

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-219928

(P2019-219928A)

(43) 公開日 令和1年12月26日(2019.12.26)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06T 15/50 (2011.01)</b>	G06T 15/50	5B080
<b>G06T 7/00 (2017.01)</b>	G06T 7/00 350C	5L096

審査請求 未請求 請求項の数 8 O L (全 22 頁)

(21) 出願番号	特願2018-117026 (P2018-117026)	(71) 出願人	504171134 国立大学法人 筑波大学 茨城県つくば市天王台一丁目1番1
(22) 出願日	平成30年6月20日 (2018.6.20)	(74) 代理人	100106909 弁理士 棚井 澄雄
		(74) 代理人	100188558 弁理士 飯田 雅人
		(74) 代理人	100169764 弁理士 清水 雄一郎
		(72) 発明者	金森 由博 茨城県つくば市天王台一丁目1番1 国立 大学法人筑波大学内
		(72) 発明者	遠藤 結城 茨城県つくば市天王台一丁目1番1 国立 大学法人筑波大学内

最終頁に続く

(54) 【発明の名称】 画像処理装置、画像処理方法、及び画像処理プログラム

(57) 【要約】

【課題】 光源情報が示す光源による照明状態に基づいて、入力画像に対応する出力画像を生成する場合に、写実性を向上できる画像処理装置、画像処理方法、及びプログラムを提供することを提供すること。

【解決手段】 画像処理装置は、入力画像を取得する取得部と、光源からの照明状態を示す情報である光源情報と、入力画像に含まれる被写体の反射率として入力画像に基づいて推定された反射率情報と、入力画像に含まれる被写体からの光の伝達状態として入力画像と光源からの光が被写体に届くか否かを示す可視情報とに基づいて推定された光伝達情報とに基づいて、入力画像に対応する出力画像を生成する生成部と、生成部が生成した出力画像を出力する出力部とを備える。

【選択図】 図2

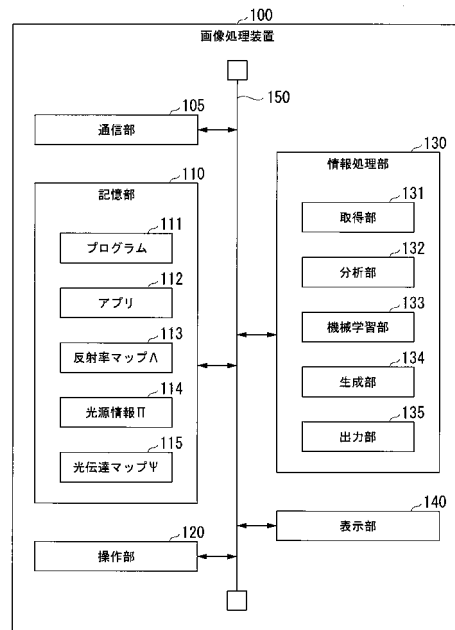


図2

**【特許請求の範囲】****【請求項 1】**

入力画像を取得する取得部と、

光源からの照明状態を示す情報である光源情報と、前記入力画像に含まれる被写体の反射率として前記入力画像に基づいて推定された反射率情報と、前記入力画像に含まれる前記被写体からの光の伝達状態として前記入力画像と前記光源からの光が前記被写体に届くか否かを示す可視情報とに基づいて推定された光伝達情報とに基づいて、前記入力画像に対応する出力画像を生成する生成部と、

前記生成部が生成した出力画像を出力する出力部と  
を備える、画像処理装置。

10

**【請求項 2】**

前記光源情報は、前記入力画像に基づいて推定されたものである、請求項 1 に記載の画像処理装置。

**【請求項 3】**

前記出力画像とは、前記光源情報が示す前記照明状態に基づいて、前記入力画像に含まれる前記被写体を再照明した再照明画像である、請求項 1 に記載の画像処理装置。

**【請求項 4】**

前記反射率情報と前記光伝達情報は、機械学習モデルによって、前記入力画像に基づいて推定されたものである、請求項 1 から請求項 3 のいずれか一項に記載の画像処理装置。

20

**【請求項 5】**

前記機械学習モデルとは、前記機械学習モデルによって、前記入力画像に基づいて推定される光源情報と、反射率情報と、光伝達情報とのいずれか又はいずれか同士の組み合わせから得られる画像と、所定の正解画像とを比較することによって学習されたものである、請求項 4 に記載の画像処理装置。

**【請求項 6】**

前記可視情報は、物体表面の複数の点の各々について、光の方向毎に、前記光が複数の前記点の各々に届くか否かに基づいて作成される、請求項 1 から請求項 5 のいずれか一項に記載の画像処理装置。

**【請求項 7】**

入力画像を取得するステップと、  
光源からの照明状態を示す情報である光源情報と、前記入力画像に含まれる被写体の反射率として前記入力画像に基づいて推定された反射率情報と、前記入力画像に含まれる前記被写体からの光の伝達状態として前記入力画像と前記光源からの光が前記被写体に届くか否かを示す可視情報とに基づいて推定された光伝達情報とに基づいて、前記入力画像に対応する出力画像を生成するステップと、

前記生成するステップで生成された出力画像を出力するステップと  
を有する、コンピュータが実行する画像処理方法。

30

**【請求項 8】**

コンピュータに、

入力画像を取得するステップと、

光源からの照明状態を示す情報である光源情報と、前記入力画像に含まれる被写体の反射率として前記入力画像に基づいて推定された反射率情報と、前記入力画像に含まれる前記被写体からの光の伝達状態として前記入力画像と前記光源からの光が前記被写体に届くか否かを示す可視情報とに基づいて推定された光伝達情報とに基づいて、前記入力画像に対応する出力画像を生成するステップと、

前記生成するステップで生成された出力画像を出力するステップと  
を実行させる、画像処理プログラム。

40

**【発明の詳細な説明】****【技術分野】**

50

## 【0001】

本発明の実施形態は、画像処理装置、画像処理方法、及び画像処理プログラムに関する。

## 【背景技術】

## 【0002】

再照明 (re-lighting) とは、被写体が撮影されたときとは異なる照明環境下での被写体の見た目を再現する技術である。人物画像の再照明が実現できれば、ポートレート写真の陰影の修正、背景画像をさしかえる切り貼り合成など、様々な応用が考えられる。

一方で、物理則に基づいて人物画像の再照明を行うには、その画像から反射率、形状および照明を推定する必要がある。これは、逆レンダリング (inverse rendering) と呼ばれる不良設定問題である。このような逆レンダリングは、主に人物の顔画像に対して、統計的な3D形状モデルや畳み込みニューラルネットワーク (convolutional neural network; CNN) を用いて実現されてきた。そして、その際の照明計算の定式化において、周囲から届く光の分布を表現するために球面調和関数 (spherical harmonics; SH) がよく用いられてきた。

10

## 【先行技術文献】

## 【非特許文献】

## 【0003】

20

【非特許文献1】S.Sengupta, A.Kanazawa, C.D.Castillo, and D.W.Jacobs, "SfSNet: Learning Shape, Reflectance and Illuminance of Faces 'in the wild'", in Conference on Computer Vision and Pattern Recognition 2018.

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0004】

SHに基づく再照明では、光の遮蔽を無視すれば、照明情報と物体表面の法線ベクトルとから、解析的な式で照明計算を行うことができる。しかし光の遮蔽を無視すると、例えば人物の脇や股など、本来なら遮蔽のために光のあまり当たらない部分が不自然に明るくなってしまふ。

30

光の遮蔽を近似的に扱う方法として、幾何的な遮蔽率をスカラー値として算出し陰影に乗算するアンビエントオクルージョンという技術もあるが、これは光源に依らず凹んだ部分を常に暗くしているため、特に再照明の際に光源を動かした場合に違和感が生じる。

## 【0005】

よりよい方法としては、リアルタイムレンダリング向けに開発された前計算放射輝度伝達 (precomputed radiance transfer; PRT) で行われたように、光の遮蔽を含めてSHの照明計算を定式化することが知られている。

しかし、PRTの枠組みを逆レンダリングに組み込もうとすると、光の遮蔽を計算するためには、推定しようとしている形状が必要になる上に、物体表面上の各点から多数のレイを飛ばして可視判定を行う必要があり、計算コストが大きい。

40

本発明は、前述した問題を解決すべくなされたもので、入力画像に対応する出力画像を生成する場合に、写実性を向上できる画像処理装置、画像処理方法、及びプログラムを提供することを目的とする。

## 【課題を解決するための手段】

## 【0006】

本発明の一実施形態は、入力画像を取得する取得部と、光源からの照明状態を示す情報である光源情報と、前記入力画像に含まれる被写体の反射率として前記入力画像に基づいて推定された反射率情報と、前記入力画像に含まれる前記被写体からの光の伝達状態として前記入力画像と前記光源からの光が前記被写体に届くか否かを示す可視情報とに基づいて推定された光伝達情報とに基づいて、前記入力画像に対応する出力画像を生成する生成

