

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6810247号
(P6810247)

(45) 発行日 令和3年1月6日(2021.1.6)

(24) 登録日 令和2年12月14日(2020.12.14)

(51) Int. Cl. F I
G06F 16/483 (2019.01) GO 6 F 16/483
G06F 16/432 (2019.01) GO 6 F 16/432
G06T 1/00 (2006.01) GO 6 T 1/00 2 0 0 E

請求項の数 20 (全 32 頁)

(21) 出願番号	特願2019-507781 (P2019-507781)	(73) 特許権者	518221025 アキフィ、インコーポレイティド
(86) (22) 出願日	平成29年8月11日 (2017. 8. 11)		アメリカ合衆国、カリフォルニア 943
(65) 公表番号	特表2019-530062 (P2019-530062A)		03, パロアルト, イーストベイショ
(43) 公表日	令和1年10月17日 (2019. 10. 17)		アロード 2225, スイート 110
(86) 国際出願番号	PCT/US2017/046642	(74) 代理人	100099759 弁理士 青木 篤
(87) 国際公開番号	W02018/031959	(74) 代理人	100123582 弁理士 三橋 真二
(87) 国際公開日	平成30年2月15日 (2018. 2. 15)	(74) 代理人	100114018 弁理士 南山 知広
審査請求日	平成31年3月26日 (2019. 3. 26)	(74) 代理人	100119987 弁理士 伊坪 公一
(31) 優先権主張番号	62/374, 598		
(32) 優先日	平成28年8月12日 (2016. 8. 12)		
(33) 優先権主張国・地域又は機関	米国 (US)		

最終頁に続く

(54) 【発明の名称】 メディアドキュメントのメタデータを自動的に生成するシステム及び方法

(57) 【特許請求の範囲】

【請求項1】

メディアドキュメントのメタデータを自動的に生成する方法であって、
 コンピュータによって、畳み込みニューラルネットワークを用いて前記メディアドキュメントの特徴ベクトルを算出することであって、前記メディアドキュメントは、3次元(3D)モデルであり、前記特徴ベクトルを算出することは、

前記コンピュータによって、前記3Dモデルの複数の2次元(2D)ビューをレンダリングするために複数の角度から前記3Dモデルをレンダリングすることと、

前記コンピュータによって、一つ以上の畳み込みニューラルネットワークを用いて、各々が前記3Dモデルの2Dビューの一つに対応する複数の単一ビュー特徴ベクトルを生成することと、

統合された特徴ベクトルを算出するために、前記コンピュータによって、前記2Dビューに対応する前記複数の単一ビュー特徴ベクトルを集約することであって、前記統合された特徴ベクトルは、前記2Dビューの数に依存しない一定の長さを有することと、

前記コンピュータによって、最終段階の畳み込みニューラルネットワークを用いて、前記統合された特徴ベクトルから前記特徴ベクトルを算出することと、
 を備えることと、

前記コンピュータによって、前記メディアドキュメントの前記特徴ベクトルに類似する対応する特徴ベクトルを有する一つ以上のマッチングメディアドキュメントの一群のメディアドキュメントを検索することであって、前記一群のメディアドキュメントの各メディア

アドキュメントがメタデータに関連することと、

前記コンピュータによって、前記一つ以上のマッチングメディアドキュメントに関連するメタデータに基づいて前記メディアドキュメントのメタデータを生成することであって、生成されたメタデータは、前記一つ以上のマッチングメディアドキュメントの数なくとも一つのしきい値のメタデータに出現する前記メタデータの部分を含むことと、

前記コンピュータによって、生成したメタデータに関連して前記メディアドキュメントを表示することと、

を備える方法。

【請求項 2】

前記一群のメディアドキュメントは、一群の 2 次元 (2 D) 画像を備え、

前記特徴ベクトルを算出することは、前記コンピュータによって、前記 3 D モデルの一つ以上の特徴ベクトルを算出することを備え、

前記一つ以上のマッチングメディアドキュメントは、前記一群の 2 D 画像の一つ以上のマッチング 2 D 画像を備え、

前記マッチング 2 D 画像の対応する特徴ベクトルの各々は、前記 3 D モデルの一つ以上の特徴ベクトルの少なくとも一つに類似する請求項 1 に記載の方法。

【請求項 3】

前記コンピュータによって、3 次元スキャナを用いて前記 3 D モデルをキャプチャすることを更に備え、前記 3 次元スキャナは、

二つ以上の赤外 (I R) カメラと、

一つ以上の I R コリメート照明と、

を備える請求項 1 に記載の方法。

【請求項 4】

前記コンピュータによって、生成したメタデータの複数の欄の一つ以上の検証を備えるユーザ入力を受け取ることを更に備える請求項 1 に記載の方法。

【請求項 5】

生成したメタデータ及び前記メディアドキュメントの各々に関連するメタデータはそれぞれ、複数の欄を備え、

前記メタデータを生成することは、

前記コンピュータによって、前記複数の欄の各欄に対して、前記一つ以上のマッチングメディアドキュメントの二つ以上のしきい値のメタデータに出現するデータを識別することと、

前記コンピュータによって、識別したデータを生成したメタデータに追加することと、を備える請求項 1 に記載の方法。

【請求項 6】

生成したメタデータは、メディアドキュメントの種類を備え、

前記コンピュータによって、前記メディアドキュメントの種類を算出するために前記メディアドキュメントの特徴ベクトルを分類器に供給することと、

前記コンピュータによって、分類に基づいて前記種類を算出することと、

を備える請求項 1 に記載の方法。

【請求項 7】

前記コンピュータによって、前記メディアドキュメント及び生成したメタデータメタデータを前記一群のメディアドキュメントに追加することを更に備える請求項 1 に記載の方法。

【請求項 8】

前記一つ以上のマッチングメディアドキュメントの一つ以上の対応する特徴ベクトルの各々は、多次元空間における前記メディアドキュメントの特徴ベクトルからの対応する距離を有し、

前記対応する距離の各々は、しきい値距離未満である請求項 1 に記載の方法。

【請求項 9】

10

20

30

40

50

前記コンピュータによって、前記メディアドキュメントに関連するメタデータを受け取ることを更に備え、

前記一群のメディアドキュメントを検索することは、前記コンピュータによって、前記メディアドキュメントに関連するメタデータの複数の欄のうちの一つにマッチングするメタデータを有する前記一群のメディアドキュメントの一つ以上のメディアドキュメントを識別することを備え、

前記マッチングメディアドキュメントは、識別した一つ以上のメディアドキュメントの少なくとも一つを備える請求項 1 に記載の方法。

【請求項 10】

前記一群のメディアドキュメントは、第 1 のセットのメディアドキュメント及び第 2 のセットのメディアドキュメントを備え、

ユーザに関連するアクセスポリシーによって、前記第 1 のセットのメディアドキュメントへのアクセスを許可するとともに前記第 2 のセットのメディアドキュメントへのアクセスを拒否し、

前記一群のメディアドキュメントを検索することは、前記第 1 のセットのメディアドキュメントに限定される請求項 1 に記載の方法。

【請求項 11】

メディアドキュメントのメタデータを自動的に生成するシステムであって、
プロセッサと、

前記プロセッサによって実行されるときに、

畳み込みニューラルネットワークを用いて前記メディアドキュメントの特徴ベクトルを算出することであって、前記メディアドキュメントは、3次元(3D)モデルであり、前記特徴ベクトルを算出することは、

前記3Dモデルの複数の2次元(2D)ビューをレンダリングするために複数の角度から前記3Dモデルをレンダリングすることと、

一つ以上の畳み込みニューラルネットワークを用いて、各々が前記3Dモデルの2Dビューの一つに対応する複数の単一ビュー特徴ベクトルを生成することと、

統合された特徴ベクトルを算出するために、前記2Dビューに対応する前記複数の単一ビュー特徴ベクトルを集約することであって、前記統合された特徴ベクトルは、前記2Dビューの数に依存しない一定の長さを有することと、

最終段階の畳み込みニューラルネットワークを用いて、前記統合された特徴ベクトルから前記特徴ベクトルを算出することと、

を備えることと、

前記メディアドキュメントの前記特徴ベクトルに類似する対応する特徴ベクトルを有する一つ以上のマッチングメディアドキュメントの一群のメディアドキュメントを検索することであって、前記一群のメディアドキュメントの各メディアドキュメントがメタデータに関連することと、

前記一つ以上のマッチングメディアドキュメントに関連するメタデータに基づいて前記メディアドキュメントのメタデータを生成することであって、生成されたメタデータは、前記一つ以上のマッチングメディアドキュメントの数なくとも一つのしきい値のメタデータに出現する前記メタデータの部分を含むことと、

生成したメタデータに関連して前記メディアドキュメントを表示することと、

を前記プロセッサによって実行させる格納された命令を有するメモリと、

を備えるシステム。

【請求項 12】

前記一群のメディアドキュメントは、一群の2次元(2D)画像を備え、

前記特徴ベクトルを算出することの命令は、前記プロセッサによって実行されるときに、前記3Dモデルの一つ以上の特徴ベクトルを算出することを前記プロセッサによって実行させる命令を備え、

前記一つ以上のマッチングメディアドキュメントは、前記一群の2D画像の一つ以上の

10

20

30

40

50

マッチング 2 D 画像を備え、

前記マッチング 2 D 画像の対応する特徴ベクトルの各々は、前記 3 D モデルの一つ以上の特徴ベクトルの少なくとも一つに類似する請求項 1 1 に記載のシステム。

【請求項 1 3】

二つ以上の赤外 (I R) カメラと、

一つ以上の I R コリメート照明と、

を備える 3 次元スキャナを更に備え、前記メモリは、前記プロセッサによって実行されるときに、3 次元スキャナを用いて前記 3 D モデルをキャプチャすることを前記プロセッサによって実行させる命令を更に格納する請求項 1 1 に記載のシステム。

【請求項 1 4】

前記メモリは、前記プロセッサによって実行されるときに、生成したメタデータの複数の欄の一つ以上の検証を備えるユーザ入力を受け取することを前記プロセッサによって実行させる命令を更に格納する請求項 1 1 に記載のシステム。

