

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5617246号
(P5617246)

(45) 発行日 平成26年11月5日(2014.11.5)

(24) 登録日 平成26年9月26日(2014.9.26)

(51) Int. Cl.			F I		
HO4N	5/00	(2011.01)	HO4N	5/00	A
HO4M	1/00	(2006.01)	HO4M	1/00	U
HO4Q	9/00	(2006.01)	HO4Q	9/00	301E
GO6T	7/00	(2006.01)	GO6T	7/00	C
GO6F	13/00	(2006.01)	GO6F	13/00	358A

請求項の数 10 (全 27 頁)

(21) 出願番号 特願2010-4170 (P2010-4170)
 (22) 出願日 平成22年1月12日 (2010.1.12)
 (65) 公開番号 特開2011-146796 (P2011-146796A)
 (43) 公開日 平成23年7月28日 (2011.7.28)
 審査請求日 平成24年11月27日 (2012.11.27)

(73) 特許権者 000002185
 ソニー株式会社
 東京都港区港南1丁目7番1号
 (74) 代理人 100095957
 弁理士 亀谷 美明
 (74) 代理人 100096389
 弁理士 金本 哲男
 (74) 代理人 100101557
 弁理士 萩原 康司
 (74) 代理人 100128587
 弁理士 松本 一騎
 (72) 発明者 福地 正樹
 東京都港区港南1丁目7番1号 ソニー株式会社内

最終頁に続く

(54) 【発明の名称】 画像処理装置、物体選択方法及びプログラム

(57) 【特許請求の範囲】

【請求項1】

実空間を映した入力画像を用いて生成される環境マップであって前記実空間内に存在する1つ以上の物体の位置を表現する当該環境マップに含まれる物体のうち、ユーザにより操作可能な物体を識別するための物体識別データに基づいて操作可能であると認識される少なくとも1つの候補物体から、操作すべき物体を選択する選択部、

を備え、

前記選択部は、ユーザインタフェースを介してユーザ入力が検知された時点で前記入力画像の中央又は当該中央の近傍に位置している候補物体を前記操作すべき物体として選択する、

画像処理装置。

【請求項2】

前記選択部は、複数の前記候補物体が前記入力画像の中央に位置している場合には、画面上での大きさが最も小さい前記候補物体を前記操作すべき物体として選択する、請求項1に記載の画像処理装置。

【請求項3】

前記画像処理装置は、

前記候補物体を操作するための制御コマンドを表示する第1の表示オブジェクトを、表示装置の画面上の前記候補物体の位置に付すように、前記表示装置への画像の出力を制御する制御部、をさらに備える、

請求項 1 又は請求項 2 に記載の画像処理装置。

【請求項 4】

前記物体識別データは、前記実空間内に存在する前記 1 つ以上の物体の各々がユーザにより操作可能であるかを示すフラグを含み、

前記選択部は、前記フラグに基づいて前記候補物体を選択する、

請求項 1 ~ 3 のいずれか 1 項に記載の画像処理装置。

【請求項 5】

前記入力画像は、前記実空間を撮影する撮像装置から取得され、

前記画像処理装置は、

前記候補物体の画面上での位置をユーザに認識させるための出力画像を前記入力画像から生成し、生成した当該出力画像を表示装置へ出力する画像出力部、

をさらに備える、請求項 1 に記載の画像処理装置。

10

【請求項 6】

前記出力画像内の各候補物体の位置は、前記環境マップ及び前記撮像装置の位置を表す撮像位置データに基づいて計算される、

請求項 5 に記載の画像処理装置。

【請求項 7】

前記画像出力部は、前記画面上で各候補物体を囲む枠を、各物体の 3 次元的な形状を表す形状データに基づいて生成し、生成した当該枠を前記入力画像に重畳することにより前記出力画像を生成する、

請求項 5 又は請求項 6 に記載の画像処理装置。

20

【請求項 8】

前記画像処理装置は、

前記撮像装置及び前記表示装置をさらに備え、

前記撮像装置の撮像面と前記表示装置の画面とは、互いに反対方向に向けて配置される、

請求項 5 ~ 7 のいずれか 1 項に記載の画像処理装置。

【請求項 9】

画像処理装置により実行される画像処理方法において、

実空間を映した入力画像を用いて生成される環境マップであって前記実空間内に存在する 1 つ以上の物体の位置を表現する当該環境マップに含まれる物体のうち、ユーザにより操作可能な物体を識別するための物体識別データに基づいて操作可能であると認識される少なくとも 1 つの候補物体から、操作すべき物体を選択すること、

を含み、

前記画像処理装置は、ユーザインタフェースを介してユーザ入力が検知された時点で前記入力画像の中央又は当該中央の近傍に位置している候補物体を前記操作すべき物体として選択する、

画像処理方法。

30

【請求項 10】

画像処理装置を制御するコンピュータを、

実空間を映した入力画像を用いて生成される環境マップであって前記実空間内に存在する 1 つ以上の物体の位置を表現する当該環境マップに含まれる物体のうち、ユーザにより操作可能な物体を識別するための物体識別データに基づいて操作可能であると認識される少なくとも 1 つの候補物体から、操作すべき物体を選択する選択部、

として機能させ、

前記選択部は、ユーザインタフェースを介してユーザ入力が検知された時点で前記入力画像の中央又は当該中央の近傍に位置している候補物体を前記操作すべき物体として選択する、

プログラム。

40

50

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像処理装置、物体選択方法及びプログラムに関する。

【背景技術】

【0002】

家庭内又はオフィス内においてユーザが操作すべき電子機器は年々増加している。また、1つの電子機器が有する機能も多様化している。そのため、これら電子機器を操作するユーザの負担は大きくなっている。そこで、1つの遠隔制御装置を用いて複数の電子機器を操作することを可能にする技術の開発が進められている。しかし、遠隔制御機能を1つの装置に集約する場合には、遠隔操作が可能な被制御機器がユーザの位置する環境内に複数存在する際に、ユーザがいずれの機器を操作しようとしているのかを特定する手段が提供されることが求められる。

10

【0003】

例えば、下記特許文献1は、音声認識技術を応用し、ユーザの発する音声から遠隔制御の対象機器及び制御内容を認識する遠隔制御装置を提案している。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開平11-345024号公報

20

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、音声から遠隔制御の対象機器及び制御内容を認識しようとする、同じ種類の機器が複数存在する場合に、個々の機器を互いに区別することが煩雑となる。例えば、リビングとリビングに隣接する部屋とにそれぞれテレビが存在する場合、「リビングの」「小さい方の」「テレビ」などのように修飾語を付さなければ、意図するテレビを適切に指定することができない。また、対象機器を指定する語が認識されなかったときには、ユーザが1つのテレビのスイッチをオフにしようとしたにも関わらず他の機器（他のテレビ又は他の種類の機器）のスイッチまでオフになってしまう恐れがある。

30

【0006】

また、例えば、ユーザのジェスチャから遠隔制御の対象機器等を認識することも考えられるが、ジェスチャのみでは、ユーザの視線上で重なる位置にある（又は互いに近傍に位置する）複数の機器を互いに区別することが難しい。

【0007】

そこで、本発明は、複数の電子機器を互いに区別して遠隔制御の対象として容易に選択することのできる、新規かつ改良された画像処理装置、物体選択方法及びプログラム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】

【0008】

ある実施形態によれば、実空間を映した入力画像を用いて生成される環境マップであって前記実空間内に存在する1つ以上の物体の位置を表現する当該環境マップに含まれる物体のうち、操作可能な物体を識別するための物体識別データに基づいて操作可能であると認識される少なくとも1つの物体を、ユーザによる操作の対象の候補である候補物体として選択する選択部、を含む画像処理装置が提供される。

40

【0009】

また、上記入力画像は、上記実空間を撮影する撮像装置から取得され、上記画像処理装置は、上記選択部により選択された上記候補物体の画面上での位置をユーザに認識させるための出力画像を上記入力画像から生成し、生成した当該出力画像を表示装置へ出力する

50

画像出力部、をさらに備えてもよい。

【0010】

また、上記出力画像内の各候補物体の位置は、上記環境マップ及び上記撮像装置の位置を表す撮像位置データに基づいて計算されてもよい。

【0011】

また、上記画像処理装置は、上記候補物体のうち操作すべき物体をユーザに指定させるためのユーザインタフェース、をさらに備えてもよい。

【0012】

また、上記ユーザインタフェースは、上記候補物体のうちユーザ入力が発知された時に上記出力画像の中央又は当該中央の近傍に位置している物体を上記操作すべき物体として決定してもよい。

【0013】

また、上記ユーザインタフェースは、上記候補物体のうちユーザ入力が発知された時に複数の上記候補物体が上記出力画像の中央に位置している場合には、上記出力画像内での大きさが最も小さい上記候補物体を上記操作すべき物体として決定してもよい。

【0014】

また、上記画像出力部は、上記画面上で各候補物体を囲む枠を、各物体の3次元的な形状を表す形状データに基づいて生成し、生成した当該枠を上記入力画像に重畳することにより上記出力画像を生成してもよい。

【0015】

また、上記画像処理装置は、上記撮像装置及び上記表示装置をさらに備え、上記撮像装置の撮像面と上記表示装置の画面とは、互いに反対方向に向けて配置されてもよい。

【0016】

また、別の実施形態によれば、画像処理装置により実行される画像処理方法において、実空間を映した入力画像を用いて生成される環境マップであって前記実空間内に存在する1つ以上の物体の位置を表現する当該環境マップに含まれる物体のうち、操作可能な物体を識別するための物体識別データに基づいて操作可能であると認識される少なくとも1つの物体を、ユーザによる操作の対象の候補である候補物体として選択すること、を含む画像処理方法が提供される。

【0017】

また、別の実施形態によれば、画像処理装置を制御するコンピュータを、実空間を映した入力画像を用いて生成される環境マップであって前記実空間内に存在する1つ以上の物体の位置を表現する当該環境マップに含まれる物体のうち、操作可能な物体を識別するための物体識別データに基づいて操作可能であると認識される少なくとも1つの物体を、ユーザによる操作の対象の候補である候補物体として選択する選択部、として機能させるための、プログラムが提供される。