50

部と、前記生成部が生成した出力画像を出力する出力部とを備える、画像処理装置である。

本発明の一実施形態は、前述の画像処理装置において、前記光源情報とは、前記入力画像に基づいて推定されたものである。

本発明の一実施形態は、前述の画像処理装置において、前記出力画像とは、前記光源情報が示す前記照明状態に基づいて、前記入力画像に含まれる前記被写体を再照明した再照明画像である。

本発明の一実施形態は、前述の画像処理装置において、前記反射率情報と前記光伝達情報とは、機械学習モデルによって、前記入力画像に基づいて推定されたものである。

本発明の一実施形態は、前述の画像処理装置において、前記機械学習モデルとは、前記機械学習モデルによって、前記入力画像に基づいて推定される光源情報と、反射率情報と、光伝達情報とのいずれか又はいずれか同士の組み合わせから得られる画像と、所定の正解画像とを比較することによって学習されたものである。

本発明の一実施形態は、前述の画像処理装置において、前記可視情報は、物体表面の複数の点の各々について、光の方向毎に、前記光が複数の前記点の各々に届くか否かに基づいて作成される。

#### 【0007】

本発明の一実施形態は、入力画像を取得するステップと、光源からの照明状態を示す情報である光源情報と、前記入力画像に含まれる被写体の反射率として前記入力画像に基づいて推定された反射率情報と、前記入力画像に含まれる前記被写体からの光の伝達状態として前記入力画像と前記光源からの光が前記被写体に届くか否かを示す可視情報とに基づいて推定された光伝達情報とに基づいて、前記入力画像に対応する出力画像を生成するステップと、前記生成するステップで生成された出力画像を出力するステップとを有する、コンピュータが実行する画像処理方法である。

#### 【0008】

本発明の一実施形態は、コンピュータに、入力画像を取得するステップと、光源からの照明状態を示す情報である光源情報と、前記入力画像に含まれる被写体の反射率として前記入力画像に基づいて推定された反射率情報と、前記入力画像に含まれる前記被写体からの光の伝達状態として前記入力画像と前記光源からの光が前記被写体に届くか否かを示す可視情報とに基づいて推定された光伝達情報とに基づいて、前記入力画像に対応する出力画像を生成するステップと、前記生成するステップで生成された出力画像を出力するステップとを実行させる、画像処理プログラムである。

#### 【発明の効果】

#### 【0009】

本発明の実施形態によれば、入力画像に対応する出力画像を生成する場合に、写実性を向上できる画像処理装置、画像処理方法、及び画像処理プログラムを提供できる。

#### 【図面の簡単な説明】

#### 【0010】

【図1】実施形態の画像処理システムの一例を示す図である。

【図2】実施形態の画像処理装置の一例を示すブロック図である。

【図3】可視関数を導出する処理の一例を示す図である。

【図4】実施形態の画像処理装置で使用される機械学習モデルの一例を示す図である。

【図5】実施形態の画像処理装置の動作の一例を示すフローチャートである。

【図6】実施形態の画像処理装置の動作の一例を示すフローチャートである。

【図7】CG人物データセットの一例を示す図である。

【図8】実施形態の画像処理装置による推定結果と従来手法との比較例を示す図である。

【図9】実施形態の画像処理装置による推定結果の一例を示す図である。

【図10】実施形態の画像処理装置による推定結果の一例を示す図である。

【図11】実施形態の画像処理装置による推定結果の一例を示す図である。

#### 【発明を実施するための形態】

## 【0011】

次に、本実施形態の画像処理装置、画像処理方法、及び画像処理プログラムを、図面を参照しつつ説明する。以下で説明する実施形態は一例に過ぎず、本発明が適用される実施形態は、以下の実施形態に限られない。

なお、実施形態を説明するための全図において、同一の機能を有するものは同一符号を用い、繰り返しの説明は省略する。

また、本願でいう「XXに基づく」とは、「少なくともXXに基づく」ことを意味し、XXに加えて別の要素に基づく場合も含む。また、「XXに基づく」とは、XXを直接に用いる場合に限定されず、XXに対して演算や加工が行われたものに基づく場合も含む。「XX」は、任意の要素（例えば、任意の情報）である。

10

## 【0012】

（実施形態）

（画像処理システム）

以下、本発明の実施形態の画像処理システムを、図面を参照して説明する。

図1は、実施形態の画像処理システムの一例を示す図である。本実施形態の画像処理システムでは、端末装置は、サーバー装置へアクセスし、マスク画像付きの人物の全身画像（以下「教師画像」という）を、サーバー装置へ送信する。画像処理装置は、端末装置が送信するマスク画像付きの教師画像を、サーバー装置から取得し、取得した教師画像に基づいて、教師あり学習を行う。

また、端末装置は、ショッピングサイトなどのウェブサイトへアクセスし、画像（再照明を行う被写体が撮像されたマスク付きの全身画像）（以下「入力画像」という）を、サーバー装置へ送信する。画像処理装置100は、端末装置が送信する入力画像を、サーバー装置から取得する。画像処理装置100は、取得した入力画像から特徴量を抽出し、抽出した特徴量と、教師あり学習の学習結果とに基づいて、入力画像に対応する出力画像を生成する。

20

実施形態の画像処理システム1は、端末装置10-1～端末装置10-n（nは、n>0の整数）と、サーバー装置70と、画像処理装置100とを含む。これらの装置は、ネットワーク50を介して互いに接続される。ネットワーク50は、例えば、無線基地局、Wi-Fiアクセスポイント、通信回線、プロバイダ、インターネットなどを含む。なお、図1に示す各装置の全ての組み合わせが相互に通信可能である必要はなく、ネットワークNWは、一部にローカルなネットワークを含んでもよい。

30

以下、端末装置10-1～端末装置10-nのうちの任意の端末装置を、端末装置10と記載する。

## 【0013】

端末装置10は、ユーザによって使用される装置である。端末装置10は、例えば、スマートフォンなどの携帯電話、タブレット端末、パーソナルコンピュータなどのコンピュータ装置である。例えば、端末装置10は、ショッピングサイトなどのウェブサイト、メールサービス、SNSサービス、情報提供サービスなどにおいてユーザIDを登録し、登録したユーザIDに関連付けて画像を送信するのに利用される。

サーバー装置70は、各種サービスを提供する。例えば、サーバー装置70は、端末装置10において起動されるウェブブラウザを介して、各種サービスを提供するためのウェブサイトを提供するウェブサーバー装置であってよい。また、サーバー装置70は、所定のアプリケーションプログラムが起動（実行）された端末装置10と通信を行うことで、各種情報の受け渡しを行うアプリケーションサーバー装置であってもよい。所定のアプリケーションプログラムが起動された端末装置10には、サーバー装置70との通信により、各種サービスを提供可能な画面が表示される。以下、説明を簡略化するために、サーバー装置70がウェブサーバー装置である場合について説明を続ける。

40

例えば、サーバー装置70は、サービスの提供前にユーザIDの認証を行い、ユーザの確認を行う。サーバー装置70は、認証の結果、既にユーザIDが登録されたユーザであれば各種サービスを提供し、ユーザIDが登録されていないユーザであれば、ユーザID

50

が未登録であることを通知したり、ユーザIDの登録を促したりする。ユーザIDが未登録であることを受けて、ユーザが新規にユーザIDの登録した場合、サーバー装置70は、新たに登録されたユーザIDを発行する。これによって、ユーザは新規にユーザIDを取得することができる。ユーザは、端末装置10を操作することによって、登録したユーザIDに関連付けて、被写体を含む画像を、サーバー装置70へ送信する。

#### 【0014】

画像処理装置100は、教師あり学習を行う。例えば、画像処理装置100は、教師画像として、マスク画像付きの人物の全身画像を取得する。例えば、画像処理装置100は、端末装置10が送信した教師画像を、サーバー装置70から取得する。画像処理装置100は、取得した教師画像から、画素ごとに反射率と、光伝達ベクトルと、光源情報とを（自動で）推定する。ここで、光伝達ベクトルには、光の遮蔽情報が含まれる。以下、光伝達ベクトルに、光の遮蔽情報が含まれる場合について説明を続ける。

10

画像処理装置100は、画素ごとに推定した反射率から反射率マップを作成し、画素ごとに推定した光伝達ベクトルから光伝達マップを作成する。画像処理装置100は、作成した反射率マップと、作成した光伝達マップと、推定した光源情報とを特徴量とし、その特徴量が教師画像に対応することを学習する。

#### 【0015】

画像処理装置100は、サーバー装置70と通信を行って、サーバー装置70が提供するサービスを利用するユーザのユーザIDを取得し、このユーザIDに関連付けて送信された入力画像（再照明を行う被写体が撮像されたマスク付きの全身画像）を取得する。画像処理装置100は、取得した入力画像から、画素ごとに反射率と、光伝達ベクトルと、光源情報とを（自動で）推定する。

