

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7128708号  
(P7128708)

(45)発行日 令和4年8月31日(2022.8.31)

(24)登録日 令和4年8月23日(2022.8.23)

(51)国際特許分類 F I  
G 0 6 T 19/00 (2011.01) G 0 6 T 19/00 6 0 0  
G 0 6 T 7/00 (2017.01) G 0 6 T 7/00 3 5 0 Z

請求項の数 10 (全17頁)

(21)出願番号	特願2018-188871(P2018-188871)	(73)特許権者	504407000
(22)出願日	平成30年10月4日(2018.10.4)		パロ アルト リサーチ センター インコーポレイテッド
(65)公開番号	特開2019-83001(P2019-83001A)		アメリカ合衆国 カリフォルニア州 9 4 3 0 4 パロ アルト カイオーテ ヒル
(43)公開日	令和1年5月30日(2019.5.30)		ロード 3 3 3 3
審査請求日	令和3年9月29日(2021.9.29)	(74)代理人	100094569
(31)優先権主張番号	62/579,000		弁理士 田中 伸一郎
(32)優先日	平成29年10月30日(2017.10.30)	(74)代理人	100088694
(33)優先権主張国・地域又は機関	米国(US)		弁理士 弟子丸 健
(31)優先権主張番号	15/826,588	(74)代理人	100067013
(32)優先日	平成29年11月29日(2017.11.29)		弁理士 大塚 文昭
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	100086771
早期審査対象出願		(74)代理人	弁理士 西島 孝喜
		(74)代理人	100109070

最終頁に続く

(54)【発明の名称】 機械学習用の訓練データの効率的な収集のための拡張現実を使用したシステム及び方法

(57)【特許請求の範囲】

【請求項1】

訓練データの効率的な収集を促進するためのコンピュータ実施方法であって、  
記録デバイスによって、3次元(3D)世界座標フレームに関連している場面にある物理オブジェクトの第1の画像を取得することと、  
前記記録デバイスのユーザによって、前記第1の画像上において、前記物理オブジェクトに関連した複数の頂点に印を付けることであって、頂点が、前記3D世界座標フレームに基づく3D座標を有する、印を付けることと、  
前記記録デバイスによって、前記場面の1つ以上の特性を変更しながら、前記場面にある前記物理オブジェクトの複数の第2の画像を取得することと、  
前記物理オブジェクトに関連した前記印付きの複数の頂点を前記それぞれの第2の画像上に投射し、  
前記物理オブジェクトに関連し、前記印付きの頂点を含む2次元(2D)境界エリアをディスプレイに示すこと、  
によって、前記場面にある前記物理オブジェクトのそれぞれの第2の画像を前記記録デバイスの前記ディスプレイに表示することと、  
訓練データ集合に前記印付きの複数の頂点を有する前記第1の画像及び前記投射された印付きの頂点を有する前記複数の第2の画像を格納することと、  
前記訓練データ集合に基づいて、畳み込みニューラルネットワークを訓練して前記物理オブジェクトの特徴を特定することと、

を含む、コンピュータ実施方法。

【請求項 2】

前記印付きの複数の頂点が、前記物理オブジェクト上の 1 つ以上の関心領域に対応し、前記印付きの頂点を投射することが、前記物理オブジェクト上に、前記 1 つ以上の関心領域に関連した 2 D 境界エリアを示すことをさらに含む、請求項 1 に記載の方法。

【請求項 3】

前記印付きの複数の頂点が、  
ポリゴンと、  
表平面の一部と、  
ポリウムと、のうちの 1 つ以上を示すことができる、請求項 1 に記載の方法。

10

【請求項 4】

前記複数の頂点に印を付けることが、  
前記それぞれの第 2 の画像上に、前記投射された印付きの頂点の前記 2 D 境界エリアをどのように示すかを決定することをさらに含む、請求項 1 に記載の方法。

【請求項 5】

前記 2 D 境界エリア及び前記それぞれの第 2 の画像が、前記記録デバイスに関連したディスプレイ上に提示され、  
前記 2 D 境界エリアが、2 D 形または 3 D ボリュームを示す、請求項 1 に記載の方法。

【請求項 6】

訓練データの効率的な収集を促進するためのコンピュータシステムであって、  
プロセッサと、  
前記プロセッサによって実行されると、前記プロセッサに、方法を行わせる命令を格納している記憶デバイスであって、前記方法が、

20

記録デバイスによって、3 次元 ( 3 D ) 世界座標フレームに関連している場面にある物理オブジェクトの第 1 の画像を取得すること、

前記記録デバイスのユーザによって、前記第 1 の画像上において、前記物理オブジェクトに関連した複数の頂点に印を付けることであって、頂点が、前記 3 D 世界座標フレームに基づく 3 D 座標を有する、印を付けること、

前記記録デバイスによって、前記場面の 1 つ以上の特性を変更しながら、前記場面にある前記物理オブジェクトの複数の第 2 の画像を取得すること、及び

30

前記物理オブジェクトに関連した前記印付きの複数の頂点を前記それぞれの第 2 の画像上に投射し、

前記物理オブジェクトに関連し、前記印付きの頂点を含む 2 次元 ( 2 D ) 境界エリアをディスプレイに示すこと、

によって、前記場面にある前記物理オブジェクトのそれぞれの第 2 の画像を前記記録デバイスの前記ディスプレイに表示すること、

訓練データ集合に前記印付きの複数の頂点を有する前記第 1 の画像及び前記投射された印付きの頂点を有する前記複数の第 2 の画像を格納すること、及び

前記訓練データ集合に基づいて、畳み込みニューラルネットワークを訓練して前記物理オブジェクトの特徴を特定すること、

40

を含む、記憶デバイスと、を備える、コンピュータシステム。

【請求項 7】

前記印付きの複数の頂点が、前記物理オブジェクト上の 1 つ以上の関心領域に対応し、前記印付きの頂点を投射することが、前記物理オブジェクト上に、前記 1 つ以上の関心領域に関連した 2 D 境界エリアを示すことをさらに含む、請求項 6 に記載のコンピュータシステム。

【請求項 8】

前記印付きの複数の頂点が、  
ポリゴンと、  
表平面の一部と、

50

ボリュームと、のうちの1つ以上を示すことができる、請求項6に記載のコンピュータシステム。

【請求項9】

前記複数の頂点に印を付けることが、

前記それぞれの第2の画像上に、前記投射された印付きの頂点の前記2D境界エリアをどのように示すかを決定することをさらに含む、請求項6に記載のコンピュータシステム。

【請求項10】

前記2D境界エリア及び前記それぞれの第2の画像が、前記記録デバイスに関連したディスプレイ上に提示され、

前記2D境界エリアが、2D形または3Dボリュームを示す、請求項6に記載のコンピュータシステム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、概して、コンピュータビジョンシステムに関する。より具体的には、本開示は、機械学習用の訓練データの効率的な収集のための拡張現実(AR: Augmented Reality)を使用したシステム及び方法に関する。

