

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7638962号
(P7638962)

(45)発行日 令和7年3月4日(2025.3.4)

(24)登録日 令和7年2月21日(2025.2.21)

(51)国際特許分類		F I	
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00 3 5 0 C
G 0 6 V	10/82 (2022.01)	G 0 6 V	10/82
G 0 6 T	7/11 (2017.01)	G 0 6 T	7/11
請求項の数 16 (全39頁)			
(21)出願番号	特願2022-510817(P2022-510817)	(73)特許権者	514108838
(86)(22)出願日	令和2年8月19日(2020.8.19)		マジック リープ, インコーポレイテッド
(65)公表番号	特表2022-544696(P2022-544696 A)		M a g i c L e a p, I n c .
(43)公表日	令和4年10月20日(2022.10.20)		アメリカ合衆国 フロリダ 3 3 3 2 2 ,
(86)国際出願番号	PCT/US2020/047046		プランテーション, ウェスト サンライズ
(87)国際公開番号	WO2021/034961		ブルバード 7 5 0 0
(87)国際公開日	令和3年2月25日(2021.2.25)		7 5 0 0 W S U N R I S E B L V D
審査請求日	令和5年7月25日(2023.7.25)		, P L A N T A T I O N, F L 3 3 3
(31)優先権主張番号	62/888,953	(74)代理人	2 2 U S A
(32)優先日	令和1年8月19日(2019.8.19)		100104824
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	弁理士 穂場 仁
(31)優先権主張番号	62/926,241	(74)代理人	100121463
(32)優先日	令和1年10月25日(2019.10.25)	(74)代理人	弁理士 矢口 哲也
最終頁に続く		(74)代理人	100137969
			弁理士 岡部 憲昭
		最終頁に続く	

(54)【発明の名称】 軸外カメラを使用した眼追跡および視線推定

(57)【特許請求の範囲】
【請求項1】

ニューラルネットワークを訓練する方法であって、前記方法は、
第1の訓練ステップを実施するステップであって、前記第1の訓練ステップは、
第1の眼の第1の画像を前記ニューラルネットワークに入力として提供することであって、
前記ニューラルネットワークは、複数のタスク特有層のセットに接続された特徴エンコーディング層のセットを含み、前記複数のタスク特有層のセットは、前記特徴エンコーディング層のセットによって生成された出力に作用する、タスク特有層の少なくとも2つのセットを含む、ことと、
前記ニューラルネットワークの前記特徴エンコーディング層のセットと前記タスク特有層のセットを使用し、入力として前記第1の眼の前記第1の画像に基づいて、前記第1の眼の複数の領域へのセグメント化を含む眼セグメント化データを生成することと、
前記特徴エンコーディング層のセットに関連付けられる加重を修正することにより、前記眼セグメント化データを使用して、前記特徴エンコーディング層のセットを訓練することとを含む、ステップと、
第2の訓練ステップを実施するステップであって、前記第2の訓練ステップは、
第2の眼の第2の画像を前記ニューラルネットワークに入力として提供することと、
前記ニューラルネットワークの前記特徴エンコーディング層のセットおよび前記複数のタスク特有層のセットのそれぞれを使用して、入力として前記第2の眼の前記第2の画像に基づいて、前記複数のタスク特有層のセットの出力を含むネットワーク出力データを生成

することと、

前記複数のタスク特有層のセットに関連付けられる加重を修正することにより、前記ネットワーク出力データを使用して、前記複数のタスク特有層のセットを訓練することと
を含む、ステップと
を含む、方法。

【請求項 2】

前記第 1 の訓練ステップは、第 1 の持続時間の間、実施され、前記第 2 の訓練ステップは、前記第 1 の持続時間の後の第 2 の持続時間の間、実施される、請求項 1 に記載の方法。

【請求項 3】

前記複数の領域は、背景領域、強膜領域、瞳孔領域、または虹彩領域のうちの 1 つまたはそれを上回るものを含む、請求項 1 に記載の方法。

10

【請求項 4】

前記第 1 の訓練ステップを実施するステップはさらに、前記眼セグメント化データを使用して、前記複数のタスク特有層のセットの単一のタスク特有層のセットを訓練するステップを含む、請求項 1 に記載の方法。

【請求項 5】

前記単一のタスク特有層のセットは、前記第 1 の訓練ステップの間に訓練される前記複数のタスク特有層のセットの唯一のタスク特有層のセットである、請求項 4 に記載の方法。

【請求項 6】

前記第 1 の訓練ステップを実施するステップはさらに、
眼セグメント化グラウンドトゥルース (G T) データを受信するステップと、
前記眼セグメント化データと前記眼セグメント化 G T データを比較するステップと
を含む、請求項 1 に記載の方法。

20

【請求項 7】

前記特徴エンコーディング層のセットは、前記第 2 の訓練ステップの間、訓練されない、請求項 1 に記載の方法。

【請求項 8】

前記ネットワーク出力データは、前記第 2 の眼に対応する 2 次元 (2 D) 瞳孔データを含む、請求項 1 に記載の方法。

【請求項 9】

前記ネットワーク出力データは、前記第 2 の眼に対応する閃光検出データを含む、請求項 1 に記載の方法。

30

【請求項 10】

前記ネットワーク出力データは、前記第 2 の眼に対応する角膜中心データを含む、請求項 1 に記載の方法。

【請求項 11】

前記ネットワーク出力データは、前記第 2 の眼に対応する瞬目予測を含む、請求項 1 に記載の方法。

【請求項 12】

前記ネットワーク出力データは、前記第 2 の眼に対応する眼表現分類を含み、前記眼表現分類は、中性、幸福、弁別、または感受性のうちの少なくとも 1 つを含む、複数の可能性のある眼表現分類の 1 つである、請求項 1 に記載の方法。

40

【請求項 13】

前記ネットワーク出力データは、前記第 2 の眼の第 2 の複数の領域への第 2 のセグメント化を含む第 2 の眼セグメント化データを含む、請求項 1 に記載の方法。

【請求項 14】

非一過性コンピュータ可読媒体であって、前記非一過性コンピュータ可読媒体は、命令を備え、前記命令は、1 つまたはそれを上回るプロセッサによって実行されると、前記 1 つまたはそれを上回るプロセッサに、ニューラルネットワークを訓練するための動作を実施させ、前記動作は、

50

第1の訓練ステップを実施するステップであって、前記第1の訓練ステップは、
第1の眼の第1の画像を前記ニューラルネットワークに入力として提供することであって、
前記ニューラルネットワークは、複数のタスク特有層のセットに接続された特徴エンコーディング層のセットを含み、前記複数のタスク特有層のセットは、前記特徴エンコーディング層のセットによって生成された出力に作用する、タスク特有層の少なくとも2つのセットを含む、ことと、
前記ニューラルネットワークの前記特徴エンコーディング層のセットと前記タスク特有層のセットを使用し、入力として前記第1の眼の前記第1の画像に基づいて、前記第1の眼の複数の領域へのセグメント化を含む眼セグメント化データを生成することと、
前記特徴エンコーディング層のセットに関連付けられる加重を修正することにより、前記眼セグメント化データを使用して、前記特徴エンコーディング層のセットを訓練することとを含む、ステップと、

10

第2の訓練ステップを実施するステップであって、前記第2の訓練ステップは、
第2の眼の第2の画像を前記ニューラルネットワークに入力として提供することと、
前記ニューラルネットワークの前記特徴エンコーディング層のセットおよび前記複数のタスク特有層のセットのそれぞれを使用して、入力として前記第2の眼の前記第2の画像に基づいて、前記複数のタスク特有層のセットの出力を含むネットワーク出力データを生成することと、
前記複数のタスク特有層のセットに関連付けられる加重を修正することにより、前記ネットワーク出力データを使用して、前記複数のタスク特有層のセットを訓練することとを含む、ステップと
を含む、非一過性コンピュータ可読媒体。

20

【請求項15】

前記第1の訓練ステップは、第1の持続時間の間、実施され、前記第2の訓練ステップは、前記第1の持続時間の後の第2の持続時間の間、実施される、請求項14に記載の非一過性コンピュータ可読媒体。

【請求項16】

システムであって、
1つまたはそれを上回るプロセッサと、
非一過性コンピュータ可読媒体であって、前記非一過性コンピュータ可読媒体は、命令を備え、前記命令は、前記1つまたはそれを上回るプロセッサによって実行されると、前記1つまたはそれを上回るプロセッサに、ニューラルネットワークを訓練するための動作を実施させ、前記動作は、

30

第1の訓練ステップを実施するステップであって、前記第1の訓練ステップは、
第1の眼の第1の画像を前記ニューラルネットワークに入力として提供することであって、
前記ニューラルネットワークは、複数のタスク特有層のセットに接続された特徴エンコーディング層のセットを含み、前記複数のタスク特有層のセットは、前記特徴エンコーディング層のセットによって生成された出力に作用する、タスク特有層の少なくとも2つのセットを含む、ことと、
前記ニューラルネットワークの前記特徴エンコーディング層のセットと前記タスク特有層のセットを使用し、入力として前記第1の眼の前記第1の画像に基づいて、前記第1の眼の複数の領域へのセグメント化を含む眼セグメント化データを生成することと、
前記特徴エンコーディング層のセットに関連付けられる加重を修正することにより、前記眼セグメント化データを使用して、前記特徴エンコーディング層のセットを訓練することとを含む、ステップと、

40

第2の訓練ステップを実施するステップであって、前記第2の訓練ステップは、
第2の眼の第2の画像を前記ニューラルネットワークに入力として提供することと、
前記ニューラルネットワークの前記特徴エンコーディング層のセットおよび前記複数のタスク特有層のセットのそれぞれを使用して、入力として前記第2の眼の前記第2の画像に基づいて、前記複数のタスク特有層のセットの出力を含むネットワーク出力データを生成

50

することと、

前記複数のタスク特有層のセットに関連付けられる加重を修正することにより、前記ネットワーク出力データを使用して、前記複数のタスク特有層のセットを訓練することと

を含む、ステップと

を含む、非一過性コンピュータ可読媒体と

を備える、システム。

【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、あらゆる目的のために、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、2019年8月19日出願され、「EYENET: A MULTI-TASK DEEP NETWORK FOR OFF-AXIS EYE GAZE ESTIMATION AND SEMANTIC USER UNDERSTANDING」と題された、米国仮出願第62/888,953号、2019年10月25日出願され、「METHOD AND SYSTEM FOR PERFORMING EYE TRACKING USING AN OFF-AXIS CAMERA」と題された、米国仮出願第62/926,241号、および2019年11月14日出願され、「METHOD AND SYSTEM FOR PERFORMING EYE TRACKING USING AN OFF-AXIS CAMERA」と題された、米国仮出願第62/935,584号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、また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、すなわち、「VR」シナリオは、典型的には、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う。拡張現実、すなわち、「AR」シナリオは、典型的には、ユーザの周囲の実際の世界の可視化に対する拡張としての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成された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性が存在する。

【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、眼追跡のためのシステムおよび方法に関する。より具体的には、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)デバイスにおける視線推定のために眼追跡を実施するためのシステムおよび方法を提供する。本開示の実施形態は、コンテンツのエネルギーおよび帯域幅効率的レンダリングの使用が、効果的かつユーザの必要性の妨げにならない様式において、多焦点ディスプレイを駆動することを可能にする。本開示は、ARデバイスを参照して説明されるが、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適用可能である。

【0005】

本発明の概要は、下記に列举される一連の実施例を参照して提供される。下記で 사용되는ように、一連の実施例の任意の参照は、それらの実施例のそれぞれを分離した参照として理解されるべきである(例えば、「実施例1-4」は、「実施例1、2、3、または4」として理解されるべきである)。

【0006】

実施例1は、特徴エンコーディング層のセットと、それぞれ、特徴エンコーディング層のセットの出力に作用する、複数のタスク特有層のセットとを有する、ニューラルネットワークを訓練する方法であって、本方法は、第1の眼の第1の画像をニューラルネットワークに提供するステップと、ニューラルネットワークを使用して、第1の画像に基づいて、眼セグメント化データを生成するステップであって、眼セグメント化データは、第1の眼の複数の領域へのセグメント化を含む、ステップと、眼セグメント化データを使用して、特徴エンコーディング層のセットを訓練するステップとを含む、第1の訓練ステップを実施するステップと、第2の眼の第2の画像をニューラルネットワークに提供するステップと、特徴エンコーディング層のセットおよび複数のタスク特有層のセットのそれぞれを使用して、第2の画像に基づいて、ネットワーク出力データを生成するステップと、ネットワーク出力データを使用して、複数のタスク特有層のセットを訓練するステップとを含む、第2の訓練ステップを実施するステップとを含む。

10

【0007】

実施例2は、実施例1に記載の方法であって、第1の訓練ステップは、第1の持続時間の間、実施され、第2の訓練ステップは、第1の持続時間の後の第2の持続時間の間、実施される。

【0008】

実施例3は、実施例1に記載の方法であって、複数の領域は、背景領域、強膜領域、瞳孔領域、または虹彩領域のうちの1つまたはそれを上回るものを含む。

20

【0009】

実施例4は、実施例1に記載の方法であって、第1の訓練ステップを実施するステップはさらに、眼セグメント化データを使用して、複数のタスク特有層のセットの単一のタスク特有層のセットを訓練するステップを含む。

【0010】

実施例5は、実施例4に記載の方法であって、単一のタスク特有層のセットは、第1の訓練ステップの間に訓練される、複数のタスク特有層のセットの唯一のタスク特有層のセットである。

【0011】

実施例6は、実施例1に記載の方法であって、第1の訓練ステップを実施するステップはさらに、眼セグメント化グラウンドトゥールズ（GT）データを受信するステップと、眼セグメント化データと眼セグメント化GTデータを比較するステップとを含む。

30

【0012】

実施例7は、実施例1に記載の方法であって、特徴エンコーディング層のセットは、第2の訓練ステップの間、訓練されない。

【0013】

実施例8は、実施例1に記載の方法であって、ネットワーク出力データは、第2の眼に対応する、2次元（2D）瞳孔データを含む。

【0014】

実施例9は、実施例1に記載の方法であって、ネットワーク出力データは、第2の眼に対応する、閃光検出データを含む。

40

【0015】

実施例10は、実施例1に記載の方法であって、ネットワーク出力データは、第2の眼に対応する、角膜中心データを含む。

【0016】

実施例11は、実施例1に記載の方法であって、ネットワーク出力データは、第2の眼に対応する、瞬目予測を含む。

【0017】

実施例12は、実施例1に記載の方法であって、ネットワーク出力データは、第2の眼に対応する、眼表現分類を含む。

50

【0018】

実施例13は、実施例1に記載の方法であって、ネットワーク出力データは、第2の眼の第2の複数の領域への第2のセグメント化を含む、第2の眼セグメント化データを含む。

【0019】

実施例14は、1つまたはそれを上回るプロセッサによって実行されると、1つまたはそれを上回るプロセッサに、実施例1-13のいずれかに記載の方法を実施させる、命令を備える、非一過性コンピュータ可読媒体である。

【0020】

実施例15は、1つまたはそれを上回るプロセッサと、1つまたはそれを上回るプロセッサによって実行されると、1つまたはそれを上回るプロセッサに、実施例1-13のいずれかに記載の方法を実施させる、命令を備える、非一過性コンピュータ可読媒体とを備える、システムである。

10

【0021】

実施例16は、ユーザの眼表現を分類するためにニューラルネットワークを訓練する方法であって、本方法は、眼の画像を捕捉するステップと、眼の画像をニューラルネットワークに提供するステップと、ニューラルネットワークを使用して、眼の画像に基づいて、眼に対応する眼表現分類を生成するステップであって、眼表現分類は、複数の可能性として考えられる眼表現分類のうちの1つである、ステップと、グラウンドトゥールズ（GT）眼表現分類を決定するステップと、眼表現分類とGT眼表現分類との間の差異に基づいて、誤差データを算出するステップと、誤差データに基づいて、ニューラルネットワークを修正するステップとを含む。

20

【0022】

実施例17は、実施例16に記載の方法であって、眼の画像は、ウェアラブルディスプレイデバイスのカメラを使用して捕捉される。

【0023】

実施例18は、実施例16に記載の方法であって、GT眼表現分類を決定するステップは、GT眼表現分類を示す、ユーザ入力を受信するステップを含む。

【0024】

実施例19は、実施例16に記載の方法であって、GT眼表現分類を決定するステップは、ユーザに通信される命令がGT眼表現分類を示すことを決定するステップを含む。

30

【0025】

実施例20は、実施例16に記載の方法であって、眼の画像を捕捉するステップに先立って、GT眼表現分類を示す、命令をユーザに通信するステップをさらに含む。

【0026】

実施例21は、実施例16に記載の方法であって、ニューラルネットワークを修正するステップは、ニューラルネットワークの加重のセットを修正するステップを含む。

【0027】

実施例22は、実施例21に記載の方法であって、加重のセットは、逆伝搬を使用して修正される。

【0028】

実施例23は、実施例16に記載の方法であって、ニューラルネットワークは、誤差データの大きさに基づいて修正される。

40

【0029】

実施例24は、実施例16に記載の方法であって、複数の赤外線（IR）発光ダイオード（LED）によって、眼の画像が複数の閃光を含むように、光を眼に向かって出力するステップをさらに含む。