20

画像処理装置100は、画素ごとに推定した反射率から反射率マップを作成し、画素ごとに推定した光伝達ベクトルから光伝達マップを作成する。画像処理装置100は、作成した反射率マップと、作成した光伝達マップと、推定した光源情報とを特徴量とし、その特徴量に基づいて、入力画像に対応する出力画像を生成する。画像処理装置100は、生成した出力画像を出力する。

#### 【0016】

（画像処理装置100）

図2は、実施形態の画像処理装置の一例を示すブロック図である。

30

画像処理装置100は、通信部105と、記憶部110と、操作部120と、情報処理部130と、表示部140と、前記各構成要素を図2に示されているように電気的に接続するためのアドレスバスやデータバスなどのバスライン150とを備える。

通信部105は、通信モジュールによって実現される。通信部105は、ネットワーク50を介して、サーバー装置70などの外部の通信装置と通信する。具体的には、通信部105は、サーバー装置70が送信した教師画像を受信し、受信した教師画像を、情報処理部130へ出力する。ここで、教師画像の一例は、市販の3Dスキャンされた人物のメッシュモデルをレンダリングして作成された画像である。市販の3D人物モデルの多くは、テクスチャとして光沢成分を含まず拡散反射成分しか持たない。このため、本実施形態では、反射率として、拡散反射成分を対象とした場合について説明する。市販されている3D人物モデルの数は限られており、実施形態で用いるモデルの数は訓練データとテストデータとを合わせても数百体しかない。衣服のパリエーションが無数にあることを考えると訓練データが足りないようにも思えるが、実際には、本実施形態で用意したデータを用いて学習することで、服の皺、脇や股などの光の遮蔽が生じやすい部位についても、光の遮蔽を考慮した陰影計算ができる。

40

また、通信部105は、サーバー装置70が送信した入力画像を受信し、受信した入力画像を、情報処理部130へ出力する。また、通信部105は、情報処理部130が出力した出力画像を取得し、取得した出力画像を、サーバー装置70へ送信する。

#### 【0017】

記憶部110は、例えば、RAM（Random Access Memory）、ROM（Read Only Memo

50

ry)、HDD(Hard Disk Drive)、フラッシュメモリ、またはこれらのうち複数組み合わせられたハイブリッド型記憶装置などにより実現される。記憶部110には、情報処理部130により実行されるプログラム111と、アプリ112と、反射率マップ113と、光源情報114と、光伝達マップ115とが記憶される。

#### 【0018】

アプリ112は、画像処理装置100に、サーバー装置70が送信した教師画像を取得させる。アプリ112は、画像処理装置100に、取得させた教師画像から、画素ごとに反射率と、光伝達ベクトルと、光源情報とを推定させる。アプリ112は、画像処理装置100に、画素ごとに推定させた反射率から反射率マップを作成させ、画素ごとに推定させた光伝達ベクトルから光伝達マップを作成させる。アプリ112は、画像処理装置100に、作成させた反射率マップと、作成させた光伝達マップと、推定させた光源情報とを特徴量とさせ、その特徴量が教師画像に対応することを学習させる。

10

#### 【0019】

アプリ112は、画像処理装置100に、サーバー装置70が送信した入力画像を取得させる。アプリ112は、画像処理装置100に、取得させた入力画像から、画素ごとに反射率と、光伝達ベクトルと、光源情報とを推定させる。アプリ112は、画像処理装置100に、画素ごとに推定させた反射率から反射率マップを作成させ、画素ごとに推定させた光伝達ベクトルから光伝達マップを作成させる。アプリ112は、画像処理装置100に、画素ごとに作成させた反射率マップと、画素ごとに作成させた光伝達マップと、推定させた光源情報とを特徴量とさせ、その特徴量に基づいて、入力画像に対応する出力画像を生成させる。

20

反射率マップ113は、教師画像の反射率を、画素ごとに表したマップである。

光源情報114は、教師画像の照明状態を示す情報である。

光伝達マップ115は、物体の幾何的な情報が記録された光伝達ベクトルを、画素ごとに表したマップである。光伝達マップ115と、光伝達ベクトルとは、光源の情報とは分離されており、光源に依存しない。物体の幾何的な情報には、光が遮蔽されるか否かを示す情報である光の遮蔽情報が含まれる。光伝達マップ115と、光伝達ベクトルとが、光源の情報とは分離されていることによって、照明計算を正しく行うことができる。

#### 【0020】

操作部120は、例えば、タッチパネルなどによって構成され、表示部140に表示される画面に対するタッチ操作を検出し、タッチ操作の検出結果を、情報処理部130へ出力する。

30

#### 【0021】

情報処理部130の全部または一部は、例えば、CPU(Central Processing Unit)などのプロセッサが記憶部110に格納されたプログラム111を実行することにより実現される機能部(以下、ソフトウェア機能部と称する)である。なお、情報処理部130の全部または一部は、LSI(Large Scale Integration)、ASIC(Application Specific Integrated Circuit)、またはFPGA(Field-Programmable Gate Array)などのハードウェアにより実現されてもよく、ソフトウェア機能部とハードウェアとの組み合わせによって実現されてもよい。

40

情報処理部130は、例えば、取得部131と、分析部132と、機械学習部133と、生成部134と、出力部135とを備える。

#### 【0022】

取得部131は、通信部105が出力した教師画像を取得し、取得した教師画像を、分析部132へ出力する。また、取得部131は、通信部105が出力した入力画像を取得し、取得した入力画像を、分析部132へ出力する。

分析部132は、取得部131が出力した教師画像を取得し、取得した教師画像を分析することによって、画素ごとに反射率と、光伝達ベクトルと、光源情報とを推定する。分析部132は、画素ごとに推定した反射率から反射率マップを作成し、画素ごとに推定した光伝達ベクトルから光伝達マップを作成する。分析部132は、作成した反射率マ

50

マップを反射率マップ 113 に記憶させ、作成した光伝達マップを光伝達マップ 115 に記憶させ、推定した光源情報を光源情報 114 に記憶させる。分析部 132 は、反射率マップと光伝達マップと光源情報とを、機械学習部 133 へ出力する。

ここで、分析部 132 が、光伝達マップを作成する処理について説明する。分析部 132 は、球面調和関数 (SH) に基づいて、光の遮蔽を考慮して、照明計算を行う。

まず、光の遮蔽を考慮しないで、照明計算を行う場合について説明する。

(1) 光の遮蔽を考慮しない場合

光の遮蔽や相互反射を考慮しない場合、物体表面上の点 p において、単位法線ベクトルを  $\mathbf{n}$  とすると、式 (1) に示されるように、放射照度  $E(\mathbf{n})$  は法線  $\mathbf{n}$  が定義する半球 ( $\Omega(\mathbf{n})$ ) のあらゆる方向  $\omega_i$  から差し込む放射輝度  $L(\omega_i)$  の積分によって計算される。

10

【0023】

【数1】

$$E(\mathbf{n}) = \int_{\Omega(\mathbf{n})} L(\omega_i) \max(\mathbf{n} \cdot \omega_i, 0) d\omega_i \quad (1)$$

【0024】

なお、位置 p に関する依存性は単純化のために省略した。入射する放射輝度の分布  $L(\omega_i)$  にコサイン減衰項  $\max(\mathbf{n} \cdot \omega_i, 0)$  をかけたものを球面調和関数に射影する。方向ベクトルと仰角と方位角とを用いて  $\omega_i = (\theta_i, \phi_i)$  とパラメータ表示すると、式 (2)、式 (3) が得られる。

20

【0025】

【数2】

$$L(\theta, \phi) = \sum_{l,m} L_{l,m} Y_{l,m}(\theta, \phi) \quad (2)$$

$$A(\theta) = \max(\cos \theta, 0) = \sum_l A_l Y_{l,0}(\theta) \quad (3)$$

30

【0026】

式 (2) において、 $Y_{l,m}$  は球面調和関数である (ただし、 $l \geq 0$ ,  $-l \leq m \leq l$ , かつ  $m \leq 2$ )。式 (2) と式 (3) とにおいて、 $L_{l,m}$  と  $A_l$  はそれぞれ、照明とコサイン減衰項の係数である。式 (1) の積分を書き換えると、式 (4) となる。

【0027】

【数3】

$$E(\theta, \phi) = \sum_{l,m} \hat{A}_l L_{l,m} Y_{l,m}(\theta, \phi) \quad (4)$$

