(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7250709号
(P7250709)

(45)発行日　令和5年4月3日(2023.4.3)

(24)登録日　令和5年3月24日(2023.3.24)

(51)国際特許分類		FI		
G06T	*7/00 (2017.01)*	G06T	7/00	350C
G06T	*7/593(2017.01)*	G06T	7/593	
G06T	*7/70 (2017.01)*	G06T	7/70	Z
G06N	*3/04 (2023.01)*	G06N	3/04	
G06N	*3/08 (2023.01)*	G06N	3/08	

請求項の数　14　（全23頁）

(21)出願番号　特願2019-571451(P2019-571451)
(86)(22)出願日　平成30年6月27日(2018.6.27)
(65)公表番号　特表2020-526818(P2020-526818A)
(43)公表日　令和2年8月31日(2020.8.31)
(86)国際出願番号　PCT/US2018/039804
(87)国際公開番号　WO2019/005999
(87)国際公開日　平成31年1月3日(2019.1.3)
審査請求日　令和3年6月16日(2021.6.16)
(31)優先権主張番号　62/526,203
(32)優先日　平成29年6月28日(2017.6.28)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　514108838
マジック リープ，インコーポレイテッド
Ｍａｇｉｃ Ｌｅａｐ，Ｉｎｃ．
アメリカ合衆国 フロリダ ３３３２２，プランテーション，ウエスト サンライズ ブールバード ７５００
７５００ Ｗ ＳＵＮＲＩＳＥ ＢＬＶＤ，ＰＬＡＮＴＡＴＩＯＮ，ＦＬ ３３３２２ ＵＳＡ
(74)代理人　100078282
弁理士　山本 秀策
(74)代理人　100113413
弁理士　森下 夏樹
(74)代理人　100181674
弁理士　飯田 貴敏

最終頁に続く

(54)【発明の名称】　畳み込み画像変換を使用して同時位置特定およびマッピングを実施する方法およびシステム

(57)【特許請求の範囲】

【請求項１】

２つの画像に基づいてホモグラフィを算出する方法であって、前記方法は、

第１のカメラ姿勢に基づく第１の画像と、第２のカメラ姿勢に基づく第２の画像とを受信することと、

前記第１の画像に基づく第１の２Ｄ点群と、前記第２の画像に基づく第２の２Ｄ点群とを生成することと、

前記第１の２Ｄ点群および前記第２の２Ｄ点群をニューラルネットワークに提供することと、

前記ニューラルネットワークによって、前記ニューラルネットワークへの入力として前記第１の２Ｄ点群および前記第２の２Ｄ点群を提供したことに応答して、前記ホモグラフィを生成することと

を含み、

前記ニューラルネットワークは、

複数の点を含む３Ｄ点群を生成することと、

２つのカメラ姿勢を取得することであって、前記２つのカメラ姿勢から前記複数の点が少なくとも部分的に視認可能であり、前記２つのカメラ姿勢は、前記第１のカメラ姿勢および前記第２のカメラ姿勢とは異なる、ことと、

前記２つのカメラ姿勢を使用して前記複数の点を２つの２Ｄ平面上に投影することにより、２つの２Ｄ点群を生成することと、

前記ニューラルネットワークによって、前記ニューラルネットワークへの入力として前記２つの２Ｄ点群を提供したことに応答して、特定のホモグラフィを生成することと、
前記２つのカメラ姿勢に基づいて、グラウンドトゥルースホモグラフィを決定することと、
前記特定のホモグラフィおよび前記グラウンドトゥルースホモグラフィを使用して前記ニューラルネットワークを修正することと
によって以前に訓練されたものである、方法。
【請求項２】
前記第１の画像は、第１の瞬間において第１のカメラによって捕捉され、前記第２の画像は、前記第１の瞬間後の第２の瞬間において前記第１のカメラによって捕捉されている、請求項１に記載の方法。
【請求項３】
前記第１の２Ｄ点群および前記第２の２Ｄ点群は、第１のニューラルネットワークを使用して生成され、前記ニューラルネットワークは、第２のニューラルネットワークである、請求項１に記載の方法。
【請求項４】
前記３Ｄ点群は、１つ以上の幾何学形状をサンプリングすることによって生成される、請求項１に記載の方法。
【請求項５】
前記２つのカメラ姿勢は、少なくとも３０％の視覚的重複を有する、請求項１に記載の方法。
【請求項６】
拡張現実（ＡＲ）デバイスであって、前記ＡＲデバイスは、
カメラと、
前記カメラに通信可能に結合されたプロセッサと
を備え、
前記プロセッサは、
前記カメラから、第１のカメラ姿勢に基づく第１の画像と、第２のカメラ姿勢に基づく第２の画像とを受信することと、
前記第１の画像に基づく第１の２Ｄ点群と、前記第２の画像に基づく第２の２Ｄ点群とを生成することと、
前記第１の２Ｄ点群および前記第２の２Ｄ点群をニューラルネットワークに提供することと、
前記ニューラルネットワークによって、前記ニューラルネットワークへの入力として前記第１の２Ｄ点群および前記第２の２Ｄ点群を提供したことに応答して、ホモグラフィを生成することと
を含む動作を実施するように構成されており、
前記ニューラルネットワークは、
複数の点を含む３Ｄ点群を生成することと、
２つのカメラ姿勢を取得することであって、前記２つのカメラ姿勢から前記複数の点が少なくとも部分的に視認可能であり、前記２つのカメラ姿勢は、前記第１のカメラ姿勢および前記第２のカメラ姿勢とは異なる、ことと、
前記２つのカメラ姿勢を使用して前記複数の点を２つの２Ｄ平面上に投影することにより、２つの２Ｄ点群を生成することと、
前記ニューラルネットワークによって、前記ニューラルネットワークへの入力として前記２つの２Ｄ点群を提供したことに応答して、特定のホモグラフィを生成することと、
前記２つのカメラ姿勢に基づいて、グラウンドトゥルースホモグラフィを決定することと、
前記特定のホモグラフィおよび前記グラウンドトゥルースホモグラフィを使用して前記ニューラルネットワークを修正することと

によって以前に訓練されたものである、ＡＲデバイス。
【請求項７】
前記第１の２Ｄ点群および前記第２の２Ｄ点群は、第１のニューラルネットワークを使用して生成され、前記ニューラルネットワークは、第２のニューラルネットワークである、請求項６に記載のＡＲデバイス。
【請求項８】
前記３Ｄ点群は、１つ以上の幾何学形状をサンプリングすることによって生成される、請求項６に記載のＡＲデバイス。
【請求項９】
前記２つのカメラ姿勢は、少なくとも３０％の視覚的重複を有する、請求項６に記載のＡＲデバイス。
【請求項１０】
命令を備えている非一過性コンピュータ読み取り可能な媒体であって、前記命令は、プロセッサによって実行されると、
第１のカメラ姿勢に基づく第１の画像と、第２のカメラ姿勢に基づく第２の画像とを受信することと、
前記第１の画像に基づく第１の２Ｄ点群と、前記第２の画像に基づく第２の２Ｄ点群とを生成することと、
前記第１の２Ｄ点群および前記第２の２Ｄ点群をニューラルネットワークに提供することと、
前記ニューラルネットワークによって、前記ニューラルネットワークへの入力として前記第１の２Ｄ点群および前記第２の２Ｄ点群を提供したことに応答して、ホモグラフィを生成することと
を含む動作を前記プロセッサに実施させ、
前記ニューラルネットワークは、
複数の点を含む３Ｄ点群を生成することと、
２つのカメラ姿勢を取得することであって、前記２つのカメラ姿勢から前記複数の点が少なくとも部分的に視認可能であり、前記２つのカメラ姿勢は、前記第１のカメラ姿勢および前記第２のカメラ姿勢とは異なる、ことと、
前記２つのカメラ姿勢を使用して前記複数の点を２つの２Ｄ平面上に投影することにより、２つの２Ｄ点群を生成することと、
前記ニューラルネットワークによって、前記ニューラルネットワークへの入力として前記２つの２Ｄ点群を提供したことに応答して、特定のホモグラフィを生成することと、
前記２つのカメラ姿勢に基づいて、グラウンドトゥルースホモグラフィを決定することと、
前記特定のホモグラフィおよび前記グラウンドトゥルースホモグラフィを使用して前記ニューラルネットワークを修正することと
によって以前に訓練されたものである、非一過性コンピュータ読み取り可能な媒体。
【請求項１１】
前記第１の画像は、第１の瞬間において第１のカメラによって捕捉され、前記第２の画像は、前記第１の瞬間後の第２の瞬間において前記第１のカメラによって捕捉されている、請求項１０に記載の非一過性コンピュータ読み取り可能な媒体。
【請求項１２】
前記第１の２Ｄ点群および前記第２の２Ｄ点群は、第１のニューラルネットワークを使用して生成され、前記ニューラルネットワークは、第２のニューラルネットワークである、請求項１０に記載の非一過性コンピュータ読み取り可能な媒体。
【請求項１３】
前記３Ｄ点群は、１つ以上の幾何学形状をサンプリングすることによって生成される、請求項１０に記載の非一過性コンピュータ読み取り可能な媒体。
【請求項１４】

前記２つのカメラ姿勢は、少なくとも３０％の視覚的重複を有する、請求項１０に記載の非一過性コンピュータ読み取り可能な媒体。

【発明の詳細な説明】

【技術分野】

【０００１】

（関連出願の相互参照）

本願は、その内容が参照することによってその全体として本明細書に組み込まれる、２０１７年６月２８日に出願され、「ＭＥＴＨＯＤ ＡＮＤ ＳＹＳＴＥＭ ＦＯＲ ＰＥＲＦＯＲＭＩＮＧ ＳＩＭＵＬＴＡＮＥＯＵＳ ＬＯＣＡＬＩＺＡＴＩＯＮ ＡＮＤ ＭＡＰＰＩＮＧ ＵＳＩＮＧ ＣＯＮＶＯＬＵＴＩＯＮＡＬ ＩＭＡＧＥ ＴＲＡＮＳＦＯＲＭＡＴＩＯＮ」と題された、米国仮特許出願第６２／５２６，２０３号の非仮出願であり、その優先権の利益を主張する。

