

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7386812号
(P7386812)

(45)発行日 令和5年11月27日(2023.11.27)

(24)登録日 令和5年11月16日(2023.11.16)

(51)国際特許分類 F I
G 0 6 T 15/50 (2011.01) G 0 6 T 15/50 6 0 0

請求項の数 18 (全25頁)

(21)出願番号	特願2020-564910(P2020-564910)	(73)特許権者	314015767 マイクロソフト テクノロジー ライセンシング, エルエルシー アメリカ合衆国 ワシントン州 9 8 0 5 2 レッドモンド ワン マイクロソフト ウェイ
(86)(22)出願日	令和1年5月13日(2019.5.13)	(74)代理人	100107766 弁理士 伊東 忠重
(65)公表番号	特表2021-524628(P2021-524628 A)	(74)代理人	100070150 弁理士 伊東 忠彦
(43)公表日	令和3年9月13日(2021.9.13)	(74)代理人	100135079 弁理士 宮崎 修
(86)国際出願番号	PCT/US2019/031907	(72)発明者	ドン, ユエ アメリカ合衆国 ワシントン州 9 8 0 5 2 - 6 3 9 9 レッドモンド ワン マイ 最終頁に続く
(87)国際公開番号	WO2019/226366		
(87)国際公開日	令和1年11月28日(2019.11.28)		
審査請求日	令和4年4月12日(2022.4.12)		
(31)優先権主張番号	201810510214.1		
(32)優先日	平成30年5月24日(2018.5.24)		
(33)優先権主張国・地域又は機関	中国(CN)		

(54)【発明の名称】 照明推定

(57)【特許請求の範囲】

【請求項1】

電子デバイスであって：

処理ユニットと；

前記処理ユニットに結合され、前記処理ユニットによる実行のための命令を記憶するメモリであって、前記命令は、前記処理ユニットによって実行されるとき、前記デバイスに：
実オブジェクトについての入力画像及び前記入力画像に対応するデプスマップを取得することと；

前記デプスマップに基づいて前記入力画像内のシーンを決定することと；

前記シーンに基づいて、前記シーン内の光源によって引き起こされる前記実オブジェクトのシェーディング及び影情報を決定することであって、前記光源は、所定の形状の画像によって表され、前記画像は、前記光源の複数の部分に対応する複数のピクセルを含み、前記複数のピクセルのそれぞれのピクセル値は、前記光源の前記複数の部分のそれぞれの放射照度強度を表し、

前記シェーディング及び影情報を決定することは：

前記画像内の前記複数のピクセルのそれぞれのピクセル値を所定の値で初期化することと；

前記シーン及び初期化された前記画像に基づいて、前記複数のピクセルに対応する第1の放射照度マップのセットであって、前記放射照度マップの各々は前記光源の対応する部分によって引き起こされる前記実オブジェクトの対応するシェーディング及び影情

10

20

報を表す、前記第1の放射照度マップのセットを生成することと；を含む、

シェーディング及び影情報を決定することと；

前記入力画像と、前記シェーディング及び影情報とに基づいて前記光源によって引き起こされる前記シーン内の照明条件を決定することとであって、前記入力画像及び前記第1の放射照度マップのセットに基づいて、前記画像内の前記複数のピクセルのそれぞれのピクセル値を決定することを含む、前記シーン内の前記照明条件を決定することと；

を含む動作を実行させる、メモリと；

を有する、電子デバイス。

【請求項2】

前記動作はさらに：

前記シーン及び前記照明条件に基づいて、前記実オブジェクト及び仮想オブジェクトを含む出力画像を生成することを含む、前記出力画像は前記入力画像の前記シーン内に前記仮想オブジェクトを合成する、

請求項1に記載のデバイス。

【請求項3】

前記シーンは、前記実オブジェクトの配置のための表面を含み、前記シーンを決定することは：

前記デプスマップを点群に変換することと；

前記点群内の複数の点を面に適合させることとであって、前記複数の点は前記表面を表す、適合させることと；

前記面の上に位置する前記点群内の前記複数の点を3次元メッシュに再構成することとであって、前記面及び前記3次元メッシュは、前記入力画像内の前記シーンを共に表す、再構成することと；を含む、

請求項1に記載のデバイス。

【請求項4】

前記光源を表す前記画像は、十字形状の画像である、

請求項1に記載のデバイス。

【請求項5】

前記画像内の前記複数のピクセルのそれぞれのピクセル値を決定することは：

前記入力画像及び前記第1の放射照度マップのセットに基づいて、訓練されたニューラルネットワークを用いて前記画像内の前記複数のピクセルのそれぞれの前記ピクセル値を決定することを含む、

請求項1に記載のデバイス。

【請求項6】

前記画像内の前記複数のピクセルのそれぞれのピクセル値を決定することは：

前記入力画像及び前記第1の放射照度マップのセットに基づいて、複数の訓練されたニューラルネットワークを用いて前記画像内の前記複数のピクセルのそれぞれの前記ピクセル値を漸進的に決定することを含む、

請求項1に記載のデバイス。

【請求項7】

前記複数の訓練されたニューラルネットワークは、少なくとも第1のニューラルネットワーク及び第2のニューラルネットワークを含み、前記画像内の前記複数のピクセルのそれぞれの前記ピクセル値を決定することは：

前記入力画像及び前記第1の放射照度マップのセットに基づいて、前記第1のニューラルネットワークを使用して、前記複数のピクセルの第1のピクセルの第1のピクセル値及び前記第1のピクセルに隣接する少なくとも1つのピクセルのピクセル値を決定することと；

前記第1のピクセル、前記少なくとも1つのピクセル及び前記シーンに基づいて第2の放射照度マップのセットを生成することとであって、前記第2の放射照度マップのセットは、前記第1のピクセルの複数のサブピクセル、前記少なくとも1つのピクセル、並びに前

10

20

30

40

50

記複数のピクセルのうちの前記第 1 のピクセル及び前記少なくとも 1 つのピクセル以外のピクセルの組み合わせそれぞれに対応する、生成することと；

前記入力画像及び前記第 2 の放射照度マップのセットに基づいて、前記第 1 のピクセルが前記第 2 のニューラルネットワークを使用して分割されるべきかどうかを決定することと；

前記第 1 のピクセル値が分割されるべきでないと決定することに応答して、前記第 1 のピクセル値を前記第 1 のピクセルの最終ピクセルとして使用することと；を含む、

請求項 6 に記載のデバイス。

【請求項 8】

前記画像内の前記複数のピクセルのそれぞれの前記ピクセル値を決定することはさらに：
前記第 1 のピクセルが分割されるべきであると決定することに応答して、前記第 1 のピクセルを複数のサブピクセルに分割することと；

10

前記複数のサブピクセルのそれぞれの子ピクセル値を決定することと；を含む、
請求項 7 に記載のデバイス。

【請求項 9】

コンピュータに実装される方法であって；

実オブジェクトに関する入力画像及び前記入力画像に対応するデプスマップを取得することと；

前記デプスマップに基づいて前記入力画像内のシーンを決定することと；

前記シーンに基づいて、前記シーン内の光源によって引き起こされる前記実オブジェクトのシェーディング及び影情報を決定することであって、前記光源は、所定の形状の画像によって表され、前記画像は、前記光源の複数の部分に対応する複数のピクセルを含み、前記複数のピクセルのそれぞれの子ピクセル値は、前記光源の前記複数の部分のそれぞれの放射照度強度を表し、

20

前記シェーディング及び影情報を決定することは；

前記画像内の前記複数のピクセルのそれぞれの子ピクセル値を所定の値で初期化することと；

前記シーン及び初期化された前記画像に基づいて、前記複数のピクセルに対応する第 1 の放射照度マップのセットであって、前記放射照度マップの各々は前記光源の対応する部分によって引き起こされる前記実オブジェクトの対応するシェーディング及び影情報を表す、前記第 1 の放射照度マップのセットを生成することと；を含む、

30

シェーディング及び影情報を決定することと；

前記入力画像と、前記シェーディング及び影情報とに基づいて、前記光源によって引き起こされる前記シーン内の照明条件を決定することであって、前記入力画像及び前記第 1 の放射照度マップのセットに基づいて、前記画像内の前記複数のピクセルのそれぞれの子ピクセル値を決定することを含む、シーン内の前記照明条件を決定することと；を含む、
方法。

【請求項 10】

前記シーン及び前記照明条件に基づいて、前記実オブジェクト及び仮想オブジェクトを含む出力画像を生成することをさらに含み、前記出力画像は、前記入力画像の前記シーン内に前記仮想オブジェクトを合成する、

40

請求項 9 に記載の方法。

【請求項 11】

前記シーンは、前記実オブジェクトの配置のための表面を含み、前記シーンを決定することは；

前記デプスマップを点群に変換することと；

前記点群内の複数の点を面に適合させることであって、前記複数の点は、前記表面を表す、適合させることと；

前記面の上に位置する前記点群内の複数の点を 3 次元メッシュに再構成することであって、前記面及び前記 3 次元メッシュは、前記入力画像内の前記シーンを共に表す、再構成

50

することと；を含む、
請求項 9 に記載の方法。

【請求項 1 2】

前記光源を表す前記画像は、十字形状の画像である、
請求項 9 に記載の方法。

【請求項 1 3】

前記画像内の前記複数のピクセルのそれぞれのピクセル値を決定することは：
前記入力画像及び前記第 1 の放射照度マップのセットに基づいて、訓練されたニューラルネットワークを用いて前記画像内の前記複数のピクセルのそれぞれの前記ピクセル値を決定することを含む、
請求項 9 に記載の方法。

10

【請求項 1 4】

命令を記憶する機械可読媒体であって、前記命令は、1 つ又は複数の機械によって実行されるとき、前記 1 つ又は複数の機械に：
実オブジェクトに関する入力画像及び前記入力画像に対応するデプスマップを取得することと；

