイ8の間に、図示しない制御基板とスピーカ6が配置されている。

【0057】

カバー12は、矩形状のベース板11に対応する直方体状に形成され、ベース板11を覆うように取り付けられている。カバー12の上述の2つのディスプレイ8に対応する互いに平行な2つの側面には、ディスプレイ8の表示画面を外部から視認可能とする窓部12aが設けられている。また、カバー12の天板には、スピーカ6に対応して開口部12bが設けられている。カバー12の開口部12bの部分には、橋状にカメラ固定部12cが設けられ、このカメラ固定部12cに全方位カメラ7が取り付けられている。なお、カバー12のマイク5に対応する位置には、1つか複数の孔を設けてもよい。

【0058】

また、この音声入出力機能付き撮像装置1aは、図2に示すように、テーブルTの互いに平行な2つの側縁にそれぞれ参加者Pが並んで座る場合に好適に用いられるように、2つのディスプレイ8を互いに反対向きに配置している。また、ディスプレイ8としては、例えば、7インチから15インチ程度の比較的画面の小さいディスプレイ8を用いており、テーブルT上に置いた場合に、互いに対向して坐っている参加者同士の視線を遮らないようになっている。また、ディスプレイ8にかかるコストを低減している。

【0059】

このような第2の実施の形態の音声入出力機能付き撮像装置1aによれば、第1の実施の形態の音声入出力機能付き撮像装置1と略同様の作用効果を得ることができる。また、全方位カメラ7の近傍にディスプレイ8が設けられており、上述のようにテーブルTの回りに参加者が着席した場合に、頭の向きを斜めにしたりすることなく、正面を向いた状態で無理なくディスプレイ8を見ることができる。

【0060】

また、参加者Pがディスプレイ8の方を向くと、ディスプレイ8の近傍でディスプレイ8の略上に全方位カメラ7があることにより、全方位カメラ7を見ることになり、参加者の略全員が他の会場の参加者を見ているような画像データを得られる。すなわち、第1の実施の形態では、主に発言者が全方位カメラ7を見て発言するように促す構造であったが、他の参加者は、全方位カメラ7と異なる場所にあるディスプレイ8を見ている可能性があり、発言者以外の参加者が全方位カメラ7を見ておらず、発言者以外の参加者がよそを向いている画像が撮像されるのを抑制することが困難であった。

10

【0061】

それに対して、第2の実施の形態では、全方位カメラ7の近傍に、ディスプレイ8を配置し、参加者がディスプレイ8を見ると参加者の顔が全方位カメラ7の方向くことになる。また、発言者もディスプレイ8を見るために、マイク5、スピーカ6、全方位カメラ7の方向から顔の向きを逸らす必要がなくなり、発言中は、全方位カメラ7に顔を向けた状態となる。

20

【0062】

次に、本発明の第3の実施の形態を説明する。

図6(a)、(b)に示すように、第3の実施の形態の音声入出力機能付き撮像装置1bは、第1の実施の形態の音声入出力機能付き撮像装置1と同様に、ベース板21、カバー22、制御基板(図1の制御基板4)、マイク5、スピーカ6、全方位カメラ7を備える。第3の実施の形態の音声入出力機能付き撮像装置1bは、第2の実施の形態の場合と同様に、ディスプレイ8を備える。

【0063】

本実施の形態において、ベース板21は、三角形の板状に形成され、その3つの隅部のそれぞれにマイク5が備えられている。また、ベース板21の3つの側縁部には、それぞれ表示画面を外側に向けてディスプレイ8が取り付けられている。また、ベース板11の3つのディスプレイ8の内側に、図示しない制御基板とスピーカ6が配置されている。

30

【0064】

カバー22は、三角形のベース板21に対応する三角柱状に形成され、ベース板21を覆うように取り付けられている。カバー22の3つの側面それぞれのディスプレイ8に対応する位置には、ディスプレイ8の表示画面を外部から視認可能とする窓部22aが設けられている。また、カバー22の天板には、スピーカ6に対応して開口部22bが設けられている。カバー22の開口部22bの部分には、Y字橋状にカメラ固定部22cが設けられ、このカメラ固定部22cに全方位カメラ7が取り付けられている。なお、カバー22のマイク5に対応する位置には、1つか複数の孔を設けてもよい。

【0065】

40

第3の実施の形態の音声入出力機能付き撮像装置1bは、基本的にディスプレイ8とマイク5の数の違いと、平面形状が四角形か三角形かの違い以外は、第2の実施の形態の音声入出力機能付き撮像装置1aと略同様の構造を有するものであり、同様の作用効果を奏する。また、第3の実施の形態では、ディスプレイ8が互いに120度離れた3方向を向いているので、テーブルTの周囲でディスプレイ8の画面が見られない死角となる方向を減らすことができる。なお、第2の実施の形態の形状で、カバー12の全ての側面にディスプレイ8を設けることで、音声入出力機能付き撮像装置1aが4つのディスプレイ8を持つものとしてもよい。