【背景技術】

【０００２】

現代のコンピューティングおよびディスプレイ技術は、いわゆる「仮想現実」または「拡張現実」体験のためのシステムの開発を促進しており、デジタル的に再現された画像またはその一部が、現実であるように見える、もしくはそのように知覚され得る様式でユーザに提示される。仮想現実または「ＶＲ」シナリオは、典型的に、他の実際の実世界の視覚的入力に対する透過性を伴わずに、デジタルまたは仮想画像情報の提示を伴い、拡張現実または「ＡＲ」シナリオは、典型的に、ユーザの周囲の実際の世界の可視化への拡張として、デジタルまたは仮想画像情報の提示を伴う。

【０００３】

これらのディスプレイ技術において行われる進歩にもかかわらず、当技術分野において、拡張現実システム、特に、ディスプレイシステムに関連する改良された方法、システム、およびデバイスの必要がある。

【発明の概要】

【課題を解決するための手段】

【０００４】

本開示は、概して、同時位置特定およびマッピング（ＳＬＡＭ）を実施するためのシステムおよび方法に関する。より具体的に、本開示の実施形態は、頭部搭載型仮想現実（ＶＲ）、複合現実（ＭＲ）、および／または拡張現実（ＡＲ）デバイスにおいて、畳み込み画像変換を使用して、ＳＬＡＭを実施するためのシステムおよび方法を提供する。本発明の実施形態は、ユーザによって装着されるデバイスによって捕捉された画像を分析し、それによって、表示される仮想コンテンツの正確度を改良することによって、ユーザ／デバイス移動の正確な検出を可能にする。本発明は、ＡＲデバイスを参照して説明され得るが、本開示は、コンピュータビジョンおよび画像ディスプレイシステムにおける種々の用途にも適用可能である。

【０００５】

本発明の第１の側面では、２つの画像に基づいてホモグラフィを算出する方法が、提供される。方法は、第１のカメラ姿勢に基づく第１の画像と、第２のカメラ姿勢に基づく第２の画像とを受信することを含み得る。方法は、第１の画像に基づく第１の点群と、第２の画像に基づく第２の点群とを生成することも含み得る。方法は、第１の点群および第２の点群をニューラルネットワークに提供することをさらに含み得る。方法は、ニューラルネットワークによって、第１の点群および第２の点群に基づいて、ホモグラフィを生成することをさらに含み得る。いくつかの実施形態では、第１の点群および第２の点群は、２次元（２Ｄ）点群である。いくつかの実施形態では、第１の画像は、第１の瞬間において第１のカメラによって捕捉されている。いくつかの実施形態では、第２の画像は、第１の瞬間後の第２の瞬間において第１のカメラによって捕捉されている。いくつかの実施形態では、第１の点群および第２の点群は、第１のニューラルネットワークを使用して生成され、ニューラルネットワークは、第２のニューラルネットワークである。

【０００６】
いくつかの実施形態では、ニューラルネットワークは、複数の点を含む１つ以上の３Ｄ点群のうちの各３次元（３Ｄ）点群のために、複数の点の閾値距離内の３Ｄ軌道を決定することと、３Ｄ軌道をサンプリングし、特定の第１のカメラ姿勢および特定の第２のカメラ姿勢を取得することであって、複数の点は、少なくとも部分的に特定の第１のカメラ姿勢および特定の第２のカメラ姿勢から視認可能である、ことと、特定の第１のカメラ姿勢に基づいて、複数の点を第１の２Ｄ平面上に投影することと、第１の２Ｄ点群を生成し、特定の第２のカメラ姿勢に基づいて、複数の点を第２の２Ｄ平面上に投影することと、第２の２Ｄ点群を生成し、特定の第１のカメラ姿勢および特定の第２のカメラ姿勢に基づいて、第１の２Ｄ点群と第２の２Ｄ点群との間のグラウンドトゥルースホモグラフィを決定することと、ニューラルネットワークによって、第１の２Ｄ点群および第２の２Ｄ点群に基づいて、特定のホモグラフィを生成することと、特定のホモグラフィをグラウンドトゥルースホモグラフィと比較することと、比較に基づいて、ニューラルネットワークを修正することとによって事前に訓練されている。いくつかの実施形態では、複数の３Ｄ点群は、１つ以上の幾何学形状をサンプリングすることによって生成される。いくつかの実施形態では、特定の第１のカメラ姿勢および特定の第２のカメラ姿勢は、少なくとも３０％重複を有する。
【０００７】
本発明の第２の側面では、ＡＲデバイスが、提供される。ＡＲデバイスは、カメラを含み得る。ＡＲデバイスは、カメラに通信可能に結合されたプロセッサも含み、プロセッサは、カメラから、第１のカメラ姿勢に基づく第１の画像と、第２のカメラ姿勢に基づく第２の画像とを受信することと、第１の画像に基づく第１の点群と、第２の画像に基づく第２の点群とを生成することと、第１の点群および第２の点群をニューラルネットワークに提供することと、ニューラルネットワークによって、第１の点群および第２の点群に基づいて、ホモグラフィを生成することとを含む動作を実施するように構成され得る。いくつかの実施形態では、第１の点群および第２の点群は、２Ｄ点群である。いくつかの実施形態では、第１の点群および第２の点群は、第１のニューラルネットワークを使用して生成され、ニューラルネットワークは、第２のニューラルネットワークである。
【０００８】
本発明の第３の側面では、非一過性コンピュータ読み取り可能な媒体が、提供される。非一過性コンピュータ読み取り可能な媒体は、命令を含み得、命令は、プロセッサによって実行されると、第１のカメラ姿勢に基づく第１の画像と、第２のカメラ姿勢に基づく第２の画像とを受信することと、第１の画像に基づく第１の点群と、第２の画像に基づく第２の点群とを生成することと、第１の点群および第２の点群をニューラルネットワークに提供することと、ニューラルネットワークによって、第１の点群および第２の点群に基づいて、ホモグラフィを生成することとを含む動作をプロセッサに実施させる。いくつかの実施形態では、第１の点群および第２の点群は、２Ｄ点群である。いくつかの実施形態では、第１の画像は、第１の瞬間において第１のカメラによって捕捉され、第２の画像は、第１の瞬間後の第２の瞬間において第１のカメラによって捕捉されている。いくつかの実施形態では、第１の点群および第２の点群は、第１のニューラルネットワークを使用して生成され、ニューラルネットワークは、第２のニューラルネットワークである。
本明細書は、例えば、以下の項目も提供する。
（項目１）
２つの画像に基づいてホモグラフィを算出する方法であって、前記方法は、
第１のカメラ姿勢に基づく第１の画像と、第２のカメラ姿勢に基づく第２の画像とを受信することと、
前記第１の画像に基づく第１の点群と、前記第２の画像に基づく第２の点群とを生成することと、
前記第１の点群および前記第２の点群をニューラルネットワークに提供することと、
前記ニューラルネットワークによって、前記第１の点群および前記第２の点群に基づい

て、前記ホモグラフィを生成することと
を含む、方法。
（項目２）
前記第１の点群および前記第２の点群は、２次元（２Ｄ）点群である、項目１に記載の方法。
（項目３）
前記第１の画像は、第１の瞬間において第１のカメラによって捕捉され、前記第２の画像は、前記第１の瞬間後の第２の瞬間において前記第１のカメラによって捕捉されている、項目２に記載の方法。
（項目４）
前記第１の点群および前記第２の点群は、第１のニューラルネットワークを使用して生成され、前記ニューラルネットワークは、第２のニューラルネットワークである、項目２に記載の方法。
（項目５）
前記ニューラルネットワークは、
複数の点を含む１つ以上の３Ｄ点群のうちの各３次元（３Ｄ）点群のために、
前記複数の点の閾値距離内の３Ｄ軌道を決定することと、
前記３Ｄ軌道をサンプリングし、特定の第１のカメラ姿勢および特定の第２のカメラ姿勢を取得することであって、前記複数の点は、少なくとも部分的に前記特定の第１のカメラ姿勢および前記特定の第２のカメラ姿勢から視認可能である、ことと、
前記特定の第１のカメラ姿勢に基づいて、前記複数の点を第１の２Ｄ平面上に投影し、第１の２Ｄ点群を生成することと、
前記特定の第２のカメラ姿勢に基づいて、前記複数の点を第２の２Ｄ平面上に投影し、第２の２Ｄ点群を生成することと、
前記特定の第１のカメラ姿勢および前記特定の第２のカメラ姿勢に基づいて、前記第１の２Ｄ点群と前記第２の２Ｄ点群との間のグラウンドトゥルースホモグラフィを決定することと、
前記ニューラルネットワークによって、前記第１の２Ｄ点群および前記第２の２Ｄ点群に基づいて、特定のホモグラフィを生成することと、
前記特定のホモグラフィを前記グラウンドトゥルースホモグラフィと比較することと、
前記比較に基づいて、前記ニューラルネットワークを修正することと
によって以前に訓練されている、項目２に記載の方法。
（項目６）
前記複数の３Ｄ点群は、１つ以上の幾何学形状をサンプリングすることによって生成される、項目５に記載の方法。
（項目７）
前記特定の第１のカメラ姿勢および前記特定の第２のカメラ姿勢は、少なくとも３０％重複を有する、項目５に記載の方法。
（項目８）
拡張現実（ＡＲ）デバイスであって、前記ＡＲデバイスは、
カメラと、
前記カメラに通信可能に結合されたプロセッサと
を備え、
前記プロセッサは、
前記カメラから、第１のカメラ姿勢に基づく第１の画像と、第２のカメラ姿勢に基づく第２の画像とを受信することと、
前記第１の画像に基づく第１の点群と、前記第２の画像に基づく第２の点群とを生成することと、
前記第１の点群および前記第２の点群をニューラルネットワークに提供することと、
前記ニューラルネットワークによって、前記第１の点群および前記第２の点群に基づい