前記デプスマップに基づいて前記入力画像内のシーンを決定することと；

前記シーンに基づいて、前記シーン内の光源によって引き起こされる前記実オブジェクトのシェーディング及び影情報を決定することであって、前記光源は、所定の形状の画像によって表され、前記画像は、前記光源の複数の部分に対応する複数のピクセルを含み、前記複数のピクセルのそれぞれのピクセル値は、前記光源の前記複数の部分のそれぞれの放射照度強度を表し、

20

前記シェーディング及び影情報を決定することは：

前記画像内の前記複数のピクセルのそれぞれの前記ピクセル値を所定の値で初期化することと；

前記シーン及び初期化された前記画像に基づいて、前記複数のピクセルに対応する第 1 の放射照度マップのセットであって、前記放射照度マップの各々は前記光源の対応する部分によって引き起こされる前記実オブジェクトの対応するシェーディング及び影情報を表す、前記第 1 の放射照度マップのセットを生成することと；を含む、

シェーディング及び影情報を決定することと；

30

前記入力画像と、前記シェーディング及び影情報とに基づいて、前記光源によって引き起こされる前記シーン内の照明条件を決定することであって、前記入力画像及び前記第 1 の放射照度マップのセットに基づいて、前記画像内の前記複数のピクセルのそれぞれの前記ピクセル値を決定することを含む、シーン内の前記照明条件を決定することと；を含む、動作を実行させる、

機械可読媒体。

【請求項 1 5】

前記動作は：

前記シーン及び前記照明条件に基づいて、前記実オブジェクト及び仮想オブジェクトを含む出力画像を生成することをさらに含み、前記出力画像は、前記入力画像の前記シーン内に前記仮想オブジェクトを合成する、

40

請求項 1 4 に記載の機械可読媒体。

【請求項 1 6】

前記シーンは、前記実オブジェクトの配置のための表面を含み、前記シーンを決定することは：

前記デプスマップを点群に変換することと；

前記点群内の複数の点を面に適合させることであって、前記複数の点は、前記表面を表す、適合させることと；

前記面の上に位置する前記点群内の複数の点を 3 次元メッシュに再構成することであって、前記面及び前記 3 次元メッシュは、前記入力画像内の前記シーンを共に表す、再構成

50

することと；を含む、

請求項 1 4 に記載の機械可読媒体。

【請求項 1 7】

前記光源を表す前記画像は、十字形状の画像である、

請求項 1 4 に記載の機械可読媒体。

【請求項 1 8】

前記画像内の前記複数のピクセルのそれぞれのピクセル値を決定することは：

前記入力画像及び前記第 1 の放射照度マップのセットに基づいて、訓練されたニューラルネットワークを用いて前記画像内の前記複数のピクセルのそれぞれの前記ピクセル値を決定することを含む、

10

請求項 1 4 に記載の機械可読媒体。

【発明の詳細な説明】

【背景技術】

【0001】

三次元仮想オブジェクトを現実のシーン (real scene) に関する画像に合成することは、拡張現実、画像編集、映画制作などの多くのアプリケーションにおいて重要なタスクである。光現実的效果を達成するために、合成画像内の仮想オブジェクト (virtual object) が、そのシーン内でシェーディング及び実オブジェクト (real object) と一致する影を示すことができ、かつ合成画像が、仮想オブジェクトと実オブジェクトとの間の投影及び/又はオクルージョン (occlusion) を正確に示すことができるように、現実のシーンの照明条件 (lighting condition) を推定する必要がある。最近のデプスカメラの進歩により、現実のシーンの三次元幾何学的構造を容易に取得することができる。しかし、現実のシーンに関する単一画像からシーンの照明条件を推定することは、依然として困難な作業である。

20

【発明の概要】

【0002】

本明細書に記載される主題の実装によれば、照明推定の解決策が提供される。この解決策では、実オブジェクトに関する入力画像と、入力画像に対応するデプスマップ (depth map) が得られる。入力画像内のシーンの幾何学的構造は、デプスマップに基づいて決定される。シーン内の光源によって引き起こされる実オブジェクトのシェーディング及び影情報 (shading and shadow information) は、決定されたシーンの幾何学的構造に基づいて決定される。そして、光源によるシーン内の照明条件が入力画像とシェーディング及び影情報とに基づいて決定される。

30

【0003】

本明細書に記載される主題による照明推定の解決策は、シーン内の実オブジェクトの材料特性に関する仮定を行わずに、入力画像並びに入力画像のシェーディング及び影情報に基づいて複数レベルのニューラルネットワークを用いて、シーン内の照明条件を漸進的に推定することができる。シーン中の光源を基準の光ベース (canonical light bases) のセットの線形結合でモデル化し、ニューラルネットワークの入力として基準の光ベースのセットからレンダリングされた放射照度マップ (irradiance maps) のセットを利用することにより、この解決策は照明推定の複雑さを大幅に単純化し、結果の精度を改善することができる。さらに、マルチツリーベースの漸進的推定を利用することにより、この解決策はニューラルネットワークの各レベルの訓練と計算コストを低減し、そのロバスト性を改善することができる。この解決策に従って得られた照明条件を用いてレンダリングされる仮想オブジェクトは、実オブジェクトと一致する現実的な効果を示すことができる。

40

【0004】

この要約は、以下の詳細な説明でさらに記述される簡略化された形式で概念の選択を導入するために提供される。この要約は、請求項に記載された主題の主要な特徴又は本質的な特徴を特定することを意図したものではなく、また、本明細書に記載された主題の範囲を限定するために使用することも意図したものではない。

50

【図面の簡単な説明】**【0005】**

【図1】本明細書に記載される主題の複数の実装を実装できるコンピューティングデバイス100のブロック図を示す。

【図2】本明細書に記載される主題のいくつかの実装による照明推定のためのシステム200を示す。

【図3】本明細書に記載される主題のいくつかの実装による、システム200内の照明条件決定サブシステムの図を示す。

【図4】本明細書に記載される主題のいくつかの実装による照明推定のための初期化ネットワークの動作図を示す。

【図5】本明細書に記載される主題のいくつかの実装による照明推定のためのリファインメント(refinement)ネットワークの動作図を示す。

【図6】本明細書に記載される主題のいくつかの実装による、複数のレベルのニューラルネットワークを使用して、シーン内の照明条件を漸進的に決定する図を示す。

【図7】本明細書に記載される主題のいくつかの実装による照明推定のためのプロセスのフローチャートを示す。

【0006】

図面を通して、同一又は類似の参照記号は、同一又は類似の要素を表す。

【発明を実施するための形態】**【0007】**

本明細書に記載する主題が、以下、いくつかの例示の実装を参照して論じられる。これらの実装は、主題の範囲に関する限定を示唆するのではなく、当業者が本明細書に記載の主題をより良く理解し、従って実施することを可能にする目的のためにのみ論じられることが理解されるべきである。

【0008】

本明細書において、用語「含む」及びその変形は、「含むが、限定されない」ことを意味するオープンな用語として読まれるべきである。用語「基づく」は、「少なくとも部分的に基づく」として読まれるべきである。用語「1つの実装」及び「実装」は、「少なくとも1つの実装」として読まれるべきである。用語「別の実装」は、「少なくとも1つの他の実装」として読まれるべきである。用語「第1の」、「第2の」などは、異なる又は同じオブジェクトを指し得る。他の定義が、明示的及び暗示的に、以下に含まれ得る。

【0009】**環境例****【0010】**

ここで、図面を参照して、本明細書に記載される主題の基本原理及び種々の例を説明する。図1は、本明細書に記載される主題の実装が実装されることができコンピューティング環境100のブロック図を示す。図1に示されるコンピューティングデバイス100は、本明細書に記載される主題の機能及び実装の範囲に対する限定をいかなる方法でも示唆することなく、単に例示的なものであることが理解されるべきである。図1に示すように、コンピューティングデバイス100は、一般的なコンピュータデバイスの形態のコンピューティングデバイス100を含む。コンピューティングデバイス100の構成要素は、1つ又は複数のプロセッサ又は処理ユニット110、メモリ120、記憶装置130、1つ又は複数の通信ユニット140、1つ又は複数の入力装置150、及び1つ又は複数の出力装置160を含むが、これらに限定されない。

【0011】

幾つかの実装では、コンピューティングデバイス100は、演算能力を有する種々のユーザ端末又はサービス端末として実装されることができ。サービス端末は、種々のサービスプロバイダによって提供されるサーバ、大規模コンピューティングデバイス等であり得る。ユーザ端末は、例えば、携帯電話、ステーション、ユニット、デバイス、マルチメディアコンピュータ、マルチメディアタブレット、インターネットノード、通信機、デス

10

20

30

40

50

クトップコンピュータ、ラップトップコンピュータ、ノートブックコンピュータ、ネットブックコンピュータ、タブレットコンピュータ、パーソナルコミュニケーションシステム（PCS）デバイス、パーソナルナビゲーションデバイス、パーソナルデジタルアシスタント（PDA）、オーディオ/ビデオプレーヤ、デジタルカメラ/ビデオプレーヤ、位置決定デバイス（positioning device）、テレビ受信機、ラジオ放送受信機、電子ブックデバイス、ゲーミングデバイス、又はこれらのデバイスのアクセサリ及び周辺機器を含むそれらの任意の組み合わせを含む、任意の種類 of 移動端末、固定端末、又は携帯端末である。さらに、コンピューティングデバイス100は、ユーザのための任意のタイプのインターフェース（「ウェアラブル」回路など）をサポートすることができると予測され得る。

【0012】

処理ユニット110は、物理プロセッサ又は仮想プロセッサであり得、メモリ120に記憶されたプログラムに基づいて種々のプロセスを実行することができる。マルチプロセッサシステムでは、複数の処理ユニットがコンピュータ実行可能命令を並列に実行し、コンピューティングデバイス100の並列処理能力を向上させる。処理ユニット110は、また、中央処理ユニット（CPU）、マイクロプロセッサ、コントローラ及びマイクロコントローラとも呼ばれる。