【0066】

次に、本発明の第4の実施の形態を説明する。

50

図 7 (a)、(b) に示すように、第 4 の実施の形態の音声入出力機能付き撮像装置 1 c は、第 1 の実施の形態の音声入出力機能付き撮像装置 1 と同様に、制御基板 (図 1 の制御基板 4)、マイク 5 a、スピーカ 6 a、全方位カメラ 7 a を備える。第 4 の実施の形態の音声入出力機能付き撮像装置 1 c は、第 2、第 3 の実施の形態の場合と同様に、ディスプレイ 8 a を備える。

【0067】

本実施の形態においては、例えば、15 インチより大きいディスプレイ 8 a として、例えば 20 ~ 32 インチ程度 (それ以上であってもよい) のディスプレイ 8 a に音声入出力機能付き撮像装置 1 c の制御基板、マイク 5 a、スピーカ 6 a を組み込み、ディスプレイ 8 a の上面の中央部に、全方位カメラ 7 a が取り付けられている。すなわち、パソコン用ディスプレイなどで、スピーカとマイクを内蔵するディスプレイに制御基板と全方位カメラ 7 a を設けた構成となっている。但し、ディスプレイ 8 a をパソコンと接続し、制御基板のデータの入出力以外の機能をパソコン PC に持たせるものとしてもよい。この場合にディスプレイ 8 a とパソコン PC の接続は、スピーカ、マイク、カメラを備えるタイプのディスプレイをパソコン PC に接続する場合と同様に行うことができる。

【0068】

図 7 (a)、(b) に示すように、第 4 の実施の形態では、ディスプレイ 8 a は、表裏面の両方に表示画面 14 a、14 b を有するものであり、図 2 に示すように、互いに対向して坐る参加者がそれぞれ別の表示画面 14 a、14 b を見るようになっている。なお、音声入出力機能付き撮像装置 1 c において、裏面側に表示画面 14 b を設けないものとして、複数台の音声入出力機能付き撮像装置 1 c を用いるものとしてもよい。この場合に、全方位カメラ 7 a が複数となるが、必ずしも複数台の全方位カメラ 7 a を必要としないので、全方位カメラ 7 a を有するタイプと、全方位カメラ 7 a が無いタイプとを組み合わせるものとしてもよい。また、ディスプレイ 8 の大きさによっては、参加者に対して全方位カメラ 7 a の位置が高くなり過ぎて、半球状の撮影範囲に参加者の胸像部分の上部しか映らない可能性があり、最悪、参加者の顔の上部しか映らない可能性がある。そこで、撮影範囲が半球より広く全球に近い撮影範囲を有する全天球カメラを全方位カメラ 7 a とすることが好ましい。

【0069】

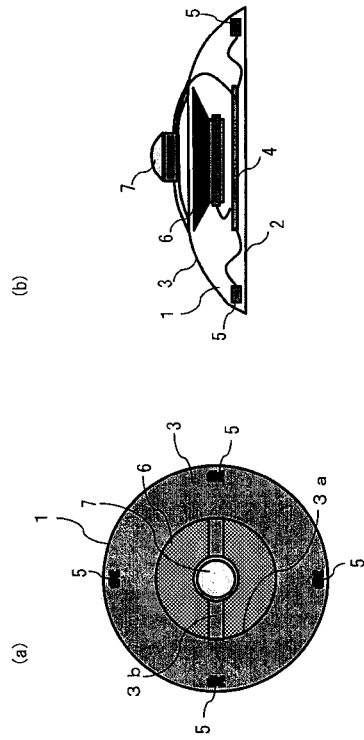
第 4 の実施の形態の音声入出力機能付き撮像装置 1 c によれば、第 1 および第 2 の実施の形態と略同様の作用効果を得ることができる。

【符号の説明】

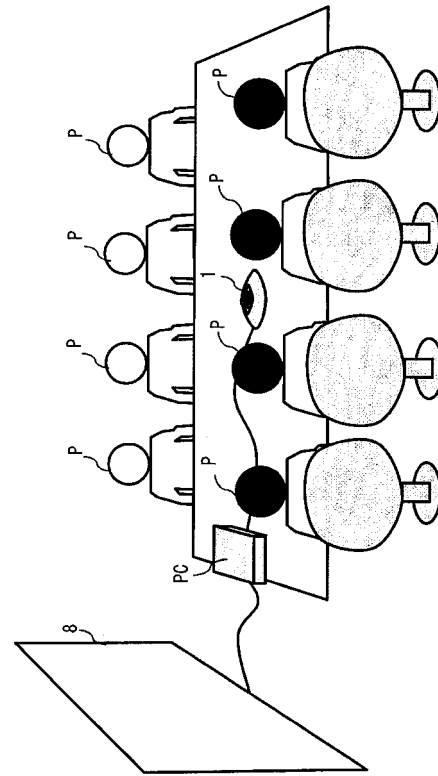
【0070】

- 1, 1 a, 1 b, 1 c 音声入出力機能付き撮像装置
- 4 制御基板 (画像認識手段、画像処理手段、音源方向認識手段、通信手段)
- 5, 5 a マイク (音声入力手段)
- 6, 6 a スピーカ (音声出力手段)
- 7, 7 a 全方位カメラ
- 8, 8 a ディスプレイ