【0030】

実施例25は、実施例16に記載の方法であって、眼の画像は、複数の赤外線（IR）発光ダイオード（LED）によって出力された光によって生産された複数の閃光を含む。

【0031】

50

実施例 26 は、実施例 16 に記載の方法であって、眼の画像は、ユーザの眼の眉毛を含まない。

【0032】

実施例 27 は、実施例 16 に記載の方法であって、複数の可能性として考えられる眼表現分類は、中性、幸福、弁別、または感受性のうちの少なくとも 1 つを含む。

【0033】

実施例 28 は、1 つまたはそれを上回るプロセッサによって実行されると、1 つまたはそれを上回るプロセッサに、実施例 16 - 27 のいずれかに記載の方法を実施させる、命令を備える、非一過性コンピュータ可読媒体である。

【0034】

実施例 29 は、1 つまたはそれを上回るプロセッサと、1 つまたはそれを上回るプロセッサによって実行されると、1 つまたはそれを上回るプロセッサに、実施例 16 - 27 のいずれかに記載の方法を実施させる、命令を備える、非一過性コンピュータ可読媒体とを備える、システム。

【0035】

実施例 30 は、視線ベクトルを算出するためにニューラルネットワークを訓練する方法であって、本方法は、眼の画像を捕捉するステップと、眼の画像を処理し、眼に対応する光学軸を生産するステップと、光学軸をニューラルネットワークに提供するステップと、ニューラルネットワークを使用して、光学軸に基づいて、眼に対応する視線ベクトルを生成するステップと、視線ベクトルグラウンドトゥールズ (GT) データを決定するステップと、視線ベクトルと視線ベクトル GT データとの間の差異に基づいて、誤差データを算出するステップと、誤差データに基づいて、ニューラルネットワークを修正するステップとを含む。

【0036】

実施例 31 は、実施例 30 に記載の方法であって、眼の画像は、ウェアラブルディスプレイデバイスのカメラを使用して捕捉される。

【0037】

実施例 32 は、実施例 30 に記載の方法であって、視線ベクトル GT データは、標的が画面上に表示される、場所に基づいて決定される。

【0038】

実施例 33 は、実施例 30 に記載の方法であって、視線ベクトル GT データを決定するステップは、視線ベクトル GT データを示す、ユーザ入力を受信するステップを含む。

【0039】

実施例 34 は、実施例 30 に記載の方法であって、視線ベクトル GT データを決定するステップは、ユーザに通信される命令が視線ベクトル GT データを示すことを決定するステップを含む。

【0040】

実施例 35 は、実施例 30 に記載の方法であって、眼の画像を捕捉するステップに先立って、視線ベクトル GT データを示す、命令をユーザに通信するステップをさらに含む。

【0041】

実施例 36 は、実施例 30 に記載の方法であって、標的を画面上のある場所に表示するステップをさらに含み、視線ベクトル GT データは、場所に基づいて決定される。

【0042】

実施例 37 は、実施例 30 に記載の方法であって、ニューラルネットワークを修正するステップは、ニューラルネットワークの加重のセットを修正するステップを含む。

【0043】

実施例 38 は、実施例 37 に記載の方法であって、加重のセットは、逆伝搬を使用して修正される。

【0044】

実施例 39 は、実施例 30 に記載の方法であって、ニューラルネットワークは、誤差デ

10

20

30

40

50

一タの大きさに基づいて修正される。

【0045】

実施例40は、実施例30に記載の方法であって、複数の赤外線（IR）発光ダイオード（LED）によって、眼の画像が複数の閃光を含むように、光を眼に向かって出力するステップをさらに含む。

【0046】

実施例41は、実施例30に記載の方法であって、眼の画像は、複数の赤外線（IR）発光ダイオード（LED）によって出力された光によって生産された複数の閃光を含む。

【0047】

実施例42は、実施例30に記載の方法であって、視線ベクトルは、少なくとも1つの角度を含む。

【0048】

実施例43は、1つまたはそれを上回るプロセッサによって実行されると、1つまたはそれを上回るプロセッサに、実施例30-42のいずれかに記載の方法を実施させる、命令を備える、非一過性コンピュータ可読媒体である。

【0049】

実施例44は、1つまたはそれを上回るプロセッサと、1つまたはそれを上回るプロセッサによって実行されると、1つまたはそれを上回るプロセッサに、実施例30-42のいずれかに記載の方法を実施させる、命令を備える、非一過性コンピュータ可読媒体とを備える、システムである。

【0050】

従来の技法に優る多数の利点が、本開示の方法によって達成される。例えば、眼画像を通したユーザの眼視線推定および同時理解が、コンテンツのエネルギーおよび帯域幅効率的レンダリング（中心窩化レンダリング）を可能にし、コンテンツのより現実的レンダリングのための多焦点ディスプレイを駆動し（遠近調節輻輳・開散運動競合を最小限にする）、ユーザ表現を理解するための効果的かつ妨げにならない方法を提供する。付加的利点は、訓練されたネットワークを使用した推定が、古典的眼追跡パイプラインと良好に連動することである。訓練されたネットワークを使用した推定は、幾何学的眼追跡システム内で利用され、その全体的ロバスト性および正確度を改良することができることが実証されている。

【0051】

加えて、本明細書に説明される多段階眼追跡モデルからの結果は、AR／VR／MRにおける他の不可欠である用途を駆動することができる。例えば、角膜予測は、中心窩化レンダリングのために使用されることができ、眼セグメント化は、アバタベースのソーシャルスイートアプリにおいて眼をレンダリングするために有用である。多数の対象に関する視線標的GTデータを収集することは、不正確かつ困難の両方であり得るが、本明細書のデータ収集は、中間予測（瞳孔および角膜推定）の訓練を最終3D視線ベクトル推定パイプラインから分断することによって、有意により単純化される。エンドツーエンド深層ネットワークにおける誤差は、解釈することが困難であり得るため、訓練されたネットワークを使用して各段階において行われる中間推定は、解釈可能性を改良する。本開示の他の利点は、当業者に容易に明白となるであろう。

【図面の簡単な説明】

【0052】

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、同一参照ラベルを有し得る。さらに、同一タイプの種々のコンポーネントは、参照ラベルの後に、ダッシュと、類似コンポーネント間で区別する、第2のラベルとが続くことによって、区別され得る。第1の参照ラベルのみが、本明細書で使用される場合、説明は、第2の参照ラベルに関係なく、同一の第1の参照ラベル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。

10

20

30

40

50

【0053】

【図1】図1は、ウェアラブルARデバイスを通して視認されるような拡張現実（AR）場面を図示する。

【0054】

【図2】図2は、ARデバイスの種々の特徴を図示する。

【0055】

【図3】図3は、ヒトの眼の標準的二重球状モデルを図示する。

【0056】

【図4】図4は、ARデバイスの概略図を図示する。

【0057】

【図5】図5は、マルチタスクニューラルネットワークを組み込む、視線ベクトルを算出するためのシステムの概略図を図示する。

【0058】

【図6】図6は、マルチタスクニューラルネットワークの概略図を図示する。

【0059】

【図7】図7は、別個の時間ステップからの特徴を使用して瞬目予測を生成するためのシステムおよび技法を図示する。

【0060】

【図8】図8は、訓練モードで動作するARデバイスの概略図を図示する。

【0061】

【図9A】図9Aおよび9Bは、マルチタスクニューラルネットワークを訓練するための順次訓練ステップの概略図を図示する。

【図9B】図9Aおよび9Bは、マルチタスクニューラルネットワークを訓練するための順次訓練ステップの概略図を図示する。

【0062】

【図10】図10は、ランタイムモードで動作するARデバイスの概略図を図示する。

【0063】

【図11】図11は、視線ベクトルニューラルネットワークの概略図を図示する。

【0064】

【図12】図12は、訓練するパイプラインを図示する。

【0065】

【図13】図13は、特徴エンコーディング層のセットと、複数のタスク特有層のセットとを有する、ニューラルネットワークを訓練する方法を図示する。

【0066】

【図14】図14は、ユーザの眼表現を分類するためのニューラルネットワークを訓練する方法を図示する。

【0067】

【図15】図15は、視線ベクトルを算出するためのニューラルネットワークを訓練する方法を図示する。

【0068】

【図16】図16は、ニューラルネットワークを使用して視線ベクトルを算出する方法を図示する。

【0069】

【図17】図17は、ニューラルネットワークを訓練する方法を図示する。

【0070】

【図18】図18は、簡略化されたコンピュータシステムを図示する。

【発明を実施するための形態】

【0071】

発明の詳細な説明

眼画像を通したユーザの眼視線推定および同時意味論理解は、仮想現実（VR）および

10

20

30

40

50

複合現実（MR）内の重要なコンポーネントであって、エネルギー効率的レンダリング、多焦点ディスプレイ、および3Dコンテンツとの効果的相互作用を可能にする。頭部搭載型VR／MRデバイスでは、眼は、ユーザの視線を遮断することを回避するために、軸外で結像され得、これは、眼関連推定を引き出すことを非常に困難にし得る。本明細書に説明される種々の実施形態では、軸外カメラ設定のための眼視線推定および意味論ユーザ理解に関連する複数の異種タスクを解決する、単一深層ニューラルネットワークが、提供される。タスクは、眼セグメント化、瞬目検出、感情表現分類、赤外線放射（IR）発光ダイオード（LED）閃光検出、および瞳孔および角膜中心推定を含んでもよい。ニューラルネットワークをエンドツーエンドで訓練するために、手動標識教師およびモデルベースの教師の両方が、採用されてもよい。

10

【0072】

正確な視線を推定するプロセスは、外観ベースの算出（セグメント化、特徴点検出、例えば、瞳孔中心、閃光）に続いて、幾何学形状ベースの算出（例えば、3次元において、角膜、瞳孔中心、および視線ベクトルを推定する）を伴う。現在の眼トラッカは、古典的コンピュータビジョン技法（学習を伴わない）を使用し、瞳孔境界／中心を推定し、次いで、それらの推定に基づいて、視線を算出する。本明細書に説明される訓練されたネットワークを使用した推定は、古典的技法より有意に正確である。本明細書に説明されるいくつかの実施形態によると、単一深層ネットワークが、ともに軸外眼画像のための眼および視線推定に関連する複数の量を推定するように訓練される。

【0073】

図1は、本明細書に説明される実施形態による、ウェアラブルARデバイスを通して視認されるような拡張現実（AR）場面を図示する。AR場面100が、描写されており、AR技術のユーザには、人々、木々、背景における建物、およびコンクリートプラットフォーム120を特徴とする、実世界公園状設定106が見える。これらのアイテムに加え、AR技術のユーザはまた、実世界プラットフォーム120上に立っているロボット像110と、マルハナバチの擬人化のように見える、飛んでいる漫画のようなアバタキャラクター102とが「見える」と知覚するが、これらの要素（キャラクター102および像110）は、実世界には存在しない。ヒト視知覚および神経系の著しい複雑性に起因して、他の仮想または実世界画像要素の中で仮想画像要素の快適で、自然な感覚で、かつ豊かな提示を促進する、VRまたはAR技術を生産することは、困難である。

20

30

【0074】

図2は、本開示のいくつかの実施形態による、ARデバイス200の種々の特徴を図示する。いくつかの実施形態では、ARデバイス200は、ユーザがユーザの環境内のある場所（例えば、1つまたはそれを上回る深度面）に位置付けられるような1つまたはそれを上回る仮想オブジェクト（例えば、キャラクター102および像110）を知覚するように、仮想画像光222（仮想コンテンツと関連付けられる、光）を接眼レンズ202上に投影するように構成される、プロジェクタ214を含んでもよい。ユーザは、これらの仮想オブジェクトを世界オブジェクト230に沿って知覚し得る。ARデバイス200はまた、軸外カメラ240と、ARデバイス200に搭載され、ユーザの眼に向かって指向される、1つまたはそれを上回るエミッタ262を含んでもよい。エミッタ262は、ユーザの眼に不可視であるが、軸外カメラ240によって検出可能な光を透過させる、IRLEDを備えてもよい。いくつかの実施形態では、エミッタ262は、軸外カメラ240がIRスペクトル内の光を検出する能力を有する必要はないように、ユーザの眼に可視である光を透過させる、LEDを備えてもよい。したがって、軸外カメラ240は、IR検出能力の有無にかかわらず、カメラであってもよい。

40

【0075】

ARデバイス200の動作の間、軸外カメラ240は、ユーザの眼に対応する視線ベクトル238の推定につながる情報を検出してもよい（例えば、画像を捕捉する）。視線ベクトル238は、画像フレーム毎に算出されてもよく、種々の実施形態では、2次元（2D）または3次元（3D）値として表されてもよい。例えば、図2に図示されるように、

50

視線ベクトル 238 は、極性角度 θ および方位角 ϕ によって、球面座標系を使用して表されてもよい。代替として、または加えて、視線ベクトル 238 は、X、Y、および Z 値によって、3D デカルト座標系を使用して表されてもよい。視線ベクトル 238 は、ユーザの眼の場所、接眼レンズ 202 の場所、および視線ベクトル 238 に基づいて計算され得る、交点 239 において、接眼レンズ 202 と交差し得る。いくつかのインスタンスでは、プロジェクタ 214 は、視野の他のエリアに関連して、交点 239 の周囲の画像明度および／または明確性を改良するために、仮想画像光 222 を調節してもよい。

【0076】

図示されるように、4つのエミッタ 262 のセットが、ディスプレイ内およびその周囲に設置され、その反射（閃光）が、軸外カメラ 240 を使用して検出される。本設定は、ユーザの左眼に関しても複製される。検出された閃光は、直接、眼カメラ画像から観察可能ではない、眼内の重要な幾何学的数量を推定するために使用される。図 2 に示されるように、ユーザの視線とカメラ軸との間に大角度が存在し得る。これは、瞳孔の増加された偏心、眼瞼および睫毛によって生じる部分的オクルージョン、および環境照明に起因して生じる閃光の乱れに起因して、眼視線推定を困難にする。

【0077】

図 3 は、ヒトの眼の標準的二重球状モデル 300 を図示する。モデル 300 によると、眼球球体 302 は、完全または部分的に、内側角膜球面 304 を包含し得る。角膜中心 306 は、角膜球面 304 の幾何学的中心であり得る。瞳孔中心 308 は、眼の瞳孔開口部または瞳孔中心に対応し得、角膜球面 304 によって包含され得る。眼の光学軸 310 は、角膜中心 306 および瞳孔中心 308 を接続することによって形成される、ベクトルであり得る。視線ベクトル 238（代替として、視軸と称される）は、眼の背面において、角膜中心 306 および中心窩 312 を接続することによって形成され得る。中心窩 312 は、概して、未知であって、推定することが困難であるため、視線ベクトル 238 は、光学軸 310 およびユーザ特有の較正角度 κ を使用して算出され得る。較正角度 κ は、1次元（1D）、2D、または3D 値であってもよく、AR デバイス 200 が、ユーザによって最初に動作されるときの較正位相の間、特定のユーザのために較正されてもよい。いったん較正角度 κ が、特定のユーザのために算出されると、固定されると仮定される。故に、角膜中心 306 および瞳孔中心 308 を使用して、光学軸 310 を推定することは、視線追跡の根底にある重要な問題であり得る。

【0078】

図 4 は、本開示のいくつかの実施形態による、AR デバイス 200 の概略図を図示する。AR デバイス 200 は、左接眼レンズ 202 A と、右接眼レンズ 202 B と、直接、左接眼レンズ 202 A の上またはその近くに取り付けられる、左正面に向いた世界カメラ 206 A と、直接、右接眼レンズ 202 B の上またはその近くに取り付けられる、右正面に向いた世界カメラ 206 B と、左側に向いた世界カメラ 206 C と、右側に向いた世界カメラ 206 D と、処理モジュール 250 とを含んでもよい。エミッタ 262 は、接眼レンズ 202 の一方または両方に搭載されてもよく、いくつかの実施形態では、直接、左接眼レンズ 202 A の上またはその近くに搭載される、左エミッタ 262 A と、直接、右接眼レンズ 202 B の上またはその近くに搭載される、右エミッタ 262 B とに分離されてもよい（例えば、AR デバイス 200 のフレームに搭載される）。いくつかのインスタンスでは、AR デバイス 200 は、中心に位置付けられた軸外カメラ 260、または図 4 に図示されるように、直接、左接眼レンズ 202 A の上またはその近くに搭載される、左軸外カメラ 260 A、および直接、右接眼レンズ 202 B の上またはその近くに搭載される、右軸外カメラ 260 B 等、単一または複数の軸外カメラ 260 を含んでもよい。

【0079】

AR デバイス 200 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、頭部に搭載されてもよい。1つの特定の実装では、図 4 に示される AR デバイス 200 のコンポーネントは全て、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）上に搭載される。別の実装では、処理モジュール 25