【0013】

コンピューティングデバイス100は、典型的には、複数のコンピュータ記憶媒体を含み、これは、揮発性及び不揮発性媒体、並びにリムーバブル及び非リムーバブル媒体を含むが、これらに限定されない、コンピューティングデバイス100によってアクセス可能な任意の利用可能な媒体であることができる。メモリ120は、揮発性メモリ（例えば、レジスタ、キャッシュ、ランダムアクセスメモリ（RAM））、不揮発性メモリ（例えば、読み出し専用メモリ（ROM）、電氣的に消去可能なプログラマブル読み出し専用メモリ（EEPROM）、フラッシュメモリ）、又はそれらの任意の組み合わせであることができる。メモリ120は、本明細書に記載される様々な実装の機能を実行するように構成されるプログラムモジュールを有する画像処理モジュール122を含む。画像処理モジュール122は、それぞれの機能を実現するために、処理ユニット110によってアクセスされ、操作されることができる。

【0014】

記憶装置130は、任意のリムーバブル又は非リムーバブル媒体であることができ、情報及び/又はデータを記憶するために使用され、コンピューティングデバイス100でアクセスされることができる機械可読媒体を含み得る。コンピューティングデバイス100はさらに、追加のリムーバブル/非リムーバブル、揮発性/不揮発性メモリ媒体を含み得る。図1には示されていないが、リムーバブル且つ不揮発性ディスクを読み書きするためのディスクドライブが提供され、リムーバブル不揮発性ディスクを読み書きするためのディスクドライブが提供されている。このような場合、各ドライブは、1つ又は複数のデータ媒体インターフェースを介してバス（図示せず）に接続される。

【0015】

通信ユニット140は、通信媒体を介してさらなるコンピューティングデバイスと通信する。加えて、コンピューティングデバイス100内のコンポーネントの機能は、通信のために通信可能に接続された単一のコンピューティングクラスタ又は複数のコンピューティングマシンによって実装されることができる。したがって、コンピューティングデバイス100は、1つ又は複数の他のサーバ、ネットワークパーソナルコンピュータ（PC）又は他の一般的なネットワークノードとの論理リンクを使用して、ネットワーク環境で動作させることができる。

【0016】

入力装置150は、マウス、キーボード、トラッキングボール、音声入力装置等の1つ又は複数の入力装置を含み得る。出力装置160は、ディスプレイ、ラウドスピーカ、プリンタなどの1つ又は複数の出力装置を含み得る。必要に応じて、コンピューティングデバイス100はまた、通信ユニット140を介して、記憶装置、表示装置などの1つ又は

10

20

30

40

50

複数の外部装置（図示せず）、ユーザがコンピューティングデバイス 100 と対話することを可能にする 1 つ又は複数のデバイス、又はコンピューティングデバイス 100 を 1 つ又は複数の他のコンピューティングデバイスと通信することを可能にする任意のデバイス（例えば、ネットワークカード、モデムなど）と通信することができる。このような通信は、入出力（I/O）インターフェース（図示せず）を介して行われる。

【0017】

コンピューティングデバイス 100 は、本明細書に記載される主題の複数の実装において照明推定を実行するために使用されることができる。したがって、以下、コンピューティングデバイス 100 を「画像処理装置 100」と呼ぶこともある。照明推定を行うとき、画像処理装置 100 は、実際に、入力装置 150 を介して、実オブジェクト（例えば、コーヒーカップ、マジックキューブ等）に関する入力画像 171 及び入力画像 171 に対応するデプスマップ（depth map）172 を受信することができる。例えば、デプスマップ 172 は、デプスカメラによってキャプチャされることができる。次に、画像処理装置 100 は、入力画像 171 及びデプスマップ 172 に基づいて複数の訓練されたニューラルネットワークを用いて、入力画像 171 内のシーンの照明条件を決定し得る。例えば、照明条件は、特定の形状の画像（図示せず）によって表され得、その画像はシーン内の光源を表す。画像は、光源の複数の部分に対応する複数のピクセルを含み得、複数のピクセルのそれぞれのピクセル値は、光源の複数の部分のそれぞれの放射照度強度（irradiance intensities）を表す。さらに、画像処理装置 100 は、シーンの決定された照明条件に基づいて、入力画像 171 のシーンに仮想オブジェクト（例えば、ハローキティドール）を合成し、出力画像 173 を生成することができる。出力画像 173 内の仮想オブジェクトは、実オブジェクトと一致する現実的な効果を示すことができる。画像処理装置 100 は、さらに、出力装置 160 を介して画像 173 を出力し得る。

【0018】

いくつかのレガシーの解決策はまた、シーン内の照明条件を実際のシーンに関する画像から復元しようとする。しかし、これらの解決策は、典型的には、いくつかの又は全ての環境情報を含むリモートビュー画像に適用可能であるが、複数のアプリケーションにおける極めて一般的なクローズビュー画像には適さない（すなわち、実オブジェクトのみが画像内に提示され、一方、オブジェクトの照明効果につながる環境情報はほとんど見えない）。クローズビュー画像に対して、いくつかの他のレガシーの解決策は、実オブジェクトのシェーディング情報から実際のシーンの照明条件を取り出す（recover）ことを提案する。しかし、これらの解決策は、典型的には、物体の材料又は物体表面の反射が既知であることを必要とする。これらの解決策は、任意の材料又は表面反射を有するオブジェクトに関する画像に基づいて、実際のシーンの照明条件を取り出すことができない。

【0019】

既存の照明推定解決策におけるいくつかの問題を上述した。本明細書に記載される主題の実装によれば、上記の問題及び 1 つ又は複数の他の潜在的な問題を解決するための照明推定の解決策が提供される。この解決策は、入力画像のシェーディング及び影情報に基づく複数レベルのニューラルネットワークを用いて、シーン内の照明条件の漸進的推定を可能にする。従って、この解決策は、シーン中の実オブジェクトの材料特性に関するいかなる仮定もすることなしに、リモートビュー画像とクローズビュー画像の両方に適用可能である。シーン中の光源を基準の光ベースのセットの線形結合でモデル化し、ニューラルネットワークの入力として基準の光ベースのセットからレンダリングされる放射照度強度マップのセットを利用することにより、この解決策は照明推定の複雑さを著しく単純化し、結果の精度を高めることができる。加えて、マルチツリーベースの漸進的推定を利用することにより、この解決策はニューラルネットワークの各レベルの訓練及び計算コストを低減し、そのロバスト性を改善することができる。この解決策による光推定の結果を用いてレンダリングされる仮想オブジェクトは、実オブジェクトと一致するシェーディング及び影の効果を示すことができる。

【0020】

10

20

30

40

50

この解決策の様々な例が、図面を参照して、以下にさらに詳細に説明される。

【 0 0 2 1 】

システムアーキテクチャ

【 0 0 2 2 】

図 2 は、本明細書に記載される主題のいくつかの実装による照明推定のためのシステム 2 0 0 のブロック図を示す。いくつかの実装では、システム 2 0 0 は、図 1 のコンピューティングデバイス 1 0 0 の画像処理モジュール 1 2 2 の少なくとも一部として実装することができ、すなわち、コンピュータプログラムモジュールとして実装することができる。代替的には、他の実装では、システム 2 0 0 は、ハードウェアデバイスによって完全に又は部分的に実装することができる。図 2 に示すように、システム 2 0 0 は、概して、シーン決定サブシステム 2 1 0 及び照明決定サブシステム 2 2 0 を含み得る。システム 2 0 0 の構造及び機能性は、本明細書に記載される主題の範囲に対するいかなる限定も示唆することなく、説明の目的のためにだけ本明細書に記載されることが理解されるべきである。本明細書に記載される主題の実装は、異なる構造及び/又は機能性で具体化することができる。

10

【 0 0 2 3 】

図 2 に示すように、システム 2 0 0 の入力は、平坦な表面上に配置された実オブジェクトに関する入力画像 2 0 1 (例えば、入力画像 2 0 1 は、図 1 に示すように入力画像 1 7 1 であることができる)と、入力画像 2 0 1 に対応するデプスマップ 2 0 2 (例えば、デプスマップ 2 0 2 は、図 1 に示すようにデプスマップ 1 7 2 であることができる)であることができる。入力画像 2 0 1 は、例えば、任意のサイズ及び/又はフォーマットであり得る。いくつかの実装では、入力画像 2 0 1 は、共通の R G B 3 チャンネル画像であり得る。デプスマップ 2 0 2 は、デプスカメラによってキャプチャされ得、各ピクセルの値は、デプスカメラと、ピクセルに対応する実際のシーン内のポイントとの間の実際の距離を反映する。さらに、デプスマップ 2 0 2 は、入力画像 2 0 1 と整列される (aligned with) ことができる。すなわち、デプスマップ 2 0 2 内のピクセルは、入力画像 2 0 1 内のピクセルと 1 対 1 で対応する。幾つかの実装では、デプスマップ 2 0 2 が実際のシーンの大まかな幾何学的構造を表すことができる限り、デプスマップ 2 0 2 内のピクセルの値によって示される奥行き (depths) は正確ではないことがある。

20

【 0 0 2 4 】

シーン決定サブシステム 2 1 0 は、デプスマップ 2 0 2 に基づいて入力画像 2 0 1 内の実際のシーン 2 2 0 を決定することができる。

30

【 0 0 2 5 】

いくつかの実装では、入力画像 2 0 1 内の実際のシーン 2 2 0 を決定するために、シーン決定サブシステム 2 1 0 は、最初に、実際のシーンを記述する座標系を決定し得る。例えば、シーン決定サブシステム 2 1 0 は、実オブジェクトの配置のための平坦な表面を、実際のシーンの座標系における X Y 平面として定義し得、一方 X Y 平面の法線は、実際のシーンの座標系における Z 軸としての役割を果たす。