て、ホモグラフィを生成することと
を含む動作を実施するように構成されている、ＡＲデバイス。
（項目９）
前記第１の点群および前記第２の点群は、２次元（２Ｄ）点群である、項目８に記載のＡＲデバイス。
（項目１０）
前記第１の点群および前記第２の点群は、第１のニューラルネットワークを使用して生成され、前記ニューラルネットワークは、第２のニューラルネットワークである、項目９に記載のＡＲデバイス。
（項目１１）
前記ニューラルネットワークは、
複数の点を含む１つ以上の３Ｄ点群のうちの各３次元（３Ｄ）点群のために、
前記複数の点の閾値距離内の３Ｄ軌道を決定することと、
前記３Ｄ軌道をサンプリングし、特定の第１のカメラ姿勢および特定の第２のカメラ姿勢を取得することであって、前記複数の点は、少なくとも部分的に前記特定の第１のカメラ姿勢および前記特定の第２のカメラ姿勢から視認可能である、ことと、
前記特定の第１のカメラ姿勢に基づいて、前記複数の点を第１の２Ｄ平面上に投影し、第１の２Ｄ点群を生成することと、
前記特定の第２のカメラ姿勢に基づいて、前記複数の点を第２の２Ｄ平面上に投影し、第２の２Ｄ点群を生成することと、
前記特定の第１のカメラ姿勢および前記特定の第２のカメラ姿勢に基づいて、前記第１の２Ｄ点群と前記第２の２Ｄ点群との間のグラウンドトゥルースホモグラフィを決定することと、
前記ニューラルネットワークによって、前記第１の２Ｄ点群および前記第２の２Ｄ点群に基づいて、特定のホモグラフィを生成することと、
前記特定のホモグラフィを前記グラウンドトゥルースホモグラフィと比較することと、
前記比較に基づいて、前記ニューラルネットワークを修正することと
によって以前に訓練されている、項目９に記載のＡＲデバイス。
（項目１２）
前記複数の３Ｄ点群は、１つ以上の幾何学形状をサンプリングすることによって生成される、項目１１に記載のＡＲデバイス。
（項目１３）
前記特定の第１のカメラ姿勢および前記特定の第２のカメラ姿勢は、少なくとも３０％重複を有する、項目１１に記載のＡＲデバイス。
（項目１４）
命令を備えている非一過性コンピュータ読み取り可能な媒体であって、前記命令は、プロセッサによって実行されると、
第１のカメラ姿勢に基づく第１の画像と、第２のカメラ姿勢に基づく第２の画像とを受信することと、
前記第１の画像に基づく第１の点群と、前記第２の画像に基づく第２の点群とを生成することと、
前記第１の点群および前記第２の点群をニューラルネットワークに提供することと、
前記ニューラルネットワークによって、前記第１の点群および前記第２の点群に基づいて、ホモグラフィを生成することと
を含む動作を前記プロセッサに実施させる、非一過性コンピュータ読み取り可能な媒体。
（項目１５）
前記第１の点群および前記第２の点群は、２次元（２Ｄ）点群である、項目１４に記載の非一過性コンピュータ読み取り可能な媒体。
（項目１６）
前記第１の画像は、第１の瞬間において第１のカメラによって捕捉され、前記第２の画

像は、前記第１の瞬間後の第２の瞬間において前記第１のカメラによって捕捉されている、項目１５に記載の非一過性コンピュータ読み取り可能な媒体。
（項目１７）
前記第１の点群および前記第２の点群は、第１のニューラルネットワークを使用して生成され、前記ニューラルネットワークは、第２のニューラルネットワークである、項目１５に記載の非一過性コンピュータ読み取り可能な媒体。
（項目１８）
前記ニューラルネットワークは、
複数の点を含む１つ以上の３Ｄ点群のうちの各３次元（３Ｄ）点群のために、
前記複数の点の閾値距離内の３Ｄ軌道を決定することと、
前記３Ｄ軌道をサンプリングし、特定の第１のカメラ姿勢および特定の第２のカメラ姿勢を取得することであって、前記複数の点は、少なくとも部分的に前記特定の第１のカメラ姿勢および前記特定の第２のカメラ姿勢から視認可能である、ことと、
前記特定の第１のカメラ姿勢に基づいて、前記複数の点を第１の２Ｄ平面上に投影し、第１の２Ｄ点群を生成することと、
前記特定の第２のカメラ姿勢に基づいて、前記複数の点を第２の２Ｄ平面上に投影し、第２の２Ｄ点群を生成することと、
前記特定の第１のカメラ姿勢および前記特定の第２のカメラ姿勢に基づいて、前記第１の２Ｄ点群と前記第２の２Ｄ点群との間のグラウンドトゥルースホモグラフィを決定することと、
前記ニューラルネットワークによって、前記第１の２Ｄ点群および前記第２の２Ｄ点群に基づいて、特定のホモグラフィを生成することと、
前記特定のホモグラフィを前記グラウンドトゥルースホモグラフィと比較することと、
前記比較に基づいて、前記ニューラルネットワークを修正することと
によって以前に訓練されている、項目１５に記載の非一過性コンピュータ読み取り可能な媒体。
（項目１９）
前記複数の３Ｄ点群は、１つ以上の幾何学形状をサンプリングすることによって生成される、項目１８に記載の非一過性コンピュータ読み取り可能な媒体。
（項目２０）
前記特定の第１のカメラ姿勢および前記特定の第２のカメラ姿勢は、少なくとも３０％重複を有する、項目１８に記載の非一過性コンピュータ読み取り可能な媒体。
【図面の簡単な説明】
【０００９】
種々の実施形態の性質および利点のさらなる理解は、以下の図を参照することによって実現され得る。添付の図では、類似コンポーネントまたは特徴は、同一参照標識を有し得る。さらに、同一タイプの種々のコンポーネントは、参照標識の後、ダッシュと、類似コンポーネント間で区別する第２の標識とが続くことによって区別され得る。第１の参照標識のみが、本明細書で使用される場合、説明は、第２の参照標識に関係なく、同一の第１の参照標識を有する類似コンポーネントのうちの任意の１つに適用可能である。
【００１０】
【図１】図１は、本発明のいくつかの実施形態による同時位置特定およびマッピング（ＳＬＡＭ）アプローチの３つのサブタスクの視覚的概要を図示する。
【００１１】
【図２】図２は、本発明のいくつかの実施形態によるＤＨ－ＳＬＡＭ追跡システムを図示する。
【００１２】
【図３】図３は、本発明のいくつかの実施形態によるＭａｇｉｃＰｏｉｎｔＮｅｔアーキテクチャを図示する。
【００１３】

【図４】図４は、本発明のいくつかの実施形態によるＭａｇｉｃＰｏｉｎｔＮｅｔ合成訓練データを図示する。
【００１４】
【図５】図５は、本発明のいくつかの実施形態による点ベースのＨｏｍｏｇｒａｐｈｙＮｅｔアーキテクチャを図示する。
【００１５】
【図６】図６は、本発明のいくつかの実施形態による点ベースのＨｏｍｏｇｒａｐｈｙＮｅｔデータ生成を図示する。
【００１６】
【図７】図７は、本発明のいくつかの実施形態によるスマートアンカＳＬＡＭシステムを図示する。
【００１７】
【図８】図８は、本発明のいくつかの実施形態による２つの画像に基づいてホモグラフィを算出する方法を図示する。
【００１８】
【図９】図９は、本発明のいくつかの実施形態によるニューラルネットワークを訓練する方法を図示する。
【００１９】
【図１０】図１０は、本発明のいくつかの実施形態によるウェアラブル拡張現実（ＡＲ）デバイスの概略図を図示する。
【発明を実施するための形態】
【００２０】
画像カテゴリ化およびオブジェクト検出等のコンピュータビジョンタスクにおける深層学習成功の多くは、ＩｍａｇｅＮｅｔおよびＭＳ－ＣＯＣＯのような大規模な注釈が付けられたデータベースの利用可能性から生じる。しかしながら、同時位置特定およびマッピング（ＳＬＡＭ）のような姿勢追跡および再構成問題に対して、代わりに、ＭｉｃｒｏｓｏｆｔＫｉｎｅｃｔに基づくＦｒｅｉｂｕｒｇ－ＴＵＭＲＧＢＤデータセット、立体視カメラおよびＩＭＵに基づくＥｕＲｏＣドローン／ＭＡＶデータセット、ならびに４つのカメラ、ＧＰＳデータ、およびＶｅｌｏｄｙｎｅレーザスキャナを装備する車両のＫＩＴＴＩ運転データセット等のより小規模なデバイス特有のデータセットのフラグメント化されたエコシステムが存在する。
【００２１】
ＳＬＡＭのためのＩｍａｇｅＮｅｔは、現在、存在しない。実際、多数の環境およびシナリオのための正確なグラウンドトゥルース（ｇｒｏｕｎｄ－ｔｒｕｔｈ）姿勢測定を取得することは、困難である。グラウンドトゥルースセンサと視覚的ＳＬＡＭシステムとの間の正確な整列を得ることは、かなりの努力を要求し、異なるカメラにおける変動にわたってスケーリングすることは容易ではない。フォトリアリスティックレンダリングは、ＳＬＡＭタスクのための全ての関連幾何学的変数が１００％正確度で記録され得るので、有用であり得る。フォトリアリスティックシーケンス上のベンチマーキングＳＬＡＭは、利点を有し得るが、そのようなレンダリングされた画像に基づく訓練は、多くの場合、ドメイン適合問題に悩まされ、多くの深層ネットは、過剰適合すると考えられる。
【００２２】
実際は、ＳＬＡＭモデルが、ロボットおよび複合現実におけるそれら等の内蔵プラットフォーム上で大規模に起動するように効率的である場合、好ましい。本発明の実施形態は、そのようなシステムが内蔵プラットフォーム上で展開され得るように、完全フレーム予測とは対照的に、幾何学的一貫性に焦点を当てる。深層学習システムによって行われる完全フレーム予測は、ある利点を有するが、多くの場合では、点を予測すること／整列させることは、メトリックレベル姿勢復元のために十分である。深層ネットの展開は、通常、ネッを可能な限り小規模にするためのカスタムオフライン最適化プロシージャを伴い、したがって、本発明のいくつかの場合では、完全視覚的フレーム予測問題は、完全に省略さ