【発明の効果】

【0018】

以上説明したように、本発明に係る画像処理装置、物体選択方法及びプログラムによれば、複数の電子機器を互いに区別して遠隔制御の対象として容易に選択することができる。

【図面の簡単な説明】

【0019】

【図1】一実施形態に係る画像処理装置の概要を説明するための模式図である。

【図2】一実施形態に係る画像処理装置の構成の一例を示すブロック図である。

10

20

30

40

50

【図3】一実施形態に係る環境マップ生成部の詳細な構成の一例を示すブロック図である。

【図4】一実施形態に係る自己位置検出処理の流れの一例を示すフローチャートである。

【図5】オブジェクト上に設定される特徴点について説明するための説明図である。

【図6】特徴点の追加について説明するための説明図である。

【図7】予測モデルの一例について説明するための説明図である。

【図8】特徴データの構成の一例について説明するための説明図である。

【図9】一実施形態に係るオブジェクト認識処理の流れの一例を示すフローチャートである。

【図10】物体識別データの構成の一例について説明するための説明図である。

10

【図11】一実施形態に係る機器認識処理の流れの一例を示すフローチャートである。

【図12】一実施形態に係る候補物体選択処理の流れの一例を示すフローチャートである。

【図13】一実施形態に係る画像出力部により生成される出力画像の一例を示す説明図である。

【図14】一実施形態に係る画像出力部により生成される出力画像の他の例を示す説明図である。

【図15】一実施形態に係る画像出力部により生成される出力画像のまた別の例を示す説明図である。

【図16】一実施形態に係る画像出力部により生成される出力画像のさらに別の例を示す説明図である。

20

【図17】一実施形態に係る画像出力部により生成される出力画像のさらに別の例を示す説明図である。

【図18】一実施形態に係る画像出力部により生成される出力画像のさらに別の例を示す説明図である。

【図19】汎用コンピュータのハードウェア構成の一例を示すブロック図である。

【図20】一変形例に係る画像処理装置の構成の一例を示すブロック図である。

【発明を実施するための形態】

【0020】

以下に添付図面を参照しながら、本発明の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付すことにより重複説明を省略する。

30

【0021】

また、以下の順序にしたがって当該「発明を実施するための形態」を説明する。

1. 一実施形態に係る画像処理装置の概要
2. 一実施形態に係る画像処理装置の構成例
 - 2-1. 撮像部
 - 2-2. 環境マップ生成部
 - 2-3. 環境マップ記憶部
 - 2-4. データ記憶部
 - 2-5. 機器認識部
 - 2-6. 選択部
 - 2-7. 画像出力部
 - 2-8. ユーザインタフェース
 - 2-9. 送信部
3. ハードウェア構成
4. まとめ
5. 変形例

40

【0022】

- < 1. 一実施形態に係る画像処理装置の概要 >

50

図1は、本発明の一実施形態に係る画像処理装置の概要を説明するための模式図である。図1には、本発明の一実施形態に係る画像処理装置100を持つユーザが存在する環境1が示されている。

【0023】

図1を参照すると、環境1の内部には、物体10a、10b、10c、10d、10e及び10fを含む複数の物体が存在している。このうち、物体10aはテレビ、物体10bはスピーカ、物体10cはDVDプレーヤ、物体10dはプリンタ、物体10eは携帯情報端末、物体10fはテーブルである。これら物体のうち、例えば、物体10a、10b、10c及び10dは、遠隔制御され得る被制御機器である。一方、例えば、物体10e及び10fは、遠隔制御され得ない物体である。

10

【0024】

画像処理装置100は、このような環境1を撮像装置を用いて撮影し、一連の入力画像を取得する。そして、画像処理装置100は、取得した入力画像に対して後に説明する一実施形態に係る画像処理を実行した後、当該画像処理の結果として出力される画像を画像処理装置100の画面上に表示する。かかる出力画像は、ユーザにより操作される候補となる候補物体の画面上での位置をユーザに認識させることのできる画像である。

【0025】

さらに、画像処理装置100は、好適には、遠隔制御装置としての機能をも有する。画像処理装置100が遠隔制御装置として機能する際には、ユーザは、画面上に表示される上述した出力画像に基づいて、操作すべき物体を指定することができる。

20

【0026】

なお、図1では画像処理装置100の一例として携帯電話端末を示しているが、画像処理装置100はかかる例に限定されない。例えば、画像処理装置100は、撮像装置を備えた携帯情報端末、ゲーム端末、デジタルビデオカメラ又はその他の種類の電子機器であってよい。また、画像処理装置100は、例えば、視覚としてのカメラを有するロボットなどであってもよい。このような画像処理装置100の構成の一例について、次節より具体的に説明する。

【0027】

< 2 . 一実施形態に係る画像処理装置の構成例 >

図2は、本発明の一実施形態に係る画像処理装置100の構成の一例を示すブロック図である。図2を参照すると、画像処理装置100は、撮像部102、環境マップ生成部110、環境マップ記憶部120、データ記憶部130、機器認識部140、選択部160、画像出力部170、ユーザインタフェース180、及び送信部190を備える。

30

【0028】

[2 - 1 . 撮像部]

撮像部102は、例えば、CCD (Charge Coupled Device) 又はCMOS (Complementary Metal Oxide Semiconductor) などの撮像素子を有する撮像装置として実現され得る。撮像部102は、図1に例示した環境1などの実空間を撮影することにより生成した画像を、入力画像として、環境マップ生成部110、機器認識部140、及び画像出力部170へ出力する。

40

【0029】

[2 - 2 . 環境マップ生成部]

環境マップ生成部110は、撮像部102から入力される入力画像、及びデータ記憶部130により記憶されている後述するオブジェクトの特徴データに基づいて、実空間内に存在する1つ以上の物体の位置等を表現する環境マップを生成する。

【0030】

図3は、本実施形態に係る環境マップ生成部の詳細な構成の一例を示すブロック図である。図3を参照すると、環境マップ生成部110は、自己位置検出部112、画像認識部114、及び構築部116を含む。

【0031】

50

(1) 自己位置検出部

自己位置検出部 112 は、撮像部 102 から入力される入力画像、及びデータ記憶部 130 により記憶されている特徴データに基づいて、入力画像を撮影した撮像装置の位置を動的に検出する。例えば、自己位置検出部 112 は、Andrew J. Davison による “Real-Time Simultaneous Localization and Mapping with a Single Camera” (Proceedings of the 9th IEEE International Conference on Computer Vision Volume 2, 2003, pp.1403-1410) に記載されている S L A M 技術を応用することにより、撮像装置が単眼カメラを有する場合にも、当該カメラの位置及び姿勢、並びにカメラの撮像面における特徴点の位置を、フレームごとに動的に決定することができる。

【0032】

まず、図 4 を用いて、自己位置検出部 112 による S L A M 技術を応用した自己位置検出処理の全体的な流れについて説明する。次に、図 5 ~ 図 7 を用いて、自己位置検出処理の詳細を説明する。

【0033】

図 4 は、自己位置検出部 112 による S L A M 技術を応用した自己位置検出処理の流れの一例を示すフローチャートである。図 4 において、自己位置検出処理が開始すると、自己位置検出部 112 は、まず、状態変数を初期化する (ステップ S102)。本実施形態において、状態変数とは、カメラの位置及び姿勢 (回転角)、当該カメラの移動速度及び角速度、並びに 1 つ以上の特徴点の位置を要素として含むベクトルである。そして、自己位置検出部 112 は、撮像部 102 から入力画像を順次取得する (ステップ S112)。ステップ 112 からステップ S118 までの処理は、各入力画像について (即ち毎フレーム) 繰り返され得る。

【0034】

ステップ S114 では、自己位置検出部 112 は、入力画像に映る特徴点を追跡する。例えば、自己位置検出部 112 は、データ記憶部 130 により予め記憶されている特徴点ごとのパッチ (例えば特徴点を中心とする $3 \times 3 = 9$ 画素の小画像) を入力画像から検出する。ここで検出されたパッチの位置、即ち特徴点の位置は、後の状態変数の更新の際に用いられる。

【0035】

ステップ S116 では、自己位置検出部 112 は、所定の予測モデルに基づいて、例えば 1 フレーム後の状態変数の予測値を生成する。また、ステップ S118 では、自己位置検出部 112 は、ステップ S116 において生成した状態変数の予測値と、ステップ S114 において検出した特徴点の位置に応じた観測値とを用いて、状態変数を更新する。自己位置検出部 112 は、ステップ S116 及び S118 における処理を、拡張カルマンフィルタの原理に基づいて実行する。

【0036】

このような処理の結果として、毎フレーム更新される状態変数の値が出力される。以下、特徴点の追跡 (ステップ S114)、状態変数の予測 (ステップ S116)、状態変数の更新 (ステップ S118) の各処理の内容について、より具体的に説明する。

【0037】

(1-1) 特徴点の追跡

本実施形態において、データ記憶部 130 は、実空間内に存在し得る物体に対応するオブジェクトの特徴を表す特徴データを予め記憶している。特徴データには、例えば、各オブジェクトの外観の特徴を表す 1 つ以上の特徴点についての小画像、即ちパッチ (Patch) が含まれる。パッチとは、例えば、特徴点を中心とする $3 \times 3 = 9$ 画素よりなる小画像であってよい。

【0038】

図 5 は、オブジェクトの 2 つの例、並びに各オブジェクト上に設定される特徴点 (F P : Feature Point) 及びパッチの例を示している。図 5 の左のオブジェクトは、テレビを表すオブジェクトである (5a 参照)。当該オブジェクト上には、特徴点 F P 1 を含む複

10

20

30

40

50

数の特徴点が設定されている。さらに、特徴点 F P 1 と関連付けて、パッチ P t h 1 が定義されている。一方、図 5 の右のオブジェクトは、丸型テーブルを表すオブジェクトである（5 b 参照）。当該オブジェクト上には、特徴点 F P 2 を含む複数の特徴点が設定されている。さらに、特徴点 F P 2 と関連付けて、パッチ P t h 2 が定義されている。

【 0 0 3 9 】

自己位置検出部 1 1 2 は、撮像部 1 0 2 から入力画像を取得すると、入力画像に含まれる部分画像と、データ記憶部 1 3 0 に予め記憶されている図 5 に例示した特徴点ごとのパッチとを照合する。そして、自己位置検出部 1 1 2 は、照合の結果として、入力画像に含まれる特徴点の位置（例えば検出されたパッチの中心画素の位置）を特定する。