【請求項 1 5】

生成したメタデータ及び前記メディアドキュメントの各々に関連するメタデータはそれぞれ、複数の欄を備え、

前記メタデータを生成することの命令は、前記プロセッサによって実行されるときに、

前記複数の欄の各欄に対して、前記一つ以上のマッチングメディアドキュメントの二つ以上のしきい値のメタデータに出現するデータを識別することと、

識別したデータを生成したメタデータに追加することと、

を前記プロセッサによって実行させる命令を備える請求項 1 1 に記載のシステム。

【請求項 1 6】

生成したメタデータは、メディアドキュメントの種類を備え、

前記メモリは、前記プロセッサによって実行されるときに、

前記メディアドキュメントの種類を算出するために前記メディアドキュメントの特徴ベクトルを分類器に供給することと、

分類に基づいて前記種類を算出することと、

を前記プロセッサによって実行させる命令を更に格納する請求項 1 1 に記載のシステム。

【請求項 1 7】

前記メモリは、前記プロセッサによって実行されるときに、前記メディアドキュメント及び生成したメタデータメタデータを前記一群のメディアドキュメントに追加することを前記プロセッサによって実行させる命令を更に格納する請求項 1 1 に記載のシステム。

【請求項 1 8】

前記一つ以上のマッチングメディアドキュメントの一つ以上の対応する特徴ベクトルの各々は、多次元空間における前記メディアドキュメントの特徴ベクトルからの対応する距離を有し、

前記対応する距離の各々は、しきい値距離未満である請求項 1 1 に記載のシステム。

【請求項 1 9】

前記メモリは、前記プロセッサによって実行されるときに、前記メディアドキュメントに関連するメタデータを受け取することを前記プロセッサによって実行させる命令を更に格納し、

前記一群のメディアドキュメントを検索することの命令は、前記プロセッサによって実行されるときに、前記メディアドキュメントに関連するメタデータの複数の欄のうち少なくとも一つにマッチングするメタデータを有する前記一群のメディアドキュメントの一つ以上のメディアドキュメントを識別することを前記プロセッサによって実行させる命令を備え、

前記マッチングメディアドキュメントは、識別した一つ以上のメディアドキュメントの少なくとも一つを備える請求項 1 1 に記載のシステム。

【請求項 2 0】

10

20

30

40

50

前記一群のメディアドキュメントは、第1のセットのメディアドキュメント及び第2のセットのメディアドキュメントを備え、

ユーザに関連するアクセスポリシーによって、前記第1のセットのメディアドキュメントへのアクセスを許可するとともに前記第2のセットのメディアドキュメントへのアクセスを拒否し、

前記一群のメディアドキュメントを検索することは、前記第1のセットのメディアドキュメントに限定される請求項11に記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

この出願は、2016年8月12日に米国特許庁に出願された米国仮出願第62/374,598号の利益を主張し、その全開示を参照によりここに組み込む。

【0002】

本発明の実施の形態の態様は、画像及び3Dモデルタグ付けを含むデジタルメディアタグ付け及び分類の分野並びに関連のユーザインタフェースシステムに関する。

【背景技術】

【0003】

画像、録音及びビデオのような種々のタイプのデジタルメディアドキュメント(digital media document)は、デジタルメディアドキュメントのテキスト記述を含むメタデータを含む。このメディアデータは、タイムスタンプ並びに関連のデジタルメディアドキュメントのコンテンツの文書による記述(narrative description)のような情報を含んでもよい。これらのメタデータを、検索クエリに関連するメディアドキュメントを識別するためにテキストベース検索を行うときに用いることができる。メタデータを、デジタルメディアドキュメントをカテゴリー化又は分類するのに用いることもできる。

【0004】

一例として、デジタル写真に関連するメタデータは、画像の(画素の)寸法及び画像の色深度(color depth)のような技術情報並びに写真のコンテンツを記述するメタデータを含むことができる。森のハイカーの写真に関連するメタデータは、「林」、「森」、「木」、「植物」、「葉(green)」、「影」、「花」、「遠足(hike)」、「ハイカー」、「ハイキング」等のような写真のコンテンツのテキスト記述を含むことができる。

【0005】

3次元(3D)モデルは、デジタルメディアドキュメントの他の形態であり、広告アプリケーション、品質管理アプリケーション、ビデオゲームアプリケーション、仮想現実アプリケーション及び拡張現実アプリケーションのようなコンテンツで一般的に用いられている。従来、3Dモデルを、例えば、コンピュータ支援設計(CAD)ツールを用いることによって手動で生成していた。これらのモデルを作成することは、一般的には、詳細なモデルを作成するときに特にたくさんの人手を要する仕事である。近年、深度カメラ及び3Dスキャナが代替的なアプローチとして設けられ、この場合、詳細な3Dモデルを、実在する対象物の形状のデジタルキャプチャを通じて作成することができる。

【0006】

CADツール、深度カメラ及び3Dスキャナを用いて生成した三次元モデルは、メタデータタグ付けから利益を得ることもできる。例えば、3Dモデルを、電子商取引ウェブサイトの商品リスティングに用いることができ、3Dモデルに関連するメタデータを、関連の商品リスティングの情報を生成又は追加するのに用いることができる。例えば、車両の3Dモデルは、車の型及び様式、内部及び外観の色、状態(例えば、へこみ又は損傷の存在)等のようなメタデータ情報を含むことができる。これらのメタデータを、買い物客が特定の色の特定のタイプの車を検索できるようにするために商品リスティングの情報に追加することができる。

10

20

30

40

50

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、各3Dモデルのこれらのメタデータを手動で生成することは、時間がかかり、一貫性がなくまとまりの悪い結果を生じさせるおそれがある。例えば、車が「赤色」のものとしてタグ付けされるのに対して、赤の色合いの実際の実名称が「ルビー」となることがある。特定の型の名称の綴りのように車両についての他の情報が不正確となることがある。

【0008】

本発明の実施の形態の態様は、メタデータによる3次元(3D)モデルの自動的な増強及びこれらの3Dモデルを収集及び生成するシステムを対象とする。

10

【課題を解決するための手段】

【0009】

本発明の一実施の形態によれば、メディアドキュメントのメタデータを自動的に生成する方法であって、方法は、畳み込みニューラルネットワークを用いてメディアドキュメントの特徴ベクトルを算出することと、メディアドキュメントの特徴ベクトルに類似する対応する特徴ベクトルを有する一つ以上のマッチング(matching)メディアドキュメントの一群のメディアドキュメントを検索することであって、一群のメディアドキュメントの各メディアドキュメントがメタデータに関連することと、一つ以上のマッチングメディアドキュメントに関連するメタデータに基づいてメディアドキュメントのメタデータを生成することと、生成したメタデータに関連してメディアドキュメントを表示することと、を備える。

20

【0010】

メディアドキュメントは、3次元(3D)モデルであってもよい。特徴ベクトルを算出することは、有界3Dモデルを生成するために3Dモデルの周辺に境界ボックスを規定することと、複数のボクセルを算出するために有界3Dモデルをボクセル化することと、ボクセルから特徴ベクトルを生成することと、を有してもよい。一群のメディアドキュメントは、一群の3次元(3D)モデルを有してもよく、一群のメディアドキュメントを検索することは、反復的な最近点技術に従って3Dモデルと一群の3Dモデルの一つの間の距離を算出することを有してもよい。特徴ベクトルを算出することは、3Dモデルの複数の2次元(2D)ビューをレンダリングするために複数の角度から3Dモデルをレンダリングすることと、一つ以上の畳み込みニューラルネットワークを用いて、各々が3Dモデルの2Dビューの一つに対応する複数の単一ビュー特徴ベクトルを生成することと、統合された特徴ベクトルを算出するために複数の単一ビュー特徴ベクトルを統合することと、最終段階の畳み込みニューラルネットワークを用いて、統合された特徴ベクトルから特徴ベクトルを算出することと、を有してもよい。一群のメディアドキュメントは、一群の2次元(2D)画像を有してもよく、特徴ベクトルを算出することは、3Dモデルの一つ以上の特徴ベクトルを算出することを有してもよく、一つ以上のマッチングメディアドキュメントは、一群の2D画像の一つ以上のマッチング2D画像を有してもよく、マッチング2D画像の対応する特徴ベクトルの各々は、3Dモデルの一つ以上の特徴ベクトルの少なくとも一つに類似してもよい。方法は、3次元スキャナを用いて3Dモデルをキャプチャすることを更に有してもよく、3次元スキャナは、二つ以上の赤外(IR)カメラと、一つ以上のIRコリメート照明と、を有してもよい。

30

40

【0011】

メディアドキュメントは、一つ以上の2次元(2D)画像を有してもよく、一群のメディアドキュメントは、一群の3次元(3D)モデルを有してもよく、特徴ベクトルを算出することは、一つ以上の2D画像の一つ以上の特徴ベクトルを算出することを有してもよく、一つ以上のマッチングメディアドキュメントは、一群の3Dモデルの一つ以上のマッチング3Dモデルを有してもよく、マッチング3Dモデルの対応する特徴ベクトルの各々は、3Dモデルの一つ以上の特徴ベクトルの少なくとも一つに類似してもよい。特徴ベク

50

トルを算出することは、一つ以上の2D画像から物の3Dモデルを算出することと、3Dモデルから特徴ベクトルを抽出することと、を有してもよい。3Dモデルから特徴ベクトルを抽出することは、有界3Dモデルを生成するために3Dモデルの周辺に境界ボックスを規定することと、複数のボクセルを算出するために有界3Dモデルをボクセル化することと、ボクセルから特徴ベクトルを生成することと、を有してもよい。

【0012】

3Dモデルから特徴ベクトルを抽出することは、3Dモデルの複数の2次元(2D)ビューをレンダリングするために複数の角度から3Dモデルをレンダリングすることと、一つ以上の畳み込みニューラルネットワークを用いて、各々が3Dモデルの2Dビューの一つに対応する複数の単一ビュー特徴ベクトルを生成することと、統合された特徴ベクトルを算出するために複数の単一ビュー特徴ベクトルを統合することと、最終段階の畳み込みニューラルネットワークを用いて、統合された特徴ベクトルから特徴ベクトルを算出することと、を有してもよい。