10

20

30

40

50

0 は、AR デバイス 200 の他のコンポーネントと物理的に別個であって、有線または無線コネクティビティによって、通信可能に結合される。例えば、処理モジュール 250 は、フレームに固定して取り付けられる、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、ヘッドホンに内蔵される、または別様に、ユーザに除去可能に取り付けられる（例えば、リュック式構成において、ベルト結合式構成において等）等、種々の構成において搭載されてもよい。

【0080】

処理モジュール 250 は、少なくとも 1 つのプロセッサ 252 と、不揮発性メモリ（例えば、フラッシュメモリ）等の関連付けられるデジタルメモリとを備えてもよく、その両方とも、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 206 および軸外カメラ 260）、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等のセンサ（例えば、AR デバイス 200 に動作可能に結合される）から捕捉されたデータを含んでもよい。例えば、処理モジュール 250 は、カメラ 206 からの画像 220、より具体的には、左正面に向いた世界カメラ 206 A からの左正面画像 220 A、右正面に向いた世界カメラ 206 B からの右正面画像 220 B、左側に向いた世界カメラ 206 C からの左側画像 220 C、および右側に向いた世界カメラ 206 D からの右側画像 220 D を受信してもよい。いくつかの実施形態では、画像 220（または軸外カメラ 260 から受信されるもの）は、単一画像、一対の画像、画像のストリームを含むビデオ、ペアリングされた画像のストリームを含むビデオ、および同等物を含んでもよい。画像 220（または軸外カメラ 260 から受信されるもの）は、AR デバイス 200 の電源がオンの間、周期的に生成され、処理モジュール 250 に送信されてもよい、または処理モジュール 250 によってカメラのうちの 1 つまたはそれを上回るものに送信される命令に応答して、生成されてもよい。

【0081】

いくつかの実施形態では、処理モジュール 250 の機能性は、別個に格納されるが、通信可能に結合される、電子ハードウェアコンポーネントの 2 つまたはそれを上回るセット（例えば、1 つまたはそれを上回るプロセッサ、記憶デバイスのセット等）によって実装されてもよい。例えば、処理モジュール 250 の機能性は、ヘッドセット内に格納される電子ハードウェアコンポーネントによって、ヘッドセットに物理的に繋がれる、コンピューティングデバイス内に格納される電子ハードウェアコンポーネント、ヘッドセットの環境内の 1 つまたはそれを上回る電子デバイス（例えば、スマートフォン、コンピュータ、周辺デバイス、スマート家電等）、1 つまたはそれを上回る遠隔に位置するコンピューティングデバイス（例えば、サーバ、クラウドコンピューティングデバイス等）、またはそれらの組み合わせと併せて行われてもよい。

【0082】

接眼レンズ 202 A および 202 B は、それぞれ、プロジェクタ 214 A および 214 B からの光を指向するように構成される、透明または半透明導波管を備えてもよい。具体的には、処理モジュール 250 は、左プロジェクタ 214 A に、左仮想画像光 222 A を左接眼レンズ 202 A 上に出力させ得、右プロジェクタ 214 B に、右仮想画像光 222 B を右接眼レンズ 202 B 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 202 はそれぞれ、異なる色および／または異なる深度面に対応する、複数の導波管を備えてもよい。

【0083】

カメラ 206 A および 206 B 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てもよい。故に、カメラ 206 A および 206 B の設置場所は、ユーザの眼の近傍であってもよいが、ユーザの視野を不明瞭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 206 A および 206 B は、それぞれ、仮想画像光 222 A および 222 B の内部結合場所と整合するように位置付けられてもよい。カメラ 206 C および 206 D 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てもよい。カメラ206Cおよび206Dを使用して捕捉された画像220Cおよび220Dは、必ずしも、カメラ206Aおよび206Bを使用して捕捉された画像220Aおよび220Bと重複する必要はない。カメラ260Aおよび260Bは、それぞれ、ユーザの左および右眼の画像を捕捉するように位置付けられてもよい。カメラ260によって捕捉された画像は、ユーザの眼をその全体として、またはユーザの眼の一部を示してもよい。

【0084】

ARデバイス200の動作の間、処理モジュール250は、マルチタスクニューラルネットワーク256を使用して、視線ベクトル238を算出してもよい。いくつかの実施形態では、マルチタスクニューラルネットワーク256は、処理モジュール250の少なくとも1つのプロセッサ252と関連付けられる、または別様にそれにアクセス可能な非一過性メモリ内に記憶されてもよい。マルチタスクニューラルネットワーク256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、または処理実施例によって徐々に「学習」し得る、任意のタイプのコンピューティングシステムであってもよい。例えば、マルチタスクニューラルネットワーク256は、グラウンドトゥールズ（GT）データを表す、手動で準備された訓練データを処理することによって、訓練されてもよい。訓練データのそれぞれを処理後、マルチタスクニューラルネットワーク256は、GTデータにより緊密に近似する、出力を生成することが可能である。

【0085】

いくつかの実施形態では、マルチタスクニューラルネットワーク256は、信号を1つのノードから別のノード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る。例えば、マルチタスクニューラルネットワーク256は、そのようなノードのいくつかの異なる層を含んでもよい。下記にさらに詳細に説明されるように、いくつかの実施形態では、マルチタスクニューラルネットワーク256は、エンコーダ層と、デコーダ層とを含んでもよい。いくつかの実施形態では、マルチタスクニューラルネットワーク256の1つまたはそれを上回るエンコーダ層は、第1のセットの1つまたはそれを上回るプロセッサと関連付けられる、非一過性メモリ内に記憶されてもよい一方、マルチタスクニューラルネットワーク256の1つまたはそれを上回るデコーダ層は、別個に格納されるが、第1のセットの1つまたはそれを上回るプロセッサに通信可能に結合される、第2のセットの1つまたはそれを上回るプロセッサと関連付けられる、非一過性メモリ内に記憶されてもよい。例えば、第1のセットの1つまたはそれを上回るプロセッサは、ヘッドセット内に格納される、1つまたはそれを上回るプロセッサを含んでもよい一方、第2のセットの1つまたはそれを上回るプロセッサは、ヘッドセットに物理的に繋がれる、コンピューティングデバイス内に格納される、1つまたはそれを上回るプロセッサ、ヘッドセットと物理的に別個である、1つまたはそれを上回る電子デバイス（例えば、スマートフォン、コンピュータ、周辺デバイス、サーバ、クラウドコンピューティングデバイス等）、またはそれらの組み合わせを含んでもよい。マルチタスクニューラルネットワーク256の訓練および使用は、下記にさらに説明される。

【0086】

図5は、マルチタスクニューラルネットワーク256を組み込む、視線ベクトルを算出するためのシステムの概略図を図示する。いくつかの実施形態では、入力画像I（x，y，c）が、軸外カメラ260によって捕捉され、マルチタスクニューラルネットワーク256への入力として提供される。入力画像I（x，y，c）は、H×W×Cの寸法を有してもよく、Hは、垂直方向におけるピクセルの数であって、Wは、水平方向におけるピクセルの数であって、Cは、画像のチャネルの数である（例えば、RGB画像に関しては、3に等しく、グレースケール画像に関しては、1に等しい）。マルチタスクニューラルネットワーク256は、入力画像I（x，y，c）を処理してもよく、入力画像I（x，y，c）に基づいて、ネットワーク出力データ264を生成してもよい。

【0087】

ARデバイス200が、ランタイムモードで動作しているとき、ネットワーク出力デー

10

20

30

40

50

タ 2 6 4 は、較正角度 κ と併用され、視線ベクトル 2 3 8 を算出し得る。いくつかの実施形態では、後処理ブロック 2 6 6 が、1 つまたはそれを上回る演算を実施し、視線ベクトル 2 3 8 を算出してよい。他の実施形態では、または同一実施形態では、較正角度 κ は、入力画像 $I(x, y, c)$ とともに、マルチタスクニューラルネットワーク 2 5 6 への入力として提供されてもよく、視線ベクトル 2 3 8 は、直接、ネットワーク出力データ 2 6 4 内に含まれてもよい、またはネットワーク出力データ 2 6 4 に基づいて算出されてもよい。

【0088】

AR デバイス 2 0 0 が、訓練モードで動作しているとき、ネットワーク出力データ 2 6 4 は、GT データ 2 6 8 と比較されてもよい。誤差データ 2 7 0 が、比較に基づいて算出されてもよく、いくつかの実施形態では、誤差データ 2 7 0 の大きさが、ネットワーク出力データ 2 6 4 と GT データ 2 6 8 との間の差異に比例し得るように、ネットワーク出力データ 2 6 4 と GT データ 2 6 8 との間の差異を表してもよい。マルチタスクニューラルネットワーク 2 5 6 は、誤差データ 2 7 0 に基づいて、修正されてもよい（例えば、修正器 2 7 2 を使用して）。いくつかの実施形態では、マルチタスクニューラルネットワーク 2 5 6 への修正の大きさは、ネットワーク出力データ 2 6 4 と GT データ 2 6 8 との間のより大きい差異が、マルチタスクニューラルネットワーク 2 5 6 へのより大きい修正に対応し得るように、誤差データ 2 7 0 の大きさに比例し得る。

【0089】

いくつかの実施形態では、訓練モードと関連付けられるように本明細書に説明される動作のいくつかまたは全ては、AR デバイス 2 0 0 から独立して実施されてもよい。例えば、そのような実施形態では、マルチタスクニューラルネットワーク 2 5 6 は、少なくとも部分的に、AR デバイス 2 0 0 の製造および／または配布に先立って訓練され、続いて、AR デバイス 2 0 0 の製造および／または配布時に、AR デバイス 2 0 0 上にロードされてもよい。これらの実施形態のうちの少なくともいくつかでは、マルチタスクニューラルネットワーク 2 5 6 は、少なくとも部分的に、対象の比較的に大母集団からのデータを用いて、かつ AR デバイス 2 0 0 と異なる 1 つまたはそれを上回るコンピューティングデバイスを用いて、訓練されてもよい。いくつかのそのような実施形態では、AR デバイス 2 0 0 は、事前ロードされたマルチタスクニューラルネットワーク 2 5 6 を AR デバイス 2 0 0 の具体的ユーザからのデータを用いてさらに訓練するように、訓練モードと関連付けられるように本明細書に説明される動作のうちの 1 つまたはそれを上回るものを実施してもよい。これは、マルチタスクニューラルネットワーク 2 5 6 の 1 つまたはそれを上回る部分が、AR デバイス 2 0 0 のユーザ毎に、個人化された状態になることを可能にし得る。いくつかの実施形態では、AR デバイス 2 0 0 は、AR デバイス 2 0 0 のユーザ毎に、マルチタスクニューラルネットワーク 2 5 6 の個人化されたバージョンを記憶してもよい。したがって、これらの実施形態では、AR デバイス 2 0 0 は、複数の異なるユーザのためにマルチタスクニューラルネットワーク 2 5 6 の複数の異なるバージョンを記憶してもよく、ランタイム時に AR デバイス 2 0 0 の現在のユーザと関連付けられる、マルチタスクニューラルネットワーク 2 5 6 のバージョンを使用してもよい。

【0090】

図 6 は、種々の層 2 5 7 から成る、マルチタスクニューラルネットワーク 2 5 6 の概略図を図示する。いくつかの実施形態では、マルチタスクニューラルネットワーク 2 5 6 は、特徴エンコーディング層 2 5 8（代替として、エンコーダ層と称される）から成る、特徴エンコーディングベースネットワークと、タスク特有層 2 9 4（代替として、デコーダ層と称される）から成る、6 つのタスク分岐とを備える。6 つのタスク分岐は、（1）2 D 瞳孔中心データ 2 7 4 を生成する、瞳孔中心推定および閃光位置特定、（2）眼セグメント化データ 2 7 6 を生成する、眼部分意味論セグメント化、（3）閃光検出データ 2 7 8 を生成する、瞳孔および閃光存在分類、（4）角膜中心データ 2 8 0 を生成する、2 D 角膜推定、（5）瞬目予測 2 9 6 を生成する、瞬目検出、および（6）眼表現分類 2 9 8 を生成する、感情表現分類に対応する。

【0091】

ネットワーク出力データ264は、図6に示されるデータのタイプのうちの1つまたはそれを上回るものを含んでもよい。ARデバイス200が訓練モードまたはランタイムモードで動作しているかどうかに基づいて、データのタイプのうちの1つまたはそれを上回るものが、後続処理において利用可能にされなくてもよい。代替として、または加えて、データのタイプのうちの1つまたはそれを上回るものは、プロセッサ使用量、電力、および／またはメモリを節約するために、マルチタスクニューラルネットワーク256によって生成されなくてもよい。代替として、または加えて、データのタイプのうちの1つまたはそれを上回るものは、ユーザ入力に基づいて、生成されなくてもよい。例えば、ARデバイス200上で動作する、あるアプリケーションは、眼表現分類298等のあるタイプのデータのみが生成されることを要求し得る。

10

【0092】

いくつかの実施形態では、特徴エンコーディング層258は、タスク分岐のそれぞれを横断して共有される、エンコーダ特徴282を生産することができる。いくつかの実装では、画像特徴抽出ネットワークおよび特徴ピラミッド(FPN)が、情報を異なるスケールから捕捉するために使用される。いくつかの実装では、エンコーダの最上位層からの特徴(例えば、サイズ20×15×256を有する)は、タスク分岐への入力として使用されてもよい。

【0093】

いくつかの実施形態では、マルチタスクニューラルネットワーク256は、(1)眼部分セグメント化と、(2)瞳孔および閃光位置特定と、(3)瞳孔および閃光存在分類とを含む、マルチタスク学習モデルにおける3つの主要な外観ベースのタスクを含む。いくつかの実施形態では、眼部分セグメント化は、入力画像I(x, y, c)内の全てのピクセルに、以下、すなわち、背景、強膜、虹彩、および瞳孔からのクラスラベルを割り当てるタスクとして定義される。本タスクに関して、エンコーダネットワーク(例えば、特徴エンコーディング層258)からの最終層特徴マップに対応する、エンコーダ特徴282が、取得され、逆畳み込み層を使用して、入力画像I(x, y, c)と同一分解能にアップサンプリングされてもよい。結果として生じる4つのチャンネル出力は、独立して、ピクセル毎に、ソフトマックス層を使用して、クラス確率に変換されてもよい。損失は、予測される確率分布と手動で注釈が付けられたグラウンドトゥールズから取得されるワンホットラベル(ワンホットラベルは、1つの値を除き、全てゼロを有する、ベクトル、例えば、[0, 0, 1, 0, 0]である)との間のクロスエントロピー損失であり得る。

20

30

【0094】

いくつかの実施形態では、以下の損失が、GTクラスcおよびk番目のクラスに関して予測される確率 $p_k(x, y)$ を伴う、ピクセルx, yに関して最小限にされる。

【化1】

$$\mathcal{L}(x, y) = - \sum_{k=1}^4 I_{x,y}[k == c] \log p_k(x, y),$$

(1)

40

式中、 $I_{x,y}[\cdot]$ は、指示関数である。全体的損失は、画像内の全てのピクセルにわたる損失の和であり得る。セグメント化タスクは、眼画像の豊かな意味論情報を捕捉するにつれて特徴エンコーダ層258を訓練するためのブートストラップ相としての役割を果たす。それ自体で、眼部分セグメント化は、閃光の検索を位置特定し(虹彩境界を使用して)、瞳孔中心を推定する(瞳孔境界を使用して)観点から、任意の古典的パイプラインの初期相に役立ち得る。いくつかの実装では、眼部分セグメント化は、デジタルアバタの眼をレンダリングするために有用であり得る。

【0095】

瞳孔および閃光位置特定分岐は、合計5つの特徴点に関する4つの閃光および瞳孔中心

50

のピクセル場所を提供する。眼部分セグメント化分岐に類似し得る、これらの2つのタスクのためのネットワークデコーダ層は、5つの特徴点に対応する出力において、5つの稠密マップのセットを予測し得る。各稠密マップは、全てのピクセルを横断して合計して1になるように正規化されてもよい。クロスエントロピー損失が、次いで、訓練の間、各マップの全てのピクセルを横断して計算されてもよい。いったん訓練されると、瞳孔中心または特定の閃光の場所は、出力における最大確率に対応する、ピクセルとなる。いくつかの実施形態では、以下の損失が、特徴点（4つの閃光および1つの瞳孔中心）毎に最小限にされる。

【化2】

$$\mathcal{L}(\text{keypoint}) = - \sum_{x,y} I[x,y] \log p_{x,y},$$

(2)

式中、 $I[\cdot]$ は、GT特徴点場所を除き、全ての場所においてゼロである、指数関数であって、 $p_{x,y}$ は、特徴点場所の予測される確率であって、総和は、画像内の全てのピクセルにわたって行われる。

【0096】

現実的設定では、閃光および／または瞳孔中心は、眼瞼の閉鎖によってオクルードされ得、擾乱反射が、閃光として現れ得、および／またはいくつかの視線角度に関して、閃光は、反射性角膜表面上に現れない場合がある。したがって、閃光および瞳孔中心の存在または不在をロバストに分類するように学習することが重要であり得る。これらの予測は、事実上、閃光が、角膜中心推定のために、同様に、3D瞳孔中心推定のために使用されるべきであるかどうかをゲート制御することができる。