【0002】

現在、コンピュータビジョンシステムを訓練することは、多くの画像または映像を手作業で収集する退屈なプロセスを通して達成されている。人間である専門家は、続いて、画像または映像フレームのそれぞれにある関心オブジェクトに注釈を付けるか、またはラベルを付けることが求められる。この非効率なプロセスは、必要とされるかなりの時間量とともに、人間の疲労に基づくラベル付け時の誤り率上昇により、多額の費用をもたらすことがある。

20

【0003】

効率的な画像のパーシング及び境界ボックスの描画を可能にするヒューマンコンピュータインターフェースの開発、ならびにスクラッチからポイントを指定する代わりに、ユーザが後で修正することができるポイントを提供するテクノロジーの使用を含む、手作業のラベル付けの負担を軽減するための様々な努力がなされてきた。例えば、その過去の軌跡に基づき、映像内のオブジェクトの場所を予想または補間するのに、カルマンフィルタが使用されることがあり、この場合、人間が、必要に応じてこれらの予想を調整する。しかしながら、これらの方法は、依然として、かなりの手作業を必要とし、訓練データを収集するのに必要なラベル付けプロセスにおける目立った軽減をもたらさない。

30

【0004】

1つの実施形態は、訓練データの効率的な収集を促進するシステムを提供する。動作中、システムは、記録デバイスによって、3次元(3D: Three-Dimensional)世界座標フレームに関連している場面にある物理オブジェクトの第1の画像を取得する。システムは、第1の画像上において、物理オブジェクトに関連した複数の頂点に印を付け、この場合、頂点は、3D世界座標フレームに基づく3D座標を有する。システムは、場面の1つ以上の特性を変更しながら、場面にある物理オブジェクトの複数の第2の画像を取得する。システムは、印付きの頂点をそれぞれの第2の画像上に投射して、物理オブジェクトに関連した2次元(2D: Two-Dimensional)境界エリアを示す。

40

【0005】

いくつかの実施形態において、印付きの複数の頂点は、物理オブジェクト上の1つ以上の関心領域に対応する。印付きの頂点を投射することは、物理オブジェクト上に、1つ以上の関心領域に関連した2D境界エリアを示すことをさらに含む。

【0006】

いくつかの実施形態において、印付きの複数の頂点は、ポリゴンと、表平面の一部と、ボリュームと、のうちの1つ以上を示すことができる。

50

## 【 0 0 0 7 】

いくつかの実施形態において、複数の頂点に印を付けることは、それぞれの第 2 の画像上に、投射された印付きの頂点の 2 D 境界エリアをどのように示すかを決定することをさらに含む。

## 【 0 0 0 8 】

いくつかの実施形態において、2 D 境界エリア及びそれぞれの第 2 の画像が、記録デバイスに関連したディスプレイ上に提示され、2 D 境界エリアが、2 D 形または 3 D ボリュームを示す。

## 【 0 0 0 9 】

いくつかの実施形態において、2 D 境界エリアは、それぞれの第 2 の画像内の投射された頂点間のコネクタのタイプ、パターン、または色と、それぞれの第 2 の画像内の投射された頂点を繋ぐことによって形成された形の色、濃淡、または占有と、2 D 境界エリアを記述するテキストと、2 D 境界エリアに対するラベルまたは記述の表示と、のうちの 1 つ以上によって示される。

10

## 【 0 0 1 0 】

いくつかの実施形態において、2 D 境界エリアは、場面の特性に対応する。

## 【 0 0 1 1 】

いくつかの実施形態において、場面の特性は、記録デバイスの構えと、場面の照明と、物理オブジェクトからの記録デバイスの距離と、物理オブジェクトに対する記録デバイスの配向と、物理オブジェクトまたは場面の背景と、物理オブジェクトの 1 つ以上の部分の閉塞と、のうちの 1 つ以上である。

20

## 【 0 0 1 2 】

いくつかの実施形態において、システムは、訓練データ集合に、印付きの複数の頂点を有する第 1 の画像を格納する。システムは、訓練データ集合に、投射された印付きの頂点を有する複数の第 2 の画像を格納する。システムは、訓練データ集合に基づき、畳み込みニューラルネットワークを訓練して、物理オブジェクトの特徴を特定する。システムは、訓練されたネットワークをモバイルコンピューティングデバイス上に展開させて、物理オブジェクトの特徴を特定する。

## 【 0 0 1 3 】

いくつかの実施形態において、記録デバイスは、拡張現実デバイスと、仮想現実デバイスと、3 D 世界座標フレーム内の頂点に対して 3 D 座標を決定する磁気センサを備えるデバイスと、カメラ及びハンドトラッキングセンサと、赤、緑、及び青 ( R G B ) を記録するカメラであって、ハンドトラッキングセンサが 3 D 世界座標フレーム内の頂点に対して 3 D 座標を決定する、カメラと、赤、緑、及び青 ( R G B ) を記録するカメラ、及び深度を記録する 3 D センサと、画像または映像を記録し、かつビジュアルキューまたは位置感知テクノロジーに基づき、3 D 世界座標フレーム内の頂点に対して 3 D 座標を決定するデバイスと、画像または映像を記録し、かつ ( 3 D ) センサを含むデバイスと、のうちの 1 つ以上を含む。

30

## 【 図面の簡単な説明 】

## 【 0 0 1 4 】

本特許包袋または本出願包袋は、カラーで実行された少なくとも 1 つの図面を含む。カラー図面を有する本特許または本特許出願公開のコピーは、請求及び必要な手数料の支払いに応じて事務局によって提供されることになる。

40

## 【 0 0 1 5 】

【 図 1 】 本発明のある実施形態による、訓練データの効率的な収集を促進するための例示的な環境を示す。

【 図 2 】 本出願のある実施形態による、訓練データの効率的な収集を促進するための方法を示すフローチャートを提示する。

【 図 3 】 本発明のある実施形態による、AR デバイスを使用して、システム特徴の角に印を付ける人間を示す。

50

【図 4 A】本発明のある実施形態による、AR デバイスを使用して、第 1 の視点からプリンタの境界ボックスを作成する例を示す。

【図 4 B】本発明のある実施形態による、AR デバイスを使用して、第 2 の視点からプリンタの境界ボックスを作成する例を示す。

【図 5 A】本発明のある実施形態による、AR デバイスを使用して、第 1 の視点からプリンタ部品の境界ボックスを作成する例を示す。

【図 5 B】本発明のある実施形態による、AR デバイスを使用して、第 2 の視点からプリンタ部品の境界ボックスを作成する例を示す。

【図 6 A】本発明のある実施形態による、Microsoft HoloLens で使用され得るラベル付けインターフェースを示す。

10

【図 6 B】本発明のある実施形態による、表面メッシュのない図 6 A のラベル付けインターフェースを示す。

【図 7 A】本発明のある実施形態による、Microsoft HoloLens インターフェースを使用した、第 1 の視点からの自動的に生成された 2 次元マーカ場所及び対応する境界ボックスの例を示す。