40

【0028】

ここで、 $\hat{A}_l$  は、以下のように表される。

【0029】



【数 4】

$$\hat{A}_l = \sqrt{\frac{4\pi}{2l+1}} A_l$$

【0030】

$Y_{l,m}$  は、法線ベクトル  $\mathbf{n} = (x, y, z)^T$  の各成分  $x, y, z$  の多項式によって表現できる。係数列  $\{L_{l,m}\}$  をベクトル  $L$  で、基底関数列  $\{\hat{A}_l Y_{l,m}\}$  をベクトル  $\hat{Y}$  で表現すると、放射照度  $E$  はベクトルの内積で計算できる。

10

【0031】

$$E = \hat{Y}^T L \quad (5)$$

【0032】

次に、光の遮蔽を考慮して照明計算を行う場合について説明する。

(2) 光の遮蔽を考慮する場合

(1) 光の遮蔽を考慮しない場合で述べた通り、本来は光が遮られて暗くなるべきところが不自然に明るくなる。式(1)で光の遮蔽を考慮するために、本実施形態では、可視関数  $V(\omega_i)$  を導入する。

【0033】

【数 5】

20

$$E(\mathbf{n}) = \int_{\Omega(\mathbf{n})} L(\omega_i) V(\omega_i) \max(\mathbf{n} \cdot \omega_i, 0) d\omega_i \quad (6)$$

【0034】

式(6)において、 $V(\omega_i)$  は、入射方向  $\omega_i$  の光が遮られれば0であり、遮られなければ1である。

図3は、可視関数  $V(\omega_i)$  を導出する処理の一例を示す図である。図3に示されるように、十分に離れたところ(無限遠)から、物体に光を放射することを考える。この場合、十分に離れたところ(無限遠)から届く光は平行光とみなすことができる。周囲から届く光(環境光源)は無数の平行光からなる。本実施形態では、平行光の入射方向  $\omega_i$  について、物体表面のすべての点  $p$  (光伝達マップのすべての画素)における、方向  $\omega_i$  に関する可視関数  $V(\omega_i, p)$  を並列で求める。つまり、光源方向を一つ選ぶたびに、物体表面のすべての点で、その方向に関する可視関数を一度にすべて求める。これを、光源方向を変えながら繰り返す。光源方向は、乱数に基づいて、ランダムに選択されてもよい。図3に示される例では、光源方向  $LD-1$  が選択され、物体表面の複数の点  $p-1, p-2, p-3, p-4, p-5$  の各々について、光源方向  $LD-1$  からの光が、複数の点  $p-1, p-2, p-3, p-4, p-5$  の各々に届くか否かを示す可視関数が求められる。

30

40

グラフィクス用ハードウェア(GPU)を用いる従来の方法では、物体表面の各点で、可視関数を計算するために物体形状を描画する必要があった。この方法では物体表面の点の数が膨大になると、物体形状の描画回数が増え、計算に時間がかかる。具体的には、仮に  $1024 \times 1024$  画素(のマスク内の)各画素で計算すると、マスクが画像の半分を覆っているとしても  $524288$  画素あり、その画素数の分だけ物体形状を描画する必要がある。また、物体表面上の各点でランダムな方向に向かってレイトラッキングして可視判定を行う場合には、効率が悪かった。

一方、本実施形態では、可視関数を評価するために物体形状を描画するのは光源方向一つ毎でよい。実際に考慮する光源方向の数を  $642$  方向と仮定した場合、上記と比べて格段に少なく、計算時間を短縮できる。図2に戻り説明を続ける。

50

P R Tの枠組みでは、この可視関数  $V(i)$  とコサイン減衰項とともに前計算し、S Hの基底関数に投影して、係数ベクトルの内積による高速な照明条件を実現する。ベクトル  $T$  を、可視関数とコサイン減衰項を基底関数に投影した際の係数からなるベクトルとすると、放射照度  $E$  は次のように計算できる。

【 0 0 3 5 】

$$E = T^T L \quad (7)$$

【 0 0 3 6 】

ここで、光伝達ベクトルはベクトル  $T$  であり、光伝達ベクトルを各画素に持つ画像を光伝達マップと呼ぶ。

また、分析部 1 3 2 は、取得部 1 3 1 が出力した入力画像を取得し、取得した入力画像を分析することによって、画素ごとに反射率と、光伝達ベクトルと、光源情報とを推定する。分析部 1 3 2 は、推定した反射率から反射率マップを作成し、推定した光伝達ベクトルから光伝達マップを作成する。分析部 1 3 2 は、反射率マップと光伝達マップと光源情報とを、生成部 1 3 4 へ出力する。

10

【 0 0 3 7 】

機械学習部 1 3 3 は、分析部 1 3 2 が出力した反射率マップと光伝達マップと光源情報とを取得し、取得した反射率マップと光伝達マップと光源情報とを特徴量とし、その特徴量が教師画像であることを学習する。本実施形態では、機械学習モデルの一例として、畳み込みニューラルネットワーク(Convolutional neural network: CNN)を使用した場合について説明を続ける。

20

【 0 0 3 8 】

ここで、機械学習部 1 3 3 が CNN による機械学習を行う際に使用する損失関数について説明する。

人物画像データセット  $D_H$  は、各人物モデルに対し、画像中で人物が写っている部分を表す二値マスク、反射率マップ、光伝達マップからなる。二値マスクと、反射率マップと、光伝達マップとは、以下のように表される。

【 0 0 3 9 】

【数 6】

$$\text{二値マスク } M_j^c \in \{0, 1\}^{N \times c}$$

(ただし  $N$  は画素数、 $c$  はチャンネル数で、 $j = 1, 2, \dots, |D_H|$ )

$$\text{反射率マップ } \Lambda_j \in \mathbb{R}^{N \times 3}$$

$$\text{光伝達マップ } \Psi_j \in \mathbb{R}^{N \times 9}$$

30

【 0 0 4 0 】

一方、光源データセット  $D_L$  は、光源ごとに照明情報を持ち、これは R G B のチャンネルごとの S H の 9 係数からなる。照明情報は、以下のように表される。

40

【 0 0 4 1 】

【数 7】

$$\text{照明情報 } \Pi_k \in \mathbb{R}^{9 \times 3} \quad (\text{ただし } k = 1, 2, \dots, |D_L|)$$

【 0 0 4 2 】

なお、二値マスクは正解データおよび CNN の出力に乗算してから損失関数を計算する

50

が（例えば  $M^3_j * j$  や  $M^9_j * j$  で、\*は要素ごとの乗算を表す）、以下の説明では単純化のために二値マスクの乗算は省略する。

本実施形態で使用するCNNは、反射率マップ と、光伝達マップ と、光源情報 とを推定する。CNNへの入力は、マスク乗算済の人物全身画像  $I_{j,k} = j * (j_k)$  である。本実施形態では、機械学習部133は、損失関数として、反射率マップ と、法線マップと、光源情報 とのそれぞれについて正解とL2損失、そしてそれらの積である人物全身画像の推定値の正解とL2損失という4種類の損失関数について、L1損失を導出し、さらに以下の損失関数を含め、合計で15種類の損失関数を計算する。

#### 【0043】

- ・反射率マップ および光伝達マップ に関するL1 total variation (TV) 損失(2通り)
  - ・陰影マップ(つまり  $j_k$ ) に関して、光伝達マップ と光源情報 を、一方を推定値とし他方を正解とした場合および両方を推定値とした場合のL1損失(3通り)
  - ・反射率マップ 、光伝達マップ および光源情報 の3つの積(つまり  $j * (j_k)$ ) に対し、そのうちの1つまたは2つを推定値にした場合のL1損失(6通り)
- なお、それぞれの損失関数の重みは全て1とした。

このように、推定値だけでなく正解も含めた積に対してL1損失を計算することは、推定値に対する重み付けを行っていることに相当する。損失関数の種類を増やすことによって、より鮮明な推定結果を得ることができる。

#### 【0044】

本実施形態で使用するCNNは、推定した反射率(アルベド)マップ と、光伝達マップ と、光源 との3つを掛け合わせなくても、掛け合わせる途中のデータについても、正解と一致しているかどうかを測ることができる。例えば、CNNでは、光伝達マップ と光源 とを掛け合わせることで、服や肌の色や模様についていない、純粋に形状だけに依存した陰影マップを取得する。

この取得した陰影マップを、正解の光伝達マップ と正解の光源 とで作ることによって、正解の陰影マップが得られる。これと推定したデータとを比較するとき、光伝達マップ と光源 とのうち、どちらか一方が正解で他方が推定したもの、あるいは両方とも推定したデータに対して、正解との誤差を測ることによって、3通りの損失関数が得られる。このようにして組み合わせを考えると、陰影マップについて3通り、入力画像の復元に関して6通り(3つとも推定した場合を除く)の組み合わせが得られる。

この他に、光伝達マップ と、反射率マップ とについてのL1 TV lossで、計2つ、反射率マップ と、光伝達マップ と、光源および入力画像をすべて推定データで復元する場合を考えると計4つ、合計で15種類の損失関数を導出する。