れる。

【００２３】

本発明の実施形態は、未知の環境における６自由度単眼カメラ位置特定のために、ＤＨ－ＳＬＡＭと称されるシステムを含み得、それは、主に、高センサ雑音、低照明、および主要なオクルージョンの存在下でホモグラフィをロバストに推定するように訓練される畳み込みニューラルネットワークを装備している。システムは、３つの深層ＣｏｎｖＮｅｔｓを含み得る。第１のネットワークであるＭａｇｉｃＰｏｉｎｔＮｅｔは、単一画像に動作し、画像内の顕著な点を抽出し（顕著な点は、設計によって、画像内で分離され、十分に分散させられている）、非最大抑制等の任意の追加の後処理を利用しないこともある。第２のネットワークであるＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔは、ＭａｇｉｃＰｏｉｎｔＮｅｔからの対の点応答マップに動作し、２つの点画像を関連させるホモグラフィを推定し得る。そして、ホモグラフィ推定は、標準的ＭＶＧ技法を使用して再推定されるか、または、場面が高度に非平面である場合、基本行列にアップグレードされ得る。第３のネットワークであるＲｅｌｏｃａｌｉｚａｔｉｏｎＮｅｔは、単一画像を信頼性があり、かつ高速な再位置特定のために使用される高速画像比較を可能にする低次元埋め込みベクトルに変換し得る。ネットワークは、単純合成データで訓練され、単純合成データは、それらが高価な外部カメラグラウンドトゥルース機器または高度なグラフィックレンダリングパイプラインに依拠しないので、ネットワークを訓練しやすくする。システムは、高速かつ効率的であり、ＣＰＵ上で３０＋ＦＰＳで起動することが可能である。

【００２４】

本発明の種々の実施形態では、３つのカスタムＳＬＡＭ畳み込みニューラルネットワークが、３つの別個の訓練ドメインとともに提示される。しかしながら、この特定の実装は、要求されず、ニューラルネットワークのうちの１つ以上のものは、いくつかの実施形態では、組み合わせられ得る。ホモグラフィが各システムの重要なコンポーネントであり得るので、ホモグラフィ的に導かれる単眼ＳＬＡＭシステムが、提示される。本発明の実施形態は、どんな種類のローカル特徴点記述子も使用しないこともあり、手動データ注釈または高価なオフライングラフィックレンダリングパイプラインを要求しないこともある。

【００２５】

図１は、本発明のいくつかの実施形態によるＳＬＡＭアプローチの３つのサブタスクの視覚的概要を図示する。サブタスク１０２では、幾何学的点プリミティブが、結像効果（センサ雑音、照明、テクスチャ、およびモーションぼけ等）を除去するために、画像から抽出される。サブタスク１０４では、対の点画像が、比較され、画像を関連させる、グローバル姿勢情報を抽出する。サブタスク１０６では、画像は、高速画像マッチングのために、低次元ベクトルに圧縮される。

【００２６】

サブタスク１０２は、画像を点のような幾何学的エンティティの組にマッピングすることに関する。サブタスク１０２の１つの目標は、照明、陰影、および全体的グローバル照明変動のような迷惑変数の損傷を元に戻すことである。角検出に類似した技法を使用して、入力画像内の信頼性がある場所を抽出することも望ましくあり得る。しかしながら、画像の角を分析的に定義する代わりに、カスタム２Ｄ形状レンダラが、角検出器を訓練するために使用される。結果として生じる２Ｄ画像場所が、場面点の準稠密被覆がロバストな変換推定のために重要であり得るので、システムから準稠密的に出力される。

【００２７】

サブタスク１０４は、一対の画像間の相対的姿勢を見出すことに関する。いくつかの場合では、ホモグラフィは、ワーピングされた自然画像の大規模データセットからサンプリングされた複数対のグレースケール画像に基づいて訓練されたＣＮＮから算出され得る。本発明のいくつかの実施形態では、ＣＮＮは、点画像、すなわち、ＭａｇｉｃＰｏｉｎｔＮｅｔによって出力された画像の種類の空間を使用して訓練される。準稠密点画像の空間は、完全ＲＧＢ画像の空間よりかなり小さくあり得るので、かなりより高い性能が、完全ＲＧＢ画像を取り扱うために必要であるものより少ない数のニューロンを使用して取得さ

れ得る。本明細書に提示される姿勢推定ネットワークは、点応答画像に動作し得、ホモグラフィモードおよび基本モードの両方において起動し得る。２つの異なる方法において３Ｄ世界を取り扱う能力は、複数視点幾何学形状の状況では、場面平面性の仮定が適用できることも、できないこともあるので、重要であり得る。平面性仮定が適用できる場合、ホモグラフィが、推定され、後に、個々の点深度を配慮せずに、（Ｒ，ｔ）推定値にアップグレードされることができる。代わりに、ソルバが、基本行列推定に基づく場合、場面平面性は、退化Ｅ推定値を提供し、全ての他の推定は、失敗し得る。故に、視認可能場面幾何学形状が高度に非平面であるとき、Ｅ、Ｒ、ｔ、および点深度は、直接、対処され得る。

【００２８】

サブタスク１０６は、画像の視覚的埋め込みを作成することに関する。他のタスクのために、姿勢に敏感であることは、重要であり得るが、埋め込みのために、姿勢に敏感でないことが、望ましくあり得る。実世界画像を使用することは、２Ｄ形状または点応答画像のようなエンジニアリングされた表現の上に、埋め込みを学習することが補助となる可能性が低くあり得るので、このタスクのために重要であり得る。代わりに、画像の大規模な自由に利用可能なデータセット（すなわち、ＩｍａｇｅＮｅｔ）が、使用され、ホモグラフィが、画像をワーピングさせるために使用される。学習中、２つの画像は、それらがホモグラフィ的に関連する場合、類似埋め込みベクトルを有するように強制され得る。

【００２９】

図２は、本発明のいくつかの実施形態によるＤＨ－ＳＬＡＭ追跡システム２００の一般的アーキテクチャを図示する。いくつかの場合では、対の画像（例えば、画像２０２および画像２０４）が、ＭａｇｉｃＰｏｉｎｔＮｅｔによって処理され、ＭａｇｉｃＰｏｉｎｔＮｅｔは、画像内の顕著な点を検出し、一対の点画像（例えば、点画像２０６および点画像２０８）を生成するように訓練される。点画像は、次いで、ＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔによって、一緒に処理され、時間Ｔにおける点画像と時間Ｔ＋１における点画像を関連させるホモグラフィＨを算出する。ネットワークからのＨ推定値は、ほぼ平面の場面に対して、多くの場合、点を互いの数ピクセル以内に投影し得、したがって、単純最近傍対応が、より精密なＨ’を再推定するために十分であるか、または、Ｈは、複数視点幾何学形状技法を使用して非平面場面をより詳細に説明する基本行列Ｆにアップグレードされ得る。そして、Ｈ’およびＦの両方は、カメラ行列Ｋを使用して、カメラの相対的姿勢に分解されることができる。

【００３０】

いくつかの場合では、ＳＬＡＭシステムは、３つのサブシステムに分解される：画像を２Ｄ点場所の規準的サブ空間の中にもたらすＭａｇｉｃＰｏｉｎｔ検出器、グローバル変換推定点ベースのＨｏｍｏｇｒａｐｈｙＮｅｔ、ＲｅｌｏｃＮｅｔと呼ばれる視覚的画像埋め込みエンジン。いくつかの実施形態では、各サブシステムは、別個の畳み込みニューラルネットワークである。ＭａｇｉｃＰｏｉｎｔＮｅｔおよびＲｅｌｏｃＮｅｔの両方が、単一画像に動作する一方、ＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔは、一対の画像に動作する。いくつかの実施形態では、ネットワークは、以下の問題を解決することが可能である：雑音の多い画像においてロバストな２Ｄ場所を検出すること、２つの画像間の相対的姿勢を算出すること、および、再位置特定。エンジニアリングされた特徴記述子（ＯＲＢまたはＳＩＦＴ）の両方に大きく依拠する従来のアプローチと異なり、本発明の実施形態は、記述子を画像内の個々の点に関連付けないこともある。これらの従来の特徴ベースのＳＬＡＭシステムと異なり、相対的姿勢推定は、無記述子方式で実施され得る。グローバル画像全体記述子に類似し得る埋め込みも、使用され得る。埋め込みは、擬似ホモグラフィ不変量であるようにエンジニアリングされ得る。設計によって、ホモグラフィによって関連付けられる２つの画像は、所与の多様体上で近接し得る。

【００３１】

いくつかの場合では、第１のステップは、画像内で顕著かつ位置特定可能な２Ｄ場所を検出することを含み得る。このステップは、ＨａｒｒｉｓまたはＦＡＳＴ等の角様応答マップを算出し、極大値を検出し、非最大抑制（ｎｏｎ－ｍａｘｉｍａｌ　ｓｕｐｐｒｅｓ

ｓｉｏｎ）を採用することによって実施され得る。追加のステップが、これらの極大値を画像全体を通して分散させるために実施され得る。このプロセスは、高度な専門領域の知識および手動エンジニアリングを伴い得、それは、一般化およびロバスト性を限定し得る。ＳＬＡＭ設定における有用性を増加させるために、点検出器によって検出された点は、フレームにわたる対応が容易であるように、画像全体を通して広く分散させられ、互いから分離され得る。システムが高センサ雑音シナリオおよび低光量において点を検出することも望ましくあり得る。いくつかの場合では、信頼度スコアが、検出された各点のために取得され得、それは、スプリアス点を排除することに役立つように使用され得る。これらの像点が画像内のローカル高勾配縁に対応する必要はないが、代わりに、ブロブの中心等の他の低レベルキューに対応し得、それが、従来の角検出器より大きい受け入れ可能野を利用し得ることに留意されたい。

【００３２】

図３は、本発明のいくつかの実施形態によるＭａｇｉｃＰｏｉｎｔＮｅｔアーキテクチャ３００を図示する。いくつかの実施形態では、ＭａｇｉｃＰｏｉｎｔＮｅｔは、グレースケール画像に動作し、入力の各ピクセルのための「点性（ｐｏｉｎｔ－ｎｅｓｓ）」確率を出力する。これは、明示的デコーダと組み合わせられたＶＧＧ式エンコーダを用いて、行われ得る。最終１５×２０×６５テンソルにおける各空間場所は、ローカル８×８領域に関する確率分布＋単一ダストビンチャネルを表し得、それは、点が検出されないことを表し得る（８×８＋１＝６５）。ネットワークは、２Ｄ形状レンダラからの点監視を使用した標準クロスエントロピ損失を使用して訓練される。