【図 7 B】本発明のある実施形態による、Microsoft HoloLens インターフェースを使用した、第 2 の視点からの自動的に生成された 2 次元マーカ場所及び対応する境界ボックスの例を示す。

【図 7 C】本発明のある実施形態による、Microsoft HoloLens インターフェースを使用した、第 2 の視点からの自動的に生成された 2 次元マーカ場所及び対応する境界ボックスの例を示す。

20

【図 8】本発明のある実施形態による、訓練データの効率的な収集を促進する例示的なコンピュータ及び通信システムを示す。

【0016】

図において、類似の参照番号は、同じ図の要素を指す。

【0017】

以下の説明は、当業者に、実施形態を作り、使用するのを可能にさせるために提示され、特定の用途及びその要件の観点から提供されている。開示された実施形態に対する様々な修正形態が当業者にはすぐに明らかになり、本明細書に定義されている一般原則が、本開示の趣旨及び範囲を逸脱しない限り、他の実施形態及び用途に適用されてもよい。したがって、本発明は、示されている実施形態に限定されるものでなく、本明細書に開示されている原則及び特徴と合致している最も広い範囲が与えられるべきである。

30

【0018】

本明細書に説明されている実施形態は、ユーザが訓練データを効率的に収集することを可能にするシステムを提供することによって、訓練データの手作業のラベル付けの負担を軽減する問題を解決する。システムは、人間が、記録デバイスを使用して、単一の画像において、関心オブジェクトを捕捉し、それに注釈を付けることを可能にする。システムは、続いて、同じ環境において変わる条件の下、多くの他の同様の画像にある同じ関心オブジェクト上に注釈を投射する。

【0019】

40

具体的には、ユーザは、記録デバイスを使用して、3次元(3D)世界座標フレームに関連している環境(例えば、「場面」)にある物理オブジェクトの画像または映像を捕捉することができる。単一の画像または映像フレームに関して記録デバイスを使用して、ユーザは、物理オブジェクト上の点に印を付けることができる。点は、3D世界座標フレームに基づく3D座標を有する頂点とすることができる。頂点の3D座標を決定することは、場面における物理オブジェクト構えを画定するビジュアルキューまたは他の位置感知テクノロジーに基づいていてもよい。場面が変更されると(例えば、ユーザ及び記録デバイスが、物理オブジェクトに対して、部屋内の異なる場所に移動すると)、システムは、物理オブジェクトに関連した境界ボックスを示すように投射された、印付きの点を、ユーザに表示することができる(変更された条件の下で捕捉された画像または映像上で)。

50

## 【 0 0 2 0 】

例えば、記録デバイスは、3Dトラッキングセンサを備えるカメラとすることができる。場面は、事務所とすることができる、関心物理オブジェクトは、事務所内のプリンタとすることができる。システムは、場面（すなわち、3D世界座標フレーム）においてカメラ構えを画定することができる一方、ユーザは、頂点に印を付け、物理オブジェクト上の関心領域を示すことができる（例えば、プリンタ上のプリンタ排紙トレイに注釈を付けるか、またはラベルを付ける）。システムは、場面の1つ以上の特性を変更しながら（例えば、カメラの場所、配向、または構えを変更する）、印付きの頂点を追跡することができる。

## 【 0 0 2 1 】

システムは、次に、様々なカメラ構えにおいて、印付きの頂点をカメラフレーム中に投射することができ、この場合、各投射は、収集された（かつラベル付けされた）訓練データの一部として使用され得る自動的に注釈が付けられた画像である。すなわち、プリンタの他の画像を様々なカメラ構えから撮ると仮定すると、システムは、他の画像上に印付きの頂点を投射して、プリンタ上に、関心領域の周りの境界エリアを示すことができ、このように、単一の注釈付きの画像にのみ基づいて、他の画像に自動的に注釈を付ける。境界エリアは、場面の特性に対応する境界ボックスまたはラベルとすることができる。

10

## 【 0 0 2 2 】

3Dセンサ（または他の位置感知テクノロジー）を備える記録デバイスを使用することによって、システムは、ユーザが、環境内の関心オブジェクトの場所、範囲、構え、または他の性質に、一度、関単にラベルを付けてから、同じ環境内の様々な条件下で撮られた非常に多くの画像または映像フレームに、このラベル付けを投射することを可能にする。このプロセスは、ほとんどユーザの手を煩わせることなく、様々な条件下で、関心オブジェクトの多数の画像または映像フレームを生成することができ、この場合、各画像またはフレームは、元の環境から投射された性質でラベルが付けられる。

20

## 【 0 0 2 3 】

したがって、本明細書に説明されている実施形態は、機械学習用の訓練データを収集する効率を高めるシステムを提供する。高められた効率は、画像内に多様な関心領域または関心オブジェクトを含む、その画像に注釈を付けるのに必要とされる手作業のラベル付け程度におけるかなりの軽減を含むことができる。有効な機械学習が、訓練データの多様性にに基づいていることから、またシステムが多様な訓練データを効率的に収集することから、本明細書に説明されている実施形態は、機械学習の技術分野を向上させる。

30

## 【 0 0 2 4 】

1つの例示的な実施形態において、ユーザは、RGB-D（赤緑青+深度）カメラ+位置推定とマッピングの同時実行（SLAM: Simultaneous Localization And Mapping）式アルゴリズムを使用するMicrosoft HoloLensシステムを使用して、部屋のモデルをビルドすることができる。ユーザは、HoloLensを使用して、部屋の3Dモデルに仮想マーカを置き、部屋内の関心オブジェクトの場所及び範囲を画定することができる。ユーザは、様々な距離及び配向における、様々な照明条件下で、また閉塞したオブジェクトの存在下で、HoloLensを用いて様々な視野からオブジェクトのいくつかの写真を撮ることができる。HoloLensは、画像ごとに、部屋の3Dモデル内のその構えを計算することができる。

40

## 【 0 0 2 5 】

システムは、部屋の3Dモデルからのユーザの元のマーカを、各画像の中に投射して、関心オブジェクトに対して境界ボックスを形成することができる。これらの自動的に注釈が付けられた画像は、ディープニューラルネットワークなどのテクノロジーを採用するオブジェクト検出及び認識システムを訓練するのに使用され得る。

## 【 0 0 2 6 】

別の例示的な実施形態において、手の動きの位置を追跡し、記録することができるデバイスが、実世界オブジェクトに注釈を付けるのに使用され得る。例えば、HTC Vive（NIR基地局を使用して、コントローラ内の多様なセンサを追跡する）の場合に含ま