【 0 0 4 0 】

なお、特徴点の追跡（図 4 のステップ S 1 1 4 ）において、追跡される全ての特徴点に関するデータが予めデータ記憶部 1 3 0 に記憶されていなくてもよい。例えば、図 6 の例では、時刻 $T = t - 1$ において、3 つの特徴点が入力画像内で検出されている（6 a 参照）。次に、時刻 $T = t$ においてカメラの位置又は姿勢が変化すると、時刻 $T = t - 1$ において入力画像に映っていた 3 つの特徴点のうち 1 つのみが入力画像内に映っている。この場合に、自己位置検出部 1 1 2 は、入力画像内で特徴的な画素のパターンを有する位置に新たに特徴点を設定し、その新たな特徴点を後のフレームにおける自己位置検出処理に用いてもよい。例えば、図 6 の例では、時刻 $T = t$ において、5 つの新たな特徴点がオブジェクト上に設定されている（6 b 参照）。かかる点は、S L A M 技術の 1 つの特長であり、それにより、予め全ての特徴点を設定しておくコストを削減できると共に、追加される多数の特徴点を用いて処理の精度を高めることができる。

【 0 0 4 1 】

（ 1 - 2 ）状態変数の予測

本実施形態において、自己位置検出部 1 1 2 は、拡張カルマンフィルタを適用すべき状態変数として、次式に示す状態変数 X を用いる。

【 0 0 4 2 】

【数 1】

$$X = \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\omega} \\ \dot{\mathbf{x}} \\ \dot{\boldsymbol{\omega}} \\ \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_N \end{pmatrix} \quad (1)$$

【 0 0 4 3 】

式（ 1 ）における状態変数 X の第 1 の要素は、次式の通り、実空間に設定される座標系である世界座標系（ x, y, z ）でのカメラの 3 次元位置を表す。

【 0 0 4 4 】

【数 2】

$$\mathbf{X} = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} \quad (2)$$

10

【 0 0 4 5 】

また、状態変数の第2の要素は、カメラの姿勢を表す回転行列に対応する四元数（クォータニオン）を要素として有する4次元ベクトルである。なお、四元数の代わりにオイラー角を用いてカメラの姿勢が表されてもよい。また、状態変数の第3及び第4の要素は、カメラの移動速度及び角速度をそれぞれ表す。

【 0 0 4 6 】

さらに、状態変数の第5及びそれ以降の要素は、次式の通り、世界座標系での特徴点 $F P_i$ ($i = 1 \dots N$) の3次元位置 p_i をそれぞれ表す。なお、上述したように、特徴点の数 N は、処理の間変化し得る。

20

【 0 0 4 7 】

【数 3】

$$\mathbf{p}_i = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \quad (3)$$

30

【 0 0 4 8 】

自己位置検出部 112 は、ステップ S102 において初期化された状態変数 X の値、又は前フレームにおいて更新された状態変数 X の値に基づいて、最新のフレームについての状態変数の予測値を生成する。状態変数の予測値は、次式に示す多次元正規分布に従った拡張カルマンフィルタの状態方程式に従って生成される。

【 0 0 4 9 】

40

【数 4】

$$\hat{\mathbf{X}} = F(\mathbf{X}, a) + \mathbf{w} \quad (4)$$

予測される状態変数

【 0 0 5 0 】

ここで、 F はシステムの状態遷移に関する予測モデル、 a は予測条件である。また、 w

50

はガウシアンノイズであり、例えばモデル近似誤差や観測誤差等を含み得る。一般的に、ガウシアンノイズ w の平均はゼロとなる。

【 0 0 5 1 】

図 7 は、本実施形態に係る予測モデルの一例について説明するための説明図である。図 7 を参照すると、本実施形態に係る予測モデルにおける 2 つの予測条件が示されている。まず、第 1 の条件として、特徴点の世界座標系における 3 次元位置は変化しないものとする。即ち、時刻 T における特徴点 $F P k$ の 3 次元位置を p_T とすると、次式の関係が成立する。

【 0 0 5 2 】

【数 5】

$$\mathbf{p}_t = \mathbf{p}_{t-1} \quad (5)$$

10

【 0 0 5 3 】

次に、第 2 の条件として、カメラの運動は等速運動であるものとする。即ち、時刻 $T = t - 1$ から時刻 $T = t$ にかけてのカメラの速度及び角速度について、次式の関係が成立する。

【 0 0 5 4 】

【数 6】

$$\dot{\mathbf{x}}_t = \dot{\mathbf{x}}_{t-1} \quad (6)$$

$$\dot{\boldsymbol{\omega}}_t = \dot{\boldsymbol{\omega}}_{t-1} \quad (7)$$

20

30

【 0 0 5 5 】

このような予測モデル及び式 (4) に示した状態方程式に基づいて、自己位置検出部 1 1 2 は、最新のフレームについての状態変数の予測値を生成する。

【 0 0 5 6 】

(1 - 3) 状態変数の更新

そして、自己位置検出部 1 1 2 は、観測方程式を用いて、例えば状態変数の予測値から予測される観測情報と、特徴点の追跡結果として得られる実際の観測情報との誤差を評価する。なお、式 (8) における \mathbf{v} がその誤差である。

【 0 0 5 7 】

【数 7】

$$\mathbf{s} = H(\hat{\mathbf{X}}) + \mathbf{v} \quad (8)$$

観測情報

$$\hat{\mathbf{s}} = H(\hat{\mathbf{X}}) \quad (9)$$

予測される観測情報

40

50

【 0 0 5 8 】

ここで、 H は観測モデルを表す。例えば、特徴点 $F P_i$ の撮像面 ($u - v$ 平面) 上の位置を、次式のように定義する。

【 0 0 5 9 】

【数 8】

$$\tilde{\mathbf{p}}_i = \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \quad (10)$$

$F P_i$ の撮像面上の位置

【 0 0 6 0 】

ここで、カメラの位置 \mathbf{x} 、カメラの姿勢、及び特徴点 $F P_i$ の3次元位置 \mathbf{p}_i は、いずれも状態変数 X の要素として与えられる。そうすると、ピンホールモデルに従い、特徴点 $F P_i$ の撮像面上の位置は次式を用いて導かれる。

【 0 0 6 1 】

【数 9】

$$\lambda \tilde{\mathbf{p}}_i = \mathbf{A} \mathbf{R}_\omega (\mathbf{p}_i - \mathbf{x}) \quad (11)$$

【 0 0 6 2 】

ここで、 λ は正規化のためのパラメータ、 \mathbf{A} はカメラ内部パラメータ、 \mathbf{R} は状態変数 X に含まれるカメラの姿勢を表す四元数に対応する回転行列である。カメラ内部パラメータ \mathbf{A} は、入力画像を撮影する撮像装置の特性に応じて、予め次式のように与えられる。

【 0 0 6 3 】

【数 10】

$$\mathbf{A} = \begin{pmatrix} -f \cdot k_u & f \cdot k_u \cdot \cot \theta & u_o \\ 0 & -\frac{f \cdot k_v}{\sin \theta} & v_o \\ 0 & 0 & 1 \end{pmatrix} \quad (12)$$

【 0 0 6 4 】

ここで、 f は焦点距離、 θ は画像軸の直交性 (理想値は 90°)、 k_u は撮像面の縦軸のスケール (世界座標系から撮像面の座標系へのスケール変化率)、 k_v は撮像面の横軸のスケール、 (u_o, v_o) は撮像面の中心位置である。

【 0 0 6 5 】

従って、式 (11) を用いて導かれる予測される観測情報、即ち各特徴点の撮像面上の

10

20

30

40

50

位置と、図4のステップS114における特徴点の追跡結果との間の誤差を最小にする状態変数Xを探索することにより、尤もらしい最新の状態変数Xを得ることができる。

【0066】

【数11】

$$\text{最新の状態変数 } X \leftarrow \hat{X} + \text{Innov}(s - \hat{s}) \quad (13)$$

10

【0067】

自己位置検出部112は、このようにSLAM技術を応用して動的に更新したカメラ（撮像装置）の位置x及び姿勢を、構築部116及び機器認識部140へ出力する。

【0068】

（2）画像認識部

画像認識部114は、データ記憶部130により記憶されている特徴データを用いて、入力画像に映っている物体がそれぞれのオブジェクトに対応するかを特定する。そこで、まず、データ記憶部130により記憶されている特徴データについて説明する。

【0069】

（2-1）特徴データ

図8は、特徴データの構成の一例について説明するための説明図である。図8を参照すると、オブジェクトの一例である物体10gについての特徴データFD1が示されている。特徴データFD1は、オブジェクト識別子（ID: Identifier）FD11、6方向から撮影した画像データFD12、パッチデータDF13、3次元形状データFD14、及びオントロジーデータFD15を含む。

20

【0070】

オブジェクト識別子FD11は、物体10gを一意に特定するための文字列である。図8の例では、オブジェクト識別子FD11は、“TV00”である。

【0071】

画像データFD12は、例えば、対応するオブジェクトを前・後・左・右・上・下の6方向からそれぞれ撮影した6つの画像データを含む。パッチデータFD13は、各オブジェクト上に設定される1つ以上の特徴点ごとの、各特徴点を中心とする小画像の集合である。画像データFD12及びパッチデータFD13は、後述する画像認識部114によるオブジェクト認識処理のために使用され得る。また、パッチデータFD13は、前述した自己位置検出部112による自己位置検出処理のために使用され得る。

30

【0072】

3次元形状データFD14は、対応するオブジェクトの形状を認識するためのポリゴン情報、及び特徴点の3次元的な位置情報を含む。3次元形状データFD14は、後述する構築部116による環境マップ構築処理、及び画像出力部170による出力画像生成処理のために使用され得る。

40

【0073】

オントロジーデータFD15は、例えば、構築部116による環境マップ構築処理を支援するために使用され得るデータである。図8の例では、オントロジーデータFD15は、テレビである物体10gが、テレビ台又は床に対応するオブジェクトに接する可能性が高いこと、及び本棚に対応するオブジェクトに接する可能性が低いことを表している。

【0074】

（2-2）オブジェクト認識処理

図9は、画像認識部114によるオブジェクト認識処理の流れの一例を示すフローチャートである。図9を参照すると、まず、画像認識部114は、撮像部102から入力画像を取得する（ステップS212）。次に、画像認識部114は、入力画像に含まれる部分