【００３３】

いくつかの場合では、ＭａｇｉｃＰｏｉｎｔＮｅｔは、カスタムＣｏｎｖＮｅｔアーキテクチャおよび訓練データパイプラインを使用して実装される。いくつかの場合では、画像Ｉを等価分解能を伴う点応答画像Ｐにマッピングすることが重要であり得、出力の各ピクセルは、入力におけるそのピクセルのための「点性」の確率に対応する。稠密予測のためのネットワーク設計は、エンコーダ－デコーダペアを伴い得、空間分解能は、プーリングまたはストライド畳み込みを介して減らされ、そして、上方畳み込み演算を介して、完全分解能に戻るようにアップサンプリングされる。アップサンプリング層は、より高い算出負担を追加し得るので、ＭａｇｉｃＰｏｉｎｔＮｅｔは、明示的デコーダを用いて実装され、モデルの算出を低減させ得る。畳み込みニューラルネットワークは、ＶＧＧ式エンコーダを使用して、画像の寸法を１２０×１６０から１５×２０セルグリッドに低減させ、各空間位置のための６５チャネルを伴い得る。いくつかの実施形態では、ＱＱＶＧＡ分解能は、算出負担を減少させるために、１２０×１６０であり得る。６５チャネルは、ピクセルのローカル非重複８×８グリッド領域＋余剰ダストビンチャネルに対応し得、それは、その８×８領域内で点が検出されないことに対応する。ネットワークは、３×３畳み込み後、ＢａｔｃｈＮｏｒｍ正規化およびＲｅＬＵ非線形性を使用して、完全に畳み込まれ得る。最終畳み込み層（ｃｏｎｖｌａｙｅｒ）は、１×１畳み込みであり得る。

【００３４】

図４は、本発明のいくつかの実施形態によるＭａｇｉｃＰｏｉｎｔＮｅｔ合成訓練データを図示する。いくつかの実施形態では、画像の着目点は、種々の視点、照明、および画像雑音変動にわたって安定している画像内の一意に識別可能な場所であり得る。スパースＳＬＡＭシステムのための事前処理ステップとして使用されるとき、所与のＳＬＡＭシステムのために良好に機能する点が、検出され得る。点検出アルゴリズムのハイパーパラメータを設計および選定することは、専門家および専門領域に特有の知識を利用し得る。

【００３５】

今日、着目点が標識された画像の大規模データベースは、存在しない。高価なデータ収集労力を回避するために、レンダラが、コンピュータビジョンライブラリを使用して実装され得る。いくつかの場合では、それらの各々のための角のグラウンドトゥルース場所が既知である三角形、四辺形、市松模様、３Ｄ立方体、および楕円形等の単純幾何学的形状が、レンダリングされる。各オブジェクトの２Ｄ面の重心も、既知であり、それらは、追

加の着目点としての役割を果たす。２Ｄ面の中心を見出す単純かつロバストな検出器を設計することが困難であろうことに留意されたい。形状がレンダリングされた後、ホモグラフィワーピングが、各画像に適用され、訓練例の数を増大させ得る。随意に、雑音、例えば、照明変化、テクスチャエンジニアリングされる雑音、ガウス雑音、ごま塩雑音、それらの組み合わせ等の形態における大量の雑音が、画像のうちの１つ以上のもの（例えば、各画像）に適用され得る。データは、オンザフライで生成され得、それは、どんな例もネットワークによって２回経験されないことを意味する。ネットワークは、１５×２０グリッドにおける各セルのためのロジットがソフトマックス関数を通して送られた後、標準クロスエントロピ損失を使用して訓練され得る。

【００３６】

いくつかの場合では、ＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔは、ＭａｇｉｃＰｏｉｎｔによって生産されるような一対の点画像を所与として、ホモグラフィを生産する。全ての画像の空間および相対的姿勢の空間（全ての画像の空間×相対的姿勢の空間）の代わりに、点画像の空間および相対的姿勢の空間（点画像の空間×相対的姿勢の空間）に動作するようにネットワークを設計することによって、照明、陰影、およびテクスチャ等のいくつかの考慮点は、重要性が低減させられるか、または無視され得る。さらに、適用できるために、測光一貫性仮定が、当てにされる必要はない。

【００３７】

図５は、本発明のいくつかの実施形態による点ベースのＨｏｍｏｇｒａｐｈｙＮｅｔアーキテクチャ５００を図示する。いくつかの場合では、対のバイナリ点画像が、連結され、そして、ＶＧＧ式エンコーダを通してフィードされる。３×３ホモグラフィＨが、完全結合層によって出力される。そして、ホモグラフィＨは、その右下要素が１であるように、正規化され得る。損失が、第２の画像内に対応を有することが既知の１つの画像内の点を第２の画像に変換し、その位置を比較することによって算出される。ＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔを訓練するための損失関数は、式（１）に示される。

【数１０】

$$L_H = \sum_{n=1}^{N} \| H x_n - x'_n \|_2 \qquad (1)$$

【００３８】

いくつかの場合では、ＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔは、直接、ＭａｇｉｃＰｏｉｎｔＮｅｔによって出力された点検出に動作するように設計される（但し、任意の従来の点検出器に動作することができる）。モデルは、対の準稠密１５×２０×６５画像上で良好に機能し得る。この小空間分解能では、ネットワークは、非常にわずかな算出を使用する。入力のチャネル毎連結後、３×３畳み込み、最大プーリング、ＢａｔｃｈＮｏｒｍ、およびＲｅＬＵアクティブ化後、２つの完全結合層から成るＶＧＧ式エンコーダは、実装され得、それは、３×３ホモグラフィＨの９つの値を出力する。

【００３９】

図６は、本発明のいくつかの実施形態による点ベースのＨｏｍｏｇｒａｐｈｙＮｅｔデータ生成を図示する。いくつかの実施形態では、ＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔを訓練するために、２つの仮想カメラの中にレンダリングされた点群の数百万の例が、生成され得る。点群は、平面、球体、および立方体を含む単純３Ｄ幾何学形状から生成され得る。２つの仮想カメラの位置は、図６に示されるように、区分線形平行移動およびランダム軸の周囲の回転から成るランダム軌道からサンプリングされる。いくつかの実施形態では、少なくとも３０％視覚的重複を有するカメラ対が、ランダムにサンプリングされる。点が、２つのカメラフレームの中に投影されると、点入力ドロップアウトが、適用され、スプリアスおよび欠測点検出に対するネットワークのロバスト性を改良する。いくつかの場合では、性能は、独立して、合致の５０％をランダムにドロップし、点の２５％をラ

ンダムにドロップすることによって改良される。
【００４０】
いくつかの考慮点が、典型的に、直接、３×３行列を出力するようにネットワークを訓練するために考慮される。いくつかの場合では、訓練は、最終ＦＣ層バイアスが、単位行列を出力するように初期化されるとき、ホモグラフィＨの座標が、範囲［－１，１］に正規化されるとき、および、ホモグラフィＨが８自由度を有し、かつ９つの要素を有するので、右下要素が１であるようにＨ数が正規化されるとき、最良に機能する。
【００４１】
いくつかの実施形態では、埋め込みネットワークの１つの目標は、グローバル１２８次元記述子を入力画像に関連付けることであり得る。いくつかの実施形態では、埋め込みが、ホモグラフィ不変であることが望ましい。例えば、ホモグラフィによって関連する２つの画像は、同一埋め込みベクトルを有するべきであり、同一場面コンテンツを描写しない（したがって、同一平面ではない）２つの画像は、異なる埋め込みベクトルを有するべきである。
【００４２】
埋め込みネットワークは、１２８　Ｌ２－正規化記述子を生産し得る。これは、ＶＧＧのようなエンコーダネットワークに加え、完全結合層によって行われ得る。いくつかの場合では、埋め込みネットワークは、ＩｍａｇｅＮｅｔデータセットからの対のホモグラフィ的に関連する画像を使用して訓練され得る。全て同一画像からの画像パッチのトリプレット（Ａ，Ａ’，Ｂ）が、サンプリングされ得、（Ａ，Ａ’）は、少なくとも３０％重複を有し、ホモグラフィによって関連付けられ、（Ａ，Ｂ）は、重複を有していない。ネットワークを訓練するために使用される、（Ａ，Ａ’）正対および（Ａ，Ｂ）負対が、生成され得る。いくつかの実施形態では、当業者に明白であり得るように、シャムネットワークが、使用されることができる一方、他の実施形態では、２タワーアプローチまたはトリプレットネットワークが、使用され得る。
【００４３】
ＭａｇｉｃＰｏｉｎｔＮｅｔが、ＦＡＳＴ角検出器およびＨａｒｒｉｓ角検出器のような従来の角検出ベースラインに対して評価された。ＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔは、実センサからの合成データおよび画像シーケンスの両方に関して、古典的ＯＲＢ＋ＲＡＮＳＡＣベースのホモグラフィ推定エンジンに対して評価された。合成ドット世界における評価の１つの利点は、点の組間のグラウンドトゥルース対応が既知であることである。雑音の量を変動させることが、２つのアルゴリズムが低下した程度を決定するために追加され得る。埋め込みネットワークを評価するために、最近傍のグリッドが、埋め込みメトリックを使用して算出される。これは、ベースＲｅｓＮｅｔアクティブ化を使用して、アクティブ化空間内の最近傍を求めることと比較された。
【００４４】
ホモグラフィＳＬＡＭシステム全体を評価するために、評価が、合致によって生産された最終（Ｒ，ｔ）推定値に関して実施された。追跡（最後のフレームに対する姿勢のみの推定）および埋め込み拡張追跡（最も近い埋め込みを伴うＫ枚の画像に対する姿勢の推定）の両方に関する数が、観察された。本発明の実施形態は、ＰＴＡＭ、ＯＲＢ－ＳＬＡＭ、およびＬＳＤ－ＳＬＡＭのようないくつかのオープンソースＳＬＡＭシステムと定質的に比較された。上で説明される評価の結果は、本発明の種々の実施形態が、従来のアプローチと比較して、より優れた性能を示すことを示した。
【００４５】
本発明の実施形態は、ＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔおよびＲｅｌｏｃＮｅｔとともに、少数の画像とともに、それらの関連付けられた点画像を使用する平面の周囲に設計される、小型拡張現実システムを含み得る。いくつかの場合では、データセット収集は、カメラ中心に向かって真っ直ぐに向いている法線［０，０，１］を伴う基準平面の頭部搭載構成から開始する短シーケンスを作成することを伴う。基準平面のそのような頭部搭載初期ビューを提供することは、ホモグラフィ分解からの２つの解の曖昧性解消ならび