【 0 0 2 6 】

いくつかの実装では、入力画像 2 0 1 内の実際のシーンを決定するために、シーン決定サブシステム 2 1 0 は、まず、デプスマップ 2 0 2 を座標系内の点群 (point cloud) に変換し得る。次に、シーン決定サブシステム 2 1 0 は、点群内の実オブジェクトの配置のための平坦な表面を表す複数の点を面に適合させ得る。いくつかの実装では、複数の平面を含むシーンについて、シーン決定サブシステム 2 1 0 は、実オブジェクトの配置のための面として、複数の平面のうち最大のものを使用し得る。点群内の実オブジェクトの配置のための平坦な表面の複数の点を面に適合させた後、シーン決定サブシステム 2 1 0 は、平坦な表面を表す複数の点を除去し得る。実オブジェクトの配置のための平坦な表面を表す複数の点を面に適合させた後、シーン決定サブシステム 2 1 0 はさらに、適合した面の上方に位置する点群内の複数の点をさらに 3 次元 (3 D) メッシュに再構成し得る。例えば、シーン決定サブシステム 2 1 0 は、ポアソン表面再構成を介して、適合した面の上方

40

50

に位置する点群内の複数の点を3次元(3D)メッシュに再構成し得る。適合した面及び再構築3Dメッシュは、入力画像201内のシーン220と一緒に表すことができる。

【0027】

シーン決定サブシステム210によって決定されたシーン220及び入力画像201は、シーンの照明推定のために照明条件決定サブシステム230に入力することができる。いくつかの実装では、シーン220内の照明条件を推定するために、照明条件決定サブシステム230は、シーン220内の環境光源をモデル化し得る。

【0028】

いくつかの実装では、照明条件決定サブシステム230は、座標系の原点を中心とし、上述した座標系の軸と整列した立方体を使用して、シーン220内の環境光源をモデル化し得る。立方体を展開することにより十字形状の画像を得ることができる。したがって、シーン220内の環境光源は、十字形画像の画像によって表すことができる。いくつかの他の実装では、照明条件決定サブシステム230は、球形又は他の形状などの3次元形状を使用して、シーン220内の環境光源をモデル化し得る。これに対応して、シーン220内の環境光源は、球形又は他の形状など、3次元形状に対応する他の形状の画像によって表すことができる。説明を容易にするために、この文脈では、十字形画像(例えば、図2の十字形画像240によって示されるように)は、シーン220における環境光源の表現の例として役立つ。これは、本明細書に記載された主題の範囲に対するいかなる限定も示唆することなく、説明の目的のためにのみ提供されることが理解されるべきである。本明細書に記載される主題は、他の表現を使用してシーン内の環境光源をモデル化することができる。

【0029】

いくつかの実装では、光源を表す十字形画像240は、シーン220内の環境光源の複数の部分にそれぞれ対応する複数のピクセルを含み得る。例えば、複数のピクセルのそれぞれのピクセル値は、シーン220内の環境光源の複数の部分のそれぞれの放射照度強度を表し得る。この文脈では、シーン220内の環境光源を表す画像240内の各ピクセルを「基準の光ベース」、シーン220内の環境光源を表す画像240を「環境マップ」と呼ぶ。いくつかの実装では、照明条件決定サブシステム230は、シーン220及び入力画像201に基づいて、環境マップ240内の複数のピクセルの各ピクセル値を決定することができる、それによってシーン220内の照明条件を決定する。

【0030】

照明条件決定サブシステム230の構成及び動作原理を以下に図3を参照して詳細に説明する。

【0031】

動作原理

【0032】

図3は、本明細書に記載される主題のいくつかの実装による、照明条件決定サブシステム230のブロック図を示す。図3に示すように、照明条件決定サブシステム230は、概して、放射照度マップ生成モジュール320、第1のニューラルネットワーク340、放射照度マップ生成モジュール360及び第2のニューラルネットワーク350を含み得る。照明条件決定サブシステム230の構造及び機能性は、本明細書に記載される主題の範囲に対するいかなる限定も示唆することなく、説明の目的のためにのみ提供されることが理解されるべきである。本明細書に記載される主題の実装は、異なる構造及び/又は機能性で具体化することができる。

【0033】

上述したように、照明条件決定サブシステム230は、シーン220内の環境光源を特定の形状の画像として表現することができる。この文脈では、環境光源を表す例として、解像度4×4×3ピクセルの十字形画像が使用される。しかし、これは、本明細書に記載される主題の範囲に対するいかなる限定も示唆することなく、一例としてのみ提供されることが理解されるべきである。いくつかの実装では、照明条件決定サブシステム230は

、シーン 2 2 0 内の環境光源を所定の値で表す十字形画像（すなわち、初期環境マップ 3 1 0）内の $4 \times 4 \times 3$ ピクセルの値を初期化し得る。例えば、初期環境マップ 3 1 0 内のピクセルの各々の値は、正規化された放射照度強度「1」で初期化することができる。

【0034】

放射照度マップ生成

【0035】

いくつかの実装では、放射照度マップ生成モジュール 3 2 0 は、シーン 2 2 0 に基づいて、シーン 2 2 0 内の光源によって引き起こされる入力画像 2 0 1 内の実オブジェクトのシェーディング及び影情報を決定し得る。具体的には、放射照度マップ生成モジュール 3 2 0 は、シーン 2 2 0 及び初期環境マップ 3 1 0 に基づいて、初期環境マップ 3 1 0 内の複数のピクセルに対応する第 1 の放射照度マップのセット 3 3 0 を生成し得る。本明細書に記載される「放射照度マップ」は、入力画像内のピクセルの各々に対する光源の放射束（radiant flux）を記録し、これは、光源によって引き起こされる入力画像のシェーディング及び影情報を反映することができる。例えば、放射照度マップの第 1 のセットの各々は、入力画像 2 0 1 と同じサイズを有し得、光源のそれぞれの部分によって引き起こされる入力画像 2 0 1 内の実オブジェクトのシェーディング及び影情報を表し得る。

10

【0036】

いくつかの実装では、光源上のある点 L （例えば、光源を表す十字形画像内のピクセル）が与えられると、それに対応する放射照度マップが、入力画像 2 0 1 中のピクセルの各々に対する点 L の放射束を記録することができる。例えば、入力画像 2 0 1 のピクセル x について、光源上の点 L によって生じる放射束は、以下のように表すことができる：

20

【数 1】

$$r(x) = V(x, l)L(l)(N(x) \cdot l) \quad (1)$$

ここで、 $N(x)$ は、ピクセル x における垂直方向を表し、 l は、光源上の点 L からピクセル x までの照明方向を表し、 $L(l)$ は、方向 l に沿った放射照度強度を表し、 $V(x, l)$ は、方向 l に沿った光源上の点 L までのピクセル x の視認度を表す。同様に、所与の光源上の複数の点の組み合わせについては、複数の点の組み合わせによって生じる放射束は、式 (1) を積分することによって、決定されることができる。例えば、入力画像 2 0 1 のピクセル x について、光源上の複数の点（例えば、「+」と表される）の組み合わせを生じさせた放射束は、以下のように表すことができる：

30

【数 2】

$$R(x) = \int_{\Omega+} V(x, l)L(l)(N(x) \cdot l) dl \quad (2)$$

【0037】

式 (1) に基づいて、放射照度マップ生成モジュール 3 2 0 は、初期環境マップ 3 1 0 内のピクセル（すなわち、シーン 2 2 0 内の環境光源の点）に対して、それに対応する放射照度マップを生成することができる。この方法では、放射照度マップ生成モジュール 3 2 0 は、初期環境マップ 3 1 0 内の複数のピクセルと 1 対 1 の対応を有する第 1 の放射照度マップのセット 3 3 0 を生成することができる。

40

【0038】

初期化ネットワーク

【0039】

図 3 に示すように、第 1 の放射照度マップのセット 3 3 0 及び入力画像 2 0 1 は、第 1 の訓練されたニューラルネットワーク 3 4 0（本明細書では「初期化ネットワーク」とも

50

呼ばれる)に入力されることができる。例えば、第1のニューラルネットワーク340は、初期環境マップ310内の複数のピクセルのそれぞれのピクセル値を決定し、したがって環境マップ310に対応する環境マップ350を生成するように訓練することができる畳み込みニューラルネットワーク(CNN)であり得る。照明推定は、ここに記載される主題の実装が、シーン中の実オブジェクトの材料特性に関するいかなる仮定も必要とせずに、リモートビュー画像及びクローズビュー画像の両方に適用可能であるように、入力画像内のシェーディング及び影情報を表す放射照度マップをニューラルネットワークに提供することによって実行される。CNNは、本明細書において第1のニューラルネットワーク340の一例として提供されるが、本明細書に記載される主題の実装は、他のタイプのニューラルネットワークにも適用可能であり、本明細書に記載される主題の範囲は、この態様に限定されないことを理解されたい。

10

【0040】

図4は、本明細書に記載される主題のいくつかの実装による照明推定のための初期化ネットワーク340の動作図である。図4は、入力画像201と、初期環境ネットワーク310に基づいて生成される第1の放射照度マップのセット330とを示す。図4に示すように、初期化ネットワーク340は、入力画像201及び第1の放射照度マップのセット330に基づいて初期化環境310内の複数のピクセルのそれぞれのピクセル値を決定することができる。したがって、環境マップ310に対応する環境マップ350を生成することができる。

【0041】

いくつかの実装では、初期化ネットワーク340内部で、入力画像201は3層エンコーダによって処理され、各放射照度マップ330は別の2層エンコーダによって別々に処理され得る。各放射照度マップ330に対するエンコーダは、同じ構造及び重みを共有することができる。エンコーダの各層は、畳み込み層、バッチ正規化層及び励起層(例えば、活性化関数として正規化線形ユニット(ReLU)を使用する)を含み得、次いで、最大プーリング層によって次のスケールにダウンサンプリングされ得る。エンコーダを通過した後、各放射照度マップ330及び入力画像201から抽出された特徴マップ(feature maps)は、連結され、次いで、別の3層のエンコーダによって処理される。最後に、初期環境マップ310内の複数のピクセルのそれぞれのピクセル値は、3層エンコーダが後に続く完全に接続された(FC)層から得ることができる。これらのピクセル値は、環境マップ310に対応する環境マップ350に編成されることができる。