【0097】

本タスクに関して、エンコーダネットワーク（例えば、特徴エンコーディング層258）からの最上位層特徴マップに対応するエンコーダ特徴282が、取得されてもよく、1つの畳み込み層が、特徴チャンネルの数を低減させるために使用されてもよく、特徴チャンネルの低減された数は、1次元アレイに再変換されてもよく、（例えば、サイズ1500×10の）1つの訓練可能全結合層が、出力（例えば、5×2サイズの出力）を生産するために追加されてもよい。各対は、4つの閃光および／または瞳孔中心のうちの1つに関する存在または不在確率を表し得る。バイナリクロスエントロピー損失が、ヒト標識グラウンドトゥールズから学習するために使用されてもよい。

【0098】

角膜中心推定に関して、角膜の中心は、眼の2D画像内で観察され得ない、3D内の幾何学的数量である。故に、瞳孔（瞳孔楕円形の中心）または閃光標識化と異なり、直接、3D角膜中心の画像上に投影された場所を手動で標識することは不可能であり得る。したがって、2ステップ方法が、マルチタスクニューラルネットワーク256のための角膜2D中心予測分岐を訓練するために採用されてもよい。最初に、周知の幾何学的制約および関連する既知の／推定される量（LED、閃光）が、角膜2D教師を生成するために使用されてもよい。次いで、2D角膜分岐が、フレーム毎に取得される、本モデルベースの教師を使用して訓練されてもよい。

【0099】

マルチタスクニューラルネットワーク256を使用して、角膜を予測することは、評価の間、幾何学的制約の使用に優る2つの主要な利点を有する。第1に、そのような予測は、深層ネットワークは、訓練の間、雑音を平均化する傾向を有し、標準的ネットワーク外最適化が、場合によって、収束をもたらし得ないため、よりロバストである。第2に、そのような予測は、角膜タスク分岐がわずかのみの全結合層から成るため、わずかつ一定の時間フィードフォワード算出のみを被り得ることである。

【0100】

顔の表情分類タスクは、入力眼画像からユーザの感情表現を分類することを伴う。タスクは、特に、大部分の感情的顔の表情分類ベンチマーク内で使用されるように、眉毛および／または顔全体ではなく、ユーザの眼領域のみが入力として利用可能であるため、困難である。いくつかの実施形態では、以下、すなわち、喜び、怒り、嫌悪、恐れ、および驚きの個々の感情的顔の表情が、検討される。これらの表現は、4 離散状態、すなわち、肯定的側面（幸福）、弁別的側面（怒りおよび嫌悪）、感受性的側面（恐れおよび驚き）、および中性的側面にグループ化されることができる。他のタスク分岐と同様に、特徴エンコーディング層 258 は、固定され、顔の表情タスク分岐（いくつかの FC 層から成る）のみが、表現分類のために訓練された。いくつかの実施形態では、本タスク分岐は、個人化されたモデルを生産するように、対象毎に訓練され、これは、対象の大母集団のための一般的モデルより良好な正確度を生産する。

10

【0101】

いくつかの実施形態では、ネットワーク出力データ 264 は、2D 瞳孔中心データ 274 を含んでもよい。いくつかの実施形態では、2D 瞳孔中心データ 274 は、2D 値として表される 2D 瞳孔中心を含んでもよい。例えば、2D 瞳孔中心は、瞳孔中心（例えば、瞳孔中心 308）の算出された場所に対応する、入力画像 $I(x, y, c)$ のフレーム内の X および Y 値を含んでもよい。代替として、または加えて、2D 瞳孔中心データ 274 は、0 または 1 のバイナリ値（1 の値は、瞳孔中心の算出された場所に対応する）を備える、 $H \times W$ の寸法を有する、行列を含んでもよい。

【0102】

20

いくつかの実施形態では、ネットワーク出力データ 264 は、眼セグメント化データ 276 を含んでもよい。眼セグメント化データ 276 は、複数の領域への眼のセグメント化を含んでもよい。1 つの特定の実装では、領域は、背景領域、強膜領域、瞳孔領域、および虹彩領域を含んでもよい。別の特定の実装では、領域は、瞳孔領域および非瞳孔領域を含んでもよい。別の特定の実装では、領域は、瞳孔領域、眼領域（瞳孔領域の一部ではない眼の部分を含む）、および背景領域を含んでもよい。

【0103】

いくつかの実施形態では、眼セグメント化データ 276 は、0、1、2、および 3（例えば、それぞれ、背景領域、強膜領域、瞳孔領域、および虹彩領域に対応する）等の値の有限セットを備える、 $H \times W$ の寸法を有する、行列を含んでもよい。いくつかの実施形態では、眼セグメント化データ 276 は、背景、強膜、瞳孔、および虹彩を含む、クラスのセットへの、入力画像 $I(x, y, c)$ の全ピクセルの割当を含み、これは、いくつかの実施形態では、（デコーダ）マルチタスクニューラルネットワーク 256 の最後の層を求め、逆畳み込みを使用して、それを入力画像 $I(x, y, c)$ と同一分解能にアップサンプリングすることによって取得され得、これは、ひいては、各特徴チャンネルがあるクラスに属するピクセルの確率を表す、特徴チャンネルを横断して、ソフトマックスクロスエントロピー損失の中にフィードされる。

30

【0104】

いくつかの実施形態では、ネットワーク出力データ 264 は、閃光検出データ 278 を含んでもよい。いくつかの実施形態では、閃光検出データ 278 は、2D または 3D 値として表される 1 つまたはそれを上回る閃光場所を含む。例えば、単一閃光場所のみが、検出される場合、閃光検出データ 278 は、単一 2D 値を含んでもよい、または 4 つの閃光場所が、検出される場合、閃光検出データ 278 は、4 つの 2D 値を含んでもよい。いくつかの実施形態では、閃光検出データ 278 は、検出された閃光の算出された場所に対応する入力画像 $I(x, y, c)$ のフレーム内の X および Y 値を含んでもよい。代替として、または加えて、閃光検出データ 278 は、0 または 1 のバイナリ値（1 の値は、検出された閃光の場所に対応する）を備える、 $H \times W$ の寸法を有する、行列を含んでもよい。

40

【0105】

いくつかの実施形態では、ネットワーク出力データ 264 は、角膜中心データ 280 を含んでもよい。いくつかの実施形態では、角膜中心データ 280 は、2D 値として表され

50

る2D角膜中心または3D値として表される3D角膜中心を含んでもよい。例えば、2D角膜中心は、角膜の中心（例えば、角膜中心306）の算出された場所に対応する入力画像 $I(x, y, c)$ のフレーム内の X および Y 値を含んでもよい。代替として、または加えて、角膜中心データ280は、0または1のバイナリ値（1の値は、角膜の中心の算出された場所に対応する）を備える、 $H \times W$ の寸法を有する、行列を含んでもよい。

【0106】

いくつかの実施形態では、ネットワーク出力データ264は、瞬目予測296を含んでもよい。いくつかの実施形態では、瞬目予測296は、0または1のバイナリ値を備える（例えば、それぞれ、開眼および瞬目の予測に対応する）。いくつかの実施形態では、瞬目予測296は、瞬目が生じたかどうかと関連付けられる、確率を備える。瞬目を検出することは、多焦点ディスプレイおよび／またはデジタルアバタを駆動するために有用である、外観ベースのタスクである。瞬目は、時間的情報が瞬目をサッカード（眼の高速側方移動）等のイベントから区別するために使用され得るように、画像のシーケンスを横断して捕捉されることができる。

【0107】

一般に、特に、30フレーム／秒の標準的フレームレートにおいて、眼が完全に閉鎖される、瞬目イベントを正確に位置特定することは、困難であり得る。他の場合では、検出と適用との間に待ち時間を低減させるために、瞬目の始まりを検出することが重要であり得る。いくつかの実施形態では、瞬目の単純定義は、上側眼瞼が瞳孔領域全体の50%にわたって被覆するときの眼の状態であり得る。これは、非専門家ヒト標識者にとって、有用な作業定義であり得る。前述の瞬目の定義を前提として、眼セグメント化等のタスクに関して訓練された特徴エンコーディング層258によって生成されたエンコーダ特徴282は、瞬目検出タスクに良好に移行する。いくつかの実施形態では、事前に訓練された特徴エンコーディングネットワーク（例えば、特徴エンコーディング層258）の最上位層（共有表現）が、瞬目検出分岐を訓練するために使用される。

【0108】

図7は、別個の時間ステップからの特徴を使用して瞬目予測296を生成するためのシステムおよび技法を図示する。図示される実施形態では、 $T-2$ 、 $T-1$ 、および T における3つの持続的時間ステップからのエンコーダ特徴が、入力 a_{T-2} 、 a_{T-1} 、および a_T として、（時間 T における）現在のフレームを瞬目または開眼として分類し、それを示す、出力 y_T を生産する、3層全結合ネットワークにフィードされる。より長い時間的ウィンドウの長さが、採用され得るが、それらは、予測正確度における収穫逡減をもたらす。再帰ニューラルネットワーク（RNN）および長短期メモリ（LSTM）は、同様の訓練および試験性能を有するが、しかしながら、ネットワーク700は、より低い算出要件を提供する。

【0109】

図8は、訓練モードで動作する、ARデバイス200の概略図を図示する。ARデバイス200が、訓練モードで動作しているとき、ネットワーク出力データ264は、眼セグメント化データ276、閃光検出データ278、および角膜中心データ280を含む。これらのネットワーク出力を生成するために使用される特定の入力画像 $I(x, y, c)$ はまた、マルチタスクニューラルネットワーク256によるネットワーク出力データ264の生成に先立って、それに続いて、またはそれと並行して、GTデータ268を準備し得る、1人またはそれを上回る個人によって手動で検査されてもよい。例えば、個人は、パーソナルコンピュータまたはスマートフォン等の電子デバイス上の入力画像 $I(x, y, c)$ の表示されるバージョンを検査してもよい。電子デバイス上のプログラムまたはアプリケーションは、個人に、入力画像 $I(x, y, c)$ に関連する質問のセットを尋ねてもよく、個人は、マウス、キーボード、タッチ画面等の入力デバイスを使用して、その応答を入力してもよい。

【0110】

入力画像 $I(x, y, c)$ を観察および検査する際、個人は、入力デバイスを使用して

、瞳孔の輪郭を識別することによって、2D瞳孔中心GTデータ283を準備してもよい。これは、個人が、楕円形境界を瞳孔にわたって設置し、設置された楕円形境界に基づいて、瞳孔中心を自動的に計算させるステップを含んでもよい。2D瞳孔中心GTデータ283は、2D瞳孔中心データ274と同一フォーマットおよび寸法（例えば、XおよびY値）を有するように準備されてもよい。加えて、入力画像I（x，y，c）を観察および検査する際、個人は、画像の第1の領域が背景領域として、第2の領域が強膜領域として、第3の領域が瞳孔領域として、および第4の領域が虹彩領域として割り当てられるべきであることを決定することによって、眼セグメント化GTデータ284を準備してもよい。眼セグメント化GTデータ284は、眼セグメント化データ276と同一フォーマットおよび寸法（例えば、異なる領域に対応する0、1、2、および3等の値の有限セットを備える、H×Wの寸法を有する、行列）を有するように準備されてもよい。

10

【0111】

加えて、入力画像I（x，y，c）を観察および検査する際、個人は、入力画像I（x，y，c）内に存在する閃光場所の数およびそれぞれの場所を決定することによって、閃光検出GTデータ286を準備してもよい。閃光検出GTデータ286は、閃光検出データ278と同一フォーマットおよび寸法（例えば、2D値のセット）を有するように準備されてもよい、またはある数の閃光場所が、検出された場合（例えば、4つ）、閃光検出GTデータ286は、その2D値の数を含んでもよい。いくつかの実施形態では、閃光検出GTデータ286は、検出された閃光の算出された場所に対応する入力画像I（x，y，c）のフレーム内にXおよびY値を含んでもよい。代替として、または加えて、閃光検出GTデータ286は、0または1のバイナリ値（1の値は、検出された閃光の場所に対応する）を備える、H×Wの寸法を有する、行列を含んでもよい。

20

【0112】

1つの特定の実装では、GTデータ268は、個人または個人のグループを、例えば、3メートルにおける近くの深度と、例えば、6メートルにおけるより遠い平面との、2つの明確に異なる深度における点の3×3グリッドに対面させることによって取得されてもよい。所与のキューに応じて、個人は、その視線をこれらの18個の3D点のうちの1つに合焦させるように求められ、これは、視線ベクトル238に関するGTデータ268がフレーム毎に収集されることを可能にする（後に全体的正確度を決定するために）。個人の眼の捕捉された画像（個人によって装着されるARデバイスのカメラを使用して）は、分析され、GTデータ268が、眼セグメント化および閃光場所情報を含むことを可能にし得る。30または60Hz記録におけるフレーム毎にセグメント化、閃光、および瞳孔中心に注釈を付ける際、収穫逓減が存在するため、ある数（例えば、200）の左または右眼画像フレームが、個人毎に、均一にサンプリングされ、セグメント化、閃光存在または不在、閃光2Dおよび瞳孔2D位置に手動で注釈を付けてもよい。1つの特定の実験工程では、87，000の注釈が付けられた画像が、マルチタスクニューラルネットワーク256を訓練し、その性能を検証するためにデータセット内で使用された。

30

【0113】

いくつかの実施形態では、誤差データ270は、2D瞳孔中心データ274と2D瞳孔中心GTデータとの間の差異に基づいて算出された、第1の誤差データ270A、眼セグメント化データ276と眼セグメント化GTデータ284との間の差異に基づいて算出された、第2の誤差データ270B、閃光検出データ278と閃光検出GTデータ286との間の差異に基づく、第3の誤差データ270C、および幾何学的制約エンジン288によって生成された、第4の誤差データ270Cを含んでもよい。幾何学的制約エンジン288への入力、角膜中心データ280、閃光検出データ278、エミッタ場所データ290、およびカメラ固有のパラメータ291のうちの1つまたはそれを上回るものを含む。エミッタ場所データ290は、エミッタ262の固定場所および／またはエミッタ262の放出方向を含んでもよい。エミッタ場所データ290は、ARデバイス200の製造の際および／または較正位相の間に決定されてもよい。カメラ固有のパラメータ291は、他の可能性の中でもとりわけ、軸外カメラ260の光学中心および／または焦点長を含

40

50

んでもよい。カメラ固有のパラメータ 291 は、軸外カメラ 260 の製造の際および／または較正位相の間に決定されてもよい。

【0114】

幾何学的制約エンジン 288 は、異なる生成されたデータ（閃光検出データ 278 および角膜中心データ 280）と較正されたデータ（エミッタ場所データ 290）との間の一貫性を評価するために、種々の演算を実施してもよく、幾何学的制約エンジン 288 の出力である、第 4 の誤差データ 270 D は、尤度または一貫性パラメータに反比例し得る。いくつかのインスタンスでは、角膜球面 304 は、閃光検出データ 278 およびエミッタ場所データ 290 を使用して再構成され、第 4 の誤差データ 270 D は、再構成された球体の中心と角膜中心データ 280 によって示されるような角膜中心との間の計算された距離に設定される。

10

【0115】

いくつかの実施形態では、マルチタスクニューラルネットワーク 256 の訓練は、異なる訓練反復の間、マルチタスクニューラルネットワーク 256 のある出力のみを使用して、順次訓練することによって改良される。第 1 の訓練ステップでは、眼セグメント化データ 276 のみが、マルチタスクニューラルネットワーク 256 を訓練するために使用される。これは、第 2 の誤差データ 270 B のみを使用して、マルチタスクニューラルネットワーク 256 を修正することによって遂行されてもよい。いったんマルチタスクニューラルネットワーク 256 が、眼セグメント化のために十分に訓練される（すなわち、十分に正確になる）と、第 2 の訓練ステップが、加えて、閃光検出データ 278 を使用して、マルチタスクニューラルネットワーク 256 を訓練することによって実施される。これは、第 3 の誤差データ 270 C のみを使用して、マルチタスクニューラルネットワーク 256 を修正することによって遂行されてもよい。いったんマルチタスクニューラルネットワーク 256 が、眼セグメント化および閃光検出のために十分に訓練されると、第 3 の訓練ステップが、加えて、角膜中心データ 280 を使用して、マルチタスクニューラルネットワーク 256 を訓練することによって実施される。これは、誤差データ 270 の全てを使用して、マルチタスクニューラルネットワーク 256 を修正することによって遂行されてもよい。いくつかのインスタンスでは、同一訓練画像および GT データが、異なる訓練ステップの間に使用されてもよい。いくつかの実施形態では、AR デバイス 200 は、正確度閾値が満たされる、または最大反復閾値が満たされる（例えば、使用される訓練画像の数が反復閾値を満たす）まで、訓練モードのままである。