10

【0013】

方法は、生成したメタデータの複数の欄の一つ以上の検証を備えるユーザ入力を受け取ることとを更に有してもよい。

【0014】

生成したメタデータ及びメディアドキュメントの各々に関連するメタデータはそれぞれ、複数の欄を有してもよく、メタデータを生成することは、複数の欄の各欄に対して、一つ以上のマッチングメディアドキュメントの二つ以上のしきい値のメタデータに出現するデータを識別することと、識別したデータを生成したメタデータに追加することと、を有してもよい。

20

【0015】

生成したメタデータは、メディアドキュメントの種類を有してもよく、方法は、メディアドキュメントの種類を算出するためにメディアドキュメントの特徴ベクトルを分類器に供給することと、分類に基づいて種類を算出することと、を有してもよい。

【0016】

方法は、メディアドキュメント及び生成したメタデータメタデータを一群のメディアドキュメントに追加することとを更に有してもよい。

【0017】

一つ以上のマッチングメディアドキュメントの一つ以上の対応する特徴ベクトルの各々は、多次元空間におけるメディアドキュメントの特徴ベクトルからの対応する距離を有してもよく、対応する距離の各々は、しきい値距離未満であってもよい。

30

【0018】

方法は、メディアドキュメントに関連するメタデータを受け取ることとを更に有してもよく、一群のメディアドキュメントを検索することは、メディアドキュメントに関連するメタデータの複数の欄のうち少なくとも一つにマッチングするメタデータを有する一群のメディアドキュメントの一つ以上のメディアドキュメントを識別することとを有してもよく、マッチングメディアドキュメントは、識別した一つ以上のメディアドキュメントの少なくとも一つを有してもよい。

40

【0019】

一群のメディアドキュメントは、第1のセットのメディアドキュメント及び第2のセットのメディアドキュメントを有してもよく、ユーザに関連するアクセスポリシーによって、第1のセットのメディアドキュメントへのアクセスを許可するとともに第2のセットのメディアドキュメントへのアクセスを拒否し、一群のメディアドキュメントを検索することは、第1のセットのメディアドキュメントに限定される。

【0020】

本発明の一実施の形態によれば、メディアドキュメントのメタデータを自動的に生成するシステムであって、プロセッサと、プロセッサによって実行されるときに、畳み込みニューラルネットワークを用いてメディアドキュメントの特徴ベクトルを算出することと、

50

メディアドキュメントの特徴ベクトルに類似する対応する特徴ベクトルを有する一つ以上のマッチングメディアドキュメントの一群のメディアドキュメントを検索することと、一群のメディアドキュメントの各メディアドキュメントがメタデータに関連することと、一つ以上のマッチングメディアドキュメントに関連するメタデータに基づいてメディアドキュメントのメタデータを生成することと、生成したメタデータに関連してメディアドキュメントを表示することと、をプロセッサによって実行させる格納された命令を有するメモリと、を有する。

【0021】

メディアドキュメントは、3次元(3D)モデルであってもよい。特徴ベクトルを算出することの命令は、プロセッサによって実行されるときに、有界3Dモデルを生成するために3Dモデルの周辺に境界ボックスを規定することと、複数のボクセルを算出するために有界3Dモデルをボクセル化することと、ボクセルから特徴ベクトルを生成することと、をプロセッサによって実行させる命令を有してもよい。一群のメディアドキュメントは、一群の3次元(3D)モデルを有してもよく、一群のメディアドキュメントを検索することの命令は、プロセッサによって実行されるときに、反復的な最近点技術に従って3Dモデルと一群の3Dモデルの一つの間の距離を算出することをプロセッサによって実行させる命令を有してもよい。特徴ベクトルを算出することの命令は、プロセッサによって実行されるときに、3Dモデルの複数の2次元(2D)ビューをレンダリングするために複数の角度から3Dモデルをレンダリングすることと、一つ以上の畳み込みニューラルネットワークを用いて、各々が3Dモデルの2Dビューの一つに対応する複数の単一ビュー特徴ベクトルを生成することと、統合された特徴ベクトルを算出するために複数の単一ビュー特徴ベクトルを統合することと、最終段階の畳み込みニューラルネットワークを用いて、統合された特徴ベクトルから特徴ベクトルを算出することと、をプロセッサによって実行させる命令を有してもよい。一群のメディアドキュメントは、一群の2次元(2D)画像を有してもよく、特徴ベクトルを算出することの命令は、プロセッサによって実行されるときに、3Dモデルの一つ以上の特徴ベクトルを算出することをプロセッサによって実行させる命令を有してもよく、一つ以上のマッチングメディアドキュメントは、一群の2D画像の一つ以上のマッチング2D画像を有してもよく、マッチング2D画像の対応する特徴ベクトルの各々は、3Dモデルの一つ以上の特徴ベクトルの少なくとも一つに類似してもよい。システムは、二つ以上の赤外(IR)カメラと、一つ以上のIRコリメート照明と、を備える3次元スキャナを更に有してもよく、メモリは、プロセッサによって実行されるときに、3次元スキャナを用いて3Dモデルをキャプチャすることをプロセッサによって実行させる命令を更に格納してもよい。

【0022】

メディアドキュメントは、一つ以上の2次元(2D)画像を有してもよく、一群のメディアドキュメントは、一群の3次元(3D)モデルを有してもよく、特徴ベクトルを算出することの命令は、プロセッサによって実行されるときに、一つ以上の2D画像の一つ以上の特徴ベクトルを算出することをプロセッサによって実行させる命令を有してもよく、一つ以上のマッチングメディアドキュメントは、一群の3Dモデルの一つ以上のマッチング3Dモデルを有してもよく、マッチング3Dモデルの対応する特徴ベクトルの各々は、3Dモデルの一つ以上の特徴ベクトルの少なくとも一つに類似してもよい。特徴ベクトルを算出することの命令は、プロセッサによって実行されるときに、一つ以上の2D画像から物の3Dモデルを算出することと、3Dモデルから特徴ベクトルを抽出することと、をプロセッサによって実行させる命令を有してもよい。3Dモデルから特徴ベクトルを抽出することの命令は、プロセッサによって実行されるときに、有界3Dモデルを生成するために3Dモデルの周辺に境界ボックスを規定することと、複数のボクセルを算出するために有界3Dモデルをボクセル化することと、ボクセルから特徴ベクトルを生成することと、をプロセッサによって実行させる命令を有してもよい。3Dモデルから特徴ベクトルを抽出することの命令は、プロセッサによって実行されるときに、3Dモデルの複数の2次元(2D)ビューをレンダリングするために複数の角度から3Dモデルをレンダリングす

10

20

30

40

50

ることと、一つ以上の畳み込みニューラルネットワークを用いて、各々が3Dモデルの2Dビューの一つに対応する複数の単一ビュー特徴ベクトルを生成することと、統合された特徴ベクトルを算出するために複数の単一ビュー特徴ベクトルを統合することと、最終段階の畳み込みニューラルネットワークを用いて、統合された特徴ベクトルから特徴ベクトルを算出することと、をプロセッサによって実行させる命令を有してもよい。

【0023】

メモリは、プロセッサによって実行されるときに、生成したメタデータの複数の欄の一つ以上の検証を備えるユーザ入力を受け取ることをプロセッサによって実行させる命令を更に格納してもよい。

【0024】

生成したメタデータ及びメディアドキュメントの各々に関連するメタデータはそれぞれ、複数の欄を有してもよく、メタデータを生成することの命令は、プロセッサによって実行されるときに、複数の欄の各欄に対して、一つ以上のマッチングメディアドキュメントの二つ以上のしきい値のメタデータに出現するデータを識別することと、識別したデータを生成したメタデータに追加することと、をプロセッサによって実行させる命令を有してもよい。

【0025】

生成したメタデータは、メディアドキュメントの種類を有してもよく、メモリは、プロセッサによって実行されるときに、メディアドキュメントの種類を算出するためにメディアドキュメントの特徴ベクトルを分類器に供給することと、分類に基づいて種類を算出することと、をプロセッサによって実行させる命令を更に格納してもよい。

【0026】

メモリは、プロセッサによって実行されるときに、メディアドキュメント及び生成したメタデータメタデータを一群のメディアドキュメントに追加することをプロセッサによって実行させる命令を更に格納してもよい。

【0027】

一つ以上のマッチングメディアドキュメントの一つ以上の対応する特徴ベクトルの各々は、多次元空間におけるメディアドキュメントの特徴ベクトルからの対応する距離を有してもよく、対応する距離の各々は、しきい値距離未満であってもよい。

【0028】

メモリは、プロセッサによって実行されるときに、メディアドキュメントに関連するメタデータを受け取ることをプロセッサによって実行させる命令を更に格納してもよく、一群のメディアドキュメントを検索することの命令は、プロセッサによって実行されるときに、メディアドキュメントに関連するメタデータの複数の欄のうちの少なくとも一つにマッチングするメタデータを有する一群のメディアドキュメントの一つ以上のメディアドキュメントを識別することをプロセッサによって実行させる命令を有してもよく、マッチングメディアドキュメントは、識別した一つ以上のメディアドキュメントの少なくとも一つを有してもよい。

【0029】

一群のメディアドキュメントは、第1のセットのメディアドキュメント及び第2のセットのメディアドキュメントを有してもよく、ユーザに関連するアクセスポリシーによって、第1のセットのメディアドキュメントへのアクセスを許可するとともに第2のセットのメディアドキュメントへのアクセスを拒否し、一群のメディアドキュメントを検索することは、第1のセットのメディアドキュメントに限定される。

【0030】

特許又は出願ファイルは、少なくとも一つのカラー図面を含む。(一つ以上の)カラー図面を有するこの特許又は出願刊行物は、請求及び必要な費用の支払いを行うことによって特許庁から提供される。

【0031】

添付図面は、明細書と共に本発明の例示的な実施の形態を説明し、明細書共に本発明の

10

20

30

40

50

原理を説明する役割を果たす。

【図面の簡単な説明】

【0032】

【図1】本発明の一実施の形態によるメタデータのタグ及び種類の自動的な追加の例である。

【図2】本発明の一実施の形態によるスキャナシステムのブロック図である。

【図3】本発明の一実施の形態によるスクリーンから切り離されているカメラを有する3Dスキャニングシステムを用いて物をスキャンするシステム及び方法をユーザの観点から示す。