20

30

【0042】

幾つかの実装では、構築された訓練データセットは初期化ネットワーク340を訓練するために使用されることができる。例えば、3Dオブジェクト及び環境マップのグループからランダムに選択された3Dオブジェクト及び環境マップは、画像をレンダリングし生成するために使用されることができる。次に、レンダリング画像及びレンダリング中に適用される環境マップは、初期化ネットワーク340の訓練データセットに収集されることができる。初期化ネットワーク340の訓練手順の間、例えば、L2損失関数が、訓練データと予測結果との間の差を測定するために使用されることができる。

【0043】

図3に戻ると、環境マップ350を生成することによって、照明条件決定サブシステム230は、シーン220内の照明条件を予備的に決定することができる。しかし、シーン内のオブジェクトの形状及びレイアウトが多様であるため、シーン内の照明条件と入力画像201及びデプスマップ202との間のマッピングは、非線形であり得る。この問題を解決するために、照明条件決定サブシステム230は、シーン220内の照明条件をより正確に決定するために、決定された環境マップ350内のピクセルをさらにリファインする(refine)ことができる。

40

【0044】

リファインメントネットワーク

【0045】

50

いくつかの実装では、照明条件決定サブシステム 230 は、環境マップ 350 内の各ピクセルをリファインし (refine) てサブピクセルにし、複数のリファインされたサブピクセルのそれぞれの値を決定することによって、シーン 220 内の照明条件をより正確に推定することができる。例えば、いくつかの実装では、照明条件決定サブシステム 230 は、4 分木 (quadtree) に基づいて環境マップ 350 内のピクセルをリファインすることができる。すなわち、照明条件決定サブシステム 230 は、環境マップ 350 内の各ピクセルを 4 つのサブピクセルにリファインし、4 つのサブピクセルの値を対応して決定することができる。いくつかの他の実装では、照明条件決定サブシステム 230 はまた、環境マップ 350 内のピクセルを他の方法で分割することもできる。例えば、各ピクセルは、9 つのサブピクセル、16 のサブピクセル等に分割することができる。

10

【0046】

図 3 に示すように、いくつかの実装において、シーン 220 内の照明条件をより正確に決定するために、環境マップ 350 及びシーン 220 は、放射照度マップ生成モジュール 360 に入力され得る。いくつかの実装では、環境マップ 350 内のピクセル (例えば、ゼロではない放射照度強度を有するピクセル) に対して、放射照度マップ生成モジュール 360 は、ピクセルに関連する第 2 の放射照度マップのセット 370 を生成し得る。第 2 の放射照度マップのセット 370 及び入力画像 201 は、第 2 の訓練されたニューラルネットワーク 380 (これは、本明細書では「リファインメントネットワーク」とも呼ばれる) に入力することができる。例えば、第 2 のニューラルネットワーク 380 は、畳み込みニューラルネットワーク (CNN) であり得、これは、環境マップ 350 内のピクセルが第 2 の放射照度マップのセット 370 及び入力画像 201 に基づいて複数のサブピクセルに分割されるべきかどうかを決定し、複数のサブピクセルのそれぞれの値を決定するように訓練されることができる。この方法では、照明条件決定サブシステム 230 は、環境マップ 350 に対応するリファインされた環境マップ 240 を生成することができる。CNN は、本明細書において第 2 のニューラルネットワーク 380 の一例として提供されるが、本明細書に記載される主題の実装は、他のタイプのニューラルネットワークにも適用可能であり、本明細書に記載される主題の範囲は、この態様に限定されないことが理解されるべきである。

20

【0047】

L^2 は解像度 $4 \times 4 \times 3$ の環境マップ 350 を表すために使用され、 L^3 は、4 分木に基づいて環境マップ 350 をリファインすることによって得られる解像度 $8 \times 8 \times 3$ の環境マップ 240 (例えば、解像度 $8 \times 8 \times 3$ の環境マップ) を表すことが仮定される。環境マップ 350 内の u^{th} (第 u) 列及び v^{th} (第 v) 行に位置するピクセルが $L^2(u, v)$ として表わされると仮定すると、環境マップ 240 内のピクセルに対応する 4 つのサブピクセルは、 $L^3(2u, 2v)$ 、 $L^3(2u+1, 2v)$ 、 $L^3(2u, 2v+1)$ 及び $L^3(2u+1, 2v+1)$ として表されることができる。

30

【0048】

いくつかの実装では、シーン 220 内の照明条件を可能な限り正確に決定するために、環境マップ 350 内のピクセル $L^2(u, v)$ に対して、放射照度マップ生成モジュール 360 は、リファインメントネットワーク 380 の入力として、次のより細かいレベルで、環境マップ 240 内のすべてのピクセルについてそれぞれの放射照度マップを生成し得る。しかし、この方法は、放射照度マップの数の著しい増加、及び、リファインメントネットワーク 380 に関連する計算及び記憶コストの著しい成長をもたらす。

40

【0049】

この問題を解決するために、いくつかの実装では、環境マップ 350 内のピクセル $L^2(u, v)$ に対して、放射照度マップ生成モジュール 360 は、リファインメントネットワーク 380 の入力として所定の数の放射照度マップを生成し得る。例えば、環境マップ 350 内のピクセル $L^2(u, v)$ に対して、放射照度マップ生成モジュール 360 は、ピクセル $L^2(u, v)$ に関連する 13 の放射照度マップを生成し得る。例えば、13 の放射照度マップは、ピクセル $L^2(u, v)$ から分割された 4 つのサブピクセル $L^3(2$

50

$u, 2v$)、 $L^3(2u+1, 2v)$ 、 $L^3(2u, 2v+1)$ 及び $L^3(2u+1, 2v+1)$) に対応する 4 つの放射照度マップ；ピクセル $L^2(u, v)$ に隣接する 8 つのピクセル（すなわち、 $L^2(u-1, v-1)$ 、 $L^2(u-1, v)$ 、 $L^2(u-1, v+1)$ 、 $L^2(u, v-1)$ 、 $L^2(u, v+1)$ 、 $L^2(u+1, v-1)$ 、 $L^2(u+1, v)$ 及び $L^2(u+1, v+1)$) に対応する 8 つの放射照度マップ；及び環境マップ 350 内の $L^2(u, v)$ 及び隣接する 8 つのピクセル以外の残りのピクセルの組み合わせに対応する放射照度マップを含み得る。

【0050】

第 2 の放射照度マップのセット 370 は、上述の式 (1) 又は (2) に基づいて生成され得る。具体的には、放射照度マップ生成モジュール 360 は、上述した式 (1) に基づいて、ピクセル $L^2(u, v)$ からそれぞれ分割された 4 つのサブピクセル $L^3(2u, 2v)$ 、 $L^3(2u+1, 2v)$ 、 $L^3(2u, 2v+1)$ 、及び $L^3(2u+1, 2v+1)$) に対応する 4 つの放射照度マップを生成し得る。同様に、放射照度マップ生成モジュール 360 は、上記の式 (1) に基づいて、ピクセル $L^2(u, v)$ に隣接する 8 つのピクセル（すなわち、 $L^2(u-1, v-1)$ 、 $L^2(u-1, v)$ 、 $L^2(u-1, v+1)$ 、 $L^2(u, v-1)$ 、 $L^2(u, v+1)$ 、 $L^2(u+1, v-1)$ 、 $L^2(u+1, v)$ 及び $L^2(u+1, v+1)$) にそれぞれ対応する 8 つの放射照度マップを生成し得る。加えて、放射照度マップ生成モジュール 360 は、上記式 (2) に基づいて、ピクセル $L^2(u, v)$ 及び環境マップ 350 内の隣接する 8 ピクセル以外の残りのピクセルの組み合わせに対応する放射照度マップを生成し得る。上述の 13 の放射照度マップは、リファインされた環境マップ 240 を生成するために使用されるように、リファインメントネットワーク 380 の入力として作用することができる。

【0051】

図 5 は、本明細書に記載される主題のいくつかの実装による照明推定のためのリファインメントネットワーク 380 の動作図を示す。図 5 は、入力画像 201 及び環境マップ 350 内のピクセル 540（本明細書では「第 1 のピクセル」とも呼ばれる）から生成された第 2 の放射照度マップのセット 370 を示す。図 5 に示すように、いくつかの実装では、図 3 に示すように、放射照度マップ生成モジュール 360 によって生成された第 2 の放射照度マップのセット 370 は、ピクセル 540 から分割された 4 つのサブピクセルに対応する 4 つの放射照度マップ 510 - 1 . . . 510 - 4、ピクセル 540 に隣接する 8 つのピクセルに対応する 8 つの放射照度マップ 520 - 1 520 - 8、及び環境 350 内のピクセル 540 及び 8 つの隣接するピクセル以外の残りのピクセルの組み合わせに対応する放射照度マップ 530 を含み得る。リファインメントネットワーク 380 は、第 2 の放射照度マップのセット 370 及び入力画像 201 に基づいて、ピクセル 540 が分割されるべきであるかどうかを決定し得る。リファインメントネットワーク 380 が、ピクセル 540 が分割されるべきであると決定する場合、リファインメントネットワーク 380 は、ピクセル 540 から分割された 4 つのサブピクセルのそれぞれの値をさらに決定し得る。このようにして、環境マップ 350 に対応するリファインされた環境マップ 240 を生成することができる。