20

30

【0116】

図 9 A および 9 B は、マルチタスクニューラルネットワーク 256 を訓練するための順次訓練ステップ 902 の概略図を図示する。図 9 A を参照すると、第 1 の訓練ステップ 902-1 が、図示される。第 1 の訓練ステップ 902-1 の間、特徴エンコーディング層 258 およびタスク特有層 294-2（眼セグメント化データ 276 を生成するためのデコーダ層に対応する）が、残りのタスク特有層 294 から独立して訓練される。例えば、訓練反復の間、入力画像 $I(x, y, c)$ が、マルチタスクニューラルネットワーク 256 に提供されてもよく、また、眼セグメント化 GT データ 284 を準備し得る、個人に提示されてもよい。眼セグメント化データ 276 と眼セグメント化 GT データ 284 との間の差異に基づいて算出された第 2 の誤差データ 270 B を使用して、修正器 272 は、第 2 の誤差データ 270 B が、眼セグメント化データ 276 と眼セグメント化 GT データ 284 との間の差異に基づいて、第 2 の誤差データ 270 B の後続算出の間に減少されるであろうように、特徴エンコーディング層 258 およびタスク特有層 294-2 と関連付けられる加重を修正してもよい（例えば、逆伝搬を使用して）。第 1 の訓練ステップ 902-1 の間、修正器 272 は、タスク特有層 294-1、294-3、294-4、294-5、または 294-6 と関連付けられる、加重を修正しない。

40

【0117】

図 9 B を参照すると、第 2 の訓練ステップ 902-2 が、図示される。いくつかの実施形態では、第 2 の訓練ステップ 902-2 は、第 1 の訓練ステップ 901-1 後に実施さ

50

れる。第2の訓練ステップ902-2の間、タスク特有層294-1、294-3、294-4、294-5、および294-6のうちの1つまたはそれを上回るものは、特徴エンコーディング層258およびタスク特有層294-2から独立して訓練される。例えば、第1の訓練反復の間、入力画像I(x, y, c)が、マルチタスクニューラルネットワーク256に提供されてもよく、また、関連GTデータ268を準備し得る、個人に提示されてもよい。ネットワーク出力データ264とGTデータ268との間の差異に基づいて算出された誤差データ270を使用して、修正器272は、誤差データ270が誤差データ270の後続算出の間に減少されるであろうように、タスク特有層294-1、294-3、294-4、294-5、および／または294-6と関連付けられる、加重を修正してもよい（例えば、逆伝搬を使用して）。第2の訓練ステップ902-2の間、修正器272は、特徴エンコーディング層258またはタスク特有層294-2と関連付けられる、加重を修正しないが、いくつかの実施形態ではタスク特有層294-2は、破線によって示されるように、第2の訓練ステップ902-2の間、微調整されてもよい。

【0118】

いくつかの実施形態では、タスク特有層294-1、タスク特有層294-2、およびタスク特有層294-3はそれぞれ、1つまたはそれを上回る畳み込み層と、1つまたはそれを上回る逆畳み込み層とを含んでもよい。いくつかの実施形態では、タスク特有層294-4およびタスク特有層294-6は、相互にアーキテクチャ的に類似する、または同じであり得るが、2つの別個の分岐として訓練されてもよい。これらの実施形態のうちの少なくともいくつかでは、タスク特有層294-4およびタスク特有層294-6はそれぞれ、1つまたはそれを上回る畳み込み層を含んでもよい。さらに、いくつかの実施形態では、タスク特有層294-5は、図7を参照して上記に説明されるようなニューラルネットワーク700のものとアーキテクチャ的に類似する、または同じであり得る。

【0119】

図6に図示されるように、マルチタスクニューラルネットワーク256のいくつかの出力が、マルチタスクニューラルネットワーク256の他の出力より少ない実施される動作を伴って、取得されてもよい。例えば、角膜中心データ280は、他の出力より少ない算出を伴って、取得されてもよく、眼セグメント化データ276は、他の出力より多い算出を伴って取得されてもよい。故に、最初に、眼セグメント化データ276を使用して、マルチタスクニューラルネットワーク256を訓練する1つの利点は、眼セグメント化データ276の算出のみのために使用される、いくつかの層が、他の出力からのフィードバックによって影響されずに、微調整され得ることである。

【0120】

図10は、ランタイムモードで動作する、ARデバイス200の概略図を図示する。ARデバイス200が、ランタイムモードで動作しているとき、ネットワーク出力データ264は、眼セグメント化データ276、閃光検出データ278、および角膜中心データ280を含んでもよい。これらの出力は、較正角度 κ と併用され、後処理ブロック266を使用して、視線ベクトル238を算出し得る。いくつかの実施形態では、後処理ブロック266は、第1の後処理ブロック266A、第2の後処理ブロック266B、および第3の後処理ブロック266Cに分離されてもよい。第1の後処理ブロック266Aは、2D瞳孔中心データ274および眼セグメント化データ276を入力として受信し、3D瞳孔中心292を算出する。第2の後処理ブロック266Bは、3D瞳孔中心292および角膜中心データ280を入力として受信し、光学軸310を算出する。第3の後処理ブロック266Cは、光学軸310および較正角度 κ を入力として受信し、視線ベクトル238を算出する。

【0121】

マルチタスクニューラルネットワーク256の正確度は、例えば、米国仮出願第62/935,584号に説明されるように実証されている。眼セグメント化の正確度の一実施例が、眼セグメント化混同行列パーセンテージ値を提供する、下記の表に示され、全ての4つのクラスに関する平均された正確度は、97.29%を上回る。

【表 1】

GT/予測	瞳孔	虹彩	強膜	BG
瞳孔	96.25	3.75	0.00	0.00
虹彩	0.04	99.03	0.93	0.00
強膜	0.00	3.27	96.71	0.02
BG	0.01	0.72	2.09	97.18

10

これらの結果は、量的および定質的評価の両方の観点から非常に正確である。これは、セグメント化境界が、特に、部分的にオクルードされる瞳孔の場合に関して、慎重に調整される楕円形適合プロシージャによって、精密な瞳孔 2 D 中心場所訓練データを生成するために使用され得るため、重要であり得る。セグメント化予測はまた、古典的幾何学的パイプラインによって使用されることができ、これは、視線推定比較のためのベースラインとして使用され得る。

【0 1 2 2】

別の実施例として、瞳孔および閃光検出の正確度は、マルチタスクニューラルネットワーク 2 5 6 (「NN 2 5 6」) および古典的パイプラインのそれぞれを使用してピクセル場所を予測するための量的結果を示す、下記の表に示される。

20

【表 2】

	ピクセル内の古典的位置特定	ピクセル内の NN256 位置特定	古典的存在/不在	NN256 存在/不在
瞳孔	0.64	0.46	92.81%	99.61%
閃光 1	1.21	0.47	90.16%	96.94%
閃光 2	1.08	0.39	90.84%	96.32%
閃光 3	0.84	0.23	92.14%	96.85%
閃光 4	0.78	0.37	91.56%	96.34%
平均値	0.86	0.38	91.72%	98.06%

30

画像が、理想的設定からのものであるとき、マルチタスクニューラルネットワーク 2 5 6 および古典的予測は全て、ゼロに近い誤差を伴って、精密である。しかしながら、画像が、深刻な反射を有する、またはユーザ視線が、中心標的から離れるとき、マルチタスクニューラルネットワーク 2 5 6 は、最初に、閃光の存在または不在を非常に正確に検出し、閃光のロバストな標識化を提供することが可能である一方、古典的アプローチは、劣った不在インジケーションおよび閃光の誤標識化に悩まされ、我々のユークリッド誤差メトリック下ではるかに高い誤差をもたらす。

40

【0 1 2 3】

図 1 1 は、校正角度 κ および光学軸 3 1 0 に基づいて、視線ベクトル 2 3 8 を生成し得る、視線ベクトルニューラルネットワーク 1 1 0 2 の概略図を図示する。いくつかの実施形態では、視線ベクトルニューラルネットワーク 1 1 0 2 は、後処理ブロック 2 6 6 C に取って代わる、またはその中に組み込まれてもよい。1 つの実装では、視線ベクトルニューラルネットワーク 1 1 0 2 は、5 つの層と、約 3 0 , 0 0 0 個のパラメータまたは加重

50

とを含む。いくつかの実施形態では、視線ベクトルニューラルネットワーク 1102 は、較正フレーム上でのみ訓練される。

【0124】

訓練の間、視線ベクトル 238 は、視線ベクトル GT データ 1104 と比較されてもよい。誤差データ 1106 は、比較に基づいて算出されてもよく、いくつかの実施形態では、誤差データ 1106 の大きさが、視線ベクトル 238 と視線ベクトル GT データ 1104 との間の差異に比例し得るように、視線ベクトル 238 と視線ベクトル GT データ 1104 との間の差異を表し得る。視線ベクトルニューラルネットワーク 1102 は、誤差データ 1106 に基づいて、修正されてもよい（例えば、修正器 1108 を使用して）。いくつかの実施形態では、視線ベクトルニューラルネットワーク 1102 の修正の大きさは、視線ベクトル 238 と視線ベクトル GT データ 1104 との間のより大きい差異が視線ベクトルニューラルネットワーク 1102 のより大きい修正に対応し得るように、誤差データ 1106 の大きさに比例してもよい。

10

【0125】

いくつかの実施形態では、視線ベクトル GT データ 1104 は、ユーザが画面上に生成された標的を見ることによって取得されてもよい。例えば、ユーザは、以前に訓練されたマルチタスクニューラルネットワーク 256 を含み得る、AR デバイス 200 を装着し得る。訓練反復の間、ユーザは、AR デバイス 200 を装着しながら、ディスプレイ上に位置する標的を見るように命令され得る。ユーザの眼の入力画像 $I(x, y, c)$ が、捕捉され、光学軸 310 を生成するために使用され得る。光学軸 310 に基づいて（および随意に、較正角度 κ に基づいて）、視線ベクトル 238 が、視線ベクトルニューラルネットワーク 1102 によって生成されてもよい。視線ベクトル GT データ 1104 は、ウェアラブルデバイスとディスプレイ上に生成された標的との間の関係に基づいて決定されてもよい。例えば、AR デバイス 200 とディスプレイとの間の配向が、カメラおよび／または慣性測定ユニット等の 1 つまたはそれを上回るセンサに基づいて決定されてもよく、決定された配向は、ユーザの眼の実際の視線ベクトルを計算するために使用されてもよい。

20

【0126】

後続訓練反復の間、標的は、ディスプレイ上の新しい場所に移動され得、ユーザの眼の新しい入力画像 $I(x, y, c)$ が、捕捉され得、視線ベクトルニューラルネットワーク 1102 は、新しく計算された誤差データ 1106 を使用して修正され得る。種々の訓練反復の間、標的は、視線ベクトルニューラルネットワーク 1102 を訓練し、広範囲の視線角度にわたって、視線ベクトルをロバストに推定するように、画面を横断して種々の場所に移動されてもよい。いくつかの実施形態では、種々の照明条件および／またはユーザ感情が、訓練プロセスの間、種々の視線ベクトルと組み合わせて採用され、ロバストに訓練されたネットワークをもたらしてもよい。

30

【0127】

視線推定の正確度の実施例は、異なる標的平面にわたって集約された 9 つの標的毎に視線誤差を示す、下記の表によって実証されることができ、全体的視線推定メトリックは、真の視線ベクトルと推定される視線ベクトルとの間の角度誤差（例えば、 arcmin 単位）として定義される。

40

【表 3】

	NN256 +NN1102平 均	古典的平均	NN256 +NN1102標準 偏差	古典的標準偏 差
左上	194.18	261.05	105.63	304.42
中央上	169.64	148.28	103.53	143.25
右上	184.95	162.57	109.16	154.25
中心左	195.15	298.44	105.78	331.17
中心中央	183.57	147.15	106.18	143.11
中心右	193.35	161.74	108.62	151.77
左下	205.55	300.94	105.56	323.00
中央下	179.15	154.19	100.55	146.39
右下	181.35	166.15	103.47	161.19

マルチタスクニューラルネットワーク 256 および視線ベクトルニューラルネットワーク 1102 を使用した推定は、全ての方向において有意に良好かつ類似することが明白である。これは、主に、視線ベクトルニューラルネットワーク 1102 の使用とともに、ロバストな閃光および角膜 2D 推定に起因し得る。

【0128】

図 12 は、本発明のいくつかの実施形態による、訓練パイプライン 1200 を図示する。いくつかのインスタンスでは、完全訓練は、フレームワークが GT を異なるソースから受信するため、かつモデルベースの教師が訓練されたネットワーク自体からの推定を使用するため、いくつかのステップをとり得る。例えば、モデルは、最初に、眼セグメント化および閃光予測を訓練し、次いで、訓練されたモデルを使用して、全ての未標識データに関する閃光を予測する。次に、これらの予測される閃光および LED の既知の場所を使用して、角膜位置が、標準的眼モデルおよび幾何学形状に基づいて推測される。モデルを訓練し、角膜予測を予測するための教師は、事前に訓練されたモデルおよび標準的眼モデルおよび幾何学形状を使用して生じるため、技法は、モデルベースの教師と称され得る。

【0129】

ステップ 1202 では、エンコーダーデコーダネットワークが、最初に、最も豊かな意味論情報を提供し、正確に訓練するための最も複雑な教師ありタスクであるため、眼セグメント化ラベル（例えば、眼セグメント化 GT データ 284）を用いて訓練される。

【0130】

ステップ 1204 では、教師ありタスクが全て、訓練される。さらに、ステップ 1204 では、ヒト標識閃光データ（例えば、閃光検出 GT データ 286）、瞳孔 2D 中心データ（2D 瞳孔中心 GT データ 283）、および眼セグメント化データ（例えば、眼セグメント化 GT データ 284）が、これらの 3 つの教師ありタスクのそれぞれと一緒に訓練するために、ともに使用され得る。いくつかのインスタンスでは、眼セグメント化から訓練された加重を用いた初期化は、ランダム初期化からより安定した訓練をもたらし得る。

【0131】

ステップ 1206 では、閃光予測（例えば、閃光検出データ 278）が、全てのフレームに関して行われ、LED の既知の場所（例えば、エミッタ場所データ 290）とともに使用され、ステップ 1210 において角膜分岐を訓練する（例えば、第 4 の誤差データ 270D を使用して）ために、ステップ 1208 において、（幾何学的制約エンジン 288 内で生成される）角膜 2D GT を生成する。角膜分岐は、訓練セット母集団全体からの

データを用いて訓練され、さらに、対象あたり較正相において個人化（微調整）されることに留意されたい。

【0132】

3D瞳孔中心が、ステップ1212において予測された後、較正フレームから予測される角膜（個人化される）および瞳孔3D中心が、ステップ1214において、光学軸を推察するために使用される。視線標的GTを使用して、視線ベクトルニューラルネットワーク1102が、ステップ1216において訓練され、光学軸を視軸に変換する。ランタイムの間、予測される角膜および瞳孔2D中心が、マルチタスクニューラルネットワーク256から取得される。これらの数量は、3Dまで高め、光学軸を取得するために使用され、これは、次いで、視線マッピングネットワークにフィードされ、予測される視線方向を推測する。

10

【0133】

瞬目および顔の表情分類タスクは、主要な特徴エンコーディング分岐の中間特徴の上で訓練される。瞬目検出は、時間的タスクであって、これは、3つの連続眼画像を捕捉し、その中間特徴を抽出することを伴う。事前に算出された特徴のセットを用いることで、瞬目検出分岐は、マルチタスクニューラルネットワーク256の主要な特徴エンコーディング分岐が固定されたまま、別個に訓練される。類似プロシージャは、ランタイム時に追従される。顔の表情分類に関して、主要な特徴エンコーディング分岐は、固定され、表現分類層のみが、表現データを使用して訓練される。表現予測は、ランタイムの間、全ての他のタスクとともに生産される。

20

【0134】

図13は、特徴エンコーディング層のセット（例えば、特徴エンコーディング層258）と、それぞれ、特徴エンコーディング層のセットの出力（例えば、エンコーダ特徴282）に作用する、複数のタスク特有層のセット（例えば、タスク特有層294）とを有する、ニューラルネットワーク（例えば、マルチタスクニューラルネットワーク256）を訓練する方法1300を図示する。方法1300のステップは、示される順序で実施される必要はなく、方法1300の1つまたはそれを上回るステップは、方法1300の実施の間、省略されてもよい。いくつかの実施形態では、方法1300の1つまたはそれを上回るステップは、処理モジュール250またはARデバイス200のある他のコンポーネントによって実施されてもよい。

30

【0135】

ステップ1302では、第1の訓練ステップ（例えば、第1の訓練ステップ902-1）が、実施される。いくつかの実施形態では、第1の訓練ステップは、第1の持続時間の間、実施される。いくつかの実施形態では、ステップ1302は、ステップ1304、1306、および／または1308を含む。