50

画像と、特徴データに含まれる各オブジェクトの1つ以上の特徴点のパッチとを照合し、入力画像に含まれる特徴点を抽出する(ステップS214)。なお、画像認識部114によるオブジェクト認識処理に用いられる特徴点と、自己位置検出部112による自己位置検出処理に用いられる特徴点とは、必ずしも同じでなくてもよい。但し、双方の処理で用いられる特徴点が共通している場合には、画像認識部114は、自己位置検出部112による特徴点の追跡結果を再利用してもよい。

【0075】

次に、画像認識部114は、特徴点の抽出結果に基づいて、入力画像に映っているオブジェクトを特定する(ステップS216)。例えば、画像認識部114は、ある領域内で1つのオブジェクトに属す特徴点が高い密度で抽出された場合には、当該領域にそのオブジェクトが映っていると認識し得る。そして、画像認識部114は、特定したオブジェクトのオブジェクト識別子及びそのオブジェクトに属す特徴点の撮像面上の位置を、構築部116へ出力する(ステップS218)。

【0076】

(3) 環境マップ構築部

構築部116は、自己位置検出部112から入力されるカメラの位置及び姿勢、画像認識部114から入力される特徴点の撮像面上の位置、及びデータ記憶部130に記憶されている特徴データを用いて、環境マップを生成する。本明細書において、環境マップとは、実空間内に存在する1つ以上の物体の位置(及び姿勢)を表現するデータの集合である。環境マップには、例えば、物体に対応するオブジェクト識別子、当該物体に属す特徴点の3次元位置、及び当該物体の形状を構成するポリゴン情報などが含まれ得る。環境マップは、例えば、画像認識部114から入力される特徴点の撮像面上の位置から、上述したピンホールモデルに従って各特徴点の3次元位置を求めることにより構築され得る。

【0077】

式(11)に示したピンホールモデルの関係式を変形すると、特徴点 FP_i の世界座標系における3次元位置 p_i は、次式により求められる。

【0078】

【数12】

$$p_i = x + \lambda \cdot R_{\omega}^T \cdot A^{-1} \cdot \tilde{p}_i = x + d \cdot R_{\omega}^T \frac{A^{-1} \cdot \tilde{p}_i}{\|A^{-1} \cdot \tilde{p}_i\|} \quad (14)$$

【0079】

ここで、 d はカメラと各特徴点との間の世界座標系における距離を表す。構築部116は、かかる距離 d を、オブジェクトごとに少なくとも4つの特徴点の撮像面上の位置、及び当該特徴点間の距離に基づいて、算出することができる。特徴点間の距離は、図8を用いて説明した特徴データに含まれる3次元形状データFD14として、予めデータ記憶部130により記憶されている。なお、式(14)における距離 d の算出処理については、特開2008-304268号公報において詳しく説明されている。

【0080】

距離 d が算出されると、式(14)における右辺の残りの変数は、自己位置検出部112から入力されるカメラの位置及び姿勢、並びに画像認識部114から入力される特徴点の撮像面上の位置であり、いずれも既知となる。そこで、構築部116は、式(14)に従って、画像認識部114から入力される各特徴点について、世界座標系における3次元位置を計算する。そして、構築部116は、算出した特徴点ごとの3次元位置に応じて最新の環境マップを構築し、構築した環境マップを環境マップ記憶部120へ出力する。な

お、このとき、構築部 116 は、図 8 を用いて説明した特徴データに含まれるオントロジーデータ FD15 を用いて、環境マップのデータの正確性を向上させてもよい。

【0081】

[2 - 3 . 環境マップ記憶部]

環境マップ記憶部 120 は、ハードディスク又は半導体メモリなどの記憶媒体を用いて、環境マップ生成部 110 の構築部 116 により構築される上述した環境マップを記憶する。さらに、環境マップ記憶部 120 は、環境マップ生成部 110 の自己位置検出部 112 により検出されるカメラの位置及び姿勢を含む撮像位置データを記憶する。なお、環境マップ記憶部 120 は、次に説明するデータ記憶部 130 と物理的に同一の記憶媒体を用いて実装されてもよい。

10

【0082】

[2 - 4 . データ記憶部]

データ記憶部 130 は、上述したように、ハードディスク又は半導体メモリなどの記憶媒体を用いて、実空間内に存在し得る物体に対応するオブジェクトの特徴を表す特徴データを予め記憶する。また、データ記憶部 130 は、実空間内に存在し得る物体のうちユーザにより操作可能な物体を識別するための物体識別データを予め記憶する。物体識別データは、後に説明する機器認識部 140 による機器認識処理において使用される。

【0083】

図 10 は、物体識別データの構成の一例について説明するための説明図である。図 10 を参照すると、物体識別データは、オブジェクト識別子 (ID)、種別、型番、物体画像、操作可能フラグ、及びコマンドリストの 6 つのデータ項目を有する。

20

【0084】

オブジェクト識別子は、実空間内の個々の物体に対応する各オブジェクトを一意に特定するための文字列である。かかるオブジェクト識別子を用いて、物体識別データと、図 8 を用いて説明した特徴データとを関連付けることができる。図 10 では、それぞれオブジェクト ID = “TV01”、“SP01”、“DV01”、“PT01”、“PD01”及び“TB01”である 6 つの物体についての物体識別データが例示されている。

【0085】

種別は、各オブジェクトの種類を表す。オブジェクト “TV01”、“SP01”、“DV01”、“PT01”、“PD01”及び“TB01”の種別は、それぞれ、テレビ、スピーカ、DVDプレーヤ、プリンタ、PDA 及びテーブルである。また、型番は、実空間内の個々の物体の型番を表す。

30

【0086】

物体画像は、後述する機器認識部 140 による機器認識処理に用いられる画像である。物体画像は、実空間内の個々の物体を少なくとも 1 方向から撮影した画像であってよい。また、例えば、物体画像は、図 8 を用いて説明した特徴データに含まれる画像データの一部であってよい。

【0087】

操作可能フラグは、実空間内の個々の物体がユーザにより操作可能な物体であるか否かを識別するためのフラグである。操作可能フラグ = “Yes” であれば、当該物体は操作可能である。また、操作可能フラグ = “No” であれば、当該物体は操作可能でない。図 10 の例では、オブジェクト “TV01”、“SP01”、“DV01” 及び “PT01” の操作可能フラグは “Yes” である。即ち、これら物体は操作可能である。一方、オブジェクト “PD01” 及び “TB01” の操作可能フラグは “No” である。即ち、これら物体は操作可能ではない。

40

【0088】

コマンドリストは、操作可能な物体について各物体を操作するための制御コマンドのリストである。例えば、オブジェクト “TV01” のコマンドリストには、“電源 ON”、“電源 OFF”、“音量 UP”、“音量 DOWN”、“チャンネル UP” 及び “チャンネル DOWN” の 6 種類の制御コマンドが含まれる。

50

【 0 0 8 9 】

[2 - 5 . 機器認識部]

機器認識部 1 4 0 は、上述した物体識別データに含まれる各物体画像を撮像部 1 0 2 から入力される入力画像と照合することにより、入力画像に映っている操作可能な物体を認識する。そして、機器認識部 1 4 0 は、操作可能な物体が存在する場合には、入力画像内の各操作可能な物体の位置、並びに自己位置検出部 1 1 2 から入力されるカメラの位置及び姿勢を用いて、式 (1 4) に従い、環境マップ 1 2 0 内での各操作可能な物体の 3 次元位置を計算する。さらに、機器認識部 1 4 0 は、計算した 3 次元位置に基づいて、新たに認識された操作可能な物体を環境マップに登録する。

【 0 0 9 0 】

図 1 1 は、機器認識部 1 4 0 による機器認識処理の流れの一例を示す説明図である。図 1 1 を参照すると、機器認識部 1 4 0 は、まず、撮像部 1 0 2 から入力画像を取得する (ステップ S 3 0 2)。次に、機器認識部 1 4 0 は、物体識別データに含まれる各物体画像を入力画像と照合することにより、入力画像に映っている操作可能な物体を認識する (ステップ S 3 0 4)。ここで照合に用いられる物体画像とは、物体識別データに含まれる物体画像のうち、操作可能フラグが “ Y e s ” であるレコードの物体画像であってよい。次に、機器認識部 1 4 0 は、入力画像内で操作可能な物体が認識されたか否かを判定する (ステップ S 3 0 6)。ここで、入力画像内で操作可能な物体が認識されなかった場合には、処理は終了する。一方、入力画像内で操作可能な物体が認識された場合には、機器認識部 1 4 0 は、認識された各操作可能な物体の 3 次元位置を計算し、当該 3 次元位置に基づいて、新たに認識された操作可能な物体を環境マップに登録する (ステップ S 3 0 8)。

【 0 0 9 1 】

このように機器認識部 1 4 0 により操作可能であると認識された物体の環境マップ内の 3 次元位置は、環境マップ記憶部 1 2 0 により継続して保持される。即ち、一度操作可能であると認識された物体は、カメラが移動して当該物体が他の物体に隠された場合 (即ち、オクルージョンを生じた場合)、又は撮像方向が変わることにより外観が変化した場合にも、環境マップ内で正確に識別され得る。

【 0 0 9 2 】

[2 - 6 . 選択部]

選択部 1 6 0 は、環境マップ記憶部 1 2 0 により記憶されている環境マップに含まれる物体のうち、物体識別データに基づいて操作可能であると認識された少なくとも 1 つの物体を、ユーザによる操作の対象の候補として選択する。選択部 1 6 0 により操作の対象の候補として選択された物体を、本明細書では、候補物体という。

【 0 0 9 3 】

選択部 1 6 0 は、例えば、環境マップに含まれる物体のうち、物体識別データにおいて操作可能フラグが “ Y e s ” である全ての物体を候補物体として選択してもよい。その代わりに、選択部 1 6 0 は、例えば、入力画像内で複数の操作可能な物体が認識される場合には、予め設定される閾値を超えない数の一部の物体のみを候補物体として選択してもよい。一部の物体とは、例えば、画面の中央又は当該中央の近傍に位置している物体であってよい。操作可能な物体の画面内の位置は、環境マップ記憶部 1 2 0 により記憶されている環境マップ並びにカメラの位置及び姿勢に基づいて、式 (1 1) に従って計算され得る。なお、操作可能な物体は、画面内で通常ある程度の大きさ (例えば画素数により表される面積など) を有する。そのため、選択部 1 6 0 は、操作可能な物体の幾何学的な重心位置を当該操作可能な物体の位置とみなして、画面の中央からの距離を決定することができる。また、選択部 1 6 0 は、例えば、複数の操作可能な物体が画面の中央に位置している場合には、画面内での大きさが最も小さい物体を候補物体として選択してもよい。