50

れている追跡式ハンドコントローラが、コントローラへのトリガ入力を使用して、実世界オブジェクトの境界をたどるのに使用され得る。第2の較正済みカメラ（例えば、HTC Vive内の埋め込み式カメラ、または外部共較正済み「複合現実」カメラ）と協同して、オブジェクトが、適切に位置合わせされた境界注釈付きで、様々な視点から画像化され得る。

【0027】

さらなる例示的な実施形態において、磁気センサを備えるデバイスが、3D座標に注釈を付け、それを追跡するのに使用され得る。例えば、Polhemusによる3D運動トラッキングシステムは、センサ（例えば、記録デバイス上の）の位置及び配向を、センサが空間を移動するのに従って、電磁場を使用することによって追跡して、遠隔オブジェクト（例えば、実世界オブジェクト）の位置及び配向を決定することができる。

10

【0028】

本明細書に説明されている実施形態は、機械学習用の訓練データを収集する効率を高めるシステムを提供する。訓練データを収集するのに必要とされる手作業の時間及び労力の程度を下げることに加えて、システムは、新しいオブジェクト検出システムの開発時間を短縮することもできる。境界ボックス座標以外に、グラウンドトゥールズ注釈は、記録デバイスから、オブジェクトの場所、配向、及び構えについての3D情報も捕捉することができる。収集されたデータは、したがって、より広いコンピュータビジョン問題セット、例えば、構え、深度、サイズ、オブジェクトクラス、及び「粗い」対「滑らか」などの性質の評価に使用され得る。

20

【0029】

さらに、システムの実施形態は、顧客によって管理されるシステムの部品の大きなラベル付きのデータセットを素早く作成することができ、ラベル付きのデータセットを使用して、コンピュータビジョンシステムを訓練することができる。訓練されたシステムは、システムの部品の管理及び修理の際にサービス技術者を手助けすることができ、また顧客が、システム部品（例えば、Digital Glove Box及びSmart Scanのアプリケーション）の修理の際にエンドユーザを手助けすることを可能にすることもできる。これらの既存のツールと提案されたテクノロジーとの間の差別化要因は、既存のツール内に入れられたコンピュータビジョンシステム用の訓練データを収集するのに求められる大量の時間である。この大量の時間は、システムの実施形態を使用して、訓練データを収集することによって、ごくわずか（例えば、1/10）に低減され得る。

30

【0030】

顧客によって管理される他の例示的なシステムは、自動車または車両（ダッシュボードまたは他の自動車部品、例えば、エンジン、または自動車のキャビン内の特徴が、顧客が補助を必要とすることがある関心領域であってもよい）と、プリンタまたは他のデバイス（給紙トレイ、排紙トレイ、制御パネル、または他の部品が、関心領域であってもよい）と、を含むことができる。補助を必要とし得る顧客（またはエンドユーザ）は、自分の携帯電話を用いてシステムの写真を撮り、その写真のある特定の「ラベル付きの」セクションについての有益な情報を取得することができる。例えば、車両のユーザが車両に関する問題に見舞われた場合、車両ユーザは、自分のモバイルデバイスを用いて、また様々なカメラ構えからのダッシュボードの以前に生成された多様なラベル付きの画像セットに基づき、車両ダッシュボードの画像を捕捉することができ、車両ユーザは、その問題にどのように対処するかを理解する上で、ユーザを手助けするのに使用されてもよいラベル付きの画像を取得することができる。

40

【0031】

システムの実施形態は、特定分野の専門家が、3Dモデル化ツールの機械学習時に専門的知識を必要とすることなく、新しい、スタンドアローンのインタラクティブな拡張現実コーチを簡単に生み出すことを可能にするディープビジョン動力式ツールも生み出すことができる。システムは、オブジェクト及び領域を指し示すためのジェスチャ、ラベルを提供するための音声テキスト化、及びオブジェクト部分セグメント化など、オーサリングに

50

関する特徴を含むことができる。システムは、記憶、索引付け及びメタデータ、ならびにクリップの基本編集などのキュレーションも含むことができる。システムは、部品認識、ラベル配置、フロー制御、及び部品条件決定など、補助に関する特徴をさらに含むことができる。システムは、背景不感性、自動照明増強、及び影生成を通して、訓練のロバスト性及び一般化を高めることができる。さらに、システムは、インタラクティブなコーチングをサポートするための発話及び行動認識とともに、アプリケーション統合及び遠隔測定接続を含むことができる。

#### 【0032】

したがって、外部トラッキングセンサを使用してデータに注釈を付けることによって、また拡張現実/仮想現実(AR/VR: Augmented Reality/Virtual Reality)を使用して、機械学習用の注釈付きの訓練データを収集することによって、本明細書に説明されている実施形態は、AR/VRの能力を活用して、場面の新しい視点ごとのラベル付けを必要とすることなく、多様な視点から訓練データを収集することができる。これにより、有効なコンピュータビジョンシステムを訓練する際のラベル付けの負担におけるかなりの軽減をもたらすことができる。

#### 【0033】

図1は、本発明のある実施形態による、訓練データの効率的な収集を促進するための例示的な環境100を示す。環境100は、デバイス104及び関連のユーザ106と、デバイス108と、物理オブジェクト120と、センサ110と、を含むことができる。デバイス104は、拡張現実デバイス(Microsoft HoloLensなど)を含むことができる。物理オブジェクト120は、関連の3D世界座標フレームを有する場面121(図示せず)の一部とすることができる。デバイス108は、データを受信、送信、かつ格納することができ、また3D世界座標フレーム内の多様な視点から撮られた画像の中に頂点を投射するためのアルゴリズムを実行することができる、サーバまたは他のコンピューティングデバイスを含むことができる。センサ110及び他のトラッキングセンサ(図示せず)は、システム内のデバイス104と協働して、画像を捕捉し、画像に注釈を付け、3D座標を決定し、注釈付きの画像を格納し、画像を投射し、かつ投射された画像を表示することができる。デバイス104、デバイス108、及びセンサ110は、ネットワーク102を介して通信することができる。

#### 【0034】

操作中、ユーザ106は、3D世界座標フレーム内の場所130からのデバイス104を使用して、物理オブジェクト120の画像134を捕捉し、注釈を付け(機能132)、画像134(ユーザ作成の注釈付き)をデバイス108に送信することができる。注釈付きの画像は、3D世界座標フレーム内の3D座標に関連している多様な印付きの頂点を含むことができる。

#### 【0035】

ユーザ106は、場所140(場所130とは異なる)からのデバイス104を使用して、物理オブジェクト120の画像144を捕捉し(機能142)、画像144をデバイス108に送信することができる。デバイス108は、画像144からの印付きの頂点を、画像146上に投射し、かつ画像146(自動作成の注釈付き)を、デバイス104上に表示されるように、ユーザ106に送信し戻すためのアルゴリズムを実行することができる。