【図4A】スキャニングシステムによってキャプチャした3次元モデルの例である。

10

【図4B】スキャニングシステムによってキャプチャした3次元モデルの例である。

【図5】本発明の一実施の形態によるメディアドキュメントのメタデータを生成する方法のフローチャートである。

【図6A】本発明の一実施の形態による3Dモデルをボクセル化することによりメディアドキュメントが3Dモデルである場合に特徴ベクトルを算出する方法を示すフローチャートである。

【図6B】ボクセル化を用いた3DモデルへのCNNの適用を視覚的に示す。

【図7A】本発明に一実施の形態による種々の視点又は方向からの3Dモデルの複数のビューをレンダリングすることによりメディアドキュメントが3Dモデルである場合に特徴ベクトルを算出する方法を示すフローチャートである。

20

【図7B】複数のビューからレンダリングした3DモデルへのCNNの適用を視覚的に示す。

【図8】本発明の一実施の形態によるモデルをキャプチャするとともにモデルのメタデータを自動的に生成する工程を示すフローチャートである。

【発明を実施するための形態】

【0033】

以下の詳細な説明において、本発明の所定の例示的な実施の形態のみを例示として示すとともに説明する。当業者が認識するように、本発明を、複数の種々の形態で実施してもよく、ここで説明する実施の形態に限定されるものと解釈すべきでない。明細書を通じて、同様な参照番号は同様な要素を表す。

30

【0034】

本発明の実施の形態の態様は、2次元(2D)画像又は3次元(3D)モデルのようなメディアドキュメントのメタデータを自動的に生成するシステム及び方法を対象とする。3Dモデルを、後に詳しく説明するように深度カメラを有するスキャナシステムによって生成してもよい。本発明の実施の形態の一部の態様は、特徴ベクトルを生成するとともに分類のセットに従ってメディアドキュメントを分類するようにトレーニングされた又は構成された分類器に特徴ベクトルを供給するためにメディアドキュメントを特徴抽出装置に供給することに関する。本発明の実施の形態の更に別の態様は、類似する分類(例えば、類似する特徴ベクトル)を有する存在するメディアドキュメントを識別するために(例えば、データベースに格納された)一群のメディアドキュメントを検索することに関する。類似する分類を有するこれらのメディアドキュメントに関連するメタデータを、供給されたメディアドキュメントのメタデータを生成するのに用いることができる。

40

【0035】

図1は、本発明の一実施の形態によるメタデータのタグ及び種類の自動的な追加の例である。図1に示すように、3Dモデル10は、クエリーとして供給されるメディアモデルとして用いられる。示す例において、3Dモデル10は、関連のメタデータを有しない。3Dモデル10を含むクエリーに基づいて、本発明の実施の形態は、クエリーに供給された3Dモデル10に類似する一連のメディアドキュメント(一連の3Dモデル)のエントリー12を識別する。識別されるエントリー12の各々は、種類14a及びタグ14bを含むメタデータ14に関連する。例えば、第1のエントリーは、class "car" 及

50

びtag “Jeep (登録商標), blue, sport” を有する青のスポーツユーティリティビークルのモデルを示す。第2のエントリーは、class “car” 及びtag “Jeep (登録商標), convertible, red” を有する赤のオープントップタイプのオフロード車のモデルを示す。第3のエントリーは、class “cap” 及びtag “black & white” を有する野球帽のモデルを示す。第4のエントリーは、class “car” 及びtag “convertible” 及び “sport” を有する緑のオープントップタイプの車のモデルを示す。識別されるエントリー12の少なくとも半分に出現するメタデータをマージすることによって、クエリーの供給された3Dモデル10に対してメタデータ16を生成することができる。

【0036】

このやり方は、スキャナシステムによって生成される3Dモデルの場合に特に有用となることができ、この場合、そのように生成された3Dモデルは、車を販売するための広告（例えば、オンラインのオークションウェブサイトの商品 리스팅）の車の3Dモデルを含むようにするために3Dスキャナシステムを用いて車の3Dモデルをキャプチャするユーザの場合のようにほとんどメタデータを有さなくてもよい。

【0037】

このやり方によって、システムは、物の画像又は3Dモデルをキャプチャするユーザがどんな物であるか又は物が何と呼ばれているかを知らないときでも物を自動的に分類するとともにタグ付けすることもできる。

【0038】

スキャナシステム

一般的には、スキャナシステムは、シーンからデータを収集するカメラのようなセンサを有するハードウェア装置を有する。スキャナシステムは、センサによって収集したデータからシーンの深度画像及び/又は3次元(3D)モデルを生成するコンピュータプロセッサ又は他の処理ハードウェアを有してもよい。

【0039】

スキャナシステムのセンサを、例えば、通常のカラカメラ、深度(又は距離)カメラ又は深度及びカラカメラの組合せを含む多様な種々のタイプのカメラとしてもよい。後者は、典型的にはRGB-Dと称され、RGBは、カラ画像を表し、Dは、深度画像を表す(この場合、各画素は、シーンの深度(又は距離)情報を作成(encode)する)。深度画像を、幾何学的方法又は電子的方法を含む種々の方法によって取得することができる。深度画像を、ポイントクラウドとして表してもよい又はポイントクラウドに変換してもよい。幾何学的方法の例は、受動型又は能動型ステレオカメラシステム及び構造化光カメラシステム(structured light camera system)を含む。電子的方法の例は、飛行時間型(TOF)カメラ、一般的なスキャニングカメラ又は固定LIDARカメラを含む。

【0040】

カメラのタイプに応じて、カメラによってキャプチャしたデータから深度画像を生成するために種々のアルゴリズムを用いてもよい。Dense Tracking and Mapping in Real Time (DTAM) と称される種類(class)のアルゴリズムは、キャプチャした画像のカラキュー(color cue)を用いるものであり、Simultaneous Localization and Mapping (SLAM) と称される他の種類のアルゴリズムは、深度(又は深度と色の組合せ)データを用いるものであり、更に別の種類のアルゴリズムは、Iterative Closest Point (ICP) 及びその派生(derivative)に基づく。

【0041】

図2に関連して更に詳しく説明するように、少なくとも一部の深度カメラシステムによって、ユーザは、物の全ての側面をキャプチャするために物の周辺でカメラを自由に動かすことができる。結合した深度画像を生成する基本アルゴリズムは、キャプチャしたデータを物又は物の部分的に構成した3Dモデルに合わせるために物に対するカメラの姿勢を

10

20

30

40

50

追跡することができる及び/又は推測することができる。3次元の物をスキャンするシステム及び方法の一例は、2017年6月22日に米国特許庁に出願された米国特許出願公開第15/630,715号明細書に記載されており、その開示の全体を参照によりここに組み込む。

【0042】

本発明の一部の実施の形態において、深度画像又は3Dモデルの構造化を、スキャナそれ自体によって局所的に行う。他の実施の形態において、有線又は無線接続(例えば、イーサネット(登録商標)接続、USB接続、セルラー式のデータ接続、ローカル無線ネットワーク接続及びBluetooth(登録商標)接続)を介してスキャナからデータを受信することができる一つ以上のローカルサーバ又はリモートサーバによって処理を行う。

10

【0043】

更に具体的な例として、スキャナをハンドヘルド3Dスキャナとしてもよい。そのようなハンドヘルド3Dスキャナは、あり得る大きい表面又は完全な物の3D表現を作成するために同一表面の複数の深度画像を記録することができるソフトウェアと共に深度カメラ(各画素によって撮像された表面要素の距離を算出するカメラ)を有してもよい。ハンドヘルド3Dスキャナのユーザは、ハンドヘルド3Dスキャナを物の周辺の種々の位置に動かすとともに物の表面の全ての位置をカバーするように(例えば、スキャナによって取得した少なくとも一つの深度画像において表面が見えるように)ハンドヘルド3Dスキャナを適応させる必要がある。さらに、各表面パッチが十分高い密度の深度測定を受け取ることが重要である(この場合、深度カメラの各画素はそのような深度測定を提供する)。深度測定の密度は、カメラによって表面パッチが見られる距離及び深度カメラの視野方向又は光軸に対する表面の角度又は傾斜に依存する。

20

【0044】

図2は、本発明の一実施の形態によるステレオ深度カメラとしてのスキャニングシステムのブロック図である。

【0045】

図2に示すスキャニングシステム100は、第1のカメラ102と、第2のカメラ104と、投影源106(又は照明源若しくは能動型投影システム)と、ホストプロセッサ108と、メモリ110と、を有し、ホストプロセッサを、例えば、グラフィックスプロセッシングユニット(GPU)、更に一般的な汎用プロセッサ(CPU)、適切に構成されたフィールドプログラマブルゲートアレイ(FPGA)又は特定用途向け集積回路(ASIC)とすることができる。第1のカメラ102及び第2のカメラ104の相対位置及び向きを十分に固定するように第1のカメラ102及び第2のカメラ104を、例えば、フレームに堅く結合してもよい。第1のカメラ102及び第2のカメラ104を共に「深度カメラ」と称してもよい。第1のカメラ102及び第2のカメラ104は、対応するイメージセンサ102a及び104aを有し、対応するイメージ信号プロセッサも有してもよい。種々の構成要素は、システムバス112を介して互いに通信を行ってもよい。スキャニングシステム100は、画像を表示することができるようにするためのディスプレイ114、他の装置と通信を行うためのネットワークアダプタ116、スキャニングシステム100の加速度を検出する(例えば、向きを決定するために重力の方向を検出するとともに位置変化を検出するために動きを検出する)ジャイロスコープのような慣性計測装置(IMU)118並びにスキャニングシステム100によって収集及び処理されるデータを格納するNANDフラッシュメモリのような永続メモリ120のような追加の構成要素を有してもよい。IMU118を、多数のモデムスマートフォンでよく見つかるタイプのものとしてもよい。イメージキャプチャシステムは、ユニバーサルシリアルバス(USB)インタフェースコントローラのような他の通信部品も有してもよい。

30

40

【0046】

一部の実施の形態において、カメラ102のイメージセンサ102a及びカメラ104のイメージセンサ104aは、RGB-IRイメージセンサである。可視光(例えば、赤-緑-青、すなわち、RGB)及び不可視光(例えば、赤外、すなわち、IR)を検出す