【 0 0 9 4 】

図 1 2 は、選択部 1 6 0 による選択処理の流れの一例を示す説明図である。図 1 2 を参照すると、選択部 1 6 0 は、まず、環境マップに含まれる物体のうち、物体識別データにおいて操作可能フラグが “ Y e s ” である物体の 3 次元座標を、式 (1 1) に従って撮像

10

20

30

40

50

面上の2次元座標に変換する(ステップS402)。例えば、物体の形状を構成するポリゴンの頂点の3次元座標を撮像面上の2次元座標に変換することにより、各物体の撮像面上での大きさ及び重心を算出することが可能となる。次に、選択部160は、式(11)により撮像面内に投影される操作可能な物体の数が予め設定される閾値よりも多いか否かを判定する(ステップS404)。予め設定される閾値とは、例えば、1であってもよく複数であってもよい。ここで、操作可能な物体の数が予め設定される閾値よりも多くない場合には、処理はステップS412へ進む。一方、操作可能な物体の数が予め設定される閾値よりも多い場合には、処理はステップS406へ進む。

【0095】

ステップS406では、選択部160は、複数の操作可能な物体が撮像面の中央で重なっているか否かを判定する(ステップS406)。例えば、カメラの位置から1つの物体の位置へ伸びる直線上に他の物体が存在している場合には、複数の物体が撮像面上で重なり得る。ここで、複数の操作可能な物体が撮像面の中央で重なっている場合には、処理はステップS408へ進む。一方、複数の操作可能な物体が撮像面の中央で重なっていない場合には、処理はステップS410へ進む。

【0096】

ステップS408では、選択部160は、撮像面の中央に位置する複数の操作可能な物体のうち撮像面上での大きさがより小さい物体(上記閾値が1である場合には最も小さい物体)を、候補物体として選択する(ステップS408)。また、ステップS410では、選択部160は、画面中央から各物体の重心までの距離が短い順に、上記閾値に相当する数の操作可能な物体を、候補物体として選択する(ステップS410)。また、ステップS412では、選択部160は、撮像面内に投影される操作可能な物体を、候補物体として選択する(ステップS412)。

【0097】

その後、選択部160は、選択した候補物体の2次元形状を表す形状データ(ポリゴンの頂点の撮像面上での2次元座標など)をオブジェクトIDと関連付けて、画像出力部170へ出力する(ステップS414)。そして、選択部160による選択処理は終了する。

【0098】

[2-7. 画像出力部]

画像出力部170は、選択部160により選択された候補物体の画面上での位置をユーザに認識させるための出力画像を、撮像部102から入力される入力画像から生成する。そして、画像出力部170は、生成した当該出力画像を表示装置へ出力する。より具体的には、例えば、画像出力部170は、画面上で各候補物体を囲む枠を当該候補物体の2次元形状データに基づいて生成し、生成した枠を入力画像に重畳することにより、出力画像を生成する。その代わりに、画像出力部170は、例えば、出力画像において各候補物体に対応する画面上の領域と他の領域との間で明暗又は色の濃淡などを変化させることにより、候補物体の画面上での位置をユーザに認識させてもよい。また、画像出力部170は、例えば、各候補物体に当該物体の名称等を表す文字列を付すことにより、候補物体の画面上での位置をユーザに認識させてもよい。

【0099】

図13は、本実施形態に係る画像出力部170により生成される出力画像の一例を示す説明図である。図13を参照すると、環境1を撮影した入力画像から生成された出力画像Im11が示されている。出力画像Im11において、4つの操作可能な物体10a、10b、10c及び10dに、各物体を囲む枠が付されている。なお、枠の形状は、円形又は矩形以外の多角形などであってもよい。図13の例では、環境マップに含まれる物体のうち、物体識別データにおいて操作可能フラグが“Yes”である4つの物体が全て候補物体として選択されたことが理解され得る。このような出力画像を参照することで、ユーザは、環境1内に存在するいずれの物体を画像処理装置100を用いて操作することができるかを知ることができる。

10

20

30

40

50

【 0 1 0 0 】

図 1 4 は、本実施形態に係る画像出力部 1 7 0 により生成される出力画像の他の例を示す説明図である。図 1 4 を参照すると、環境 1 を撮影した入力画像から生成された出力画像 I m 1 2 が示されている。出力画像 I m 1 2 において、画面の中央の最も近くに位置する操作可能な物体 1 0 a に当該物体を囲む枠が付されている。即ち、図 1 4 の例では、環境マップに含まれる物体のうち 1 つの物体のみが候補物体として選択されたことが理解され得る。また、出力画像 I m 1 2 には、画面の中央を指し示す照準 1 2 も重畳されている。このような出力画像によれば、ユーザは、画像処理装置 1 0 0 (又は撮像部 1 0 2) を動かすことで操作対象としたい機器に照準 1 2 を合わせて、容易に操作対象を指定することができる。

10

【 0 1 0 1 】

図 1 5 は、本実施形態に係る画像出力部 1 7 0 により生成される出力画像のまた別の例を示す説明図である。図 1 5 を参照すると、環境 1 を撮影した入力画像から生成された出力画像 I m 1 3 が示されている。出力画像 I m 1 3 において、画面の中央の最も近くに位置する操作可能な物体 1 0 a に当該物体を囲む枠が付されている。また、出力画像 I m 1 3 には、画面の中央を指し示す照準 1 2 も重畳されている。さらに、出力画像 I m 1 3 において、物体 1 0 a には、当該物体を操作するための制御コマンド (図 1 4 の例では “電源 ON”) を表示するコマンド欄 1 4 が付されている。かかる制御コマンドは、例えば、図 1 0 を用いて説明した物体識別データに含まれるコマンドリストから取得される。このような出力画像によれば、ユーザは、画像処理装置 1 0 0 (又は撮像部 1 0 2) を動かすことにより容易に操作対象を指定できると共に、操作対象へ送信すべき制御コマンドを容易に選択することができる。

20

【 0 1 0 2 】

図 1 6 は、本実施形態に係る画像出力部 1 7 0 により生成される出力画像のさらに別の例を示す説明図である。図 1 6 を参照すると、左にテンポラリ画像 I m 2 1、右に出力画像 I m 2 2 が示されている。テンポラリ画像 I m 2 1 には、操作可能な物体 1 0 c 及び 1 0 d が映っている。但し、テンポラリ画像 I m 2 1 の中央において、物体 1 0 c が部分的に物体 1 0 d に隠されている。即ち、この場合、複数の操作可能な物体が撮像面の中央で重なっている。これに対し、出力画像 I m 2 2 では、物体 1 0 c 及び 1 0 d のうち撮像面上での大きさがより小さい物体 1 0 c に、候補物体であることを示す枠が付されている。このような出力画像によれば、操作可能な小さな物体が他の操作可能な物体と重なっていることにより当該小さな物体を選択し得なくなることが防がれる。

30

【 0 1 0 3 】

図 1 7 は、本実施形態に係る画像出力部 1 7 0 により生成される出力画像のさらに別の例を示す説明図である。図 1 7 を参照すると、左に出力画像 I m 3 1、右に出力画像 I m 3 2 が示されている。出力画像 I m 3 1 には、操作可能な物体 1 0 b が映っている。かかる物体 1 0 b が操作可能であることは、機器認識部 1 4 0 による認識の結果として環境マップへ登録される。出力画像 I m 3 2 は、その後、画像処理装置 1 0 0 が環境 1 の内部で移動した後に生成される出力画像である。出力画像 I m 3 2 においても、物体 1 0 b が操作可能な物体として認識され、候補物体として選択されている。このように、実空間内に存在する物体の位置を表現する環境マップに基づいて候補物体を選択し表示することにより、物体の外観が撮像位置及び撮像方向に応じて変化する場合であっても、柔軟に操作対象を選択することができる。

40

【 0 1 0 4 】

図 1 8 は、本実施形態に係る画像出力部 1 7 0 により生成される出力画像のさらに別の例を示す説明図である。図 1 8 を参照すると、左に出力画像 I m 4 1、右に出力画像 I m 4 2 が示されている。出力画像 I m 4 1 には、操作可能な物体 1 0 d が映っている。かかる物体 1 0 d が操作可能であることは、機器認識部 1 4 0 による認識の結果として環境マップへ登録される。出力画像 I m 4 2 は、その後、画像処理装置 1 0 0 が環境 1 の内部で移動した後に生成される出力画像である。出力画像 I m 4 2 においては、物体 1 0 b はテ

50

ーブルの背後に隠れていて視認できない。しかしながら、出力画像 I m 4 2 においても、テーブルの背後に隠れた物体 1 0 d が候補物体として選択されている。このように、実空間内に存在する物体の位置を表現する環境マップに基づいて候補物体を選択し、当該物体を識別可能な枠を表示することにより、物体が他の物体の背後に隠れている場合であっても、柔軟にその隠れた物体を操作対象として選択することができる。

【 0 1 0 5 】

なお、画像出力部 1 7 0 は、操作可能な物体が存在しない場合には、撮像部 1 0 2 から入力される入力画像をそのまま出力画像として表示装置へ出力してよい。また、画像出力部 1 7 0 により生成される出力画像が表示される表示装置の画面と、撮像部 1 0 2 に相当する撮像装置の撮像面とは、互いに反対方向に向けて（即ち、背面を合わせるように）配置されるのが好適である。それにより、シースルー方式のディスプレイが構成されるため、ユーザにとって、環境 1 内で画像処理装置 1 0 0 をかざして出力画像を見ながら、操作すべき物体を指定することがさらに容易となる。

【 0 1 0 6 】

[2 - 8 . ユーザインタフェース]