#### 【0036】

さらに、ユーザ106は、様々な場面特性変更160(例えば、他の場所、異なるカメラ構え、異なる照明条件など)に基づいてデバイス104を使用し、画像164を、ネットワーク102を介してデバイス108に送信することができる。デバイス108は、印付きの頂点(画像144において特定または登録されたような)を画像166上に投射し、かつ画像166(自動作成の注釈付き)を、デバイス104上に表示されるように、ユーザ106に送信し戻すためのアルゴリズムを実行することができる。

#### 【0037】

10

20

30

40

50



デバイス 108 は、場面 121 に対応し、かつ物理オブジェクト 120 を含む環境を描写する世界座標フレーム 150 と、画像 134 (ユーザ作成の注釈付き) と、ユーザ 106 によって捕捉されたような画像 144 と、画像 146 (自動作成の注釈付き) と、画像 166 (自動作成の注釈付き) と、などのデータを格納することができる。デバイス 108 は、画像 134、144、及び 166 を含むことができる訓練データ 170 も、データ集合として格納することができる。

#### 【0038】

図 2 は、本発明のある実施形態による、訓練データの効率的な収集を促進するための方法を示すフローチャート 200 を提示する。動作中、システムは、記録デバイスによって、3次元(3D)世界座標フレームに関連している場面にある物理オブジェクトの第1の画像を取得する(動作202)。システムは、第1の画像上において、物理オブジェクトに関連した複数の頂点に印を付け、この場合、頂点は、3D世界座標フレームに基づく3D座標を有する(動作204)。システムは、場面の1つ以上の特性を変更しながら、場面にある物理オブジェクトの複数の第2の画像を取得する(動作206)。システムは、印付きの頂点をそれぞれの第2の画像上に投射して、物理オブジェクトに関連した2次元(2D)境界エリアを示す(動作208)。システムは、印付きの複数の頂点を有する第1の画像と、投射された印付きの頂点を有する複数の第2の画像とを、訓練データ集合に格納する(動作210)。システムは、格納された訓練データ集合に基づき、ニューラルネットワークを訓練する(動作212)。

10

#### 【0039】

本明細書に説明されている実施形態は、単一の画像(またはフレーム)において、多様な関心領域に、一度に注釈を付けることができ(例えば、多様な複数の頂点/点に印を付けることによって)、このように、多様な印付きの複数の頂点を有する単一のユーザ注釈付きの画像は、様々なカメラ構えからの画像の生成をもたらすことができ、この場合、各画像は、注釈付きの多様な関心領域を表示する。すなわち、システムは、多様な関心領域に、同時にラベルを付けることができる。

20

#### 【0040】

さらに、本明細書に説明されている実施形態は、ユーザが、最初に多くの画像を捕捉し、次に1つの画像に注釈を付けることを可能にし、その時点で、システムは、ユーザ注釈付きの画像に基づき、以前に捕捉された画像に自動的に注釈を付けることができる。例えば、ユーザが30秒間、オブジェクトの周りを歩き、記録デバイスまたはシステムを用いて100枚の画像を捕捉すると仮定する。ある時間歩き、いくつかの画像を捕捉した後(例えば、7秒間歩き、24枚の画像を捕捉した後)、ユーザは、ある画像(例えば、25番目の画像)にラベルを付けるか、または印を付けることができ、システムに、以前に捕捉された24枚の画像と、その後捕捉された75枚の画像との両方に自動的に注釈を付けさせることができる。

30

#### 【0041】

本明細書に説明されている実施形態は、境界ボックスを示すことができる形などのポリゴンに印を付けることによって、表平面の注釈も提供することができる。ポリゴンは、凸状または非凸状のいずれであってもよい。非凸状ポリゴンに印を付けることは、追加の情報を必要とすることがあり、また、例えば、点が印付けされる順序に基づくことができる。さらに、システムの実施形態は、複数の点に印を付けることなど、ボリュームの注釈を提供して、プリンタ自体全体の周りの立方体を示すことができる。システムは、注釈付きのボリュームに基づき、自動的に注釈が付けられる後続の画像または映像フレーム上に、ボリューム全体を投射することができる。

40

#### 【0042】

図3、図4A及び4B、図5A及び5B、図6A及び6B、ならびに図7A~7Cは、訓練データの効率的な収集を促進するための例示的な画像を示す。図3は、本発明のある実施形態による、ARデバイスを使用して、システム特徴の角に印を付ける人間を示す。図3における環境は、プリンタを含むいくつかのオブジェクトがある部屋である。ARデ

50

バイスは、環境（例えば、部屋）に対して3D世界座標フレームを画定することができるシステムである、Microsoft HoloLensとすることができる。

【0043】

図4Aは、本発明のある実施形態による、ARデバイスを使用して、第1の視点からプリンタの境界ボックスを作成する例を示す。ユーザは、部屋内のある場所に立って、画像を取得することができる。すなわち、ユーザは、ある特定の視点から、プリンタを含む部屋の画像を捕捉することができ、この場合、視点は、プリンタに対するARデバイスの構え、場所、配向などに基づいている。ARデバイスを使用して、ユーザは、この第1の視点から撮られた第1の画像内に緑のマーカを置くことができ、システムは、この画像上に、緑のマーカによって画定された境界エリアを表示することができる。図4Aでは、境界エリアは、プリンタの周りに青の輪郭を有するボックスとして現れる。システムは、したがって、物理オブジェクトに関連した複数の頂点に印を付ける。

10

【0044】

図4Bは、本発明のある実施形態による、ARデバイスを使用して、第2の視点からプリンタの境界ボックスを作成する例を示す。図4Bでは、第2の視点からのARデバイスを使用して、ユーザは、部屋の第2の画像を捕捉することができる。第2の視点は、図4Aの第1の視点とは異なる（すなわち、ARデバイスがプリンタに対して、異なる構え、場所、配向などにある）。システムは、印付きの頂点を緑のマーカとして、第2の画像上に投射し、また、投射された緑のマーカによって画定された境界エリアも、第2の画像上に投射することができる。

20

【0045】

図4A及び4Bは、場面内の単一の関心領域に印を付けることを示す。上に説明されているように、ユーザは、物理オブジェクトの多様な複数の頂点に印を付けることによって、単一のフレームにおいて、多様な関心領域に一度に注釈を付けることができる。そのように、多様な注釈付きの関心領域を有する単一のユーザ注釈付きのフレームは、様々なカメラ構えからの画像の生成をもたらすことができ、この場合、各画像は、注釈付きの多様な関心領域を表示する。すなわち、システムは、多様な関心領域に同時にラベルを付けることができる。