#### 【0045】

生成部134は、分析部132が出力した反射率マップ と光伝達マップ と光源情報 とを取得し、取得した反射率マップ と光伝達マップ と光源情報 とを特徴量とし、その特徴量に基づいて、出力画像を生成する。生成部134は、生成した出力画像を、出力部135へ出力する。

出力部135は、生成部134が出力した出力画像を取得し、取得した出力画像を、通信部105へ出力する。

#### 【0046】

次に、本実施形態の画像処理装置100で用いる機械学習モデルについて説明する。

図4は、本実施形態の画像処理装置で使用する機械学習モデルの一例を示す図である。

本実施形態の画像処理装置100で使用する機械学習モデル200の一例は、前述したように、CNNである。

機械学習モデル200は、乗算部201と、エンコーダ202と、反射率マップデコーダ203と、光伝達マップデコーダ206と、畳み込み部(Conv.)210とによって表される。反射率マップデコーダ203は、ResNetブロック204と、逆畳み込み部(Dconv.)205を含む。光伝達マップデコーダ206は、Re

10

20

30

40

50

sNetブロック207と、逆畳み込み部(Dconv)208とを含む。

なお、畳み込み部210と、ResNetブロック204と、逆畳み込み部205と、ResNetブロック207と、逆畳み込み部208との各々の直後には、データが入出力される最初と最後の層を除きバッチ正規化(batch normalization)およびReLUを適用する。また、逆畳み込み部205と、逆畳み込み部208との各々の最初の逆畳み込み層各3層の直後には確率0.5のドロップアウトを適用する。

#### 【0047】

入力画像211として、人物全身画像I212と、二値マスクM213とが乗算部201へ出力される。乗算部201は、人物全身画像I212と二値マスクM213とを乗算し、乗算した結果を、エンコーダ202へ出力する。エンコーダ202と逆畳み込み部205との間と、エンコーダ202と逆畳み込み部208との間とは、スキップコネクション(skip-connection)によって連結されている。

エンコーダ202の一例は、畳み込み層7層で構成される。エンコーダ202は、乗算部201が出力した人物全身画像I212と二値マスクM213とを乗算した結果に対して、フィルタを適用する。エンコーダ202は、フィルタを適用した結果を、反射率マップデコーダ203と、光伝達マップデコーダ206と、連結部209とへ出力する。

#### 【0048】

反射率マップデコーダ203では、エンコーダ202が出力したフィルタを適用した結果は、ResNetブロック204へ出力される。ResNetブロック204は、2つのResNetブロックを含み、2つのResNetブロックの各々は畳み込み計算を行う。ResNetブロック204は、畳み込み計算を行った結果を、連結部209と、逆畳み込み部205とへ出力する。

逆畳み込み部205は、ResNetブロック204が出力した畳み込み計算を行った結果を取得し、取得した畳み込み計算を行った結果に対して7層の逆畳み込みを行うことによって、反射率マップ221を導出する。このように構成することによって、反射率マップデコーダ203は、反射率マップを推定できる。

#### 【0049】

光伝達マップデコーダ206では、エンコーダ202が出力したフィルタを適用した結果は、ResNetブロック207へ出力される。ResNetブロック207は、2つのResNetブロックを含み、2つのResNetブロックの各々は畳み込み計算を行う。ResNetブロック207は、畳み込み計算を行った結果を、連結部209と、逆畳み込み部208とへ出力する。

逆畳み込み部208は、ResNetブロック207が出力した畳み込み計算を行った結果を取得し、取得した畳み込み計算を行った結果に対して7層の逆畳み込みを行うことによって、光伝達マップ223を導出する。このように構成することによって、光伝達マップデコーダ206は、光伝達マップを推定できる。

#### 【0050】

連結部209は、エンコーダ202が出力したフィルタを適用した結果と、ResNetブロック204が出力した畳み込み計算を行った結果と、ResNetブロック207が出力した畳み込み計算を行った結果とを取得し、取得したフィルタを適用した結果と、畳み込み計算を行った結果と、畳み込み計算を行った結果とを連結する。連結部209は、フィルタを適用した結果と、畳み込み計算を行った結果と、畳み込み計算を行った結果とを連結した結果を、畳み込み部210へ出力する。

畳み込み部210は、連結部209が出力した連結した結果を、畳み込む。具体的には、畳み込み部210は、空間解像度が1×1、チャンネル数が27になるまで畳み込むことによって、光源情報222を導出する。

#### 【0051】

逆畳み込み部205は、導出した反射率マップ221を、乗算部224へ出力する。畳み込み部210は、導出した光源情報222を、乗算部224へ出力する。逆畳

10

20

30

40

50

み込み部 208 は、導出した光伝達マップ $\sim$  223 を、乗算部 224 へ出力する。乗算部 224 は、逆畳み込み部 205 が出力した反射率マップ $\sim$  と、畳み込み部 210 が出力した光源情報 $\sim$  222 と、逆畳み込み部 208 が出力した光伝達マップ $\sim$  223 とを取得し、取得した反射率マップ $\sim$  221 と、光源情報 $\sim$  222 と、光伝達マップ $\sim$  223 とを乗算することによって、人物全身画像 $\sim$  I を導出する。このように構成することによって、画像処理装置 100 は、入力画像 211 を再現した画像を導出できる。

#### 【0052】

(画像処理装置 100 の動作)

本実施形態の画像処理装置 100 の動作について、教師あり学習を行う場合と、教師あり学習の結果に基づいて、入力画像に対応する出力画像を生成する場合とに分けて説明する。

10

図 5 は、本実施形態の画像処理装置の動作の一例を示すフローチャートである。図 5 に示される例では、教師あり学習を行う処理について示される。

(ステップ S1)

画像処理装置 100 の通信部 105 は、サーバー装置 70 が送信した教師画像を受信し、受信した教師画像を、情報処理部 130 へ出力する。情報処理部 130 の取得部 131 は、通信部 105 が出力した教師画像を取得し、取得した教師画像を、分析部 132 へ出力する。

(ステップ S2)

分析部 132 は、取得部 131 が出力した教師画像を取得し、取得した教師画像から、画素ごとに反射率と、光伝達ベクトルと、光源情報とを推定する。

20

(ステップ S3)

分析部 132 は、画素ごとに推定した反射率から反射率マップ $\sim$  を作成し、画素ごとに推定した光伝達ベクトルから光伝達マップ $\sim$  を作成する。分析部 132 は、作成した反射率マップ $\sim$  と、光伝達マップ $\sim$  と、推定した光源情報 $\sim$  とを、機械学習部 133 へ出力する。

(ステップ S4)

機械学習部 133 は、分析部 132 が出力した反射率マップ $\sim$  と、光伝達マップ $\sim$  と、光源情報 $\sim$  とを取得し、取得した反射率マップ $\sim$  と、光伝達マップ $\sim$  と、光源情報 $\sim$  とを特徴量とし、その特徴量が教師画像に対応することを学習する。

30

#### 【0053】

図 6 は、本実施形態の画像処理装置の動作の一例を示すフローチャートである。図 6 に示される例では、教師あり学習の結果に基づいて、入力画像に対応する出力画像を生成する処理について示される。

(ステップ S11)

画像処理装置 100 の通信部 105 は、サーバー装置 70 が送信した入力画像を受信し、受信した入力画像を、情報処理部 130 へ出力する。情報処理部 130 の取得部 131 は、通信部 105 が出力した入力画像を取得し、取得した入力画像を、分析部 132 へ出力する。

(ステップ S12)

分析部 132 は、取得部 131 が出力した入力画像を取得し、取得した入力画像から、画素ごとに反射率と、光伝達ベクトルと、光源情報とを推定する。

40

(ステップ S13)

分析部 132 は、画素ごとに推定した反射率から反射率マップ $\sim$  を作成し、画素ごとに推定した光伝達ベクトルから光伝達マップ $\sim$  を作成する。分析部 132 は、作成した反射率マップ $\sim$  と、光伝達マップ $\sim$  と、推定した光源情報 $\sim$  とを、生成部 134 へ出力する。

(ステップ S14)

生成部 134 は、分析部 132 が出力した反射率マップ $\sim$  と、光伝達マップ $\sim$  と、光源情報 $\sim$  とを取得し、取得した反射率マップ $\sim$  と、光伝達マップ $\sim$  と、光源情報 $\sim$

50

とに基づいて、入力画像に対応する出力画像を生成する。

#### 【0054】

(本実施形態の手法と従来手法との比較)

光伝達ベクトルに、光の遮蔽情報を含めた場合(以下「本手法」という)と、含めない場合(以下「従来手法」という)とについて、比較を行った。

最初に、訓練データ(教師画像)およびテストデータ(入力画像)として用いる、3D人物モデルを用いたCG人物画像データセットと環境光源データセットとを作成した。

CG人物画像データセットは、GPUベースレンダラによって生成した。CG人物画像データベースは、各3D人物モデルについて、二値マスク、反射率マップ、法線マップ、光伝達マップからなる。本手法では法線マップは必要ないが、法線マップは、従来手法での訓練データとテストデータとに使用する。