ユーザインタフェース 1 8 0 は、候補物体のうち操作すべき物体をユーザに指定させるためのインタフェースを提供する。ユーザインタフェース 1 8 0 は、例えば、ボタン、タッチパネル、スイッチ又はキーパッドなどの、ユーザ入力を検出可能な任意の入力手段であってよい。例えば、図 1 4 の出力画像 I m 1 2 又は図 1 6 の出力画像 I m 2 2 などのように 1 つの候補物体のみが表示されている場合には、ユーザ入力（例えばボタンの押下など）が検知された時点で表示されている候補物体が、操作すべき物体であるものと決定され得る。また、図 1 3 の出力画像 I m 1 1 のように複数の候補物体が表示されている場合には、例えば、ユーザ入力（例えばボタンの押下など）が検知された時点で表示されている出力画像の中央又は当該中央の近傍に位置している物体が、操作すべき物体であるものとして決定され得る。また、ユーザ入力（例えばボタンの押下など）が検知された時に複数の候補物体が出力画像の中央に位置している場合には、例えば、ユーザインタフェース 1 8 0 は、出力画像内での大きさが最も小さい候補物体を操作すべき物体であるものとして決定してもよい。また、ユーザインタフェース 1 8 0 は、G U I (graphical User Interface) を含んでもよい。例えば、図 1 5 の出力画像 I m 1 3 内のコマンド欄 1 4 は、G U I の一種である。例えば、コマンド欄 1 4 に表示された制御コマンドをユーザが画面上で選択した場合に、次に説明する送信部 1 9 0 は、当該制御コマンドを含む制御信号を、コマンド欄 1 4 が付された操作対象の物体へ送信することができる。

【 0 1 0 7 】

[2 - 9 . 送信部]

送信部 1 9 0 は、ユーザインタフェース 1 8 0 に対するユーザ入力に基づいて決定した操作対象の物体への制御コマンドを含む制御信号を生成し、生成した当該制御信号を送信する。制御信号は、例えば、I r D A (Infrared Data Association) に準拠した赤外線信号として送信されてもよい。その代わりに、制御信号は、例えば、無線 L A N (Local Area Network) などによる無線信号として送信されてもよい。かかる制御信号を操作対象の物体（被制御機器）が受信することにより、ユーザの望む操作が当該物体において行われる。

【 0 1 0 8 】

< 3 . ハードウェア構成 >

上述した画像処理装置 1 0 0 による一連の処理をハードウェアで実現するかソフトウェアで実現するかは問わない。一連の処理又はその一部をソフトウェアで実行させる場合には、ソフトウェアを構成するプログラムが、専用のハードウェアに組み込まれたコンピュータ、又は例えば図 1 9 に示した汎用コンピュータなどを用いて実行される。

【 0 1 0 9 】

図 1 9 において、C P U (Central Processing Unit) 9 0 2 は、汎用コンピュータ

10

20

30

40

50

の動作全般を制御する。ROM (Read Only Memory) 904には、一連の処理の一部又は全部を記述したプログラム又はデータが格納される。RAM (Random Access Memory) 906には、処理の実行時にCPU 902により用いられるプログラムやデータなどが一時的に記憶される。

【0110】

CPU 902、ROM 904、及びRAM 906は、バス910を介して相互に接続される。バス910にはさらに、入出力インタフェース912が接続される。

【0111】

入出力インタフェース912は、CPU 902、ROM 904、及びRAM 906と、入力装置920、表示装置922、記憶装置924、撮像装置926、及び送信装置930とを接続するためのインタフェースである。

10

【0112】

入力装置920は、例えばボタン、タッチパネル、スイッチ又はキーパッドなどの入力手段を介して、ユーザ入力を受け付ける。表示装置922は、例えばCRT (Cathode Ray Tube)、液晶ディスプレイ、OLED (Organic Light Emitting Diode) などを用いて構成され、その画面上に出力画像を表示させる。

【0113】

記憶装置924は、例えばハードディスクドライブ又は半導体メモリなどにより構成され、プログラムやデータを記憶する。撮像装置926は、上述した撮像部102のハードウェアに相当し、CCD又はCMOSなどの撮像素子を用いて、実空間を撮影する。送信装置930は、例えばIrDAに準拠した赤外線ポートなどにより構成され、電子機器を遠隔制御するための制御コマンドを含む制御信号を赤外線信号などに変調して送信する。

20

【0114】

< 4 . まとめ >

ここまで、図1～図19を用いて、本発明の一実施形態に係る画像処理装置100について説明した。画像処理装置100によれば、実空間内に存在する1つ以上の物体の位置を表現する環境マップに含まれる物体のうち、物体識別データに基づいて操作可能であると認識される少なくとも1つの物体が、ユーザによる操作の対象の候補である候補物体として選択される。従って、撮像装置が移動して入力画像内の物体の外観が変化した場合、又は操作すべき物体が他の物体に隠された場合などにおいても、これら物体を互いに区別して遠隔制御の対象として容易に選択することができる。また、シースルー方式のディスプレイを用いた直感的な操作対象の選択が可能となる。さらに、電子機器に個別に操作用のディスプレイを設けることなく、あるいは個別の専用の遠隔制御装置を用意することなく、1つの画像処理装置100を用いて各電子機器を操作することが可能となる。

30

【0115】

また、出力画像を生成する際に利用される撮像装置の位置及び姿勢は、SLAM技術を用いて動的に検出される。それにより、撮像装置の位置及び姿勢がフレームごとに変化する場合であっても、出力画像に表示すべき候補物体を囲む枠などを、撮像装置の位置及び姿勢に応じて適切に生成することができる。

【0116】

< 5 . 変形例 >

なお、上述した画像処理装置100の一変形例として、環境マップを構築することなく遠隔制御装置として動作可能な画像処理装置を構成することもできる。図20は、そのような一変形例に係る画像処理装置200の構成の一例を示すブロック図である。図20を参照すると、画像処理装置200は、撮像部102、データ記憶部130、機器認識部240、選択部260、画像出力部270、ユーザインタフェース180、及び送信部190を備える。

40

【0117】

機器認識部240は、データ記憶部130に記憶されている物体識別データに含まれる各物体画像を撮像部102から入力される入力画像と照合することにより、入力画像に映

50

っている操作可能な物体を認識する。そして、機器認識部 240 は、操作可能な物体が存在する場合には、入力画像内の各操作可能な物体の位置（各物体に対応する画素群の位置など）をオブジェクト ID と関連付けて、選択部 260 及び画像出力部 270 へ出力する。

【0118】

選択部 260 は、機器認識部 240 により物体識別データに基づいて操作可能であると認識された少なくとも 1 つの物体を、候補物体として選択する。選択部 260 は、例えば、物体識別データにおいて操作可能フラグが “Yes” である全ての物体を候補物体として選択してもよい。その代わりに、選択部 260 は、例えば、入力画像内で複数の操作可能な物体が認識される場合には、予め設定される閾値を超えない数の一部の物体のみを候補物体として選択してもよい。一部の物体とは、例えば、画面の中央又は当該中央の近傍に位置している物体であってよい。また、選択部 260 は、例えば、複数の操作可能な物体が画面の中央に位置している場合には、画面内での大きさが最も小さい物体を候補物体として選択してもよい。選択部 260 は、選択した候補物体のオブジェクト ID を画像出力部 270 へ出力する。

10

【0119】

画像出力部 270 は、選択部 260 により選択された候補物体の画面上での位置をユーザに認識させるための出力画像を、撮像部 102 から入力される入力画像から生成する。そして、画像出力部 270 は、生成した当該出力画像を表示装置へ出力する。より具体的には、例えば、画像出力部 270 は、画面上で各候補物体を囲む枠を機器認識部 240 から入力される当該候補物体の位置に基づいて生成し、生成した枠を入力画像に重畳することにより、出力画像を生成する。ユーザは、かかる出力画像を見ながら、ユーザインタフェース 180 を介して操作対象を指定することができる。

20

【0120】

なお、画像出力部 270 は、上述した画像出力部 170 と同様、操作可能な物体が存在しない場合には、撮像部 102 から入力される入力画像をそのまま出力画像として表示装置へ出力してよい。画像出力部 270 により生成される出力画像が表示される表示装置の画面と、撮像部 102 に相当する撮像装置の撮像面とは、互いに反対方向に向けて配置されるのが好適である。

【0121】

かかる画像処理装置 200 によれば、物体識別データに基づいて操作可能であると認識される入力画像内の少なくとも 1 つの物体が、ユーザによる操作の対象の候補である候補物体として選択され、出力画像内に表示される。従って、ユーザは、直感的に操作対象を選択することができる。この場合にも、電子機器に個別に操作用のディスプレイを設けることなく、あるいは個別の専用の遠隔制御装置を用意することなく、1 つの画像処理装置 200 を用いて各電子機器を操作することが可能となる。

30

【0122】

以上、添付図面を参照しながら本発明の好適な実施形態について詳細に説明したが、本発明はかかる例に限定されない。本発明の属する技術の分野における通常の知識を有する者であれば、特許請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本発明の技術的範囲に属するものと了解される。