【0052】

いくつかの実装では、リファインメントネットワーク 380 の内部構造は、初期化ネットワーク 340 の内部構造と同様であり得る。具体的には、入力画像 201 及び各放射照度マップ 370 は、異なるエンコーダによって別々に処理され得る。第 2 の放射照度マップのセット 370 及び入力画像 201 が別々のエンコーダによって処理された後、抽出された特徴を連結し、別の 3 層のエンコーダに供給することができる。最後に、出力結果（すなわち、ピクセルが分割されるかどうか、及びピクセルから分割されるサブピクセルのそれぞれの値）は、3 層エンコーダに続く完全接続 (FC) 層から得ることができる。特に、リファインメントネットワーク 380 では、4 つの放射照度マップ 510 - 1 510 - 4 に対する 4 つのエンコーダが重みを共有することができる。同様に、8 つの放射照度マップ 520 - 1 520 - 8 に対する 8 つのエンコーダが重みを共

10

20

30

40

50

有することができる。ピクセルが分割されるべきかどうかの決定に関して、リファインメントネットワーク380は、それが分割されるべき確率及びそれが分割されるべきでない確率を出力することができる。次いで、ソフトマックス層が、それらを最終決定に変換し得る。4つのサブピクセルの値に対して、それらの放射照度強度の間により強いコントラストが存在することがある。幾つかの実装では、出力値の範囲は、複数の間隔に一様に分割することができ、各サブピクセルの値は、離散化された多項分類で出力することができる。

【0053】

いくつかの実装では、構築されたデータセットを使用して、リファインメントネットワーク380を訓練することができる。例えば、画像は、3Dオブジェクトと、3Dオブジェクトのセット及び照明条件からランダムに選択された照明条件を使用してレンダリングすることができる。ランダムに選択された各照明条件に対して、予め定義されたポリシーを使用して、その最適な十字形画像表現を決定することができる。その後、リファインメントネットワーク380は、入力画像から最適な十字形画像表現へのマッピングを近似するように訓練される。いくつかの実装では、例えば、クロスエントロピー損失関数を使用して、分割決定及びサブピクセル値の両方に対してリファインメントネットワーク380を訓練することができる。

【0054】

図3は、2つのレベルのニューラルネットワーク（すなわち、初期化ネットワーク340及びリファインメントネットワーク380）を有する照明条件決定サブシステム230の例を示すのみであるが、本明細書に記載される主題の実装は、より少ないレベル又はより多いレベルのニューラルネットワークを使用する照明条件の推定に適用可能であることが理解されるべきである。例えば、いくつかの実装では、計算コストを低減するために、1つのレベルのニューラルネットワーク（例えば、初期化ネットワーク340）のみを使用することができる。あるいは、いくつかの他の実装では、より正確な結果を得るために、より多いレベルのリファインメントネットワーク（例えば、2つ以上のレベルのリファインメントネットワーク）を使用することができる。

【0055】

図6は、本明細書に記載される主題のいくつかの実装による、複数のレベルのニューラルネットワーク（例えば、初期化ネットワーク及び2つのレベルのリファインメントネットワーク）を使用して、シーン内の照明条件を漸進的に決定する図を示す。図6は、初期化ネットワークによって決定される環境マップ610を示す。第1のレベルのリファインメントネットワークは、環境マップ610内のピクセル601、602...605の各々に対して、ピクセルがさらに分割されるべきかどうかを決定し、分割されるべきと決定されるピクセル601、603及び605の各々に対して、ピクセルから分割される4つのサブピクセルのそれぞれの値を決定する。第2のレベルのリファインメントネットワークは、第1のレベルのリファインメントネットワークから出力される環境マップ内のピクセル606、607、...611の各々に対して、ピクセルがさらに分割されるべきかどうかを決定し、分割されるべきと決定されるピクセル606、607及び610の各々に対して、ピクセルから分割される4つのサブピクセルの各々の値を決定する。このようにして、最終的な環境マップ620（例えば、16×16×3ピクセルの解像度の）を得ることができる。

【0056】

本明細書に記載された主題の実装による照明推定の解決策が、上述されている。以上の説明から、本明細書に記載される主題による照明推定の解決策は、シーン内の実オブジェクトの材料特性を仮定することなく、入力画像と入力画像のシェーディング及び影情報に基づいて、複数のレベルのニューラルネットワークを用いて、シーン内の照明条件を漸進的に推定することができることが分かる。シーン内の光源を基準の光ベースのセットの線形結合でモデル化し、ニューラルネットワークの入力として基準の光ベースのセットからレンダリングされる放射照度マップのセットを利用することにより、この解決策は照明

10

20

30

40

50

推定の複雑さを大幅に単純化し、結果の精度を改善することができる。さらに、マルチツリーベースの漸進的推定を利用することにより、この解決策は各レベルのニューラルネットワークの訓練と計算コストを低減し、そのロバスト性を改善することができる。この解決策に従って得られた照明条件を用いてレンダリングされる仮想オブジェクトは、実オブジェクトと一致する現実的な効果を示すことができる。

【0057】

例示のプロセス

【0058】

図7は、本明細書に記載される主題のいくつかの実装による照明推定のためのプロセス700のフローチャートを示す。プロセス700は、例えば、図1に示すように、画像処理モジュール122によって実現することができる。プロセス700は、図示されていない追加のブロックをさらに含むことができ、及び/又は図示されているブロックを省略することができることが理解されるべきである。本明細書に記載される主題の範囲は、この態様に限定されない。

10

【0059】

ブロック710において、画像処理モジュール122は、実オブジェクトに関する入力画像及び入力画像に対応するデプスマップを取得する。

【0060】

ブロック720では、画像処理モジュール122は、デプスマップに基づいて入力画像内のシーンを決定する。

20

【0061】

ブロック730では、画像処理モジュール122は、シーンに基づいて、シーン内の光源によって引き起こされる実オブジェクトのシェーディング及び影情報を決定する。

【0062】

ブロック740では、画像処理モジュール122は、入力画像と、シェーディング及び影情報とに基づいて、光源によって引き起こされるシーン内の照明条件を決定する。

【0063】

いくつかの実装では、プロセス700はさらに：シーン及び照明条件に基づいて、実オブジェクト及び仮想オブジェクトを有する出力画像を生成することを含み、出力画像は、入力画像のシーン内に仮想オブジェクトを合成する。

30

【0064】

いくつかの実装では、シーンは、実オブジェクトの配置のための表面を含む。シーンを決定することは：デプスマップを点群に変換することと；前記点群内の複数の点を面に適合させることと；複数の点は表面を表す、適合させることと；前記面の上に位置する点群内の複数の点を3次元メッシュに再構成することと；面及び3次元メッシュは、入力画像内のシーンを共に表す、再構成することとを含む。

【0065】

いくつかの実装では、光源は、所定の形状の画像によって表され、画像は、光源の複数の部分に対応する複数のピクセルを含み、複数のピクセルのそれぞれのピクセル値は、光源の複数の部分のそれぞれの放射照度強度を表す。いくつかの実装では、シェーディング及び影情報を決定することは：画像内の複数のピクセルのそれぞれのピクセル値を所定の値で初期化することと；シーン及び初期化された画像に基づいて、複数のピクセルに対応する第1の放射照度マップのセットを生成することとを含み、放射照度マップの各々は、光源の対応する部分によって引き起こされる実オブジェクトの対応するシェーディング及び影情報を表す。幾つかの実装では、シーン内の照明条件を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、画像内の複数のピクセルのそれぞれのピクセル値を決定することを含む。

40

【0066】

いくつかの実装では、光源を表す画像は、十字形状の画像である。

【0067】

50

幾つかの実装では、画像内の複数のピクセルのそれぞれのピクセル値を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、訓練されたニューラルネットワークを用いて画像内の複数のピクセルのそれぞれのピクセル値を決定することを含む。

【0068】

幾つかの実装では、画像内の複数のピクセルのそれぞれのピクセル値を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、複数の訓練されたニューラルネットワークを用いて画像内の複数のピクセルのそれぞれのピクセル値を漸進的に決定することを含む。

【0069】

いくつかの実装では、複数の訓練されたニューラルネットワークは、少なくとも第1のニューラルネットワーク及び第2のニューラルネットワークを含む。画像内の複数のピクセルのそれぞれのピクセル値を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、第1のニューラルネットワークを使用して、複数のピクセルの第1のピクセルの第1のピクセル値及び第1のピクセルに隣接する少なくとも1つのピクセルのピクセル値を決定することと；第1のピクセル、少なくとも1つのピクセル及びシーンに基づいて第2の放射照度マップのセットを生成することと；第2の放射照度マップのセットは、第1のピクセルの複数のサブピクセル、少なくとも1つのピクセル、並びに複数のピクセルのうちの第1のピクセル及び少なくとも1つのピクセル以外のピクセルの組み合わせそれぞれに対応する、生成することと；入力画像及び第2の放射照度マップのセットに基づいて、第1のピクセルが第2のニューラルネットワークを使用して分割されるべきかどうかを決定することと；第1のピクセル値が分割されるべきでないことを決定することに応答して、第1のピクセル値を第1のピクセルの最終ピクセルとして使用することと；を含む。

【0070】

幾つかの実装では、画像内の複数のピクセルのそれぞれのピクセル値を決定することはさらに：第1のピクセルが分割されるべきであるという決定に応答して、第1のピクセルを複数のサブピクセルに分割することと、複数のサブピクセルのそれぞれのサブピクセル値を決定することとを含む。

【0071】

例示の実装

【0072】

本明細書に記載される主題のいくつかの例示的な例示の実装が、以下に挙げられる。

【0073】

第1の態様では、本明細書に記載される主題は：処理ユニットと；処理ユニットに結合され、処理ユニットによる実行のための命令を記憶するメモリと；を有する電子デバイスを提供する。命令は、処理ユニットによって実行されるとき、デバイスに：実オブジェクトについての入力画像及び入力画像に対応するデプスマップを取得することと；デプスマップに基づいて入力画像内のシーンを決定することと；シーンに基づいて、シーン内の光源によって引き起こされる実オブジェクトのシェーディング及び影情報を決定することと；入力画像と、シェーディング及び影情報とに基づいて光源によって引き起こされるシーン内の照明条件を決定することと；を含む動作を実行させる。