図7は、CG人物データセットの一例を示す図である。図7の(1)は反射率マップの一例を示し、図7の(2)は二値マスクの一例を示し、図7の(3)は法線マップの一例を示し、図7の(4)は光伝達マップの一例を示す。

データの取得先の一例は、BUFFデータセットおよび商用ウェブサイトである。ここでは、345体分を用意し、そのうち276体分を訓練データ、69体分をテストデータとした。なお、入手した3D人物モデルの多くは光沢反射成分のテクスチャが提供されていないため、ここでは、拡散反射を扱った場合について説明を続ける。ただし、3Dスキャン時に細かい皺などによる陰影が完全には除去されていないため、正解反射率マップにも遮蔽による陰影が含まれている。

#### 【0055】

本実施形態では、訓練データのばらつきを抑えて、CNNの推定精度を高めるため、3D人物モデルは、立った姿勢とし、座った姿勢のものは除外した。そして、描画位置は画像の中央とし、上下に画像の縦幅5%分の余白を空けて描画した。なお、推定を行う際も、入力画像の二値マスクを利用して、上下に画像の縦幅5%分の余白が空くように整形してからCNNに入力する。

光源データセットは、Laval Indoor HDRデータセットから取得したパノラマHDR形式の環境マップを、SHの係数に変換することで作成した。環境マップによっては暗すぎたり明るすぎたりするものがあるため、明るさの調整を行った。まず、視点座標系の法線ベクトル(0, 0, 1)<sup>T</sup>から解析的に光伝達ベクトルを計算し、それと各環境マップのSHの係数を乗算して、参照用の輝度値を計算する。その参照用の輝度値が閾値0.2よりも暗すぎる環境マップは除外した上で、[0.7, 0.9]の範囲に収まるように環境マップの輝度値をスケールした。さらに、バリエーションを増やすため、環境マップを、鉛直方向を軸として10度ずつ35回回転させた。その上で、冗長度を減らすため、k-meansクラスタリングによって数を減らし、不自然な光源を取り除いた。最終的に、合計50個の光源を用意し、そのうちランダムに選んだ40個を訓練データ、10個をテストデータとした。

#### 【0056】

本手法を、Pythonおよびchainerを用いて実装し、教師あり学習を行い、教師あり学習の結果に基づいて、推論することによって、入力画像に対応する出力画像を生成した。最適化には、Adamを使用し、学習率は0.0002に固定し、バッチサイズは1とした。訓練に要した時間は、1つのGPUを使用して、1024×1024画素のデータを入力した場合、1エポック当たり約3時間であった。本手法として、20エポックまで学習したものを使用した。推論に要した時間は、1024×1024画素の入力1つ当たり0.43秒程度である。

#### 【0057】

光伝達ベクトルに、光の遮蔽情報が含まれることによる効果を明らかにするために、本手法と、光伝達マップの代わりに、法線マップを推定した従来手法との比較を行った。

従来手法の機械学習モデルは、本手法の機械学習モデルとは異なるが、本手法の機械学習モデルの方が、入力画像の解像度、層数、中間層のチャンネル数のいずれについても従

10

20

30

40

50

来手法の機械学習モデルよりも規模が大きく、推定の精度が劣らないと想定される。

また、2つ目の比較対象として、本手法で最適化する損失関数を15種類から4種類に減少させたものも用意した。定量的比較として、全テストデータに対する各手法の推定結果の、マスク内の平均二乗誤差平方根(Root Mean Square Error: RMSE)を、図8に示す。法線マップについてのRMSEは、従来手法の結果で計算している。

#### 【0058】

光源情報を除くと、共通する要素ではいずれも従来手法の結果よりも、本手法(損失関数4種類)がよく、さらに本手法(損失関数15種類)が最もよい。定性的比較として、テストデータに対する各手法での推定結果を図9に示す。

図9は、実施形態の画像処理装置による推定結果の一例を示す図である。図9に示される例では、CGテストデータを用いた場合の、各手法において、正解と推定結果との比較が示される。

図9の(1)は入力画像の一例であり、図9の(2)は正解反射率マップの一例であり、図9の(3)は従来手法による反射率マップの一例であり、図9の(4)は本手法(損失関数4種類)による反射率マップの一例であり、図9の(5)は本手法(損失関数15種類)による反射率マップの一例である。図9の(6)は光源であり、上から順に正解、従来手法による光源、本手法(損失関数4種類)による光源、本手法(損失関数15種類)による光源である。図9の(7)は正解法線マップの一例であり、図9の(8)は従来手法による法線マップの一例であり、図9の(9)は正解陰影マップの一例であり、図9の(10)は従来手法による陰影マップの一例であり、図9の(11)は本手法(損失関数4種類)による陰影マップの一例であり、図9の(12)は本手法(損失関数15種類)による陰影マップの一例である。

図9によれば、定量的比較と符号するように、反射率マップと陰影マップとにおいて、従来手法、本手法(損失関数4種類)、本手法(損失関数15種類)の順でよくなっていることが分かる。従来手法の陰影マップでは、首や脇などの部分の遮蔽を法線マップで再現できない分、反射率マップの該当箇所が暗くなっている。本手法(損失関数4種類)と、本手法(損失関数15種類)とを比較すると、本手法(損失関数15種類)の方が、陰影マップがより先鋭になっていることが分かる。

#### 【0059】

また、同様に、実写の人物画像を入力として、前述した3つの手法を用いて推定した結果を、図10に示す。

図10は、実施形態の画像処理装置による推定結果の一例を示す図である。図10に示される例では、実写画像を用いた場合の、各手法において、正解と推定結果との比較が示される。図10の(1)は入力画像であり、図10の(2)は従来手法による反射率マップの一例であり、図10の(3)は本手法(損失関数4種類)による反射率マップの一例であり、図10の(4)は本手法(損失関数15種類)による反射率マップの一例である。図10の(5)は光源であり、上から順に、従来手法による光源、本手法(損失関数4種類)による光源、本手法(損失関数15種類)による光源である。図10の(6)は従来手法による陰影マップの一例であり、図10の(7)は本手法(損失関数4種類)による陰影マップの一例であり、図10の(8)は本手法(損失関数15種類)による陰影マップの一例である。

図10の結果については、テストデータを用いた図9の結果と同様に、従来手法で推定した陰影マップでは平坦なレリーフに凹凸がついているように見える一方、本手法(損失関数4種類)と、本手法(損失関数15種類)とでは、遮蔽による陰影が観察できる。

また、2つの人物画像についてそれぞれ同時推定を行い、光源情報を入れ替えて再照明することで、それぞれ互いの照明環境下での見た目を再現した。

図11は、実施形態の画像処理装置による推定結果の一例を示す図である。図10に示される例では、推定した光源情報を互いに入れ替えて再照明を行った結果を示す。図10によれば、推定した反射率マップ、光伝達マップ、光源情報の精度が従来よりも向上でき

10

20

30

40

50

るため、全て正解データを用いて計算した結果と見比べてもあまり違いが見られない。

【0060】

前述した実施形態では、端末装置10からサーバー装置70と経由して、画像処理装置100へ、画像が送信される場合について説明したが、この限りでない。例えば、画像が、画像処理装置100へ直接入力されてもよい。

前述した実施形態において、端末装置10と画像処理装置100とが同一の装置であってもよいし、サーバー装置70と画像処理装置100とが同一の装置であってもよい。

前述した実施形態では、端末装置10が送信した入力画像（再照明を行う被写体が撮影されたマスク付きの全身画像）を、画像処理装置100が取得する場合について説明したが、この限りでない。例えば、端末装置10は、再照明を行う被写体が撮影された全身画像を送信し、画像処理装置100は、端末装置10が送信した再照明を行う被写体が撮影された全身画像を取得し、取得した全身画像に基づいて、マスクを推定してもよい。

10

前述した実施形態では、反射率として、拡散反射成分を対象とした場合について説明したが、この限りでない。例えば、拡散反射成分だけでなく光沢を含む鏡面反射成分を対象とした場合についても適用できる。この場合、鏡面反射に関する反射率マップについて教師付き学習が行われ、教師付き学習の学習結果に基づいて、入力画像の鏡面反射成分が推定される。

前述した実施形態では、機械学習モデルの一例として、畳み込みニューラルネットワークを使用した場合について説明したが、この例に限られない。例えば、再起型ニューラルネットワークなどの一般的な機械学習モデルを使用してもよい。

20

前述した実施形態では、光源および光伝達ベクトルの表現方法として、SHの2次の9係数を使用した場合について説明したが、この例に限られない。例えば、SHの1次以下、3次以上でもよく、係数のすべてを使用してもよく、係数の一部のみを使用してもよい。