40

【符号の説明】

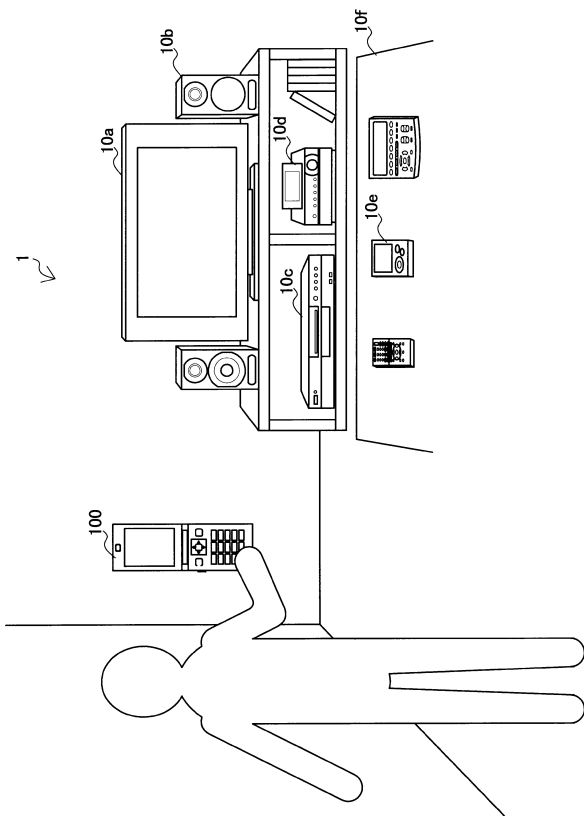
【0123】

- 100 画像処理装置
- 102 撮像部
- 110 環境マップ生成部
- 120 環境マップ記憶部
- 130 データ記憶部
- 140 機器認識部

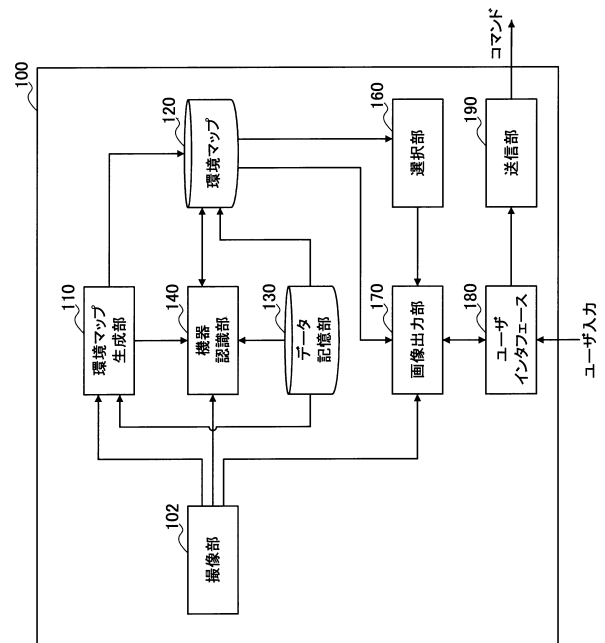
50

- 160 選択部
- 170 画像出力部
- 180 ユーザインタフェース
- 190 送信部

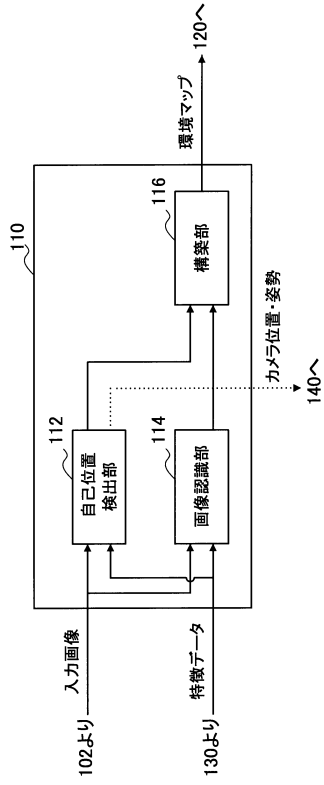
【図1】



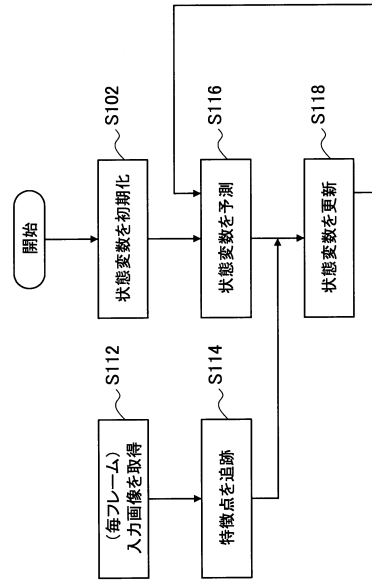
【図2】



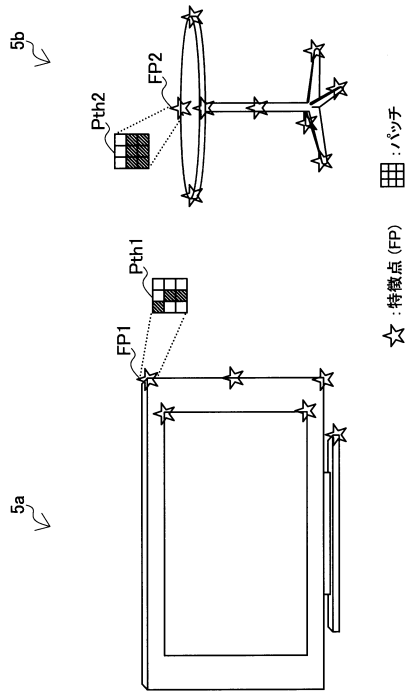
【図3】



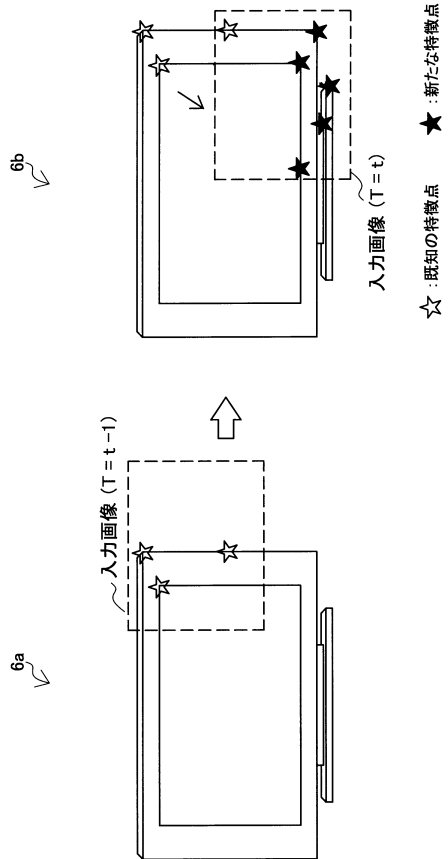
【図4】



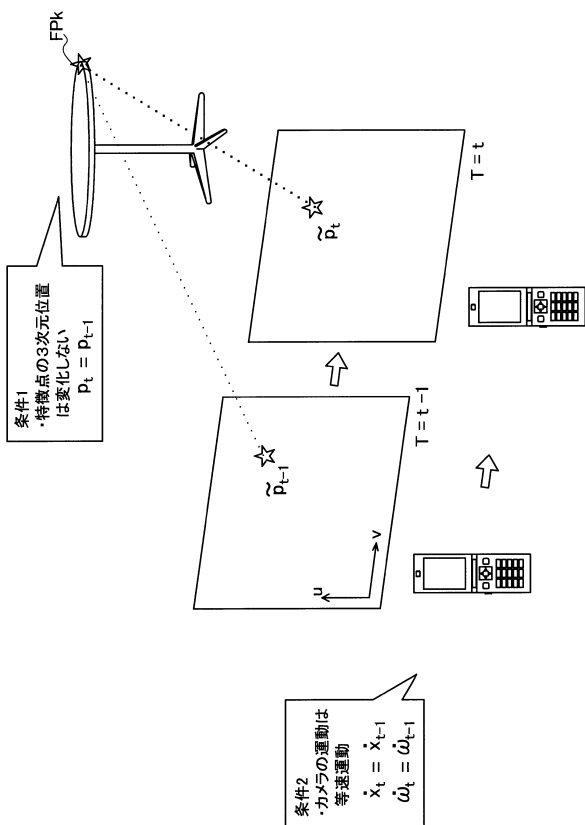
【図5】



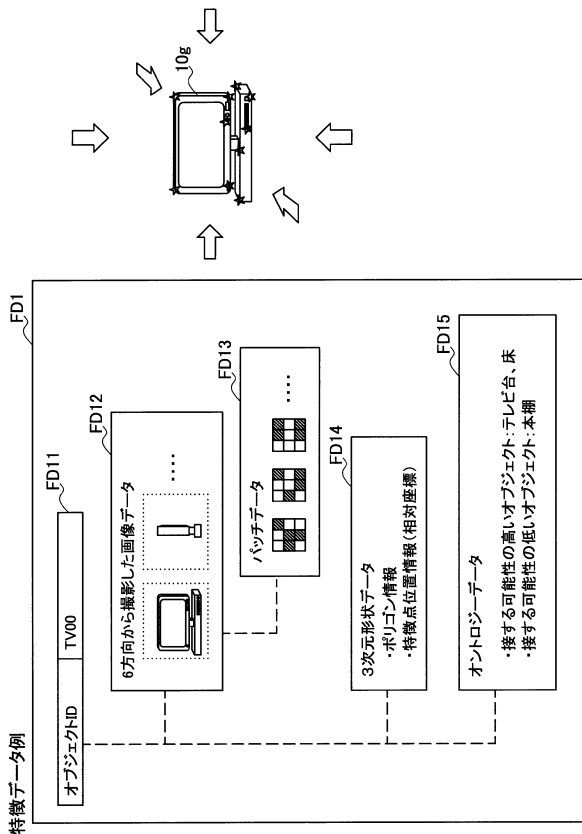
【図6】



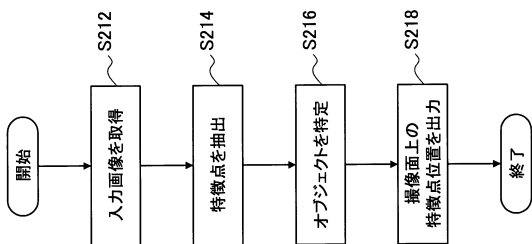
【 図 7 】



【 図 8 】



【 図 9 】

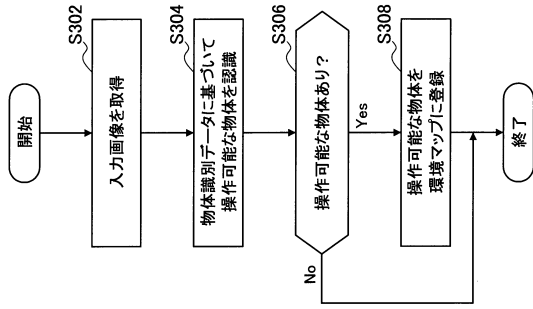


【 図 10 】

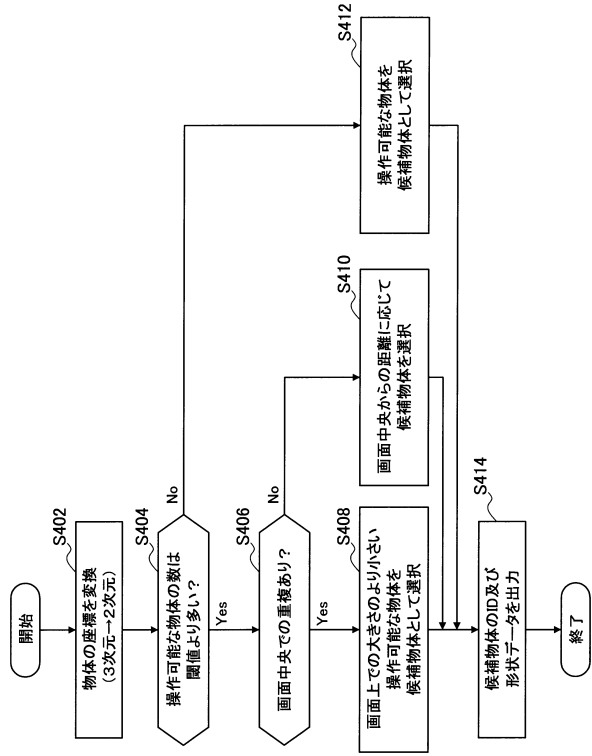
物体識別データ例

ID	種別	型番	物体画像	操作可能フラグ	コマンドリスト
TV01	テレビ	TV-SB-J11		Yes	電源ON, 電源OFF, 音量UP, 音量DOWN, チャンネルUP, チャンネルDOWN
SP01	スピーカー	SP-SS-K12		Yes	電源ON, 電源OFF, 音量UP, 音量DOWN
DV01	DVDプレーヤー	DV-SS-L13		Yes	電源ON, 電源OFF, 音量UP, 音量DOWN 再生, 停止, 早送り, 巻き戻し
PT01	プリンタ	PT-OM-N14		Yes	電源ON, 電源OFF 画像印刷
PD01	PDA	PD-PP-N15		No	-
TB01	テーブル	-		No	-