に各２Ｄ目印のための初期スケールを可能にする。

【００４６】

図７は、本発明のいくつかの実施形態によるスマートアンカＳＬＡＭシステム７００を図示する。いくつかの実施形態では、ユーザは、最初に、屋内環境内のほとんど平面の表面の写真を撮影することによって、「スマートアンカ」の組を構築する。これは、対の点画像のユーザのキーフレームインデックスおよび埋め込み（E_0，E_1，E_2，・・・）を取り込む。インデックスが構築されると、システムは、追跡モードで起動されることができる。ＭａｇｉｃＰｏｉｎｔＮｅｔおよびＲｅｌｏｃＮｅｔは、時間Ｔにおける入力画像を処理し、点画像P_Tおよび埋め込みベクトルE_Tを生産し得る。いくつかの実施形態では、E_Tのドット積が、次いで、他の（例えば、１つおきの）埋め込み（E_0，E_1，E_2，・・・）のうちの１つ以上のもので算出され、埋め込み多様体上の最近傍を見出す。図７に示される特定の実施形態では、E_2が、選択される。E_2に対応する点画像P_2が、次いで、P_TとともにＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔの中にフィードされ、ホモグラフィＨを算出し得、それは、P_Tにおける点をP_2に変換する。ホモグラフィＨは、最後に、回転Ｒ、平行移動ｔ、および主平面ｎに分解され得る。最後に、（P_2，E_2）アンカに対応するコンテンツが、ワーピングされ、ＡＲオーバーレイとして入力画像内に表示され得る。

【００４７】

図８は、２つの画像に基づいてホモグラフィを算出する方法８００を図示する。方法８００のステップは、示されるものと異なる順序で実施され得、方法８００の１つ以上のステップは、方法８００の実施中、省略され得る。方法８００の１つ以上のステップは、非一過性コンピュータ読み取り可能な媒体内に含まれる命令を実行するように構成されるプロセッサによって、実施および／または開始され得る。

【００４８】

ステップ８０２では、第１の画像および第２の画像が、受信される。第１の画像は、第１のカメラ姿勢に基づき得、第２の画像は、第２のカメラ姿勢に基づき得、第２のカメラ姿勢は、第１のカメラ姿勢と異なる。いくつかの実施形態では、第１の画像および第２の画像は、同一カメラによって捕捉され得（それぞれ、第１の瞬間および第２の瞬間において、第２の瞬間は、第１の瞬間後に生じる）、および他の実施形態では、同時または２つの瞬間に、第１の画像が、第１のカメラによって捕捉され得、第２の画像が、第２のカメラによって捕捉され得る。

【００４９】

ステップ８０４では、第１の点群が、第１の画像に基づいて生成され、第２の点群が、第２の画像に基づいて生成される。いくつかの実施形態では、第１のニューラルネットワークが、点群を生成するために使用され、すなわち、第１の画像は、第１のニューラルネットワークへの入力として提供され、第１の点群は、第１の画像に基づいて、第１のニューラルネットワークによって生成され、第２の画像は、第１のニューラルネットワークへの入力として提供され、第２の点群は、第２の画像に基づいて、第１のニューラルネットワークによって生成される。第１のニューラルネットワークは、本明細書に説明されるＭａｇｉｃＰｏｉｎｔＮｅｔであり得、それは、画像に基づいて、２Ｄ点群を生成し得る。

【００５０】

ステップ８０６では、第１の点群および第２の点群は、第２のニューラルネットワークへの入力として提供される。第２のニューラルネットワークは、本明細書に説明されるＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔであり得、それは、点群に基づいて、ホモグラフィを生成し得る。いくつかの実施形態では、第１のニューラルネットワークは、２つのネットワークの機能性が単一システム内で組み合わせられ得るように、第２のニューラルネットワークと組み合わせられ得る。

【００５１】

ステップ８０８では、ホモグラフィは、第２のニューラルネットワークを使用して、第１の点群および第２の点群に基づいて生成される。いくつかの実施形態では、生成された

ホモグラフィは、行列（例えば、３×３）を含み、それから、第１のカメラ姿勢と第２のカメラ姿勢との間の相対的回転および相対的平行移動（すなわち、相対的姿勢）が、抽出され得る。

【００５２】

図９は、ニューラルネットワークを訓練する方法９００を図示する。方法９００のステップは、示されるものと異なる順序で実施され得、方法９００の１つ以上のステップは、方法９００の実施中、省略され得る。方法９００を参照して説明されるニューラルネットワークは、方法８００を参照して説明される第２のニューラルネットワークであり得、それは、本明細書に説明されるＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔであり得る。方法９００の１つ以上のステップは、非一過性コンピュータ読み取り可能な媒体内に含まれる命令を実行するように構成されるプロセッサによって、実施および／または開始され得る。

【００５３】

ステップ９０２では、１つ以上の３Ｄ点群が、生成される。３Ｄ点群の各々は、複数の３Ｄ点を含み得る。いくつかの実施形態では、１つ以上の３Ｄ点群は、他の可能性の中でもとりわけ、平面、球体、立方体等の１つ以上の幾何学形状をランダムにサンプリングすることによって生成される。例えば、特定の幾何学形状（例えば、球体または立方体）の表面が、ランダムにサンプリングされ、複数の３Ｄ点を生産し得る。代替として、特定の幾何学形状の縁が、ランダムにサンプリングされ得るか、または、表面および縁の両方が、ランダムにサンプリングされ得る。いくつかの実施形態では、特定の幾何学形状の体積全体が、ランダムにサンプリングされ得る。

【００５４】

いくつかの実施形態では、ステップ９０４－９１６の各々は、１つ以上の３Ｄ点群の各３Ｄ点群のために実施され得る。ステップ９０４では、３Ｄ軌道が、複数の点の近傍で決定され得る。例えば、３Ｄ軌道の全ては、複数の点の閾値距離内にあり得る。いくつかの実施形態では、３Ｄ軌道は、ランダム開始場所およびランダム終了場所を決定することによって形成される線形軌道である。他の実施形態では、または、同一実施形態では、３Ｄ軌道は、非線形軌道（例えば、湾曲）であるか、または、３Ｄ軌道は、ランダム開始場所、ランダム終了場所、および１つ以上の中間場所を決定することによって形成される一連の線形軌道である。

【００５５】

ステップ９０６では、３Ｄ軌道が、サンプリングされ、特定の第１のカメラ姿勢および特定の第２のカメラ姿勢を取得し得る。いくつかの実施形態では、複数の点は、少なくとも部分的に特定の第１のカメラ姿勢および特定の第２のカメラ姿勢から視認可能である。例えば、取得されるカメラ姿勢は、複数の点の少なくとも２５％、５０％、７５％、または１００％を視認するそれらのカメラ姿勢に制限され得る。カメラ姿勢が、所定の閾値（例えば、複数の点の少なくとも５０％が視認可能である）を満たさない場合、カメラ姿勢は、破棄され、３Ｄ軌道は、再サンプリングされ、別のカメラ姿勢を取得する。いくつかの実施形態では、取得されるカメラ姿勢は、互いに視覚的重複の少なくともある閾値（例えば、３０％）を有するように制限される。いくつかの実施形態では、視覚的重複は、特定の第１のカメラ姿勢および特定の第２のカメラ姿勢の両方によって視認可能な、複数の点のうちの点のパーセンテージに対応し得る。他の実施形態では、視覚的重複が、取得される姿勢間の共有視野に基づいて計算され得る。

【００５６】

ステップ９０８では、複数の点は、特定の第１のカメラ姿勢に基づいて、第１の２Ｄ平面上に投影され、第１の２Ｄ点群を生成し、同様に、複数の点は、特定の第２のカメラ姿勢に基づいて、第２の２Ｄ平面上に投影され、第２の２Ｄ点群を生成する。いくつかの実施形態では、第１の２Ｄ平面は、特定の第１のカメラ姿勢の向きベクトルに直交し得、第２の２Ｄ平面は、特定の第２のカメラ姿勢の向きベクトルに直交し得る。

【００５７】

ステップ９１０では、第１の２Ｄ点群と第２の２Ｄ点群との間のグラウンドトゥルース

ホモグラフィが、特定の第１のカメラ姿勢および特定の第２のカメラ姿勢に基づいて決定される。いくつかの実施形態では、グラウンドトゥルースホモグラフィは、最初に、特定の第１のカメラ姿勢と特定の第２のカメラ姿勢との間の相対的回転および相対的平行移動を決定し、そして、ニューラルネットワークによって生成されたホモグラフィと構造が一貫するホモグラフィ（行列）を形成することによって決定される。

【００５８】

ステップ９１２では、第１の２Ｄ点群および第２の２Ｄ点群は、入力としてニューラルネットワークに提供され得、特定のホモグラフィが、第１の２Ｄ点群および第２の２Ｄ点群に基づいて、ニューラルネットワークによって生成され得る。

【００５９】

ステップ９１４では、特定のホモグラフィは、グラウンドトゥルースホモグラフィと比較され、例えば、誤差信号を生成し得る。いくつかの実施形態では、誤差信号の大きさは、特定のホモグラフィとグラウンドトゥルースホモグラフィとの間の差異の大きさに比例し得る。１つの特定の実施形態では、誤差信号は、行列の対応する要素が互いから減算される標準行列減算方法を使用して計算される。他の実施形態では、または、同一実施形態では、誤差信号は、相対的回転における差異に対応する第１の成分と、相対的平行移動における差異に対応する第２の成分とを含む。いくつかの実施形態では、誤差信号は、相対的姿勢における差異に対応する単一成分を含む。