【0061】

前述した実施形態では、光源および光伝達ベクトルの表現方法として、SHについて説明したが、この限りでない。例えば、Haarウェーブレット、球面ガウス関数などで、光源および光伝達ベクトルを表現してもよい。

前述した実施形態では、可視関数 $V(i)$ が、物体表面の複数の点 $p-1$ 、 $p-2$ 、 $p-3$ 、 $p-4$ 、 $p-5$ の各々について、光源方向 $LD-1$ からの光が、複数の点 $p-1$ 、 $p-2$ 、 $p-3$ 、 $p-4$ 、 $p-5$ の各々に届くか否かに基づいて導出される場合について、説明したがこの例に限られない。例えば、可視関数 $V(i)$ が、物体表面の点（画素）毎且つ光源 $LS$ が出射する光の方向毎に、光が届くか否かを判定することによって導出されてもよい。

30

前述した実施形態では、光源情報 $\tilde{\quad}$ が入力画像に含まれる被写体に基づいて、推定される場合について説明したが、この例に限られない。例えば、光源情報 $\tilde{\quad}$ が、背景の情報（建物の陰影など）に基づいて、推定されてもよい。つまり、光源情報 $\tilde{\quad}$ が、背景を含む入力画像に基づいて、推定されてもよい。

前述した実施形態では、機械学習モデルとは、機械学習モデルによって、入力画像に基づいて推定される光源情報 $\tilde{\quad}$ と、反射率マップ $\tilde{\quad}$ と、光伝達マップ $\tilde{\quad}$ とのいずれか又はいずれか同士の組み合わせから得られる画像と、入力画像とを比較することによって学習されたものである場合について説明したが、この例に限られない。例えば、機械学習モデルによって、入力画像に基づいて推定される光源情報 $\tilde{\quad}$ と、反射率マップ $\tilde{\quad}$ と、光伝達マップ $\tilde{\quad}$ とのいずれか又はいずれか同士の組み合わせから得られる画像と、入力画像以外の所定の正解画像とを比較することによって学習されたものであってもよい。具体的には、光源と光伝達マップをかけ合わせると（反射率マップに含まれるような色・模様を含まない）陰影マップが得られ、これをCGで生成し、機械学習モデルの出力（同じように推定された光源と光伝達マップを掛け合わせたもの）とを比較してもよい。

40

前述した実施形態では、画像処理装置100が、取得した教師画像から、画素ごとに反射率と、光伝達ベクトルと、光源情報とを推定する場合について説明したが、この例に限

50



られない。例えば、画像処理装置100が、取得した教師画像から、複数の画素ごとに反射率と、光伝達ベクトルと、光源情報とを推定するようにしてもよい。

前述した実施形態では、画像処理装置100が、取得した入力画像から光源情報を推定する場合について説明したが、この例に限られない。例えば、画像処理装置100は、外部から光源情報を取得し、取得した光源情報に基づいて、出力画像を生成してもよい。このように構成することによって、入力画像に含まれる被写体を再照明した再照明画像を生成できる。

【0062】

前述した実施形態では、15種類の損失関数について説明したが、この例に限られない。例えば、前述した15種類の損失関数に加え、又は前述した15種類の損失関数の代わり

10

【0063】

$$|x - x'| + |y - y'| \quad (A)$$

【0064】

$$|x - x'| + |y - y'| \quad (B)$$

【0065】

また、 $x$ や $y$ を計算する前に、 $3 \times 3$ などの平滑化フィルタを掛けた上で、 $x$ や $y$ を計算してもよい。具体的には、平滑化フィルタを $H$ とした場合に、式(C)と式(D)とを導出する。

20

【0066】

$$|xH - x'H| + |yH - y'H| \quad (C)$$

【0067】

$$|xH - x'H| + |yH - y'H| \quad (D)$$

【0068】

この場合、実際の計算の順序としては、 $xH$ などを掛け合わせたもの、つまり平滑化カーネルの $x$ 、 $y$ 方向の微分を計算してから畳み込むことになる。

本実施形態の画像処理装置100によれば、画像処理装置100は、単視点で撮影された人物の全身画像の再照明という新しい問題に対し、反射率と光源情報とに加え、再照明可能な形での光の遮蔽情報を推定できる。具体的には、光の遮蔽情報を、球面調和関数(SH)の係数として表現し、球面調和関数の係数として表現された光の遮蔽情報を推定する。このように構成することによって、光の遮蔽を考慮しない場合と比較して、写実的な再照明を実現できる。

30

また、画素単位でカラーチャンネルごとに9次元ベクトルの内積計算を行うのみなので高速である。実写の人物画像を入力した場合も含め、尤もらしい再照明を行うことができる。

【0069】

<構成例>

40

一構成例として、入力画像を取得する取得部と、光源からの照明状態を示す情報である光源情報(実施形態では、光源情報)と、入力画像に含まれる被写体の反射率として入力画像に基づいて推定された反射率情報(実施形態では、反射率マップ)と、入力画像に含まれる被写体からの光の伝達状態として入力画像と光源からの光が被写体に届くか否かを示す可視情報(実施形態では、可視情報 $V(i)$ )とに基づいて推定された光伝達情報(実施形態では、光伝達マップ)とに基づいて、入力画像に対応する出力画像を生成する生成部と、生成部が生成した出力画像を出力する出力部とを備える、画像処理装置。

一構成例として、光源情報とは、入力画像に基づいて推定されたものである。

一構成例として、出力画像とは、光源情報が示す照明状態に基づいて、入力画像に含ま

50

れる被写体を再照明した再照明画像である。

一構成例として、反射率情報と光伝達情報とは、機械学習モデル（実施形態では、CNN）によって、入力画像に基づいて推定されたものである。

一構成例として、機械学習モデルとは、機械学習モデルによって、入力画像に基づいて推定される光源情報と、反射率情報と、光伝達情報とのいずれか又はいずれか同士の組み合わせから得られる画像と、所定の正解画像とを比較することによって学習されたものである。

一構成例として、可視情報は、物体表面の複数の点の各々について、光の方向毎に、光が複数の前記点の各々に届くか否かに基づいて作成される。

#### 【0070】

以上、本発明の実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更、組合せを行うことができる。これら実施形態及びその変形例は、発明の範囲や要旨に含まれると同時に、特許請求の範囲に記載された発明とその均等の範囲に含まれるものである。

なお、前述の画像処理装置100は内部にコンピュータを有している。そして、前述した各装置の各処理の過程は、プログラムの形式でコンピュータ読み取り可能な記録媒体に記憶されており、このプログラムをコンピュータが読み出して実行することによって、上記処理が行われる。ここでコンピュータ読み取り可能な記録媒体とは、磁気ディスク、光磁気ディスク、CD-ROM、DVD-ROM、半導体メモリなどをいう。また、このコンピュータプログラムを通信回線によってコンピュータに配信し、この配信を受けたコンピュータが当該プログラムを実行するようにしてもよい。

また、上記プログラムは、前述した機能の一部を実現するためのものであってもよい。さらに、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるもの、いわゆる差分ファイル（差分プログラム）であってもよい。

#### 【符号の説明】

#### 【0071】

1 ... 画像処理システム、 10、10-1、10-2、・・・、10-n ... 端末装置、  
 50 ... ネットワーク、 70 ... サーバ装置、 100 ... 画像処理装置、 105 ... 通信部、  
 110 ... 記憶部、 111 ... プログラム、 112 ... アプリ、 113 ... 反射率マップ、  
 114 ... 光源情報、 115 ... 光伝達マップ、 120 ... 操作部、 130 ... 情報処理部、  
 131 ... 取得部、 132 ... 分析部、 133 ... 機械学習部、 134 ... 生成部、  
 135 ... 出力部、 140 ... 表示部、 202 ... エンコーダ、 203 ... 反射率マップデコーダ、  
 204 ... ResNetブロック、 205 ... 逆畳み込み部、 206 ... 光伝達マップデコーダ、  
 207 ... ResNetブロック、 208 ... 逆畳み込み部、 209 ... 連結部、  
 210 ... 畳み込み部、 211 ... 入力画像、 212 ... 人物全身画像I、  
 213 ... 二値マスクM、 214 ... 正解データ、 215 ... 反射率マップ、  
 216 ... 光源情報、 217 ... 光伝達マップ、 220 ... 推定値、  
 221 ... 反射率マップ、 222 ... 光源情報、 223 ... 光伝達マップ、 224 ... 乗算部、  
 225 ... 人物全身画像I