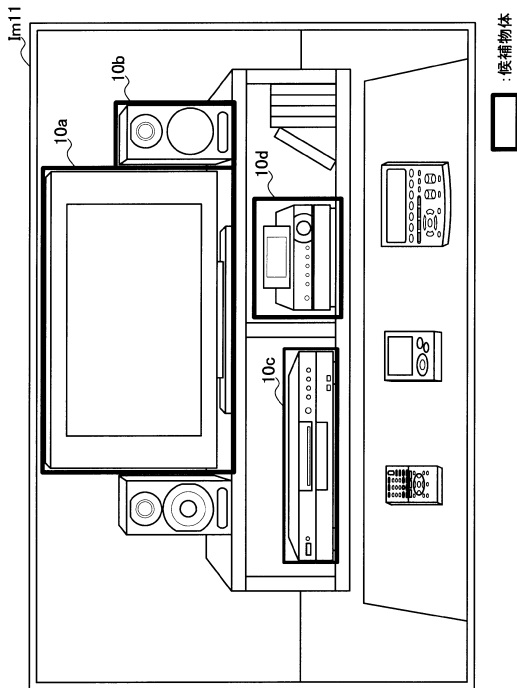
【図 1 1】



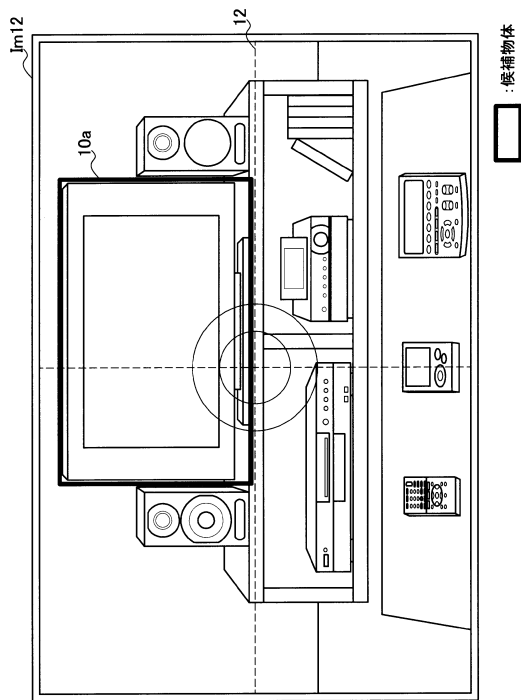
【図 1 2】



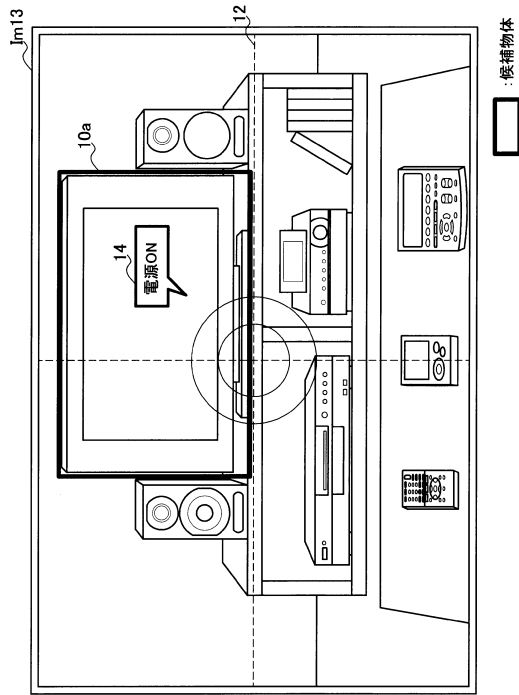
【図 1 3】



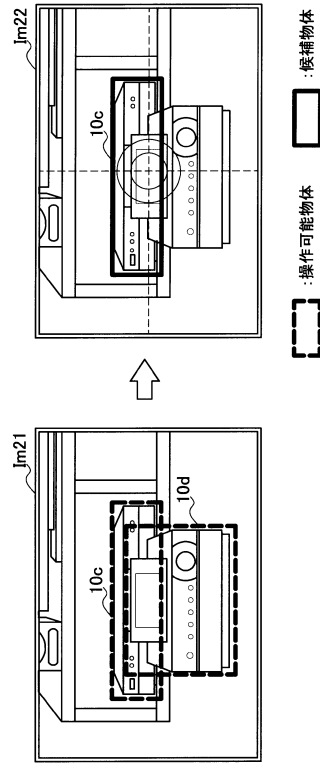
【図 1 4】



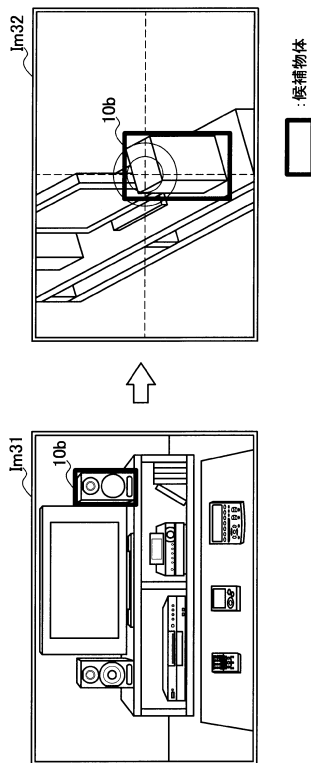
【図15】



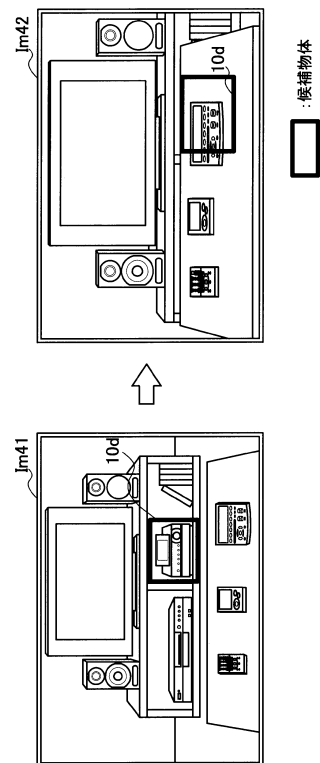
【図16】



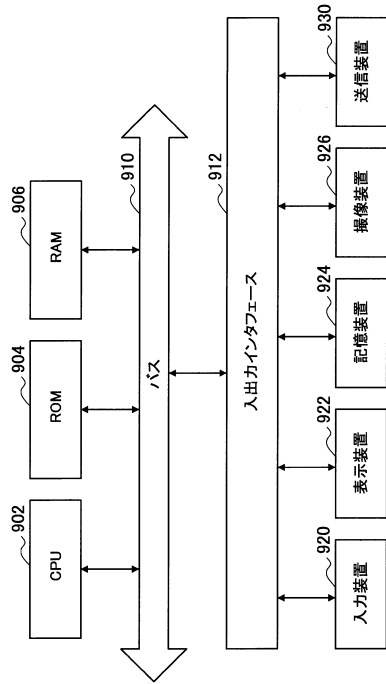
【図17】



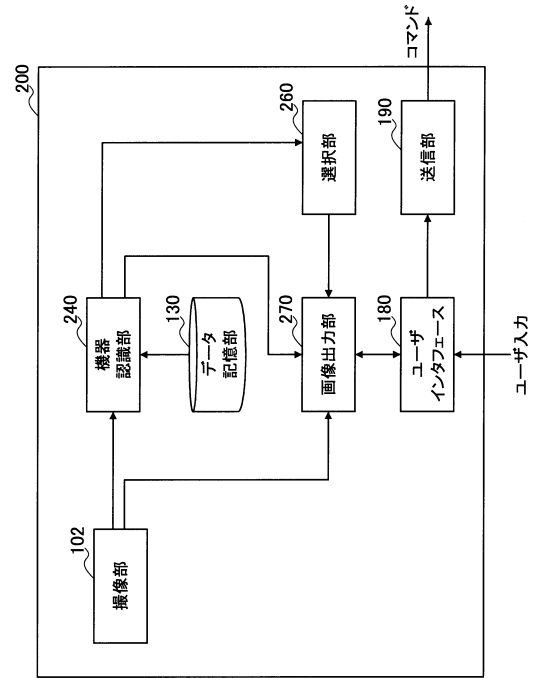
【図18】



【図19】



【図20】



フロントページの続き

- (72)発明者 松田 晃一
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 周藤 泰広
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 多井 堅一郎
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 郭 晶晶
東京都港区港南1丁目7番1号 ソニー株式会社内

審査官 大室 秀明

- (56)参考文献 国際公開第2009/110204(WO, A1)
特開2009-105574(JP, A)
特開2008-304268(JP, A)
特開2003-018672(JP, A)
特開2004-320209(JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F13/00
G06T 7/00 - 7/60
H03J 9/00 - 9/06
H04M 1/00
H04M 1/24 - 1/82
H04M99/00
H04N 5/00
H04Q 9/00 - 9/16