【0074】

いくつかの実装では、動作はさらに：シーン及び照明条件に基づいて、実オブジェクト及び仮想オブジェクトを含む出力画像を生成することを含み、出力画像は入力画像のシーン内に仮想オブジェクトを合成する。

【0075】

いくつかの実装では、シーンは、実オブジェクトの配置のための表面を含む。シーンを決定することは：デプスマップを点群に変換することと；点群内の複数の点を面に適合させることと；複数の点は表面を表す、適合させることと；面の上に位置する点群内の複数の点を3次元メッシュに再構成することと；面及び3次元メッシュは、入力

10

20

30

40

50

画像内のシーンを共に表す、再構成することと；を含む。

【 0 0 7 6 】

いくつかの実装では、光源は、所定の形状の画像によって表され、画像は、光源の複数の部分に対応する複数のピクセルを含み、複数のピクセルのそれぞれのピクセル値は、光源の複数の部分のそれぞれの放射照度強度を表す。いくつかの実装では、シェーディング及び影情報を決定することは：画像内の複数のピクセルのそれぞれのピクセル値を所定の値で初期化することと；シーン及び初期化された画像に基づいて、複数のピクセルに対応する第1の放射照度マップのセットを生成することとを含み、放射照度マップの各々は、光源の対応する部分によって引き起こされる実オブジェクトの対応するシェーディング及び影情報を表す。いくつかの実装では、シーン内の照明条件を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、画像内の複数のピクセルのそれぞれのピクセル値を決定することを含む。

10

【 0 0 7 7 】

いくつかの実装では、光源を表す画像は、十字形状の画像である。

【 0 0 7 8 】

いくつかの実装では、画像内の複数のピクセルのそれぞれのピクセル値を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、訓練されたニューラルネットワークを用いて画像内の複数のピクセルのそれぞれのピクセル値を決定することを含む。

【 0 0 7 9 】

いくつかの実装では、画像内の複数のピクセルのそれぞれのピクセル値を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、複数の訓練されたニューラルネットワークを用いて画像内の複数のピクセルのそれぞれのピクセル値を漸進的に決定することを含む。

20

【 0 0 8 0 】

いくつかの実装では、複数の訓練されたニューラルネットワークは、少なくとも第1のニューラルネットワーク及び第2のニューラルネットワークを含む。画像内の複数のピクセルのそれぞれのピクセル値を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、第1のニューラルネットワークを使用して、複数のピクセルの第1のピクセルの第1のピクセル値及び第1のピクセルに隣接する少なくとも1つのピクセルのピクセル値を決定することと；第1のピクセル、少なくとも1つのピクセル及びシーンに基づいて第2の放射照度マップのセットを生成することとであって、第2の放射照度マップのセットは、第1のピクセルの複数のサブピクセル、少なくとも1つのピクセル、並びに複数のピクセルのうちの第1のピクセル及び少なくとも1つのピクセル以外のピクセルの組み合わせそれぞれに対応する、生成することと；入力画像及び第2の放射照度マップのセットに基づいて、第1のピクセルが第2のニューラルネットワークを使用して分割されるべきかどうかを決定することと；第1のピクセル値が分割されるべきでないことを決定することに応答して、第1のピクセル値を第1のピクセルの最終ピクセルとして使用することと；を含む。

30

【 0 0 8 1 】

いくつかの実装では、画像内の複数のピクセルのそれぞれのピクセル値を決定することはさらに：第1のピクセルが分割されるべきであるということに決定することとに回答して、第1のピクセルを複数のサブピクセルに分割することと；複数のサブピクセルのそれぞれのサブピクセル値を決定することとを含む。

40

【 0 0 8 2 】

第2の態様では、本明細書に記載の主題は、コンピュータに実装される方法を提供する。本方法は：実オブジェクトに関する入力画像及び入力画像に対応するデプスマップを取得することと；デプスマップに基づいて入力画像内のシーンを決定することと；シーンに基づいて、シーン内の光源によって引き起こされる実オブジェクトのシェーディング及び影情報を決定することと；入力画像と、シェーディング及び影情報とに基づいて、光源によって引き起こされるシーン内の照明条件を決定することと；を含む。

50

【 0 0 8 3 】

いくつかの実装では、本方法はさらに：シーン及び照明条件に基づいて、実オブジェクト及び仮想オブジェクトを含む出力画像を生成することをさらに含み、出力画像は、入力画像のシーン内に仮想オブジェクトを合成する。

【 0 0 8 4 】

いくつかの実装では、シーンは、実オブジェクトの配置のための表面を含む。シーンを決定することは：デプスマップを点群に変換することと；点群内の複数の点を面に適合させることであって、複数の点は、表面を表す、適合させることと；面の上に位置する点群内の複数の点を3次元メッシュに再構成することであって、面及び3次元メッシュは、入力画像内のシーンを共に表す、再構成することと；を含む。

10

【 0 0 8 5 】

いくつかの実装では、光源は、所定の形状の画像によって表され、画像は、光源の複数の部分に対応する複数のピクセルを含み、複数のピクセルのそれぞれのピクセル値は、光源の複数の部分のそれぞれの放射照度強度を表す。いくつかの実装では、シェーディング及び影情報を決定することは：画像内の複数のピクセルのそれぞれのピクセル値を所定の値で初期化することと；シーン及び初期化された画像に基づいて、複数のピクセルに対応する第1の放射照度マップのセットを生成することと；を含み、放射照度マップの各々は、光源の対応する部分によって引き起こされる実オブジェクトの対応するシェーディング及び影情報を表す。いくつかの実装では、シーン内の照明条件を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、画像内の複数のピクセルのそれぞれのピクセル値を決定することを含む。

20

【 0 0 8 6 】

いくつかの実装では、光源を表す画像は、十字形状の画像である。

【 0 0 8 7 】

いくつかの実装では、画像内の複数のピクセルのそれぞれのピクセル値を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、訓練されたニューラルネットワークを用いて画像内の複数のピクセルのそれぞれのピクセル値を決定することを含む。

【 0 0 8 8 】

いくつかの実装では、画像内の複数のピクセルのそれぞれのピクセル値を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、複数の訓練されたニューラルネットワークを用いて画像内の複数のピクセルのそれぞれのピクセル値を漸進的に決定することを含む。

30

【 0 0 8 9 】

いくつかの実装では、複数の訓練されたニューラルネットワークは、少なくとも第1のニューラルネットワーク及び第2のニューラルネットワークを含む。画像内の複数のピクセルのそれぞれのピクセル値を決定することは：入力画像及び第1の放射照度マップのセットに基づいて、第1のニューラルネットワークを使用して、複数のピクセルの第1のピクセルの第1のピクセル値及び第1のピクセルに隣接する少なくとも1つのピクセルのピクセル値を決定することと；第1のピクセル、少なくとも1つのピクセル及びシーンに基づいて第2の放射照度マップのセットを生成することであって、第2の放射照度マップのセットは、第1のピクセルの複数のサブピクセル、少なくとも1つのピクセル、並びに複数のピクセルのうちの第1のピクセル及び少なくとも1つのピクセル以外のピクセルの組み合わせそれぞれに対応する、生成することと；入力画像及び第2の放射照度マップのセットに基づいて、第1のピクセルが第2のニューラルネットワークを使用して分割されるべきかどうかを決定することと；第1のピクセル値が分割されるべきでないと決定することに応答して、第1のピクセル値を第1のピクセルの最終ピクセル値として使用することと；を含む。

40

【 0 0 9 0 】

いくつかの実装では、画像内の複数のピクセルのそれぞれのピクセル値を決定することはさらに：第1のピクセルが分割されるべきであると決定することに応答して、第1のピ

50

クセルを複数のサブピクセルに分割することと；複数のサブピクセルのそれぞれのサブピクセル値を決定することと；を含む。

【0091】

第3の態様では、本明細書に記載される主題は、非一過性コンピュータ記憶媒体に実体的に記憶され、デバイスによって実行されるとき、デバイスに本明細書に記載される主題の第2の態様における方法を実施させるコンピュータ実行可能命令を含む、コンピュータプログラム製品を提供する。

【0092】

第4の態様では、本明細書に記載される主題は、その上に記憶されるコンピュータ実行可能命令を有するコンピュータ読取可能媒体を提供し、コンピュータ実行可能命令は、デバイスによって実行されるとき、デバイスに本明細書に記載される主題の第2の態様における方法を実行させる。

【0093】

本明細書に記載される機能は、少なくとも部分的に、1つ又は複数のハードウェア論理コンポーネントによって実行することができる。例えば、限定されるものではないが、使用することができるハードウェア論理コンポーネントの例示的なタイプは、フィールドプログラマブルゲートアレイ（FPGA）、特定用途向け集積回路（ASIC）、特定用途向け標準製品（ASSP）、システムオンチップシステム（SOC）、複雑なプログラマブルロジックデバイス（CPLD）などを含む。

【0094】

本明細書に記載される主題の方法を実行するためのプログラムコードは、1つ又は複数のプログラミング言語の任意の組み合わせで書かれ得る。これらのプログラムコードは、汎用コンピュータ、専用コンピュータ、又は他のプログラマブルデータ処理装置のプロセッサ又はコントローラに提供され得、その結果、プログラムコードは、プロセッサ又はコントローラによって実行されるとき、フローチャート及び/又はブロック図に指定された機能/動作を実行させる。プログラムコードは、完全にマシン上で、部分的にマシン上で、スタンドアロンソフトウェアパッケージとして、部分的にマシン上で且つ部分的にリモートマシン上で、又は完全にリモートマシン又はサーバ上で、実行し得る。