【0136】

ステップ1304では、第1の眼の第1の画像（例えば、入力画像I（x，y，c））が、ニューラルネットワークに提供される。いくつかの実施形態では、第1の画像は、カメラ（例えば、軸外カメラ260）によって捕捉され、および／またはそこから受信される。いくつかの実施形態では、方法1300は、カメラを使用して、第1の眼の第1の画像を捕捉するステップを含む。いくつかの実施形態では、方法1300は、第1の眼の第1の画像をカメラから処理モジュール（例えば、処理モジュール250）に送信するステップを含む。

40

【0137】

ステップ1306では、眼セグメント化データ（例えば、眼セグメント化データ276）が、ニューラルネットワークを使用して、第1の画像に基づいて生成される。いくつかの実施形態では、眼セグメント化データは、第1の眼の複数の領域へのセグメント化を含む。

【0138】

ステップ1308では、特徴エンコーディング層のセットが、眼セグメント化データを

50

使用して、訓練される。いくつかの実施形態では、複数のタスク特有層のセットの単一のタスク特有層のセットもまた、第1の訓練ステップの間、眼セグメント化データを使用して訓練される。いくつかの実施形態では、誤差データ（例えば、誤差データ270B）が、眼セグメント化データと眼セグメント化GTデータ（例えば、眼セグメント化GTデータ284）との間の差異に基づいて算出される。いくつかの実施形態では、誤差データは、特徴エンコーディング層のセットを訓練するために使用される。

【0139】

ステップ1310では、第2の訓練ステップ（例えば、第2の訓練ステップ902-2）が、実施される。いくつかの実施形態では、第2の訓練ステップは、第2の持続時間の間、実施される。いくつかの実施形態では、第2の持続時間は、第1の持続時間後である。いくつかの実施形態では、ステップ1310は、ステップ1312、1314、および／または1316を含む。

【0140】

ステップ1312では、第2の眼の第2の画像（例えば、入力画像I（x，y，c））が、ニューラルネットワークに提供される。第2の眼は、第1の眼と同一である、または異なってもよい。いくつかの実施形態では、第2の画像は、カメラによって捕捉され、および／またはそこから受信される。いくつかの実施形態では、方法1300は、カメラを使用して、第2の眼の第2の画像を捕捉するステップを含む。いくつかの実施形態では、方法1300は、第2の眼の第2の画像をカメラから処理モジュールに送信するステップを含む。

【0141】

ステップ1314では、ネットワーク出力データ（例えば、ネットワーク出力データ264）が、特徴エンコーディング層のセットおよび複数のタスク特有層のセットのそれぞれを使用して、第2の画像に基づいて生成される。

【0142】

ステップ1316では、複数のタスク特有層のセットが、ネットワーク出力データを使用して訓練される。いくつかの実施形態では、特徴エンコーディング層のセットは、第2の訓練ステップの間、訓練されない。いくつかの実施形態では、誤差データ（例えば、誤差データ270）が、ネットワーク出力データとGTデータ（例えば、GTデータ268）との間の差異に基づいて算出される。いくつかの実施形態では、誤差データは、複数のタスク特有層のセットを訓練するために使用される。

【0143】

図14は、ユーザの眼表現を分類するためのニューラルネットワーク（例えば、マルチタスクニューラルネットワーク256）を訓練する方法1400を図示する。方法1400のステップは、示される順序で実施される必要はなく、方法1400の1つまたはそれを上回るステップは、方法1400の実施の間、省略されてもよい。いくつかの実施形態では、方法1400の1つまたはそれを上回るステップは、処理モジュール250またはARデバイス200のある他のコンポーネントによって実施されてもよい。

【0144】

ステップ1402では、眼の画像（例えば、入力画像I（x，y，c））が、捕捉される。いくつかの実施形態では、第1の画像は、カメラ（例えば、軸外カメラ260）によって捕捉され、および／またはそこから受信される。いくつかの実施形態では、方法1400は、カメラを使用して、眼の画像を捕捉するステップを含む。いくつかの実施形態では、方法1400は、眼の画像をカメラから処理モジュール（例えば、処理モジュール250）に送信するステップを含む。

【0145】

ステップ1404では、眼の画像が、ニューラルネットワークに提供される。いくつかの実施形態では、眼の画像をニューラルネットワークに提供するステップは、眼の画像を表すデータを、入力として、ニューラルネットワークを実装する、動作のセットに提供するステップを含んでもよい。

10

20

30

40

50

【0146】

ステップ1406では、眼に対応する、眼表現分類（例えば、眼表現分類298）が、ニューラルネットワークによって生成される。いくつかの実施形態では、眼表現分類は、複数の可能性として考えられる眼表現分類のうちの1つである。

【0147】

ステップ1408では、GT眼表現分類（例えば、GTデータ268）が、決定される。いくつかの実施形態では、GT眼表現分類を決定するステップは、GT眼表現分類を示す、ユーザ入力を受信するステップを含む。例えば、ユーザは、彼らが入力デバイスを通して「幸福」表現を呈したことを示し得る。いくつかの実施形態では、GT眼表現分類を決定するステップは、ユーザに通信される命令がGT眼表現分類を示すことを決定するステップを含む。例えば、ユーザは、ディスプレイデバイスを通して、「幸福」の顔の表情を呈するように命令されてもよい。

10

【0148】

ステップ1410では、誤差データ（例えば、誤差データ270）が、眼表現分類とGT眼表現分類との間の差異に基づいて算出される。

【0149】

ステップ1412では、ニューラルネットワークが、誤差データに基づいて修正される。いくつかの実施形態では、ニューラルネットワークを修正するステップは、ニューラルネットワークの加重のセットを修正するステップを含む。いくつかの実施形態では、加重のセットは、逆伝搬を使用して修正されてもよい。いくつかの実施形態では、ニューラルネットワークのタスク特有層のセット（例えば、タスク特有層294-6）が、誤差データに基づいて修正されてもよい。

20

【0150】

図15は、視線ベクトル（例えば、視線ベクトル238）を算出するためのニューラルネットワーク（例えば、視線ベクトルニューラルネットワーク1102）を訓練する方法1500を図示する。方法1500のステップは、示される順序で実施される必要はなく、方法1500の1つまたはそれを上回るステップは、方法1500の実施の間、省略されてもよい。いくつかの実施形態では、方法1500の1つまたはそれを上回るステップは、処理モジュール250またはARデバイス200のある他のコンポーネントによって実施されてもよい。

30

【0151】

ステップ1502では、眼の画像（例えば、入力画像I(x, y, c)）が、捕捉される。いくつかの実施形態では、第1の画像は、カメラ（例えば、軸外カメラ260）によって捕捉され、および／またはそれから受信される。いくつかの実施形態では、方法1500は、カメラを使用して、眼の画像を捕捉するステップを含む。いくつかの実施形態では、方法1500は、眼の画像をカメラから処理モジュール（例えば、処理モジュール250）に送信するステップを含む。

【0152】

ステップ1504では、眼の画像が、処理され、眼に対応する、光学軸を生産する。いくつかの実施形態では、眼の画像を処理するステップは、マルチタスクニューラルネットワーク（例えば、マルチタスクニューラルネットワーク256）を使用して、2D瞳孔中心データ（例えば、2D瞳孔中心データ274）、眼セグメント化データ（例えば、眼セグメント化データ276）、および／または角膜中心データ（例えば、角膜中心データ280）を生成するステップを含んでもよい。

40

【0153】

ステップ1506では、光学軸が、ニューラルネットワークに提供される。いくつかの実施形態では、光学軸をニューラルネットワークに提供するステップは、光学軸を表すデータを、入力として、ニューラルネットワークを実装する、動作のセットに提供するステップを含んでもよい。

【0154】

50

ステップ1508では、眼に対応する、視線ベクトルが、ニューラルネットワークによって生成される。いくつかの実施形態では、視線ベクトルは、少なくとも1つの角度を含む。

【0155】

ステップ1510では、視線ベクトルGTデータ（例えば、視線ベクトルGTデータ1104）が、決定される。いくつかの実施形態では、視線ベクトルGTデータは、標的が画面上に表示される、場所に基づいて決定される。いくつかの実施形態では、視線ベクトルGTデータを決定するステップは、視線ベクトルGTデータを示す、ユーザ入力を受信するステップを含む。例えば、ユーザは、画面上に表示される複数の標的の特定の標的を見て、ユーザが見ている標的として入力を提供してもよい。

10

【0156】

ステップ1512では、誤差データ（例えば、誤差データ1106）が、視線ベクトルと視線ベクトルGTデータとの間の差異に基づいて算出される。

【0157】

ステップ1514では、ニューラルネットワークが、誤差データに基づいて修正される。いくつかの実施形態では、ニューラルネットワークを修正するステップは、ニューラルネットワークの加重のセットを修正するステップを含む。いくつかの実施形態では、加重のセットは、逆伝搬を使用して修正されてもよい。

【0158】

図16は、ニューラルネットワークを使用して視線ベクトルを算出する方法1600を図示する。方法1600のステップは、示される順序で実施される必要はなく、方法1600の1つまたはそれを上回るステップは、方法1600の実施の間、省略されてもよい。いくつかの実施形態では、方法1600の1つまたはそれを上回るステップは、処理モジュール250またはARデバイス200のある他のコンポーネントによって実施されてもよい。

20

【0159】

ステップ1602では、ユーザの眼の入力画像（例えば、入力画像I（x，y，c））が、受信される。いくつかの実施形態では、入力画像は、カメラ（例えば、軸外カメラ260）から受信される。カメラは、光学デバイスに搭載されてもよく、および／または光学デバイスのコンポーネントであってもよい。いくつかの実施形態では、方法1600は、カメラを使用して、ユーザの眼の入力画像を捕捉するステップを含む。いくつかの実施形態では、方法1600は、カメラからの入力画像を処理モジュール（例えば、処理モジュール250）に送信するステップを含む。

30

【0160】

ステップ1604では、眼の入力画像が、ニューラルネットワーク（例えば、マルチタスクニューラルネットワーク256）に提供される。いくつかの実施形態では、入力画像は、ニューラルネットワークを実装する、プロセッサに提供される。プロセッサは、ニューラルネットワークによって一般に実施されるある演算（例えば、畳み込み、行列乗算）が汎用プロセッサを用いた場合より高速で実施されることを可能にする、アーキテクチャを有する、特殊目的プロセッサ（例えば、ニューラルネットワークプロセッサ）であってもよい。例えば、特殊目的プロセッサは、並行して、または同時に、種々の算術演算を入力画像の異なるピクセル上で実施するための複数の処理要素を有する、シストリックアレイを含んでもよい。

40

【0161】

ステップ1606では、ネットワーク出力データ（例えば、ネットワーク出力データ264）が、ニューラルネットワークを使用して生成される。ネットワーク出力データは、ニューラルネットワークの全体的出力およびニューラルネットワークの中間層の出力に対応するデータを含んでもよい。例えば、ネットワーク出力データは、ニューラルネットワークの全体的出力から導出される、あるデータ（例えば、眼セグメント化データ276）と、ニューラルネットワークの中間層の出力から導出される、あるデータ（例えば、瞬目

50

予測 2 9 6 および角膜中心データ 2 8 0) とを含んでもよい。加えて、または代替として、ネットワーク出力データは、ニューラルネットワークの異なる中間層およびニューラルネットワークの全体的出力の処理に関わらない 1 つまたはそれを上回る付加的層の出力から導出される、あるデータ (例えば、閃光検出データ 2 7 8 および 2 D 瞳孔中心データ 2 7 4) を含んでもよい。

【0 1 6 2】

ステップ 1 6 0 8 では、3 D 瞳孔中心 (例えば、3 D 瞳孔中心 2 9 2) が、ネットワーク出力データに基づいて算出される。いくつかの実施形態では、3 D 瞳孔中心は、2 D 瞳孔データおよび眼セグメント化データに基づいて算出される。

【0 1 6 3】

ステップ 1 6 1 0 では、ユーザの眼と関連付けられる光学軸 (例えば、光学軸 3 1 0) が、ネットワーク出力データに基づいて算出される。いくつかの実施形態では、光学軸は、3 D 瞳孔中心およびネットワーク出力データのあるデータ (例えば、角膜中心データ 2 8 0) に基づいて算出される。

【0 1 6 4】

ステップ 1 6 1 2 では、眼に対応する視線ベクトル (例えば、視線ベクトル 2 3 8) が、ネットワーク出力データに基づいて算出される。いくつかの実施形態では、視線ベクトルは、ネットワーク出力データ (例えば、2 D 瞳孔中心データ 2 7 4、眼セグメント化データ 2 7 6、および角膜中心データ 2 8 0) のあるコンポーネントのみを使用して算出される一方、ネットワーク出力データの他のコンポーネント (例えば、閃光検出データ 2 7 8) は、算出において使用されない。いくつかの実施形態では、視線ベクトルを算出するステップは、1 つまたはそれを上回る後処理ステップを含んでもよい。例えば、3 D 瞳孔中心 (例えば、3 D 瞳孔中心 2 9 2) が、最初に、ネットワーク出力データの 1 つまたはそれを上回るコンポーネント (例えば、2 D 瞳孔中心データ 2 7 4 および眼セグメント化データ 2 7 6) に基づいて算出され得る。第 2 に、光学軸 (例えば、光学軸 3 1 0) が、3 D 瞳孔中心およびネットワーク出力データの付加的コンポーネント (例えば、角膜中心データ 2 8 0) に基づいて算出され得る。次に、視線ベクトルが、光学軸およびユーザに対応する校正角度に基づいて算出され得る。

【0 1 6 5】

図 1 7 は、ニューラルネットワークを訓練する方法 1 7 0 0 を図示する。方法 1 7 0 0 のステップは、示される順序で実施される必要はなく、方法 1 7 0 0 の 1 つまたはそれを上回るステップは、方法 1 7 0 0 の実施の間、省略されてもよい。いくつかの実施形態では、方法 1 7 0 0 の 1 つまたはそれを上回るステップは、処理モジュール 2 5 0 または A R デバイス 2 0 0 のある他のコンポーネントによって実施されてもよい。

【0 1 6 6】

ステップ 1 7 0 2 では、複数の訓練入力画像 (例えば、入力画像 $I(x, y, c)$) が、受信される。複数の訓練入力画像は、カメラ (例えば、軸外カメラ 2 6 0) から受信されてもよい、または訓練の目的のために、人工的に生成される、または読み出されてもよい。複数の訓練画像はそれぞれ、眼の画像であってもよい。ステップ 1 7 0 2 は、ステップ 1 6 0 2 に類似してもよい。

【0 1 6 7】

ステップ 1 7 0 4 - 1 7 1 2 は、複数の訓練入力画像の訓練入力画像毎に実施されてもよい。ステップ 1 7 0 4 では、訓練入力画像は、ニューラルネットワーク (例えば、マルチタスクニューラルネットワーク 2 5 6) に提供される。ステップ 1 7 0 4 は、ステップ 1 6 0 4 に類似してもよい。

【0 1 6 8】

ステップ 1 7 0 6 では、訓練ネットワーク出力データ (例えば、ネットワーク出力データ 2 6 4) は、ニューラルネットワークを使用して生成される。ステップ 1 7 0 6 は、ステップ 1 6 0 6 に類似してもよい。

【0 1 6 9】

ステップ1708では、GTデータ（例えば、GTデータ268）が、ユーザ入力デバイスから受信される。GTデータは、訓練ネットワーク出力データの1つまたはそれを上回るコンポーネントに対応する、1つまたはそれを上回るコンポーネント（例えば、2D瞳孔中心GTデータ283、眼セグメント化GTデータ284、閃光検出GTデータ286）を含んでもよい。

【0170】

ステップ1710では、誤差データ（例えば、誤差データ270）が、訓練ネットワーク出力データとGTデータとの間の差異に基づいて算出される。誤差データは、GTデータおよび／または訓練ネットワーク出力データの1つまたはそれを上回るコンポーネントに対応する、1つまたはそれを上回るコンポーネント（例えば、第1の誤差データ270A、第2の誤差データ270B、第3の誤差データ270C、第4の誤差データ270D）を含んでもよい。

【0171】

ステップ1712では、ニューラルネットワークは、誤差データに基づいて修正される。いくつかの実施形態では、ニューラルネットワークへの修正の大きさは、訓練ネットワーク出力データとGTデータとの間のより大きい差異が、ニューラルネットワークへのより大きい修正に対応し得るように、誤差データの大きさに比例する。いくつかの実施形態では、ニューラルネットワークは、ニューラルネットワークの加重への1つまたはそれを上回る加重更新を計算する、逆伝搬アルゴリズムを使用して訓練されてもよい。

【0172】

図18は、本明細書に説明される実施形態による、簡略化されたコンピュータシステム1800を図示する。図18に図示されるようなコンピュータシステム1800は、本明細書に説明されるようなARデバイス200等のデバイスの中に組み込まれてもよい。図18は、種々の実施形態によって提供される方法のステップの一部または全部を実施し得る、コンピュータシステム1800の一実施形態の概略化された例証を提供する。図18は、種々のコンポーネントの一般化された例証を提供するためだけに意図され、そのいずれかまたは全てが、必要に応じて利用されてもよいことに留意されたい。図18は、したがって、広義には、個々のシステム要素が比較的に分離された様式または比較的により統合された様式において実装され得る状況を図示する。