50

ることができるイメージセンサを、例えば、電荷結合素子（ＣＣＤ）センサ又は相補型金属酸化膜半導体（ＣＭＯＳ）センサとしてもよい。一般的には、通常のＲＧＢカメラセンサは、５０％が緑、２５％が赤及び２５％が青である「ベイヤー配列」、すなわち、「ＲＧＢＧ配列」で配置された画素を有する。バンドパスフィルタ（又は「マイクロフィルタ」）を、ベイヤー配列に従って緑、赤及び青のそれぞれに対して個別のフォトダイオードの前に（例えば、フォトダイオードとカメラに関連する光学系の間に）配置する。一般的には、通常のＲＧＢカメラセンサは、電磁スペクトルのＩＲ部分の信号を遮断する（レンズの一部としての又はイメージセンサチップの全体のコーティングとしての）赤外（ＩＲ）フィルタ又はＩＲカットフィルタも有する。

【００４７】

ＲＧＢ－ＩＲセンサは、通常のＲＧＢセンサに略類似するが、種々のカラーフィルタを有してもよい。例えば、ＲＧＢ－ＩＲセンサにおいて、四つのフォトダイオードの各グループの緑フィルタの一つを、２５％の緑、２５％の赤、２５％の青及び２５％の赤外である配列を形成するためにＩＲバンドパスフィルタ（又は「マイクロフィルタ」）と共に配置し、この場合、赤外画素を可視光画素の間に混ぜる。さらに、ＩＲカットフィルタをＲＧＢ－ＩＲセンサから省略してもよい、ＩＲカットフィルタを、赤色光、緑色光及び青色光を検出する画素の上にのみ配置してもよい、又は、ＩＲフィルタを、可視光及び特定の波長間隔（例えば、８４０～８６０ｎｍ）の光を通過させるように設計することができる。電磁スペクトルの複数の部分、帯域又はスペクトル帯の光（例えば、赤色光、青色光、緑色光及び赤外光）をキャプチャすることができるイメージセンサを、ここでは、「マルチチャンネル」イメージセンサと称する。

【００４８】

本発明の一部の実施の形態において、イメージセンサ１０２ａ及び１０４ａは、通常の可視光センサである。本発明の一部の実施の形態において、システムは、一つ以上の可視光カメラ（例えば、ＲＧＢカメラ）と、それとは別の一つ以上の不可視光カメラ（例えば、赤外カメラ、この場合、ＩＲバンドパスフィルタが全て画素の上に配置される）と、を有する。本発明の他の実施の形態において、イメージセンサ１０２ａ及び１０４ａは、赤外（ＩＲ）光センサである。

【００４９】

一般的に言えば、立体深度カメラシステムは、互いに離間するとともに堅いフレームのような共有構造に堅く取り付けられた少なくとも二つのカメラを有する。カメラは、略同一の方向に向けられ（例えば、カメラの光軸を略平行にしてもよい）、重複する視野を有する。これらの個別のカメラを、例えば、（例えば、一つ以上のレンズを有する）光学系を有する相補型金属酸化膜半導体（ＣＭＯＳ）イメージセンサ又は電荷結合素子（ＣＣＤ）イメージセンサを用いて実現することができる。光学系は、光をイメージセンサに向ける又はイメージセンサ上に光の焦点を合わせるように構成される。光学系は、例えば、光学系が「広角」レンズ、望遠レンズ又は中間のもの（something in between）を実現するか否かに基づいてカメラの視野を決定することができる。

【００５０】

以下の考察において、深度カメラシステムの画像取得システムを、「マスター」カメラ及び一つ以上の「スレーブ」カメラと称することができる少なくとも二つのカメラを有するものとして説明する。一般的に言えば、推定される深度又は視差マップは、マスターカメラの視点から算出されるが、カメラのいずれかをマスターカメラとして用いればよい。ここで用いられるように、マスター／スレーブ、左／右、上／下、第１／第２及びＣＡＭ１／ＣＡＭ２は、特に断りのない限り互いに用いられる。換言すれば、カメラのいずれか一つをマスターカメラ又はスレーブカメラとすることができ、右のカメラに対する考察を、他の方向において対称的に左のカメラに適用することができる。さらに、後に説明する考察は、便宜上種々の数のカメラに対して有効であるが、考察を、一般的には、二つのカメラを有するシステムの状況で説明する。例えば、深度カメラシステムは、三つのカメラを有してもよい。そのようなシステムにおいて、カメラの二つは、不可視光（赤外）カメ

10

20

30

40

50

ラであってもよく、第3のカメラは、可視光カメラ（例えば、赤／青／緑のカラーカメラ）であってもよい。三つ全てのカメラを互いに光学的に調整（例えば、較正）してもよい。三つのカメラを有する深度カメラシステムの一例は、2016年5月5日に米国特許庁に出願された米国特許出願公開第15/147,879号明細書の「深度投影3眼カメラシステム（Depth Perspective Trinocular Camera System）」に記載されており、その開示の全体を参照によりここに組み込む。

【0051】

カメラによって撮影されたシーンの特徴の深度を検出するために、深度カメラシステムは、カメラによってキャプチャされた画像の各々の特徴の画素位置を決定する。二つの画像の特徴の間の距離は、視差と称され、それは、物の距離又は深度に反比例する。（これは、比較の際に物を1回に一つの目で見たとときに物が「変化する」程度の影響であり、変化の大きさは、物が観察者の目からどの程度離れているかに依存し、物が近づくに従って変化が大きくなり、物が遠くなるに従って変化が小さくなり、遠方の物は、検出できる変化がほとんどない。）視差を用いて深度を算出する技術は、例えば、R. Szeliskiによる“Computer Vision: Algorithms and Applications”, Springer, 2010 pp. 467 et seqに記載されている。

【0052】

マスターカメラとスレーブカメラの間の視差の大きさは、カメラのピクセル解像度（pixel resolution）、カメラの間の距離及びカメラの視野のような深度カメラシステムの物理的な特性に依存する。したがって、正確な深度測定を生成するために、深度カメラシステム（又は深度投影カメラシステム）は、これらの物理的特性に基づいて較正される。

【0053】

一部の深度カメラシステムにおいて、カメラのイメージセンサの画素の水平行が略平行になるようにカメラを配置してもよい。画像修正技術を、カメラのレンズの形状及びカメラの向きの変化による画像の歪みを調整するのに用いることができる。

【0054】

更に詳しくは、カメラ較正情報は、入力画像を修正するために情報を提供することができ、その結果、等価カメラシステムのエピポーラ線は、修正された画像のスキャンラインに一致する。そのような場合、シーンの3Dポイントは、マスター画像及びスレーブ画像の同一のスキャンラインインデックスに投影される。 U_m 及び U_s をマスターカメラ及びスレーブ等価カメラ（slave equivalent camera）の同一の3Dポイント p の画像のスキャンラインの座標とし、この場合、各カメラにおいて、これらの座標は、主点（焦点面の光軸の交差点）を中心とするとともに修正された画像のスキャンラインに平行な水平軸を有する軸システム（axis system）を参照する。差 $U_m - U_s$ は、視差と称され、 d によって表される。視差は、修正されたカメラに対する3Dポイントの直行距離（すなわち、いずれかのカメラの光軸に対するポイントの直角投影の距離）に反比例する。

【0055】

立体アルゴリズム（stereoscopic algorithms）は、視差のこの特性を利用する。これらのアルゴリズムは、視差を推定することと等価である左右のビューで検出されるポイント（又は特徴）のマッチングを行うことによって3D再構成を実現する。ブロックマッチング（BM）は、一般的に用いられる立体アルゴリズムである。マスターカメラ画像の画素が与えられると、アルゴリズムは、この画素をスレーブカメラ画像の他の任意の画素にマッチングするためのコストを算出する。このコスト関数は、マスター画像の画素を包囲する小窓内の画像コンテンツとスレーブ画像の画素の間の相違として規定される。あるポイントの最適視差は、最小マッチングコストの引数（argument）として最終的に推定される。この手順は、一般的には勝者独り勝ち（WTA）と称される。これらの技術は、例えば、R. Szeliskiの“Computer Vi

10

20

30

40

50

sion: Algorithms and Applications”, Springer, 2010で更に詳しく記載されている。BMのような立体アルゴリズムは外観類似に依存するので、視差の算出は、スレーブ画像の二つ以上の画素が同一の局部外観を有する場合に困難になり、これらの画素の全てがマスター画像の同一の画素と類似するので、不明確な視差推定になる。これが生じうる典型的な状況は、平坦な壁のような一定の輝度を有するシーンを視覚化するときである。

【0056】

2016年7月12日に発行された米国特許第9,392,262号明細書の「複数のマルチチャンネルカメラを用いて3D再構成のシステム及び方法(System and Method for 3D Reconstruction Using Multiple Multi-Channel Cameras)」に記載されているような小さい3Dの詳細をキャプチャすることができるブロックマッチングアルゴリズムの実行を向上させる又は最適化させるように設計されたパターンを投影することによって追加の照明を提供する方法が存在し、その開示の全体を参照によりここに組み込む。他の手法は、シーンにテクスチャを提供するとともに同時に出現するシーンの部分を除去する(disambiguating)ことによってテクスチャの少ない領域(texture-less regions)の深度推定を特に向上させるために単に用いられるパターンを投影する。

10

【0057】

本発明の実施の形態による投影源106を、カメラ102及び104によって撮影されたシーンに可視光(例えば、人間及び/又は他の動物の目に見えるスペクトル内の光)又は不可視光(例えば、赤外光)を投影するように構成してもよい。換言すれば、投影源は、カメラ102及び104の光軸に略平行な光軸を有してもよく、投影源を、カメラ102及び104の視野の方向に光を出射するように構成してもよい。一部の実施の形態において、投影源106は、複数の個別の照明器を有してもよく、各照明器は、他の(一つ以上の)照明器の(一つ以上の)光軸から離間するとともにカメラ102及び104の光軸から離間した光軸を有する。