【0095】

本開示の文脈において、機械可読媒体は、命令実行システム、装置、若しくはデバイスによる使用するための又はこれらと関連して使用するためのプログラムを含み得る又は記憶し得る任意の有形媒体であり得る。機械可読媒体は、機械可読信号媒体又は機械可読記憶媒体であり得る。機械可読媒体は、電子、磁気、光学、電磁、赤外線、もしくは半導体システム、装置、もしくはデバイス、又は前述の任意の適切な組み合わせを含み得るが、これらに限定されない。機械可読媒体のより具体的な例としては、1つ又は複数のワイヤを有する電気接続、ポータブルコンピュータディスク、ハードディスク、ランダムアクセスメモリ（RAM）、読取り専用メモリ（ROM）、消去可能プログラマブル読取り専用メモリ（EPROM又はフラッシュメモリ）、光ファイバ、ポータブルコンパクトディスク読取り専用メモリ（CD-ROM）、光記憶装置、磁気記憶装置、又はこれらの適切な組み合わせが挙げられる。

【0096】

さらに、動作が特定の順序で示されているが、これは、このような動作が、図示された特定の順序で又は順番に行われること、又は、望ましい結果を達成するために、図示されたすべての動作が行われることを要求するものとして理解されるべきではない。特定の状況下では、マルチタスク及び並列処理が有利であり得る。同様に、いくつかの特定の実装の詳細が上記の議論に含まれているが、これらは、本明細書に記載される主題の範囲に対する限定としてではなく、むしろ特定の实装に固有であり得る特徴の説明として解釈されるべきである。別々の実装のコンテキストで記述される特定の特征是、単一の実装において組み合わせて実装されてもよい。逆に、単一の実装のコンテキストで記載される種々の特征是、複数の実装において別々に又は任意の適切なサブコンビネーションで実装されて

10

20

30

40

50

もよい。

【 0 0 9 7 】

主題は、構造的特徴及び／又は方法論的動作に特有の言語で記載されているが、添付の請求項に記載された主題は、必ずしも上記の特定の特徴又は動作に限定されるものではないことを理解されたい。むしろ、上述の特定の特徴及び動作は、特許請求の範囲を実施する例示的な形態として開示される。

10

20

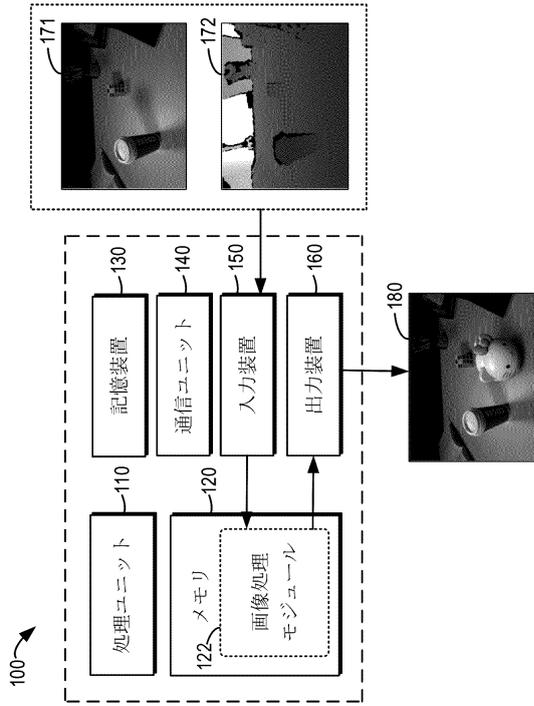
30

40

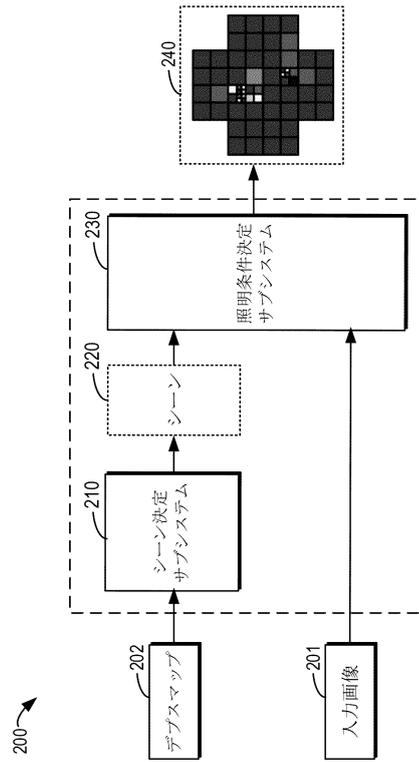
50

【図面】

【図 1】



【図 2】



10

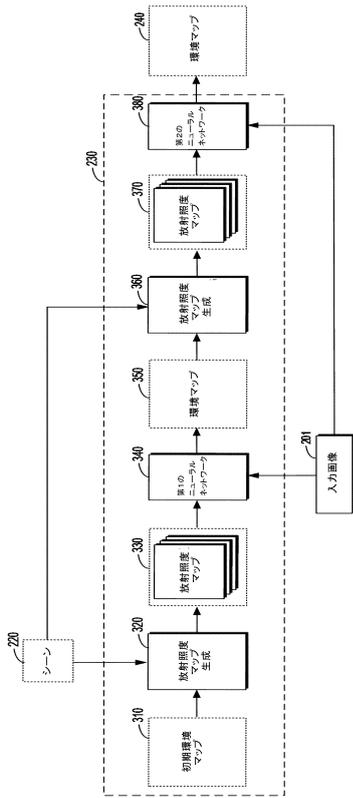
20

30

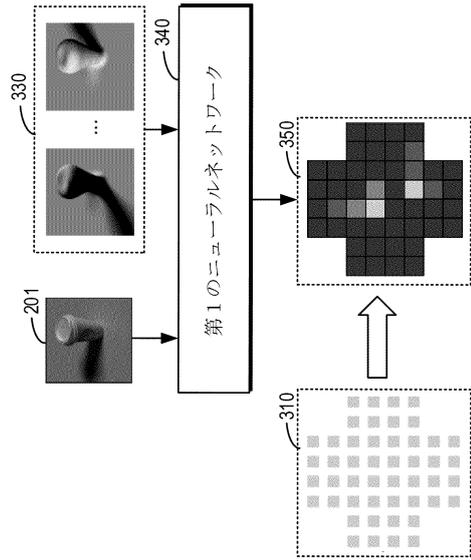
40

50

【図 3】



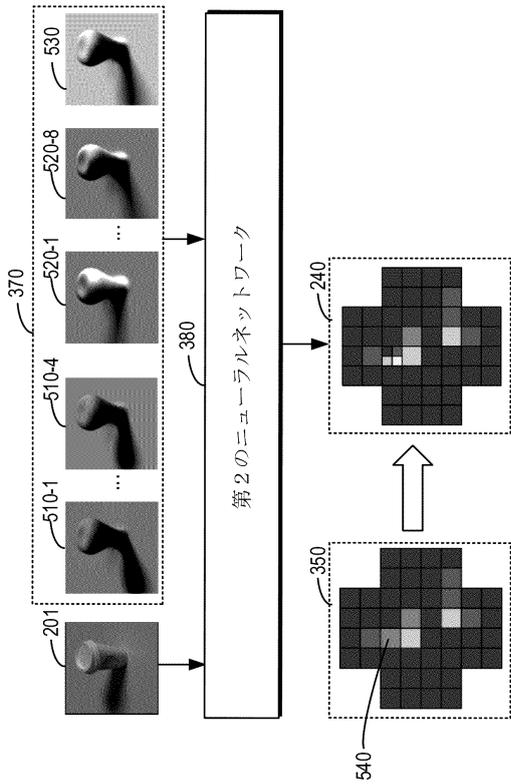
【図 4】



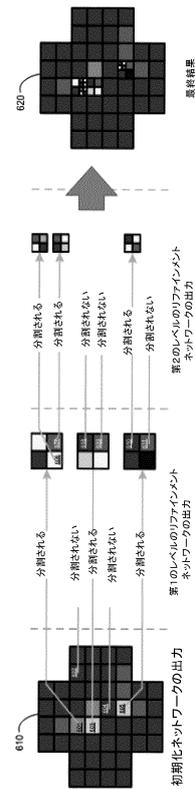
10

20

【図 5】



【図 6】

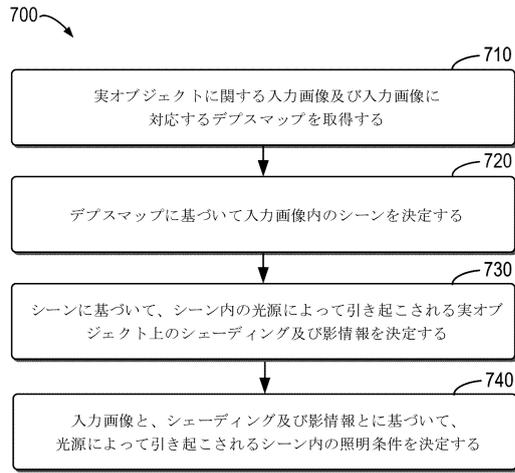


30

40

50

【 図 7 】



10

20

30

40

50

フロントページの続き

- クロソフト ウェイ マイクロソフト テクノロジー ライセンシング, エルエルシー (番地無し)
(72)発明者 チェン, グオジュン
アメリカ合衆国 ワシントン州 98052 - 6399 レッドモンド ワン マイクロソフト ウェイ
マイクロソフト テクノロジー ライセンシング, エルエルシー (番地無し)
- (72)発明者 トン, シン
アメリカ合衆国 ワシントン州 98052 - 6399 レッドモンド ワン マイクロソフト ウェイ
マイクロソフト テクノロジー ライセンシング, エルエルシー (番地無し)
- 審査官 中田 剛史
- (56)参考文献 特開2015 - 185176 (JP, A)
特開2009 - 163610 (JP, A)
特開2013 - 156670 (JP, A)
特開2014 - 164497 (JP, A)
平原 匡哲, RGB - Dカメラを用いた確率的探索による光源推定, 映像情報メディア学会
技術報告 Vol. 37 No. 36 ITE Technical Report, 日本, (一社)映像情報メデ
ィア学会 The Institute of Image Information and Television Engineers, 2013年08月12日
, 第37巻, p.9-12
- (58)調査した分野 (Int.Cl., DB名)
G06T 15/50