【0173】

コンピュータシステム1800は、バス1805を介して電氣的に結合されることができ、または必要に応じて別様に通信し得る、ハードウェア要素を備えるように示される。ハードウェア要素は、限定ではないが、デジタル信号処理チップ、グラフィック加速プロセッサ、および／または同等物等の、1つまたはそれを上回る汎用プロセッサおよび／または1つまたはそれを上回る特殊目的プロセッサを含む、1つまたはそれを上回るプロセッサ1810と、限定ではないが、マウス、キーボード、カメラ、および／または同等物を含むことができる、1つまたはそれを上回る入力デバイス1815と、限定ではないが、ディスプレイデバイス、プリンタ、および／または同等物を含むことができる、1つまたはそれを上回る出力デバイス1820とを含んでもよい。

【0174】

コンピュータシステム1800はさらに、限定ではないが、ローカルおよび／またはネットワークアクセス可能記憶装置を備えることができ、および／または、限定ではないが、プログラム可能である、フラッシュ更新可能である、および／または同等物であることができる、ディスクドライブ、ドライブアレイ、光学記憶デバイス、ランダムアクセスメモリ（「RAM」）等のソリッドステート記憶デバイス、および／または読取専用メモリ（「ROM」）を含むことができる、1つまたはそれを上回る非一過性記憶デバイス1825を含む、および／またはそれと通信してもよい。そのような記憶デバイスは、限定ではないが、種々のファイルシステム、データベース構造、および／または同等物を含む、任意の適切なデータ記憶を実装するように構成されてもよい。

【0175】

コンピュータシステム1800はまた、限定ではないが、Bluetooth（登録商標）デバイス、802.11デバイス、WiFiデバイス、WiMaxデバイス、セルラ通信設備等、および／または同等物等のモデム、ネットワークカード（無線または有線）、赤外線通信デバイス、無線通信デバイス、および／またはチップセットを含むことができる、通信サブシステム1830を含み得る。通信サブシステム1830は、1つまたはそれを上回る入力および／または出力通信インターフェースを含み、データが、一実施例として挙げるために以下に説明されるネットワーク、すなわち、他のコンピュータシステム、テレビ、および／または本明細書に説明される任意の他のデバイス等のネットワークと交換されることを可能にしてもよい。所望の機能性および／または他の実装懸念に応じて、ポータブル電子デバイスまたは類似デバイスは、通信サブシステム1830を介して、画像および／または他の情報を通信してもよい。他の実施形態では、ポータブル電子デバイス、例えば、第1の電子デバイスは、コンピュータシステム1800、例えば、電子デバイスの中に入力デバイス1815として組み込まれてもよい。いくつかの実施形態では、コンピュータシステム1800はさらに、作業メモリ1835を備え、これは、上記に説明されるようなRAMまたはROMデバイスを含むことができる。

【0176】

コンピュータシステム1800はまた、種々の実施形態によって提供されるコンピュータプログラムを備え得る、および／または本明細書に説明されるような他の実施形態によって提供される方法を実装し、および／またはシステムを構成するように設計され得る、1つまたはそれを上回るアプリケーションプログラム1845等のオペレーティングシステム1840、デバイスドライバ、実行可能ライブラリ、および／または他のコードを含む、作業メモリ1835内に現在位置するものとして示される、ソフトウェア要素を含むことができる。単に、一例として、上記に議論される方法に関して説明される1つまたはそれを上回るプロシージャは、コンピュータまたはコンピュータ内のプロセッサによって実行可能なコードおよび／または命令として実装され得、ある側面では、次いで、そのようなコードおよび／または命令は、説明される方法に従って1つまたはそれを上回る動作を実施するように汎用コンピュータまたは他のデバイスを構成および／または適合するために使用されることができる。

【0177】

これらの命令および／またはコードのセットは、上記に説明される記憶デバイス1825等の非一過性コンピュータ可読記憶媒体上に記憶されてもよい。ある場合には、記憶媒体は、コンピュータシステム1800等のコンピュータシステム内に組み込まれ得る。他の実施形態では、記憶媒体は、コンピュータシステムと別個である、例えば、コンパクトディスク等の可搬性媒体である、および／または記憶媒体が、汎用コンピュータをその上に記憶される命令／コードを用いてプログラム、構成、および／または適合するために使用され得るように、インストールパッケージ内に提供され得る。これらの命令は、コンピュータシステム1800によって実行可能である、実行可能コードの形態をとり得る、および／または、例えば、種々の概して利用可能なコンパイラ、インストールプログラム、圧縮／解凍ユーティリティ等のいずれかを使用したコンピュータシステム1800上へのコンパイルおよび／またはインストールに応じて、次いで、実行可能コードの形態をとる、ソースおよび／またはインストール可能コードの形態をとり得る。

【0178】

実質的な変形例が、具体的要件に従って構成されてもよいことが、当業者に明白となるであろう。例えば、カスタマイズされたハードウェアもまた、使用され得る、および／または特定の要素が、ハードウェア、アプレット等のポータブルソフトウェアを含む、ソフトウェア、または両方内に実装され得る。さらに、ネットワーク入力／出力デバイス等の他のコンピューティングデバイスへの接続も、採用されてもよい。

【0179】

上記に述べられたように、一側面では、いくつかの実施形態は、コンピュータシステム1800等のコンピュータシステムを採用し、本技術の種々の実施形態による方法を実施

10

20

30

40

50

してもよい。一式の実施形態によると、そのような方法のプロシージャの一部または全部は、プロセッサ1810が、オペレーティングシステム1840の中に組み込まれ得る、1つまたはそれを上回る命令の1つまたはそれを上回るシーケンス、および／または作業メモリ1835内に含有される、アプリケーションプログラム1845等の他のコードを実行することに応答して、コンピュータシステム1800によって実施される。そのような命令は、記憶デバイス1825のうちの1つまたはそれを上回るもの等の別のコンピュータ可読媒体から作業メモリ1835の中に読み取られてもよい。単に、一例として、作業メモリ1835内に含有される命令のシーケンスの実行は、プロセッサ1810に、本明細書に説明される方法の1つまたはそれを上回るプロシージャを実施させ得る。加えて、または代替として、本明細書に説明される方法の一部は、特殊ハードウェアを通して実行されてもよい。

10

【0180】

用語「機械可読媒体」および「コンピュータ可読媒体」は、本明細書で使用されるとき、機械を具体的方式で動作させるデータを提供することに関わる、任意の媒体を指す。コンピュータシステム1800を使用して実装される、ある実施形態では、種々のコンピュータ可読媒体は、実行のための命令／コードをプロセッサ1810に提供する際に関わり得る、および／またはそのような命令／コードを記憶および／または搬送するために使用され得る。多くの実装では、コンピュータ可読媒体は、物理的および／または有形記憶媒体である。そのような媒体は、不揮発性媒体または揮発性媒体の形態をとってもよい。不揮発性媒体は、例えば、記憶デバイス1825等の光学および／または磁気ディスクを含む。揮発性媒体は、限定ではないが、作業メモリ1835等の動的メモリを含む。

20

【0181】

一般的形態の物理的および／または有形コンピュータ可読媒体は、例えば、フロッピー（登録商標）ディスク、可撓性ディスク、ハードディスク、磁気テープ、または任意の他の磁気媒体、CD-ROM、任意の他の光学媒体、パンチカード、紙テープ、孔のパターンを伴う任意の他の物理的媒体、RAM、PROM、EPROM、FLASH-EPROM、任意の他のメモリチップまたはカートリッジ、またはコンピュータが命令および／またはコードを読み取ることができる、任意の他の媒体を含む。

【0182】

種々の形態のコンピュータ可読媒体が、実行のための1つまたはそれを上回る命令の1つまたはそれを上回るシーケンスをプロセッサ1810に搬送する際に関わってもよい。単に、一例として、命令は、最初に、遠隔コンピュータの磁気ディスクおよび／または光学ディスク上で搬送されてもよい。遠隔コンピュータは、命令をその動的メモリの中にロードし、コンピュータシステム1800によって受信および／または実行される伝送媒体を経由して、命令を信号として送信し得る。

30

【0183】

通信サブシステム1830および／またはそのコンポーネントは、概して、信号を受信し、バス1805が、次いで、信号および／または信号によって搬送されるデータ、命令等を作業メモリ1835に搬送し得、そこから、プロセッサ1810が、命令を読み出し、実行する。作業メモリ1835によって受信された命令は、随意に、プロセッサ1810による実行前または後のいずれかにおいて、非一過性記憶デバイス1825上に記憶されてもよい。

40

【0184】

上記に議論される方法、システム、およびデバイスは、実施例である。種々の構成は、必要に応じて、種々のプロシージャまたはコンポーネントを省略、代用、または追加してもよい。例えば、代替構成では、本方法は、説明されるものと異なる順序で実施されてもよく、および／または種々の段階は、追加される、省略される、および／または組み合わせられてもよい。また、ある構成に関して説明される特徴は、種々の他の構成において組み合わせられてもよい。構成の異なる側面および要素は、類似様式で組み合わせられてもよい。また、技術は、進歩するものであって、したがって、要素の多くは、実施例であっ

50

て、本開示の範囲または請求項を限定するものではない。

【0185】

具体的詳細が、実装を含む、例示的構成の完全な理解を提供するために説明に与えられる。しかしながら、構成は、これらの具体的詳細を伴わずに実践されてもよい。例えば、周知の回路、プロセス、アルゴリズム、構造、および技法は、構成を曖昧にすることを回避するために、不必要な詳細を伴わずに示されている。本説明は、例示的構成のみを提供し、請求項の範囲、可用性、または構成を限定するものではない。むしろ、構成の前述の説明は、当業者に説明される技法を実装するための有効な説明を提供するであろう。種々の変更が、本開示の精神または範囲から逸脱することなく、要素の機能および配列に行われてもよい。

10

【0186】

また、構成は、概略フローチャートまたはブロック図として描写される、プロセスとして説明され得る。それぞれ、シーケンシャルプロセスとして動作を説明し得るが、動作の多くは、並行して、または同時に実施されることができる。加えて、動作の順序は、再配列されてもよい。プロセスは、図内に含まれない付加的ステップを有してもよい。さらに、本方法の実施例は、ハードウェア、ソフトウェア、ファームウェア、ミドルウェア、マイクロコード、ハードウェア記述言語、または任意のそれらの組み合わせによって実装されてもよい。ソフトウェア、ファームウェア、ミドルウェア、またはマイクロコード内に実装されるとき、必要タスクを実施するためのプログラムコードまたはコードセグメントは、記憶媒体等の非一過性コンピュータ可読媒体内に記憶されてもよい。プロセッサは、説明されるタスクを実施してもよい。

20

【0187】

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、使用されてもよい。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ってもよく、他のルールが、本技術の用途に優先する、または別様にそれを修正してもよい。また、いくつかのステップは、前述の要素が検討される前、間、または後に行われてもよい。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。

【0188】

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つまたはそれを上回る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。

30

【0189】

また、単語「comprise（～を備える）」、「comprising（～を備える）」、「contains（～を含有する）」、「containing（～を含有する）」、「include（～を含む）」、「including（～を含む）」、および「includes（～を含む）」は、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つまたはそれを上回る他の特徴、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。

40

【0190】

また、本明細書に説明される実施例および実施形態は、例証目的のみのためのものであって、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および添付の請求項の範囲内に含まれることを理解されたい。