20

【0058】

不可視光投影源は、対象が(テレビ会議システムにおけるような)人々である状況により適している。その理由は、不可視光が対象の見る能力を妨害しないからである。それに対し、可視光投影源は、対象の目に対して不快になるように照射するおそれがある又はシーンにパターンを加えることにより認識に不所望な悪影響を及ぼすおそれがある。不可視投影源を有するシステムの例は、例えば、2015年6月30日に米国特許庁に出願された米国特許出願公開第14/788,078号明細書の「複数の露出設定に基づくマルチチャンネル撮影のシステム及び方法(System and Methods for Multi-Channel Imaging Based on Multiple Exposure Settings)」に記載されており、その開示の全体を参照によりここに組み込む。

30

【0059】

能動型投影源を、投影静止パターン、例えば、経時的に変化しないパターンと、動的パターン、例えば、経時的に変化するパターンと、に分類することができる。両方の場合において、パターンの一態様は、照明レベルの投影パターンである。これは関連性がある。その理由は、深度カメラシステムの深度ダイナミックレンジ(depth dynamic range)に影響を及ぼし得るからである。例えば、光照明がハイレベルである場合、深度測定を(例えば、距離の逆2乗に比例する因子によって物までの距離に亘る光照明の減衰を克服するために)明るい周囲光の状況の下で遠くの物から行うことができる。しかしながら、高い光照明レベルによって、クローズアップされたシーンの部分に飽和が生じることがある。それに対し、低い照明レベルによって、近くの物の測定を可能にするが、遠くの物の測定ができない。

40

【0060】

50

状況次第で、深度カメラシステムは、二つの構成要素：取り外し可能なスキャニング部及び表示部を有する。一部の実施の形態において、表示部は、スマートフォン、タブレット、携帯端末又は他の同様なシステムのようなコンピュータシステムである。分離可能なスキャニング部及び表示部を用いるスキャニングシステムは、例えば、2016年12月16日に米国特許庁に出願された米国特許出願公開第15/382,210号明細書の「スクリーンから分離可能なスキャニングセンサを有するスキャニング装置(Scanning Apparatus including Scanning Sensor Detachable from Screen)」に更に詳しく記載されており、その開示の全体を参照によりここに組み込む。

【0061】

本発明の実施の形態をここでは立体深度カメラシステムに関連して説明するが、本発明の実施の形態は、それに限定されず、構造化された光飛行時間型カメラ及びLIDARカメラのような他の深度カメラシステムと共に用いられてもよい。

【0062】

カメラの選択に応じて、3Dモデルを生成するために種々の技術を用いてもよい。例えば、Dence Tracking and Mapping in Real Time (DTAM)は、スキャニングのためにカラーキューを用いるものであり、Simultaneous Localization and Mappingは、3Dモデルを生成するために深度データ(又は深度とカラーの組合せのデータ)を用いるものである。

【0063】

図3は、本発明の一実施の形態によるスクリーンから切り離されているスキャニングセンサ(又はカメラ)を有する3Dスキャニングシステム50を用いて固定位置の物20をスキャンするシステム及び方法をユーザの観点から示す。特に、図3は、八つの異なる時点で物20の周辺の八つの異なる物理的位置の単一のスキャニングセンサモジュール100を示す。スキャニングセンサモジュール100とスキャニング装置200の間の破線は、二つのハードウェアの間のデータ接続を表し、この場合、スキャニング装置200は、データ接続を介してコマンドをスキャニングセンサモジュール100に送信することができ、スキャニングセンサモジュール100は、画像を含むデータをスキャニング装置200に送信することができる。

【0064】

図3でわかるように、ユーザ22は、(右手として図3に示すように)一方の手にスキャニングセンサモジュール100(例えば、カメラ)を保持してもよく、(左手として図3に示すように)他方の手にスキャニング装置200(例えば、スマートフォン、タブレットコンピュータ、携帯端末又はディスプレイを有する他の携帯装置)を保持してもよい。図3に示すように、ユーザは、スキャニング装置200の位置を変えずにスキャニングセンサを物の前から物の後ろに動かしてもよい。ユーザがスキャニングセンサを用いて物20(図3に示すようなティッシュボックス)をスキャンする間、スキャニング装置200は、ビュー202(例えば、リアルタイムビュー)、すなわち、スキャニングセンサモジュールによってキャプチャした画像の表現を表示する。図3に示すように、ユーザは、物20のトップダウンビューを更に有してもよいが、ユーザ22は、物20のサイドビューを有するようにするためにスキャニングセンサモジュール100を配置してもよい。図3に示すスキャニング装置200のビュー202は、位置100vのスキャニングセンサモジュール100によってキャプチャしたサイドビュー画像に対応する。これにより、ユーザ22は、スキャニング装置200のディスプレイのビュー202の眺めを維持するためにユーザの身体をねじる必要なく物20を更に容易かつ楽にスキャンすることができる。したがって、ユーザは、スキャニングセンサによってキャプチャしたデータに関するリアルタイムフィードバックを受け取り、これによって、トップを含む物の全てのサイドがスキャニングセンサによって撮影されることを保証するのを助ける。

【0065】

このフィードバックがない場合、ユーザは、物の有用な画像をキャプチャし損なう方向

10

20

30

40

50

にスキャニングセンサをうっかり向けることがある。例えば、ユーザは、天井又は床の画像をうっかりキャプチャすることがある又は背景の画像をキャプチャすることがある。さらに、カメラ及びディスプレイが互いに堅く固定された比較撮像システム (comparative imaging system) において、カメラの視野と撮像システムのディスプレイの観察方向の間の一定の関係によって、ユーザは、物の全てのサイドをスキャニングしながらスクリーンの観察を維持するのが困難になることがある。

【 0 0 6 6 】

一旦、スキャンされたものの3次元形状がスキャニングシステムによって取得及び処理されると、クリーンな3次元モデルを取得するために一部の修正を行うことができる。3次元モデルは、テクスチャ情報 (例えば、カラー情報) に関連してもよい。そのようなモデルをメッシュとして表現してもよい。テクスチャモデルを作成することは、非テクスチャモデルを作成することより複雑である。その理由の一部は、テクスチャモデルの観察者がフォトリアリズム及びテクスチャをメッシュの表面に合わせることの追加の課題のために更に高い予測を有することがあるからである。測光誤差最小化及びグラフカット最適化に基づく種々の技術を、テクスチャモデルに適用してもよく、一部の実施の形態において、そのような技術を、スキャナそれ自体ではなくサーバによって実行してもよい。

10

【 0 0 6 7 】

図 4 A 及び図 4 B は、スキャニングシステムによってキャプチャした3次元モデルの例である。図 4 A は、茶色の靴のテクスチャモデルであり、R G G - I R 深度カメラシステムによってキャプチャされたものである。図 4 B は、ダッフルバックの非テクスチャモデルであり、テクスチャ化を行うことなく (例えば、可視光又は R G B センサなしで I R センサのみを有する深度カメラシステムによって) キャプチャされたものである。生成したモデルを、" p l y " フォーマット又は " o b j " フォーマットのような標準的なフォーマットで格納してもよく、ビューアソフトウェアを用いて表示装置に表示することができる。一部の実施の形態において、ビューイングソフトウェアを、T h e e J S ビューアを用いる場合のように (例えば、ウェブブラウザによって実行される) ウェブベースのものとしてもよい。

20

【 0 0 6 8 】

3次元モデルの格納

本発明の実施の形態の一態様によれば、3次元 (3 D) モデルのような一連の既存のメディアドキュメントがデータベースに格納されている。(3 D モデルのような) 供給されたメディアドキュメントのメタデータを生成するとき、供給されたメディアドキュメントに類似するメディアドキュメントについてデータベースに対するクエリーが行われ、類似するエントリーが結果として戻される。

30

【 0 0 6 9 】

データベース構成

本発明の実施の形態の一態様によれば、データベースは、K個の異なる種類に分けられたN個のエントリー (N個のメディアドキュメント) を格納する。エントリーの数 (N) は数百万程度であってもよく、種類の数 (K) は、数千又は数万程度であってもよい。K個の種類を階層的なツリー状構造に分けてもよく、その一部を表 1 に示す。

40

【 0 0 7 0 】

【表 1】

表 1

➤ Alcohol	
➤ Antiques	
➤ Art	
▼ Baby	
▼ Baby Carriers	10
▼ Baby Carriers Slings	
➤ Tapestries	
➤ Backpacks	
➤ Baby Clothing	
➤ Baby Shoes	
▼ Bathing	
➤ Grooming	20
➤ Bath Tubs	
➤ Bathing Accessories	
➤ Health	
➤ Other Bathing	
➤ Skincare	
▼ Towels	
➤ Washcloths	30
➤ Bedding	
➤ Car Seats	

【0071】

一実施の形態によれば、データベースの各エントリーは、名前、一つ以上の種類、一つ以上のタグ、テキスト記述及び視覚情報を含むメタデータに関連する。

【0072】

名前欄は、エントリーのメディアドキュメントの特定の名前である。例えば、メディアドキュメントが特定のタイプの車の3Dモデルであるとき、車のフルネーム（例えば、年式、製造、型及びストリングに連結したトリム（trim concatenated into a string））を特定することができる。 40

【0073】

種類欄は、特定のエントリーに関連するK個の種類の一つ以上の種類を識別する。特に、K個の種類がツリー状構造に階層的に分けられる場合、エントリーがある種類に属するときには、当該エントリーは、ツリーの特定の種類より上からツリーのルートまでの種類の全てに属する。表1に示す“Washcloths”は、そのすぐ前の親の種類“Towels”並びに更に上の種類の“Bathing”及び“Baby”に属する。

【0074】

一つ以上のタグは、特定のエントリーに関連するキーワード（例えば、非階層的な単語又は語句）に対応する。これらは、例えば、製品の色の記述（例えば、赤、白及び黄）、 50

パターン（例えば、格子柄、縞模様及びギンガム）、材料（例えば、綿、ポリエステル、混合物）、サイズ、又は、種類のような階層に適合しない他の記述を含んでもよい。

【0075】

テキスト記述欄は、物語形式で型を記述してもよく、人間が読めるように書き込まれていてもよいが、エントリーのテキスト検索に用いてもよい。

【0076】

メディア情報欄は、メディアドキュメントそれ自体（又は、更に広く見れば、特定の画像、特定の3Dモデル、特定の録音又は特定のビデオのようなエントリーに関連する特定のデジタルメディアドキュメント）に対応してもよい。