10

20

30

40

【 図 1 】

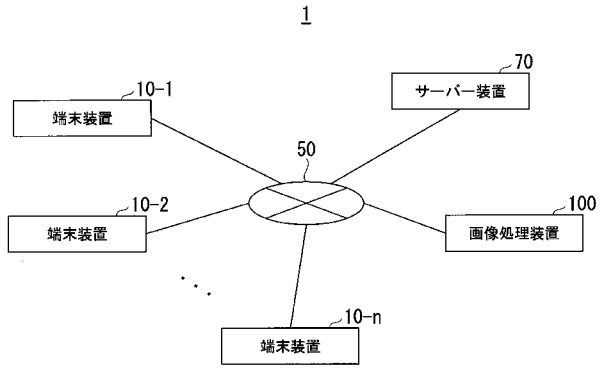


図 1

【 図 2 】

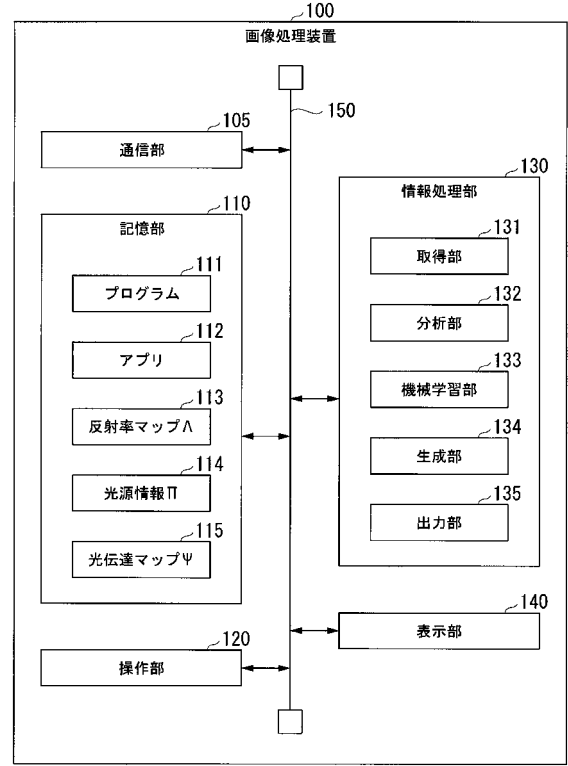


図 2

【 図 3 】

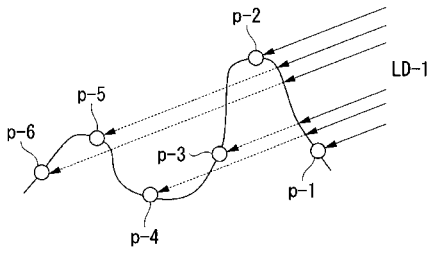


図 3

【 図 4 】

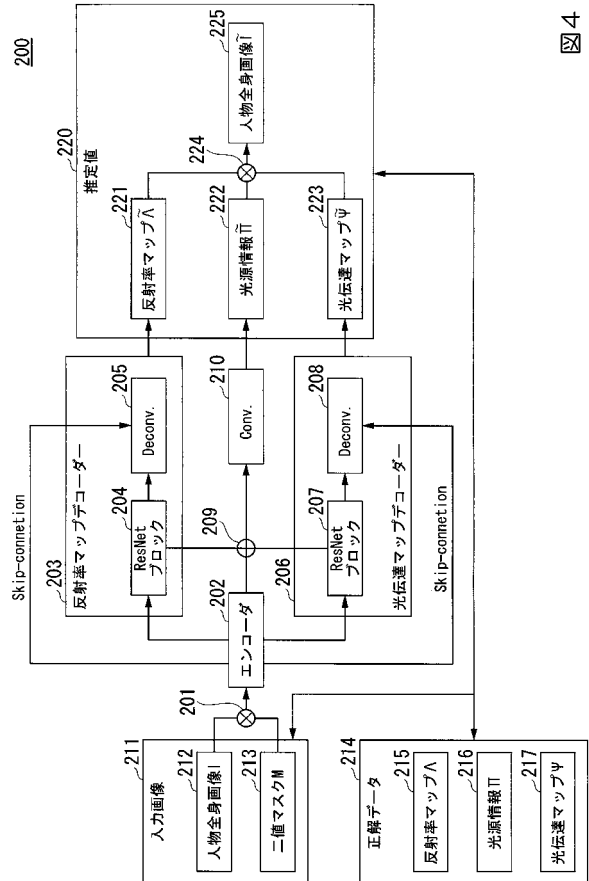


図 4

【 図 5 】

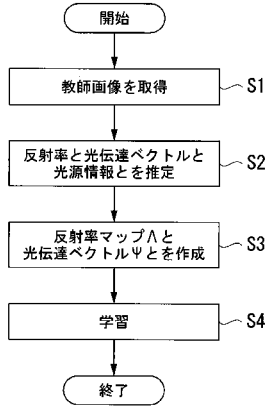


図 5

【 図 6 】

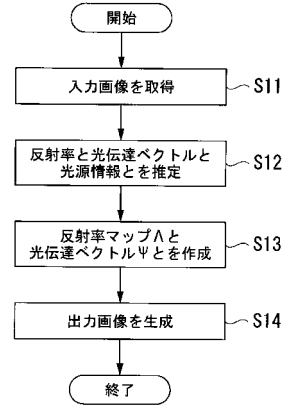


図 6

【 図 7 】



図 7

【 図 8 】

各手法の推定結果に対するマスク内のRMSE

	反射率マップ	法線マップ	光伝達マップ	光源情報	陰影マップ	人物全身画像
従来手法	0.168±0.063	0.360±0.052	0.522±0.053	0.202±0.076	0.289±0.123	0.104±0.057
本手法 (損失関数4種類)	0.167±0.048	N/A	0.442±0.068	0.210±0.068	0.242±0.113	0.088±0.039
本手法 (損失関数15種類)	0.162±0.049	N/A	0.433±0.065	0.203±0.075	0.233±0.111	0.084±0.044

図 8

【 図 9 】

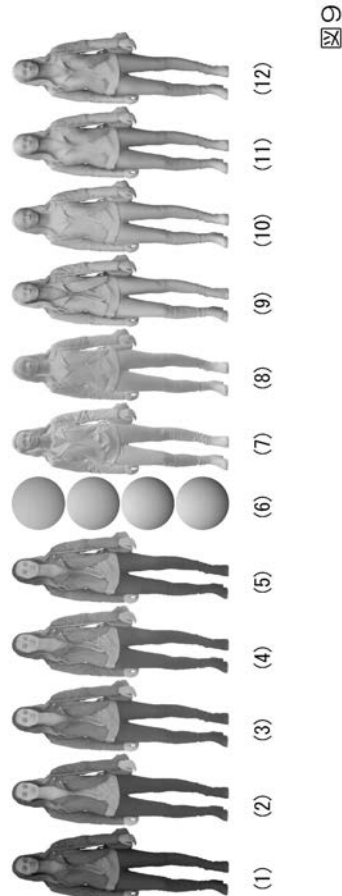


図 9

【 図 1 0 】

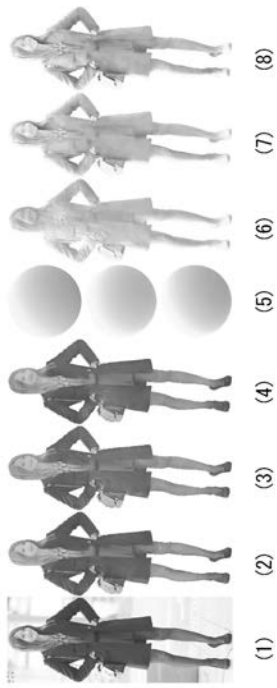


図10

【 図 1 1 】

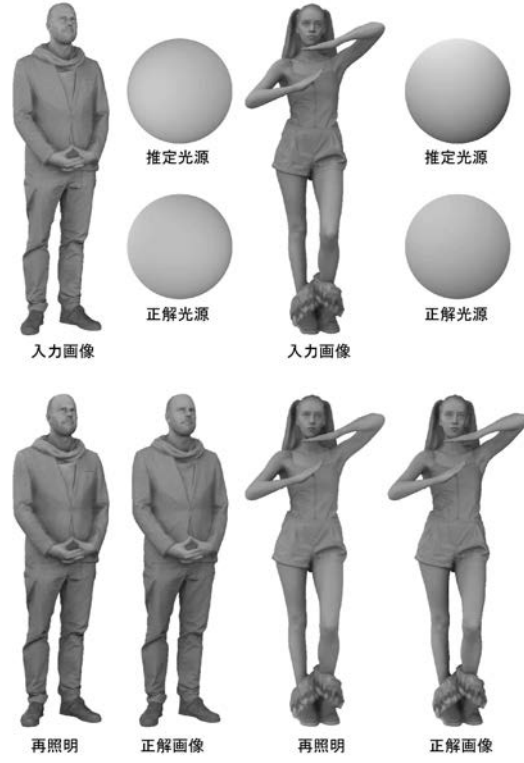


図 1 1

フロントページの続き

Fターム(参考) 5B080 AA19 CA00 DA07 DA08 GA13  
5L096 DA01 GA49 HA11 MA03