【0046】

図5Aは、本発明のある実施形態による、ARデバイスを使用して、第1の視点からプリンタ部品の境界ボックスを作成する例を示す。ARデバイスを使用して、ユーザは、この第1の視点から撮られた第1の画像にある多様な関心領域上に、緑のマーカ（図示せず）を置くことができ、システムは、この画像上に、緑のマーカによって画定された多様な境界エリア、例えば、排紙トレイに対応する青の境界ボックス、制御パネルに対応する緑の境界ボックス、及び用紙引き出しに対応する赤の境界ボックスを表示することができる。いくつかの実施形態において、図5Aは、以前に印が付けられた頂点に基づく画像の表示を示し、このように、システムは、この第1の画像上に、以前に印が付けられた頂点に基づき、多様な境界ボックスを表示する。

30

【0047】

図5Bは、本発明のある実施形態による、ARデバイスを使用して、第2の視点からプリンタ部品の境界ボックスを作成する例を示す。図5Bでは、第2の視点からのARデバイスを使用して、ユーザは、部屋の第2の画像を捕捉することができる。システムは、この第2の画像上に、印付きの複数の頂点を緑のマーカ（図示せず）として投射することができ、また、以前に印が付けられた複数の頂点によって画定されたような多様な境界エリアも、この第2の画像上に投射することができる。図5Aと同様に、青の境界ボックスは、排紙トレイに対応し、緑の境界ボックスは、制御パネルに対応し、赤の境界ボックスは、用紙引き出しに対応する。

40

【0048】

図6Aは、本発明のある実施形態による、Microsoft HoloLensで使用され得るラベル付けインターフェースを示す。Microsoft HoloLens

50

を使用して、ユーザは、環境の制御ボタン（例えば、緑の楕円及び青の円）、ならびに表面メッシュ（例えば、白のポリゴン）を見ることができる。ユーザは、Microsoft HoloLensを装着しながら手の動きを使用して、いくつかの仮想マーカ（緑、赤、及び青の四角）を置くことによって、見た画像に注釈を付けることができる。マーカは、プリンタに関連した様々な関心領域に対応することができる。例えば、ユーザは、緑のマーカが用紙引き出しに対応し、赤のマーカが手作業の入力エリアに対応し、青のマーカが制御パネルに対応することに、注釈を付けることができる。

【0049】

図6Bは、本発明のある実施形態による、表面メッシュのない図6Aのラベル付けインターフェースを示す。

【0050】

図7A～7Cは、図6A及び6Bにおけるユーザによって入れられた注釈に基づいた、自動的に生成されたマーカ場所及び境界ボックスを示す。図7A～7Cにおける境界エリアの色が図6Aにおける対応する境界エリアとは異なり、システムが、以前に注釈が付けられた多様な関心領域を表示することに留意されたい。それに加え、ユーザは、境界エリアの指標を、マーカ場所間のコネクタのいずれのタイプ、パターン、または色にも設定することができる。

【0051】

図7Aは、本発明のある実施形態による、Microsoft HoloLensインターフェースを使用した、第1の視点からの自動的に生成された2次元マーカ場所及び対応する境界ボックスの例を示す。図7Aでは、緑の境界ボックスは、制御パネルに対応し、青の境界ボックスは、手作業の入力エリアに対応し、赤の境界ボックスは、用紙引き出しに対応する。

【0052】

図7Bは、本発明のある実施形態による、Microsoft HoloLensインターフェースを使用した、第2の視点からの自動的に生成された2次元マーカ場所及び対応する境界ボックスの例を示す。図7Bでは、緑の境界ボックスは、制御パネルに対応し、青の境界ボックスは、手作業の入力エリアに対応し、赤の境界ボックスは、用紙引き出しに対応する。

【0053】

図7Cは、本発明のある実施形態による、Microsoft HoloLensインターフェースを使用した、第2の視点からの自動的に生成された2次元マーカ位置及び対応する境界ボックスの例を示す。図7Cでは、緑の境界ボックスは、制御パネルに対応し、青の境界ボックスは、手作業の入力エリアに対応し、赤の境界ボックスは、用紙引き出しに対応する。

【0054】

図8は、本発明のある実施形態による、訓練データの効率的な収集を促進する例示的なコンピュータ及び通信システム800を示す。システム800は、ネットワーク（図示せず）を介して通信することができるコンピュータシステム802と記録デバイス842とを含む。コンピュータシステム802と記録デバイス842とは、それぞれ、図1のデバイス108とデバイス104とに対応することができる。

【0055】

コンピュータシステム802は、プロセッサ804、メモリ806、及び記憶デバイス808を含む。メモリ806は、マネージドメモリとして働く揮発性メモリ（例えば、RAM）を含むことができ、1つ以上のメモリプールを格納するのに使用され得る。さらに、コンピュータシステム802は、表示デバイス810、キーボード812、及びポインティングデバイス814に結合され得る。記憶デバイス808は、オペレーティングシステム816、コンテンツ処理システム818、及びデータ828を格納することができる。

【0056】

コンテンツ処理システム818は、コンピュータシステム802によって実行されると

10

20

30

40

50

、コンピュータシステム 802 に、本開示に説明されている方法及び／またはプロセスを行わせることができる命令を含むことができる。具体的には、コンテンツ処理システム 818 は、コンピュータネットワークにわたって、他のネットワークノードにノからデータパケットを送信する、かつノまたは受信するノ取得するための命令を含んでもよい（通信モジュール 820）。データパケットは、画像と、映像と、映像フレームと、頂点の 3D 座標と、場面または場面にある物理オブジェクトについての情報と、を含むことができる。

【0057】

コンテンツ処理システム 818 は、3D 世界座標フレームに関連している場面にある物理オブジェクトの第 1 の画像を受信するための、また場面の 1 つ以上の変更された特性に基づき、場面にある物理オブジェクトの複数の第 2 の画像を受信するための命令を含むことができる（通信モジュール 820）。コンテンツ処理システム 818 は、物理オブジェクトに関連した印付きの複数の頂点を登録するための命令を含むことができる（投射決定モジュール 822）。コンテンツ処理システム 818 は、それぞれの第 2 の画像上への印付きの頂点の投射を決定するための命令を含むことができる（投射決定モジュール 822）。コンテンツ処理システム 818 は、物理オブジェクトに関連した 2D 境界エリアを示すための命令を含むことができる（境界エリア管理モジュール 824）。コンテンツ処理システム 818 は、第 1 の画像及び第 2 の画像を訓練データ集合に格納するための、また畳み込みニューラルネットワークを訓練して、物理オブジェクトの特徴を特定するための命令を含むことができる（ネットワーク訓練モジュール 826）。

【0058】

記録デバイス 842 は、プロセッサ 844、メモリ 846、及び記憶デバイス 848 を含む。メモリ 846 は、マネージドメモリとして働く揮発性メモリ（例えば、RAM）を含むことができ、1 つ以上のメモリプールを格納するのに使用され得る。記憶デバイス 848 は、コンテンツ処理システム 858 及びデータ 868 を格納することができる。