【００６０】

ステップ９１６では、ニューラルネットワークは、ステップ９１４において実施される特定のホモグラフィとグラウンドトゥルースホモグラフィとの間の比較に基づいて、例えば、ニューラルネットワークの１つ以上の重みまたは係数を調節することによって、修正される。いくつかの実施形態では、ニューラルネットワークは、より大きい誤差信号がより大きい修正をニューラルネットワークに生じさせるように、ホモグラフィ間の計算された差異（すなわち、誤差信号）に基づいて、修正され得る。一般に、ニューラルネットワークを修正することは、ニューラルネットワークがより正確になるようにし、それによって、特定のホモグラフィとグラウンドトゥルースホモグラフィとの間の差異を減少させる。

【００６１】

図１０は、本明細書に説明される実施形態のうちの１つ以上のものを採用し得るウェアラブルＡＲデバイス１０００の概略図を図示する。ＡＲデバイス１０００は、左接眼レンズ１００２Ａと、右接眼レンズ１００２Ｂと、直接、左接眼レンズ１００２Ａ上またはその近傍に取り付けられる左正面に面した世界カメラ１００６Ａと、直接、右接眼レンズ１００２Ｂ上またはその近傍に取り付けられる右正面に面した世界カメラ１００６Ｂと、左側に面した世界カメラ１００６Ｃと、右側に面した世界カメラ１００６Ｄと、処理モジュール１０５０とを含み得る。ＡＲデバイス１０００のコンポーネントの一部または全部は、投影された画像がユーザによって視認され得るように、頭部搭載型であり得る。１つの特定の実装では、図１０に示されるＡＲデバイス１０００のコンポーネントの全ては、ユーザによって装着可能な単一デバイス（例えば、単一ヘッドセット）ウェアラブル上に搭載される。別の実装では、処理モジュール１０５０は、ＡＲデバイス１０００の他のコンポーネントと物理的に別個であり、有線または無線接続性によって、それに通信可能に結合される。例えば、処理モジュール１０５０は、フレームに固定して取り付けられる構成、ユーザによって装着されるヘルメットまたは帽子に固定して取り付けられる構成、ヘッドホンに内蔵される構成、または別様に、ユーザに除去可能に取り付けられる構成（例えばリュック式構成、ベルト結合式構成等において）等、種々の構成において搭載され得る。

【００６２】

処理モジュール１０５０は、プロセッサ１０５２と、不揮発性メモリ（例えば、フラッシュメモリ）等のデジタルメモリとを備え得、両方は、データの処理、キャッシュ、および記憶を補助するために利用され得る。データは、画像捕捉デバイス（例えば、カメラ１００６）、マイクロホン、慣性測定ユニット、加速度計、コンパス、ＧＰＳユニット、無線デバイス、および／またはジャイロスコープから捕捉されたデータを含み得る。例えば

、処理モジュール１０５０は、カメラ１００６からの画像１０２０、より具体的に、左正面に面した世界カメラ１００６Ａからの左正面画像１０２０Ａ、右正面に面した世界カメラ１００６Ｂからの右正面画像１０２０Ｂ、左側に面した世界カメラ１００６Ｃからの左側画像１０２０Ｃ、および右側に面した世界カメラ１００６Ｄからの右側画像１０２０Ｄを受信し得る。いくつかの実施形態では、画像１０２０は、単一画像、一対の画像、画像のストリームを備えているビデオ、ペアリングされた画像のストリームを備えているビデオ等を含み得る。画像１０２０は、ＡＲデバイス１０００が電源オンである間、周期的に、生成され、処理モジュール１０５０に送信され得るか、または、処理モジュール１０５０によってカメラのうちの１つ以上のものに送信される命令に応答して生成され得る。

【００６３】

接眼レンズ１００２Ａおよび１００２Ｂは、プロジェクタ１０１４Ａおよび１０１４Ｂからの光を向けるように構成される透明または半透明導波管を備え得る。具体的に、処理モジュール１０５０は、左プロジェクタ１０１４Ａに、左投影画像１０２２Ａを左接眼レンズ１００２Ａの中に出力させ得、右プロジェクタ１０１４Ｂに、右投影画像１０２２Ｂを右接眼レンズ１００２Ｂの中に出力させ得る。いくつかの実施形態では、接眼レンズ１００２の各々は、各々が異なる色および／または異なる深度平面に対応する複数の導波管を備え得る。

【００６４】

カメラ１００６Ａおよび１００６Ｂは、それぞれ、ユーザの左および右眼の視野と実質的に重複する画像を捕捉するように位置付けられ得る。故に、カメラ１００６Ａおよび１００６Ｂの場所は、ユーザの眼の近傍であり得るが、ユーザの視野を曖昧にするほど近傍ではない。代替として、または加えて、カメラ１００６Ａおよび１００６Ｂは、それぞれ、投影された画像１０２２Ａおよび１０２２Ｂの内部結合場所と整合するように位置付けられ得る。カメラ１００６Ｃおよび１００６Ｄは、ユーザの側面、例えば、ユーザの周辺視覚内またはユーザの周辺視覚外の画像を捕捉するように位置付けられ得る。カメラ１００６Ｃおよび１００６Ｄを使用して捕捉された画像１０２０Ｃおよび１０２０Ｄは、必ずしも、カメラ１００６Ａおよび１００６Ｂを使用して捕捉された画像１０２０Ａおよび１０２０Ｂと重複する必要はない。

【００６５】

ＡＲデバイス１０００の動作中、処理モジュール１０５０は、訓練されたネットワーク１０５６を使用して、カメラ１００６の任意のものによる２つの捕捉された画像に基づいて、ホモグラフィを算出し得る。推定されたホモグラフィは、プロセッサ１０５２によって使用され、ユーザの移動に起因するユーザの視野の変化をより正確に反映する仮想コンテンツをレンダリングし得る。ネットワーク１０５６は、人工ニューラルネットワーク、畳み込みニューラルネットワーク、深層ネットワーク、または例を処理することによって徐々に「学習」し得る任意のタイプのネットワークもしくはシステムであり得る。いくつかの実施形態では、ネットワーク１０５６は、信号を１つのものから別のものに伝送することが可能である接続されたノードの集合を備えている。プロセッサ１０５２は、単一ネットワーク１０５６と通信し得るか、またはいくつかの実施形態では、プロセッサ１０５２は、第１のネットワーク（例えば、ＭａｇｉｃＰｏｉｎｔＮｅｔに対応する）、第２のネットワーク（例えば、ＰｏｉｎｔＨｏｍｏｇｒａｐｈｙＮｅｔに対応する）、および第３のネットワーク（例えば、ＲｅｌｏｃＮｅｔに対応する）等の複数のニューラルネットワークと通信し得る。

【００６６】

いくつかの例示的構成が説明されたが、種々の修正、代替構造、および均等物が、本開示の精神から逸脱することなく、使用され得る。例えば、前述の要素は、より大きいシステムのコンポーネントであり得、他のルールが、本技術の用途に優先するか、または別様にそれを修正し得る。いくつかのステップは、前述の要素が検討される前、間、または後にも行われ得る。故に、前述の説明は、請求項の範囲を束縛するものではない。

【００６７】

本明細書および添付の請求項で使用されるように、単数形「ａ」、「ａｎ」、および「ｔｈｅ」は、文脈によって明確に別様に示されない限り、複数参照を含む。したがって、例えば、「ユーザ」の言及は、複数のそのようなユーザを含み、「プロセッサ」の言及は、１つ以上のプロセッサおよび当業者に公知のその均等物等の言及を含む。

【００６８】

単語「ｃｏｍｐｒｉｓｅ（～を備えている）」、「ｃｏｍｐｒｉｓｉｎｇ（～を備えている）」、「ｃｏｎｔａｉｎｓ（～を含む）」、「ｃｏｎｔａｉｎｉｎｇ（～を含む）」、「ｉｎｃｌｕｄｅ（～を含む）」、「ｉｎｃｌｕｄｉｎｇ（～を含む）」、および「ｉｎｃｌｕｄｅｓ（～を含む）」も、本明細書および以下の請求項で使用されるとき、述べられた特徴、整数、コンポーネント、またはステップの存在を規定するために意図されるが、それらは、１つ以上の他の特徴、整数、コンポーネント、ステップ、行為、またはグループの存在または追加を除外するものではない。

【００６９】

本明細書に説明される例および実施形態が、例証目的のためだけのものであり、それに照らして、種々の修正または変更が、当業者に示唆され、本願の精神および権限ならびに添付される請求項の範囲内に含まれるものであることも理解されたい。