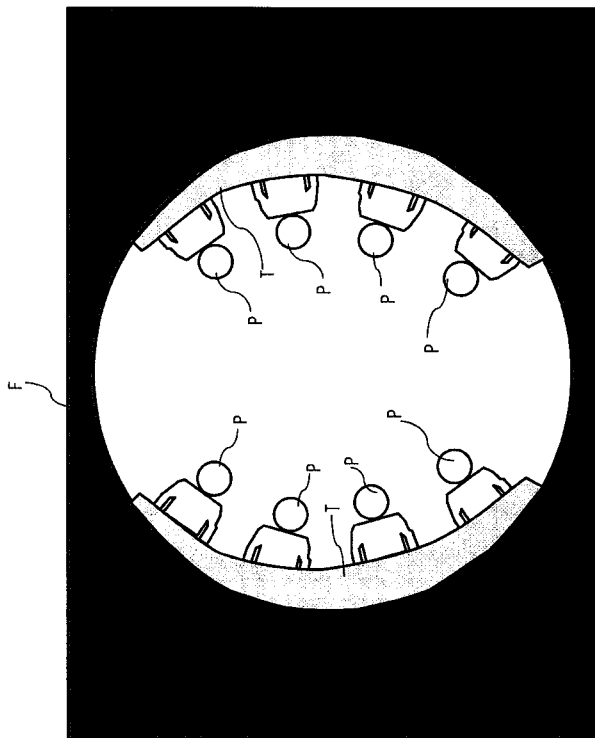
【図 1】



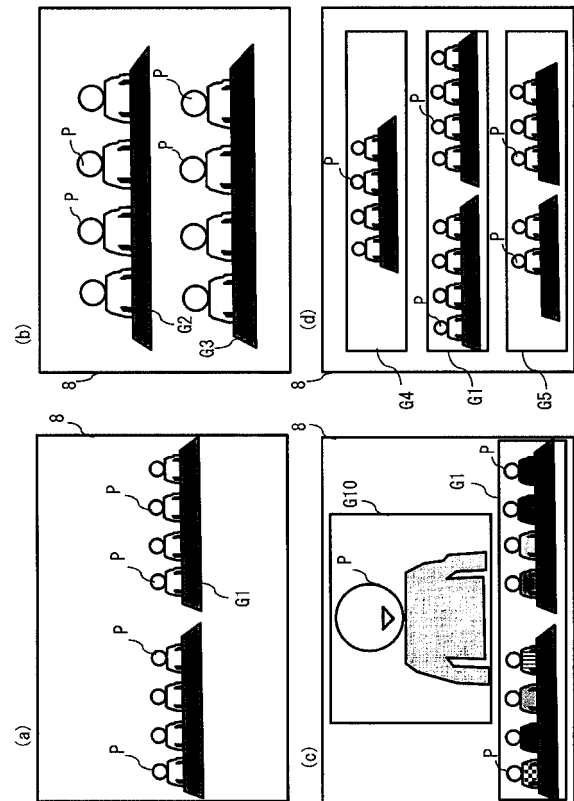
【図 2】



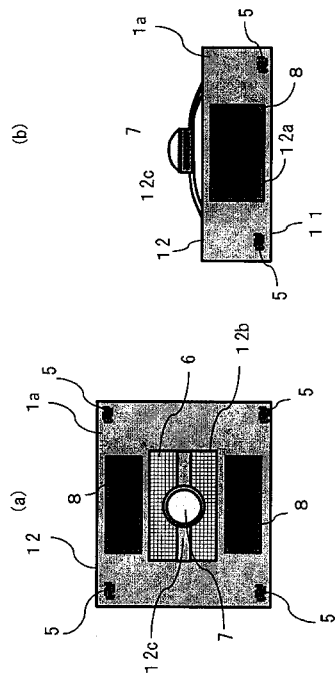
【図 3】



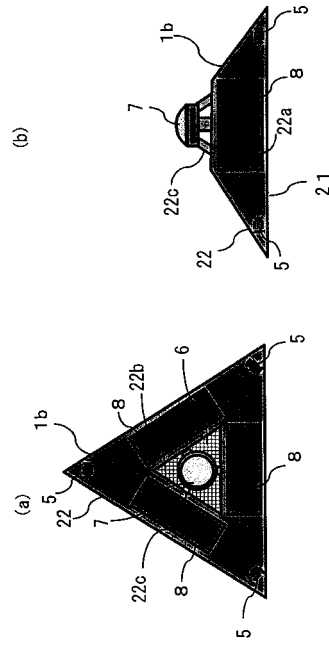
【図 4】



【図 5】



【図 6】



【図 7】