【0059】

コンテンツ処理システム 858 は、記録デバイス 842 によって実行されると、記録デバイス 842 に、本開示に説明されている方法及び／またはプロセスを行わせることができる命令を含むことができる。具体的には、コンテンツ処理システム 858 は、コンピュータネットワークにわたって、他のネットワークノードにノからデータパケットを送信する、かつノまたは受信するノ取得するための命令を含んでもよい（通信モジュール 860）。データパケットは、画像と、映像と、映像フレームと、頂点の 3D 座標と、場面または場面にある物理オブジェクトについての情報と、を含むことができる。

【0060】

コンテンツ処理システム 858 は、3次元（3D）世界座標フレームに関連している場面にある物理オブジェクトの第 1 の画像を取得するための命令を含むことができる（画像取得モジュール 862）。コンテンツ処理システム 858 は、第 1 の画像上において、物理オブジェクトに関連した複数の頂点に印を付けるための命令を含むことができ、この場合、頂点は、3D 世界座標フレームに基づく 3D 座標を有する（頂点印付けモジュール 864）。コンテンツ処理システム 858 は、場面の 1 つ以上の特性を変更しながら、場面にある物理オブジェクトの複数の第 2 の画像を取得するための命令を含むことができる（画像取得モジュール 862）。コンテンツ処理システム 858 は、印付きの頂点をそれぞれの第 2 の画像上に投射して、物理オブジェクトに関連した 2次元（2D）境界エリアを示すための命令を含むことができる（画像表示モジュール 866）。

【0061】

データ 828 及び 868 は、本開示に説明されている方法及び／またはプロセスによって、入力として必要とされるか、または出力として生成されるいずれのデータも含むことができる。具体的には、データ 828 及び 868 は、少なくとも、データと、画像と、物理オブジェクトの画像と、3D 世界座標フレームと、頂点と、頂点に対する 3D 座標と、場面と、場面の特性と、物理オブジェクト上の関心領域の指標と、印付きの頂点と、印付きの頂点の投射と、ポリゴンと、表平面の一部と、ボリュームと、2D 形と、3D ボリュ

10

20

30

40

50

ームと、2D境界エリアと、注釈と、ラベルと、画像内の投射された頂点間のコネクタのタイプ、パターン、または色と、画像内の投射された頂点を繋ぐことによって形成された形の色、濃淡、及び占有と、2D境界エリアを描写するテキストと、2D境界エリアに対するラベルまたは描写の表示と、記録デバイスの構えと、場面の照明と、物理オブジェクトからの記録デバイスの距離と、物理オブジェクトに対する記録デバイスの配向と、物理オブジェクトまたは場面の背景と、物理オブジェクトの1つ以上の部分の閉塞と、訓練データ集合と、訓練されたネットワークと、ユーザ作成の注釈付きの画像と、システム作成または自動生成の注釈付きの画像と、を格納することができる。

【0062】

この詳細な説明に説明されているデータ構造及びコードは、コンピュータシステムによる使用のためのコード及び/またはデータを格納することができるいずれのデバイスまたは媒体であってもよい、コンピュータ可読記憶媒体上に通常格納されている。コンピュータ可読記憶媒体は、揮発性メモリ、不揮発性メモリ、またディスクドライブ、磁気テープ、CD（コンパクトディスク）、DVD（デジタルバーサタイルディスクまたはデジタルビデオディスク）などの磁気及び光記憶デバイス、または現在知られているかまたは後に開発されるコンピュータ可読媒体を格納することができる他の媒体を含むが、それらに限定されるものではない。

10

【0063】

詳細な説明の項に説明されている方法及びプロセスは、上に説明されているようなコンピュータ可読記憶媒体に格納され得る、コード及び/またはデータとして具体化され得る。コンピュータシステムが、コンピュータ可読記憶媒体上に格納されているコード及び/またはデータを読み取り、実行すると、コンピュータシステムは、データ構造及びコードとして具体化され、かつコンピュータ可読記憶媒体内に格納されている方法及びプロセスを行う。

20

【0064】

さらに、上に説明されている方法及びプロセスは、ハードウェアモジュールまたは装置に含まれ得る。ハードウェアモジュールまたは装置は、特定用途向け集積回路（ASIC：Application-Specific Integrated Circuit）チップ、フィールドプログラマブルゲートアレイ（FPGA：Field-Programmable Gate Array）、特定の時点で特定のソフトウェアモジュールまたは1つのコードを実行する専用または共有のプロセッサ、及び現在知られているかまたは後に開発される他のプログラマブルロジックデバイスを含むことができるが、それらに限定されるものではない。ハードウェアモジュールまたは装置が起動されると、それらは、それら内に含まれている方法及びプロセスを行う。