【図面】

【図１】

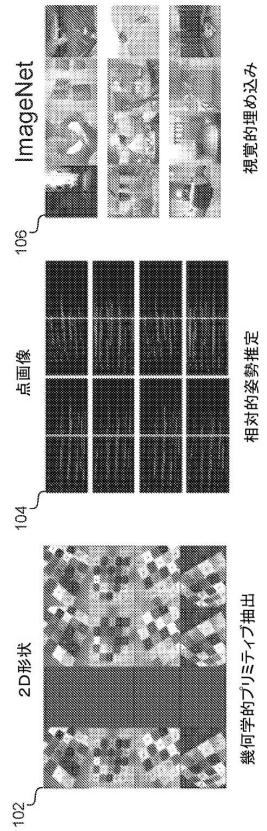

FIG. 1

【図２】

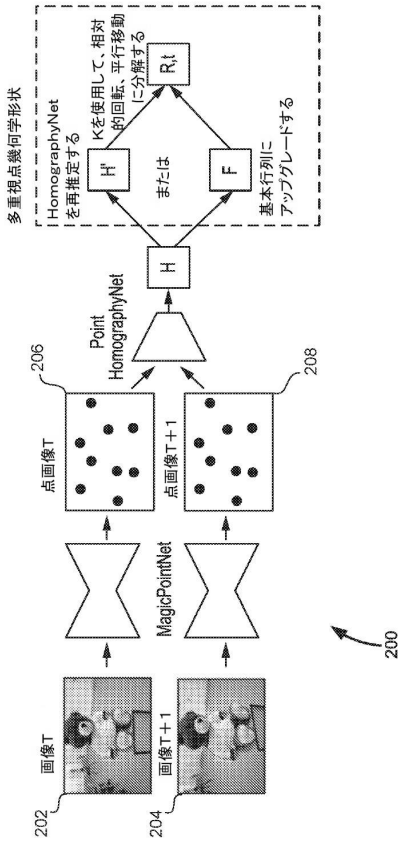

FIG. 2

【図３】

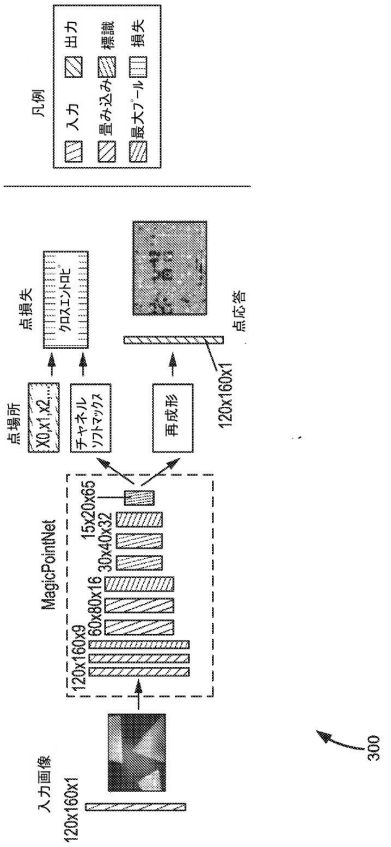

FIG. 3

【図４】

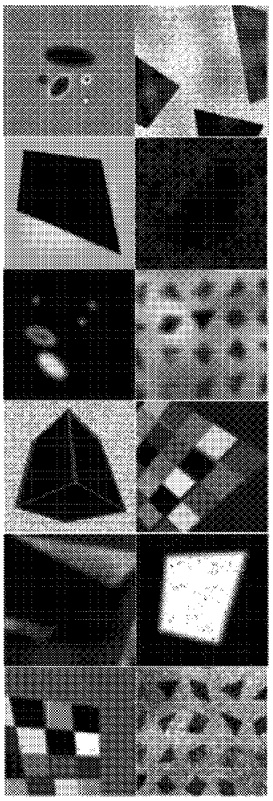

FIG. 4

【図５】

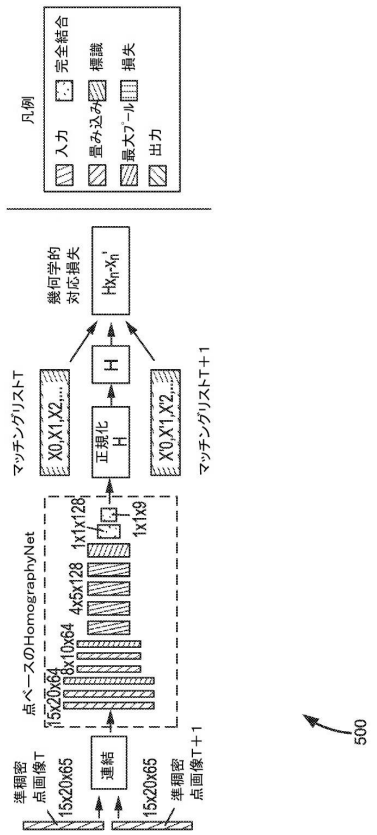

FIG. 5

【図６】

FIG. 6

【図７】

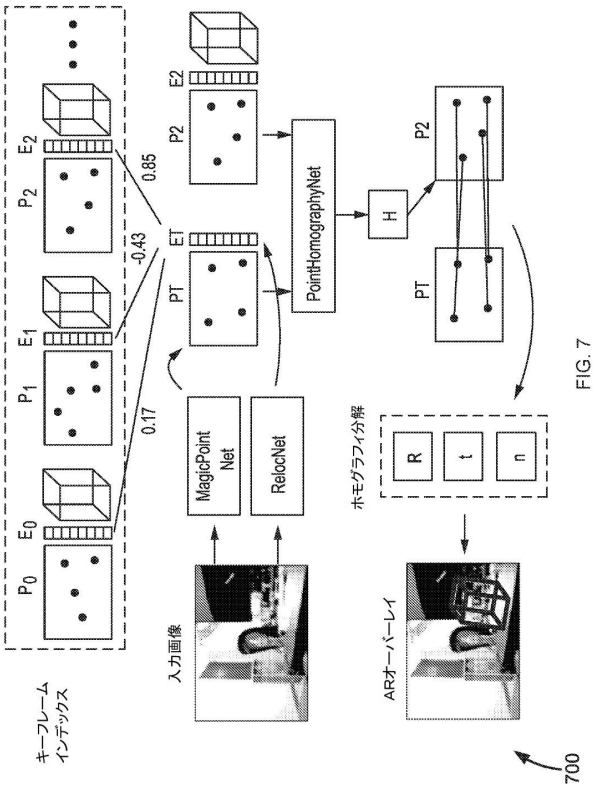

FIG. 7

【図８】

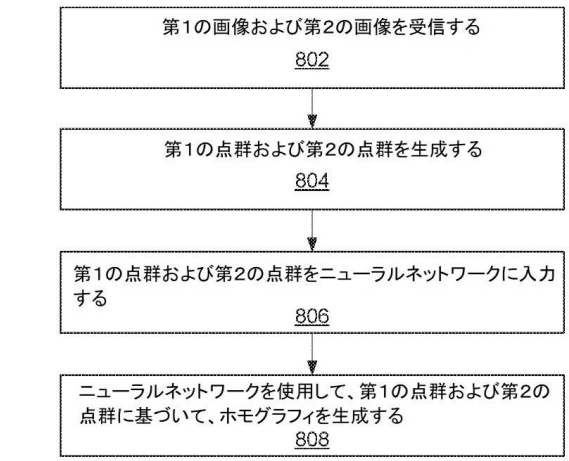

FIG. 8

【図９】

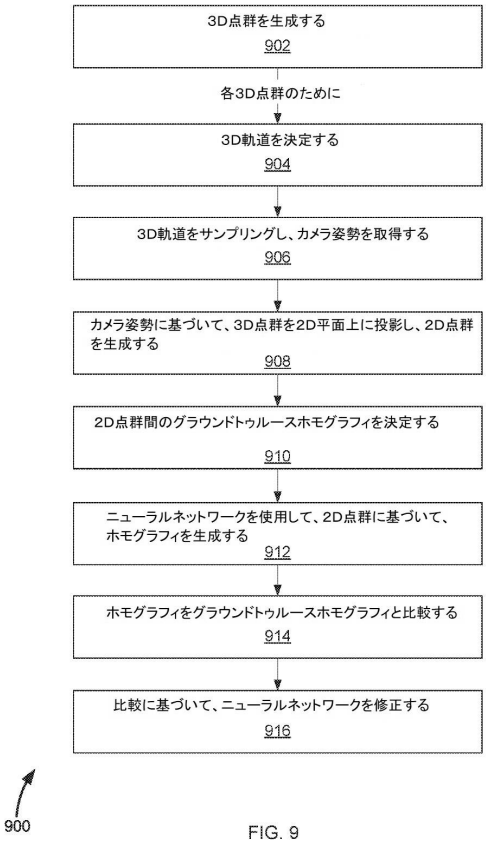

FIG. 9

【図１０】

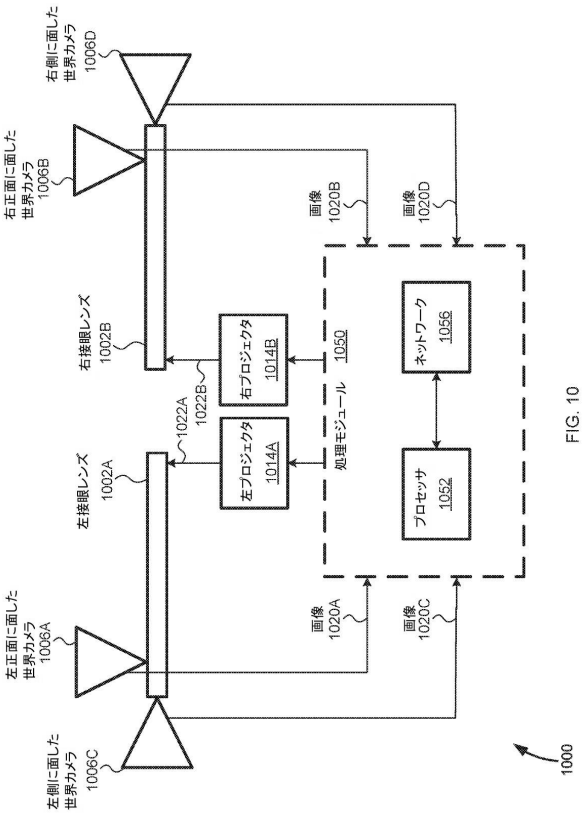

FIG. 10

フロントページの続き

(74)代理人　100181641
弁理士　石川 大輔
(74)代理人　230113332
弁護士　山本 健策
(72)発明者　デトン， ダニエル エル．
アメリカ合衆国 フロリダ ３３３２２， プランテーション， ダブリュー． サンライズ ブールバード ７５００
(72)発明者　マリシーウィッツ， トマシュ ジャン
アメリカ合衆国 フロリダ ３３３２２， プランテーション， ダブリュー． サンライズ ブールバード ７５００
(72)発明者　ラビノビッチ， アンドリュー
アメリカ合衆国 フロリダ ３３３２２， プランテーション， ダブリュー． サンライズ ブールバード ７５００

審査官　小池 正彦

(56)参考文献　Hani Altwaijry, et al.，Learning to Detect and Match Keypoints with Deep Architectures，vision.cornell.edu，米国，2016年08月01日，https://vision.cornell.edu/se3/wp-content/uploads/2016/08/learning-detect-match.pdf
Daniel DeTone, et al.，Deep Image Homography Estimation，arxiv.org，米国，CORNELL UNIVERSIT，2016年06月13日，https://arxiv.org/pdf/1606.03798.pdf

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｔ　７／００
Ｇ０６Ｔ　７／５９３
Ｇ０６Ｔ　７／７０
Ｇ０６Ｎ　３／０４
Ｇ０６Ｎ　３／０８