【0077】

一部の実施の形態によれば、名前、種類及び視覚情報のようなこれらのメタデータ欄の一部は必須であるのに対して、タグ及びテキスト記述のような他のメタデータ欄は任意である。所定のエントリーに対するメタデータ欄の一部又は全てを、ユーザによって手動で入力してもよい又は後に詳しく説明するように自動的に生成してもよい。

【0078】

自動的なメタデータ追加のためのデータベース検索

本発明の実施の形態の一態様は、所定の入力メディアドキュメントのメタデータを生成するために所定の入力メディアドキュメントに類似するエントリーの上述したデータベースの検索の実行を対象とする。

【0079】

本開示の文脈において、クエリー及びエントリーの用語を以下で考察する。エントリーは、データベースに既に存在するとともにメタデータが利用できるメディアドキュメント（例えば、3Dモデル、画像、録音又はビデオ）である（又はそれに関連する）。それに対し、クエリーは、データベースに存在しないとともデータベースに既に存在するエントリー又はモデルのメタデータからの情報を組み合わせることによってメタデータ欄の一部又は全てが満たされることが所望されるメディアドキュメント（例えば、3次元モデル）である。追加されるメタデータ欄がクエリーの種類である場合を、クエリーのメディアドキュメントの分類と称する。

【0080】

画像のデータベースにクエリーを行う技術は、一般的には、二つの異なる形式：（一つ以上の種類を画像に割り当てる）画像分類及び（クエリー画像に関するデータベースの最も類似する画像入力を識別する）画像検索に含まれる。一つの共通画像データベースは、数百万の画像及び数千の異なる種類を有するImageNet（例えば、J. Deng, W. Dong, R. Socher, L. - J. Li, and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009 参照）である。画像分類及び検索を行う技法は、畳み込みニューラルネットワーク（例えば、A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet Classification with deep convolutional neural networks", Advances in Neural Information Processing Systems, 2012, Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, "Backpropagation applied to handwritten zip code recognition", Neural Computation, 1989 及び C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions", CVPR, 2015 参照）及び階層ベイズモデル（例えば、L. Fei-Fei, P. Perona, "A Bayesian hier

10

20

30

40

50

archical model for learning natural scene categories", CVPR, 2005)を用いる技術を有する。

【0081】

さらに、録音と、ビデオの音声成分とを、録音を画像に変換するとともに音声の一つ以上のスペクトログラム（例えば、音声の各チャンネルに対する一つのスペクトログラム）を作成することによる特徴抽出のために畳み込みニューラルネットワークに供給することができる。

【0082】

便宜上、限定を行うことなく、CNNを、入力画像が与えられたときに2次元(2D)畳み込み、非線形マッピング、Max-pooling集約(Max-pooling aggregations)及び(一般的に特徴ベクトルと称される)値又は特徴のベクトルを算出するための接続のような動作のセットを実行するシステムとみなすことができる。特に、クエリー画像に対する予測される特徴ベクトルは、大次元(例えば、4096の値)の(一般的には浮動小数点又は固定小数点の数値としてコード化される)値のベクトルである。特徴ベクトルを、入力画像の「シグネチャー」とみなすことができ、特徴ベクトルは、同じ種類の物の特徴ベクトルが特徴ベクトル空間に近接するポイントであるという特性を有する(例えば、類似の特徴ベクトルを有する)。特徴ベクトルを、入力画像のメタデータの一つ以上の種類の推定を取得するために分類器(例えば、ソフトマックス分類器)によって用いることができる。したがって、分類器によって算出される分類、どの種類が所定の画像に関連するかを識別するために用いることができる。

【0083】

分類を算出するために特徴ベクトルを分類器に供給することに加えて、特徴ベクトルを、画像検索に用いることもできる。類似する画像が類似する特徴ベクトルを有するので、データベースのエントリーは、エントリーの特徴ベクトルが特徴ベクトル空間のクエリー画像の特徴ベクトルに近接する場合にはクエリー画像に類似し、「近接」又は「類似」を、 L^1 又は L^2 メトリック(L^1 or L^2 metric)のような標準メトリック(standard metric)に関連して算出することができる。大次元の特徴ベクトル空間が与えられる場合、特徴ベクトルは、主成分分析(PCA)又は線形判別分析(LDA)のような次元減少の技術を適用することによって更に計算的に扱いやすくすることができる。

【0084】

2次元画像と類似するやり方で、3次元モデルを、テクスチャ化されているかテクスチャ化されていないかに関係なく、畳み込みニューラルネットワーク又は分類器を用いる分類又は検索に関連して考察することもできる。

【0085】

したがって、本発明の実施の形態の一態様は、メタデータを取得するために類似のメディアドキュメントに関連するエントリーを検索するためのクエリーのメディアドキュメントからの情報を用いることに関する。特に、本発明の実施の形態の一態様は、畳み込みニューラルネットワークを用いてメディアドキュメントから特徴ベクトルを抽出するためにメディアドキュメントを作成することに関する。抽出した特徴ベクトルを、メディアドキュメントを分類するために分類器に供給することができる又は特徴ベクトル空間の類似のメディアドキュメントを検索するのに用いてもよい。

【0086】

図5は、本発明の一実施の形態によるメディアドキュメントのメタデータを生成する方法500のフローチャートである。図5に示す方法を、スキャナ100のホストプロセッサ108、サーバ(例えば、スキャナ100にローカルなコンピュータ又は「クラウド」のデータセンタのコンピュータ)又は(一部の動作がスキャナのホストプロセッサ108によって実行されるとともに他の動作がサーバによって実行される場合の)その組合せによって実行してもよい。便宜上、動作を、「プロセッサ」によって実行されるものとして説明し、プロセッサがホストプロセッサ108であるかサーバのプロセッサであるかにつ

10

20

30

40

50

いての限定はない。

【0087】

動作510において、プロセッサは、プロセッサが受け取ったメディアドキュメントの特徴ベクトルを算出する。場合によっては、メディアドキュメントは、スキャニング処理の間にスキャナ100によってキャプチャされた3Dモデルであってもよい。他の場合において、メディアドキュメントを、第三者（例えば、異なるスキャニング装置によってキャプチャされた3Dモデル）から検索してもよい、コンピュータ支援設計ツールを用いて手動で生成してもよい、又は、コンピュータ支援設計ツールを用いて修正したスキャンモデルとしてもよい。

【0088】

上述したように、クエリーのメディアドキュメントが2D画像であり、データベースのメディアドキュメントも2D画像である場合、2D画像を、トレーニングされた（例えば、2D画像でトレーニングされた）CNNに直接供給してもよい。メディアドキュメントが録音である場合又はメディアドキュメントが音声成分を含む録画である場合、プロセッサは、音声を一つ以上のスペクトログラムに変換し、入力としての一つ以上のスペクトログラムを、トレーニングされた（例えば、録音のスペクトログラムでトレーニングされた）CNNに供給する。

【0089】

クエリーのメディアドキュメントが3Dモデルであり、データベースが一連の2D画像を格納する場合、一つ以上の特徴ベクトルを3Dモデルから抽出することができる。例えば、一実施の形態において、3Dモデルは、複数の2D画像を生成するために複数の視点からレンダリングされ、2D画像の各々を、3Dモデルに関連する（2Dビューから抽出される）複数の特徴ベクトルを生成するために、トレーニングされたCNNに供給してもよい。したがって、3Dモデルを、2D画像のデータベースに対してクエリーを行うのに用いることができる。図6Aは、本発明の一実施の形態による3Dモデルをボクセル化することによりメディアドキュメントが3Dモデルである場合に特徴ベクトルを算出する方法を示すフローチャートであり、図6Bは、ボクセル化を用いた3DモデルへのCNNの適用を視覚的に示す。図6A及び図6Bを参照すると、動作610において、プロセッサは、例えば、rotating calipers アルゴリズムを3Dモデルに適用することによって3Dモデルの周辺の境界ボックス612を規定する。

【0090】

動作630において、プロセッサは、3Dモデルを表すボクセルのセットを生成するためにモデルをボクセル化する。一実施の形態において、ボクセル化処理の間に、プロセッサは、境界ボックスを、ボクセルと称するサブユニットに分割する。例えば、境界ボックスを、 $256^3 = 16,777,216$ ボクセルの総数に対して一辺に256ボクセルを有する立方体としてもよいが、本発明の実施の形態は、それに限定されず、十分に大きくしてもよい。各ボクセルは、（例えば、2進数ボクセル化で）3Dモデルの一部がボクセル内に含まれるか否かを表す値に関連してもよく、他の場合において、各ボクセルは、3Dモデルによって占有されるボクセルの比又は割合（例えば、整数値又は浮動小数点ボクセル化）を表す値（例えば、8進数）に関連してもよい。図6Bは、ボクセル化された3Dモデル632を示す。

【0091】

本発明の一実施の形態によれば、動作630は、ボクセル化を行う前に「好適なビュー」を識別するために特徴を推定すること又は主成分分析を行うことを更に有する。ボクセル化の前にモデルの一貫した好適なビューを識別することは、略類似する物（又は同一の物）の二つの異なる3Dモデルが（例えば、略同一の座標空間にそって規定されるボクセルを用いて）同一の視点からボクセル化される可能性を上げ、これによって、回転不変性（例えば、回転した物のモデルを認識するシステムの能力）を提供する。

【0092】

動作650において、プロセッサは、動作630で生成されたボクセルから特徴ベクト

10

20

30

40

50

ルを生成する。本発明の一実施の形態によれば、特徴ベクトル 6 5 4 は、トレーニングされた畳み込みニューラルネットワークにボクセルを供給することによって算出される。ボクセル表示を 3 D テンソルとみなすことができるので、ボクセルを入力として CNN に直接供給することができ、この場合、CNN は、ボクセル化された 3 D モデルに基づいてトレーニングされる。図 6 B に示す破線の立方体は、畳み込み動作を表す。図 6 B に示すように、特徴ベクトル 6 5 4 を、ボクセル化された 3 D モデル 6 1 2 の分類 6 5 6 を生成するために分類器に供給することもできる。