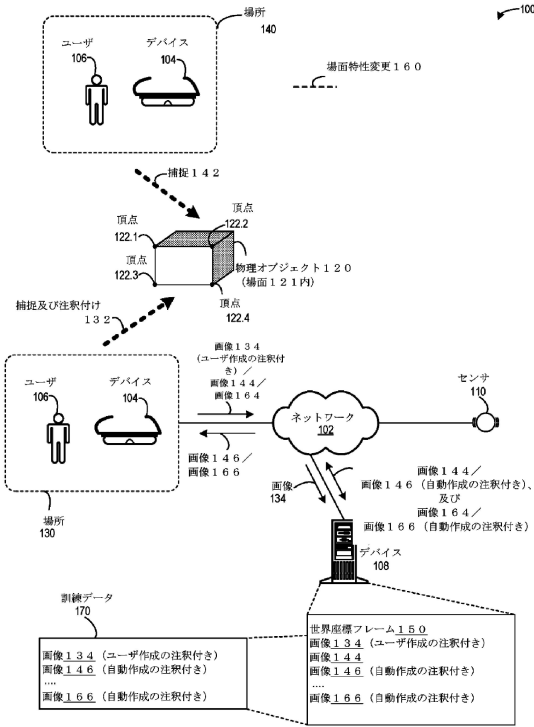
30

40

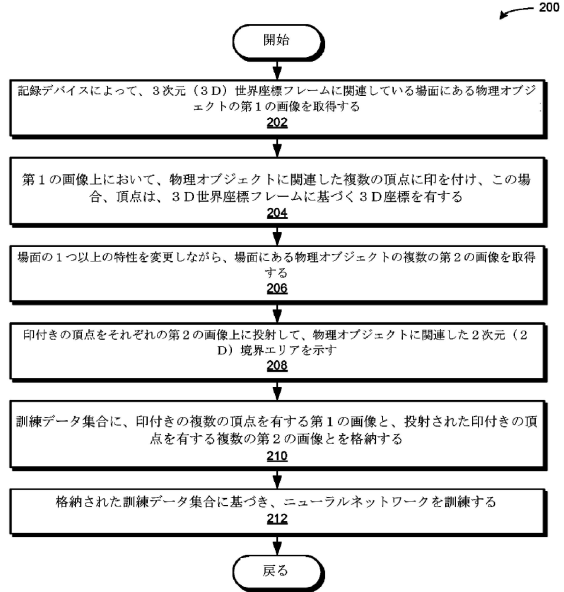
50

【図面】

【図 1】



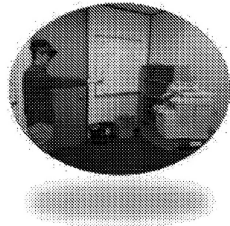
【図 2】



10

20

【図 3】



【図 4 A】

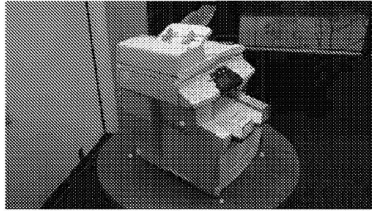


30

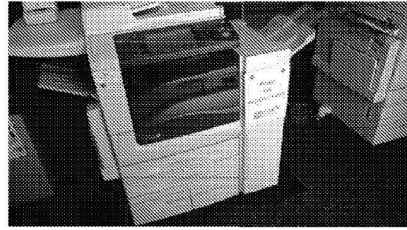
40

50

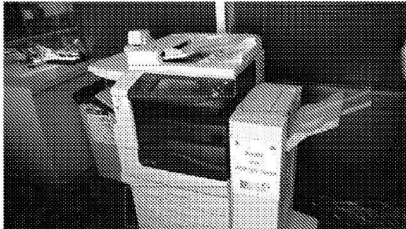
【 図 4 B 】



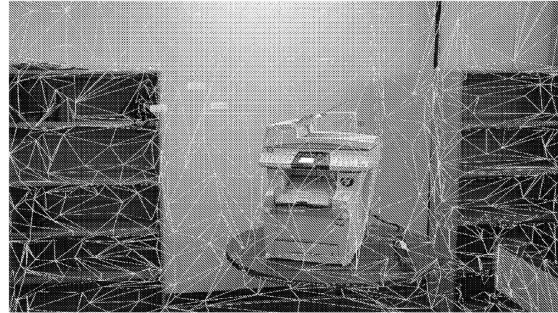
【 図 5 A 】



【 図 5 B 】



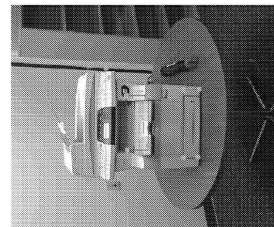
【 図 6 A 】



【 図 6 B 】



【 図 7 A 】



10

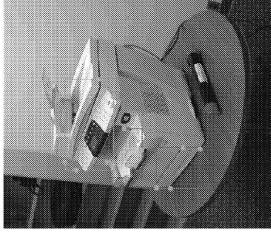
20

30

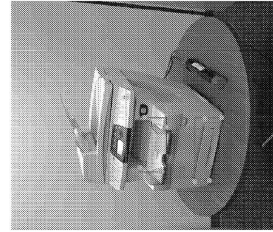
40

50

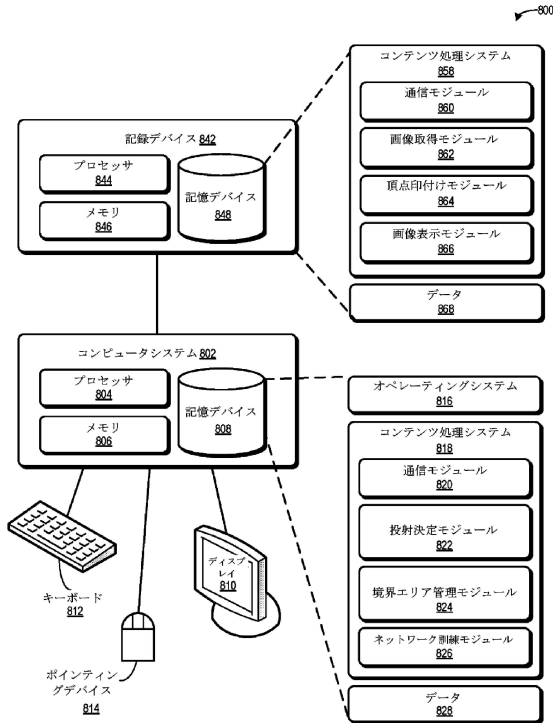
【図7B】



【図7C】



【図8】



10

20

30

40

50



## フロントページの続き

- 弁理士 須田 洋之  
 (74)代理人 100109335  
 弁理士 上杉 浩  
 (74)代理人 100120525  
 弁理士 近藤 直樹  
 (74)代理人 100139712  
 弁理士 那須 威夫  
 (72)発明者 マシュー・エイ・シュリーブ  
 アメリカ合衆国 カリフォルニア州 9 4 0 4 3 マウンテン・ビュー ロック・ストリート 2 2 1 0  
 (72)発明者 スリチャラン・カルーア・パリ・クマール  
 アメリカ合衆国 カリフォルニア州 9 4 0 4 1 マウンテン・ビュー ハイドン・アベニュー 1 8  
 0 2 アパートメント 4  
 (72)発明者 ジン・サン  
 アメリカ合衆国 カリフォルニア州 9 4 0 6 5 レッドウッド・シティ シャノン・ウェイ 5 3 0  
 ナンバー 4 1 0 2  
 (72)発明者 ガウラン・アール・ガヴァイ  
 アメリカ合衆国 カリフォルニア州 9 4 1 1 4 サンフランシスコ アルバラード・ストリート 8 2 1  
 (72)発明者 ロバート・アール・プライス  
 アメリカ合衆国 カリフォルニア州 9 4 3 0 6 パロ・アルト エドリー・アベニュー 3 5 0  
 (72)発明者 ホダ・エム・エイ・エルダーディリー  
 アメリカ合衆国 カリフォルニア州 9 4 0 7 0 サン・カルロス チェスナット・ストリート 3 2 0  
 審査官 板垣 有紀  
 (56)参考文献 特開 2 0 1 4 - 1 7 8 9 5 7 ( J P , A )  
 特開 2 0 1 2 - 0 8 8 7 8 7 ( J P , A )  
 Pat Marison, et. al. , LabelFusion: A Pipeline for Generation Ground Truth Labels for Real R  
 GBD Data of Cluttered Scenes , arXive , 2017年09月26日 , <URL:https://arxiv.org/pdf/1  
 707.04796.pdf > , [ 検索日 : 2022年2月25日 ]  
 (58)調査した分野 (Int.Cl. , D B 名)  
 G 0 6 T 1 9 / 0 0  
 G 0 6 T 7 / 0 0 - 7 / 9 0