50

【図面】
【図 1】

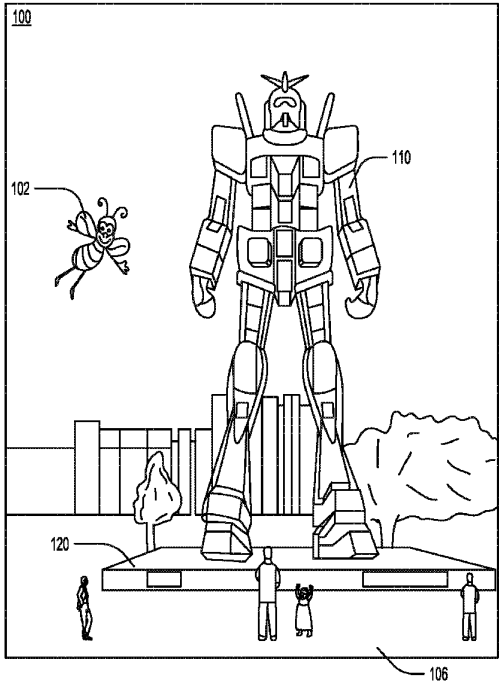


FIG. 1

【図 2】

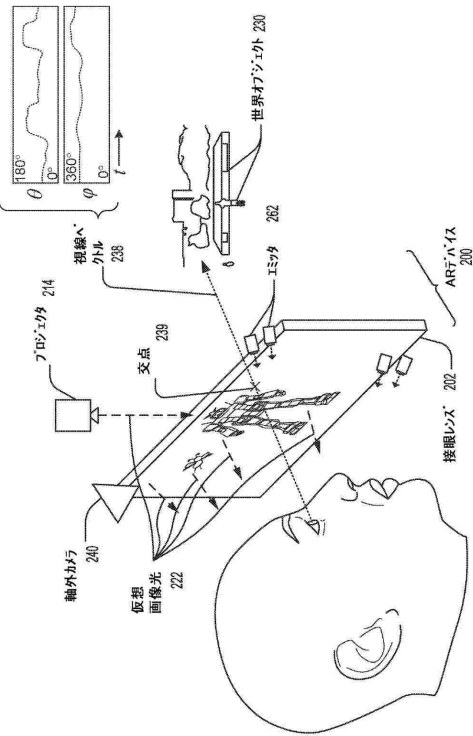


FIG. 2

【図 3】

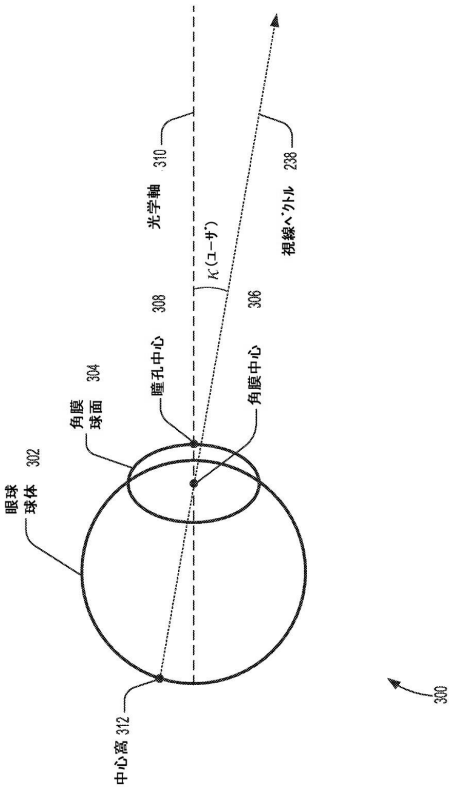


FIG. 3

【図 4】

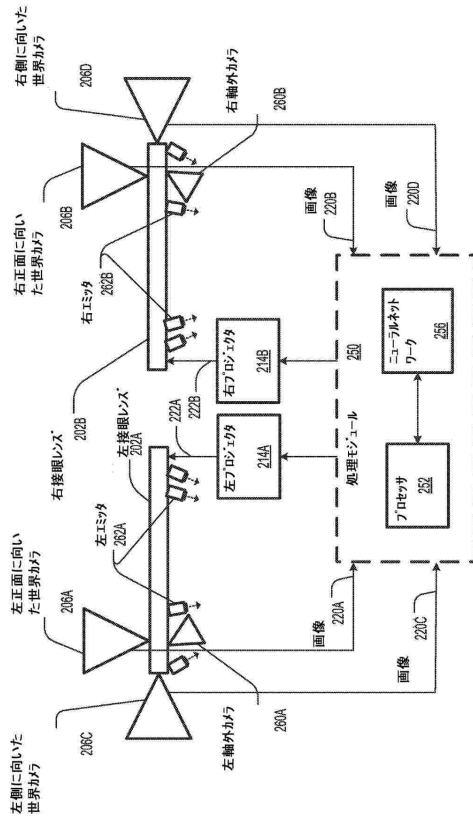


FIG. 4

10

20

30

40

50

【図 5】

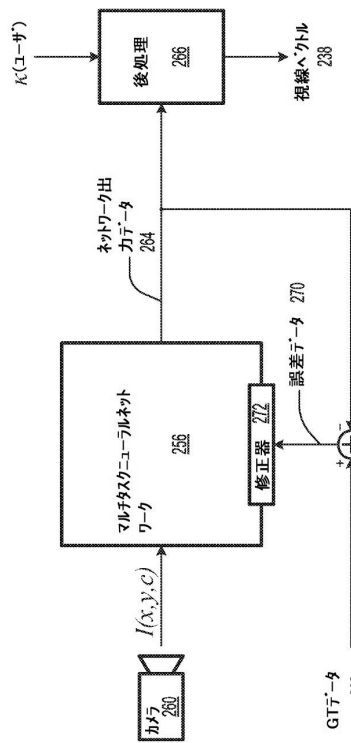


FIG. 5

【図 6】

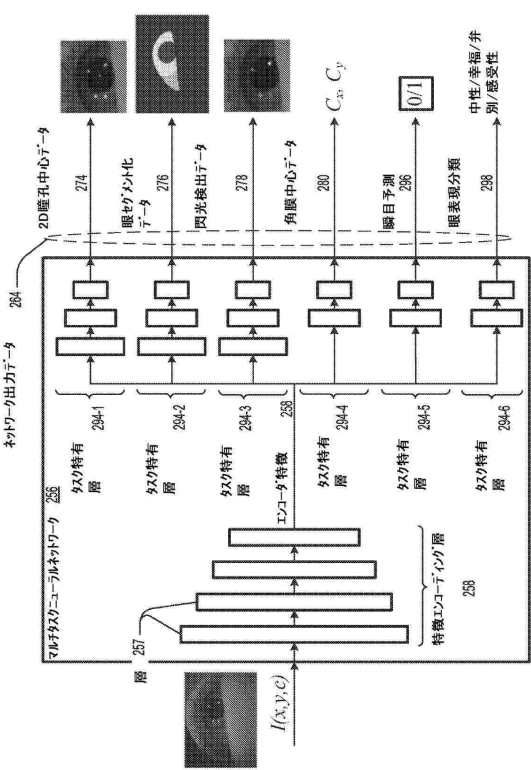


FIG. 6

【図 7】

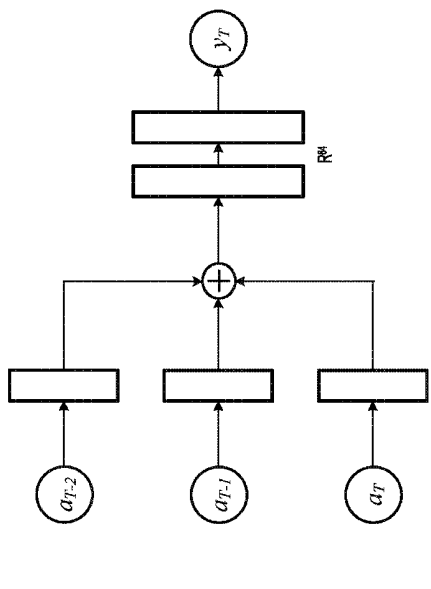


FIG. 7

【図 8】

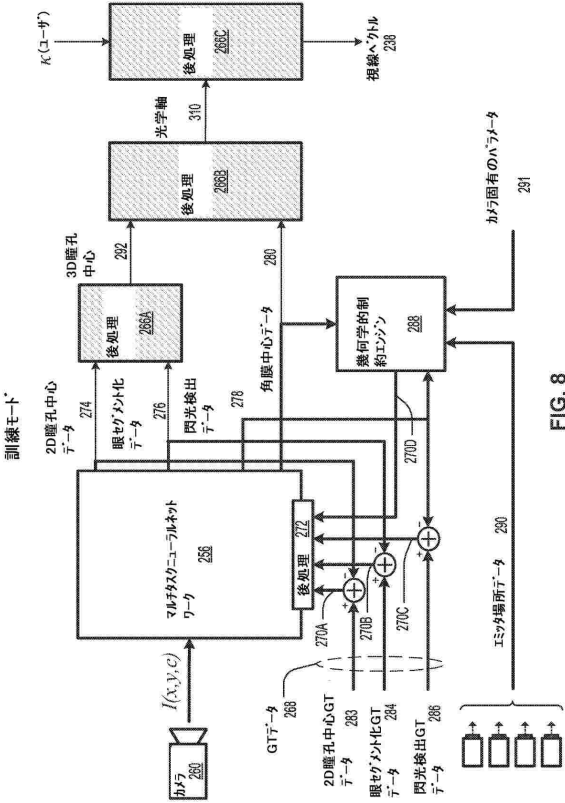


FIG. 8

10

20

30

40

50

【図 9 A】

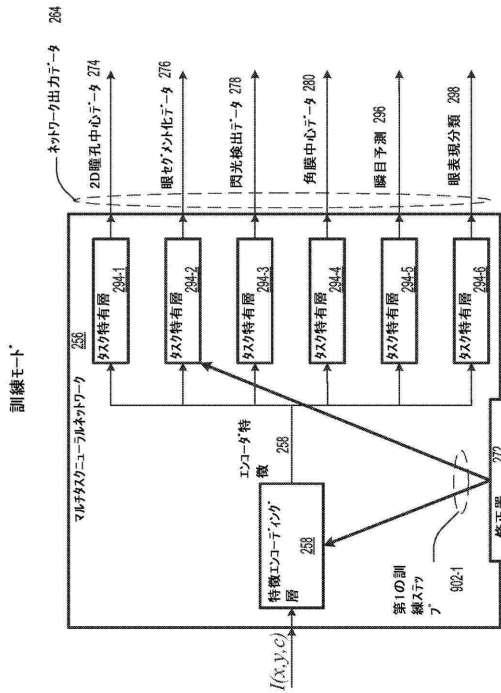


FIG. 9A

【図 9 B】

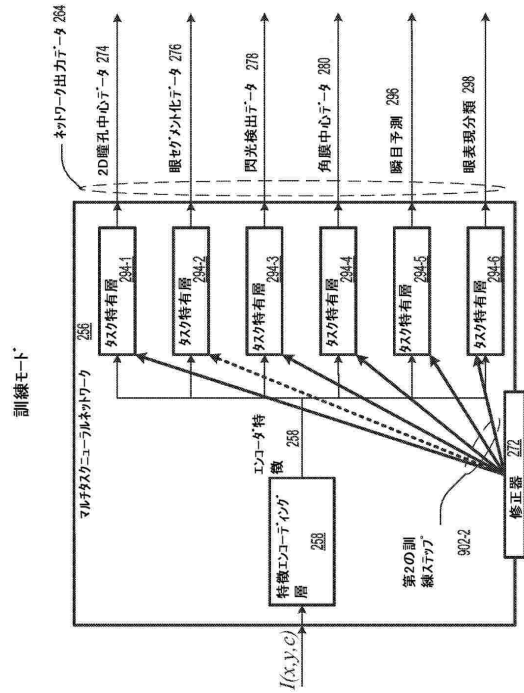


FIG. 9B

【図 10】

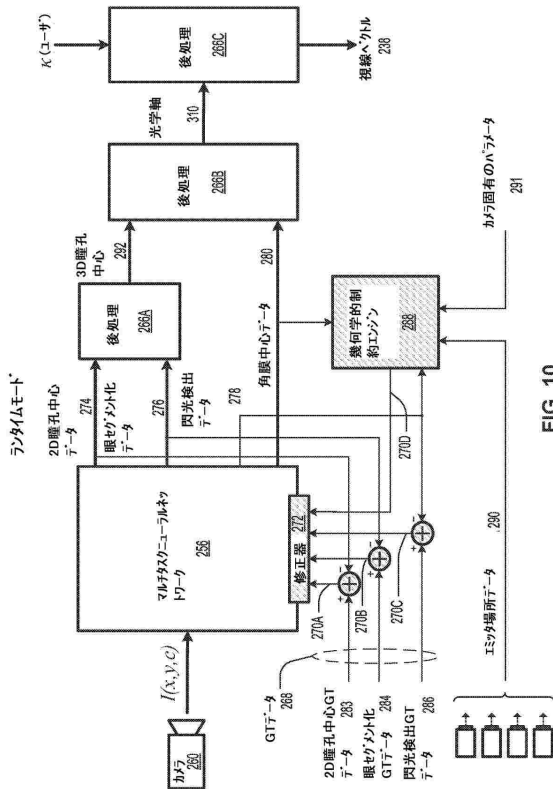


FIG. 10

【図 11】

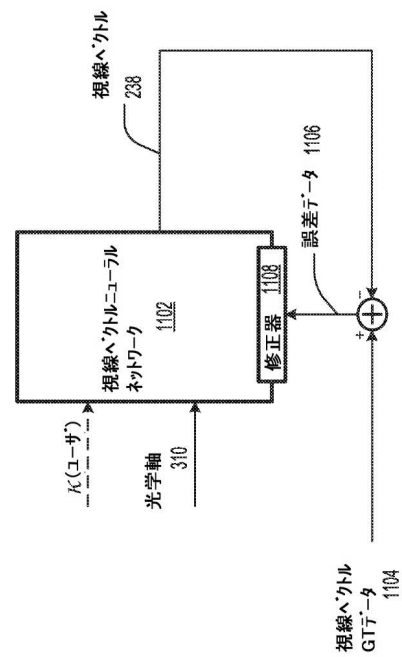


FIG. 11

10

20

30

40

50

【図 1 2】

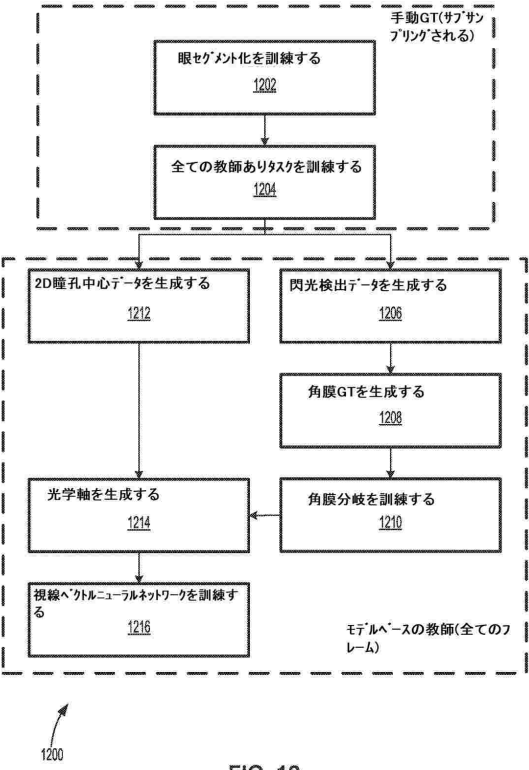


FIG. 12

【図 1 3】

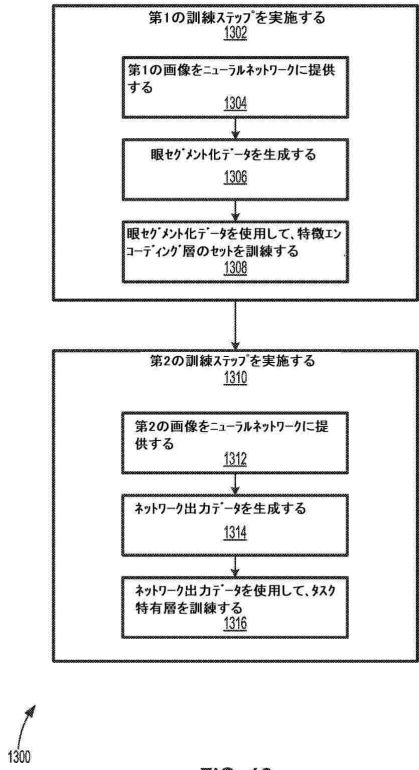


FIG. 13

【図 1 4】

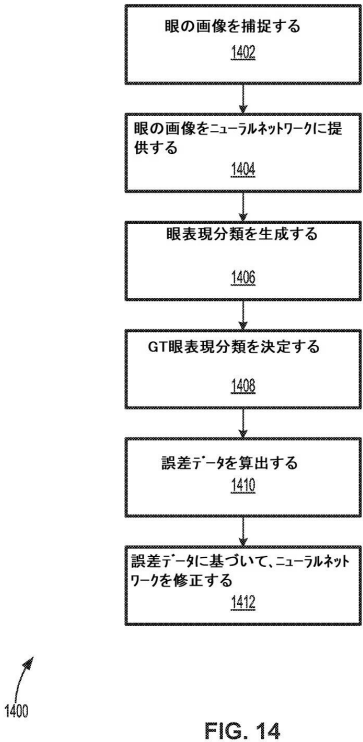


FIG. 14

【図 1 5】

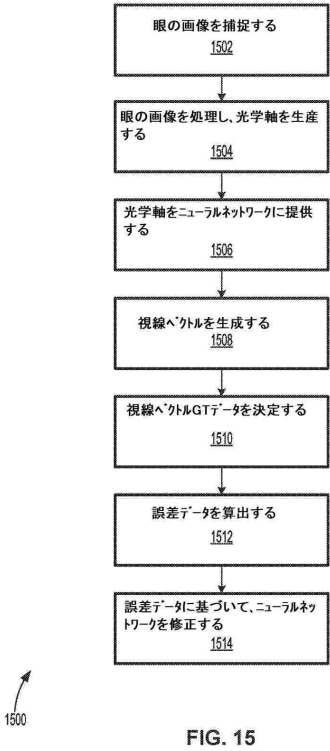


FIG. 15

10

20

30

40

50

【図 1 6】

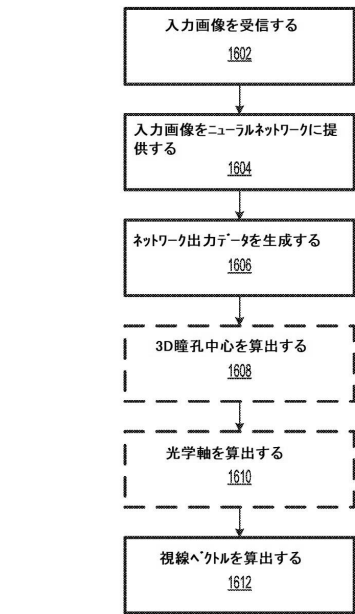


FIG. 16

【図 1 7】

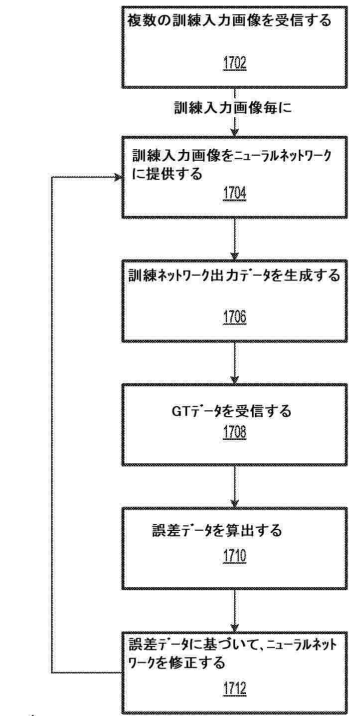


FIG. 17

【図 1 8】

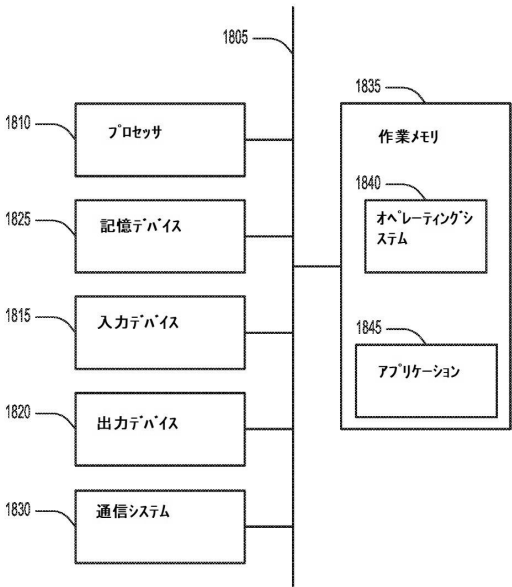


FIG. 18

10

20

30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/935,584

(32)優先日 令和1年11月14日(2019.11.14)

(33)優先権主張国・地域又は機関

米国(US)

(72)発明者 ウー, ジェンヤン

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, マジック リープ, インコーポレイテッド 気付

(72)発明者 ラジェンドラン, スリヴィグネシュ

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, マジック リープ, インコーポレイテッド 気付

(72)発明者 ヴァン アズ, タレンス

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, マジック リープ, インコーポレイテッド 気付

(72)発明者 ジマーマン, ジョエル

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, マジック リープ, インコーポレイテッド 気付

(72)発明者 バドリナラヤナン, ヴィジャイ

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, マジック リープ, インコーポレイテッド 気付

(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, マジック リープ, インコーポレイテッド 気付

審査官 山田 辰美

(56)参考文献 国際公開第2018/063451 (WO, A1)

国際公開第2018/039269 (WO, A1)

国際公開第2018/035531 (WO, A1)

Clement Ployat; Renaud Duval; Farida Cheriet, A Novel Weakly Supervised Multitask Arch
itecture for Retinal Lesions Segmentation on Fundus Images, IEEE Transactions on Medical
Imaging, 米国, I E E E, 2019年03月20日, Volume: 38, Issue: 10, p.2434-p.2444,
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8672120>松井 涼, 山下 隆義, 藤吉 弘亘, Separation Multi-task Netwo
rksによる顔器官点と顔属性の同時推定, 電子情報通信学会技術研究報告, 日本, 2018
年12月06日, Vol. 118 No. 362, p.39-p.44

(58)調査した分野 (Int.Cl., DB名)

G06T 7/00-7/90

G06V 10/82

G06T 7/11

I E E E X p l o r e