【 0 0 9 3 】

本発明の他の実施の形態によれば、3 D モデルについての形状及び文字情報を、複数の方向又は視点からモデルをレンダリングすることによってコード化する。図 7 A は、本発明に一実施の形態による種々の視点又は方向からの 3 D モデルの複数のビューをレンダリングすることによりメディアドキュメントが 3 D モデルである場合に特徴ベクトルを算出する方法を示すフローチャートであり、図 7 B は、複数のビューからレンダリングした 3 D モデルへの CNN の適用を視覚的に示す。

【 0 0 9 4 】

図 7 A 及び図 7 B を参照すると、動作 7 1 0 において、プロセッサは、3 D モデルの複数の 2 次元 (2 D) ビュー 7 1 2 を生成するために複数の角度から 3 D モデルをレンダリングする。類似特徴推定又は主成分分析技術を、3 D モデルの複数の 2 D ビューを生成する 3 D モデルの「好適なビュー」を識別するために適用してもよい。動作 7 3 0 において、プロセッサは、複数の対応する単一ビュー特徴ベクトル (2 D ビューの単一のものに対応する特徴ベクトル) を生成するために 2 D ビューの各々を畳み込みニューラルネットワーク 7 3 2 に供給する。一部の実施の形態において、図 7 B に示すように、各ビューに対して個別の畳み込みニューラルネットワークが存在する。他の実施の形態において、ビューの各々を、同一の畳み込みニューラルネットワークに供給する。動作 7 5 0 において、ビュー追加層 7 5 2 は、2 D ビュー又は画像から算出した単一ビュー特徴ベクトルを集める。ビュー追加層 7 5 2 は、システムの追加の柔軟性を可能にする。その理由は、ビュー追加層 7 5 2 によって追加の特徴ベクトルの長さが 3 D モデルのビューの数に依存しなくなるからである。

【 0 0 9 5 】

動作 7 7 0 において、プロセッサは、特徴ベクトル 7 7 4 を算出するために最終段階 CNN 7 7 2 を適用する。特徴ベクトル 7 7 4 を、3 D モデルのビューの分類 7 7 6 を生成するために分類器に供給してもよい。

【 0 0 9 6 】

クエリーのメディアドキュメントが同一の物の一つ以上の 2 D 画像を有するとともにデータベースが一連の 3 D モデルである場合、一連の 3 D モデルの特徴ベクトルと比較することができる特徴ベクトルを生成するために、一つ以上の特徴ベクトルを 2 D 画像から抽出してもよい。本発明の一実施の形態において、例えば、既に簡潔に説明するとともに R . S z e l i s k i . “ C o m p u t e r V i s i o n : A l g o r i t h m s a n d A p p l i c a t i o n s ” , S p r i n g e r , 2 0 1 0 で更に詳しく説明されているブロックマッチングのような立体アルゴリズムを用いて 3 D モデルを合成するために、2 D 画像が用いられる。この場合、図 6 A 、図 6 B 、図 7 A 及び図 7 B に関連して上述した技術のようにクエリーのメディアドキュメントが 3 D モデルである手法と略類似する手法で、特徴ベクトルを、合成された 3 D モデルから抽出することができる。

【 0 0 9 7 】

動作 5 3 0 において、プロセッサは、入力メディアドキュメントについて算出した特徴ベクトルに類似する特徴ベクトルを有するデータベースのメディアドキュメントを検索する。上述したように、特徴ベクトルの「近接」又は「類似」を、 L^p 空間の p -ノルム (p -norm) のような多次元距離メトリック (例えば、二つのベクトル x_1 及び x_2 に対して、

10

20

30

40

【数 1】

$$x = (x_1, x_2), \|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$$

）に関連して算出することができる。例えば、 L^1 は、「タクシー (taxicab)」空間又は「マンハッタン空間」の距離に対応し、 L^2 は、ユークリッド空間の距離に対応する。一実施の形態において、プロセッサは、二つの特徴ベクトル（例えば、クエリーモデルの特徴ベクトル及びデータベースのエントリーの一つのモデルの特徴ベクトル）の間の距離（又は相違）を表すメトリックがしきい値未満であるときに二つの特徴ベクトルは類似することを決定する。

10

【0098】

本発明の一実施の形態において、クエリーのメディアドキュメントの特徴ベクトルとデータベースの各エントリーの特徴ベクトルとを比較することによって、類似するメディアドキュメントがデータベースで識別される。本発明の他の実施の形態によれば、類似するエントリーは、特徴ベクトルの類似に基づいてデータベースにおいてグループ化（又はビンニング）される。最初の検索は、クエリーのメディアドキュメントに類似するエントリーの一つ以上のピンを識別してもよく、この場合、エントリーの識別されたピンの各々のエントリーの全てを、クエリーのメディアドキュメント特徴ベクトルに類似するものとみなしてもよい。クエリーのメディアドキュメントの特徴ベクトルと識別されたピンの各々の各エントリーの各特徴ベクトルとを比較することによって、検索を更に精緻化することができる。

20

【0099】

本発明の一実施の形態において、エントリーのメディアドキュメントが3Dモデルであるとともにエントリーのデータベースが一連の3Dモデルを含む場合、類似する3Dモデルは、Iterative Closest Point (ICP) 技術を適用することによってデータベースにおいて識別される。一般的には、Iterative Closest Pointは、（例えば、3Dモデルに対応する）二つのポイントクラウドを調整することを試みる。ICP技術の一つの副産物は、あるポイントクラウドのポイントと他のポイントクラウドのポイントの間の距離のマップである。したがって、一実施の形態において、類似メトリックは、ポイントクラウドのポイント間の平均距離のような距離の統計値に対応する。

30

【0100】

クエリーのメディアドキュメントが同一の物の一つ以上の2D画像を有するとともにデータベースが一連の3Dモデルである一実施の形態において、動作510において抽出した特徴ベクトルを、一つ以上の2D画像の各々に対して個別の特徴ベクトルとしてもよく、複数の異なる視点から3Dモデルの各々をレンダリングするとともに、各々が異なる視点に対応する一つ以上の特徴ベクトルを生成するために3Dモデルの個別のレンダリングしたビューの各々を畳み込みニューラルネットワークに供給することによって、複数の特徴ベクトルを、データベースの3Dモデルの各々に対して生成することができる。この場合、クエリーの2D画像と3Dモデルの間の類似を計算するために、クエリーの2D画像から抽出した特徴ベクトルを（例えば、上述した L^1 又は L^2 メトリックを用いて）3Dモデルの2Dビューの特徴ベクトルと比較してもよい。

40

【0101】

メタデータを推論するのに用いることができる検索した物の更に関連するセットにつながるために、クエリー結果の検索したエントリーのセットをユーザに表示するとともにユーザによって手動で精緻化することもできる。

【0102】

本発明の実施の形態の一態様によれば、クエリーが部分的に満たされたメタデータを有する場合において、動作530の類似するエントリーについてのデータベースを検索する

50

ことは、クエリーメタデータの部分的に満たされた欄に類似するメタデータを含むエントリーを検索することを含んでもよい。類似するエントリーのメタデータを、クエリーのメタデータの満たされていない部分のメタデータを生成するために用いることができる。

【0103】

例えば、クエリーは、メタデータ<title = "black BMW m3 2016", class = "car", tags = "", text = ">を有することができる。そのようなクエリーにおいて、“title”エントリー及び“class”エントリーが満たされており、“tags”エントリー及び“text”エントリーが満たされていない。“tags”及び“text”の満たされていないエントリー並びに潜在的な更に詳しい種類を、“title”及び“class”の利用できる欄に基づいて類似するエントリーに対してデータベースを検索することによって自動的に生成することができる。特に、データベースを、クエリーのメタデータのtitleに類似するtitleを有するclass “car”のエントリーに対して検索することができる。例えば、エントリーのタイトルがクエリーのタイトルの少なくとも75%を含む場合、データベースのエントリーのタイトルは、十分に類似しているものとみなしてもよい。他の例として、レーベンシュタイン距離又は他の距離メトリック（例えば、編集距離メトリック）を、タイトルの比較を行うのに用いてもよく、この場合、しきい値距離より下の距離を有するエントリーを類似しているものとみなしてもよい。

10

【0104】

動作550において、類似するエントリーのタグ及びテキストメタデータ欄が判断され、そのコンテンツは、クエリーのメタデータの満たされていない欄に追加を行うタグ及びテキスト欄を自動的に生成するために併合される。一実施の形態において、検索したエントリーの全てのタグを列挙するとともに少なくとも二つのアイテムが出現するタグのみを保持することによってタグを取得することができ、意味的テキスト分析技術によってテキストを取得することができる。一実施の形態において、生成したメタデータの種類欄は、分類器によって特徴ベクトルに対して算出された分類（例えば、図6Bに示す分類656又は図7Bに示す分類776）を用いることによって満たされる。

20

【0105】

例えば、再び図1を参照すると、検索したエントリー12並びにそれに対応する種類及びタグメタデータを、以下の表2に示す。

30

【0106】

【表2】

検索したエントリー	種類	タグ
1	car	Jeep, blue, sport
2	car	Jeep, convertible, red
3	cap	black & white
4	car	convertible, sport

【0107】

本発明の一部の実施の形態によれば、生成したメタデータに少なくともしきい値の回数が出現するタグを追加するという規則を適用する。表2に示すように、タグ“Jeep”、“convertible”及び“sport”の全ては、結果のタグに2回出現し、その結果、しきい値が2であると仮定される。これらの三つのタグを、生成したメタデータのタグに追加する。タグ“blue”、“red”及び“black & white”は、結果に1回しか出現せず、したがって、これらは、生成したメタデータのタグに追加されない。タグ“black & white”がクエリーの3Dモデルの色を正確に反映するとしても、結果は、このタグを生成したメタデータに含めるには十分でない。

40

【0108】

検索の結果に少なくとも2回出現するメタデータを生成したメタデータに追加する例を

50

ここで説明するが、本発明の実施の形態はそれに限定されない。例えば、プロセッサを、多数の結果（例えば、最も類似する10個のエントリー）を検索するように構成してもよく、結果に出現する種類又はタグを追加するためのしきい値を少なくとも四つの結果としてもよい。他の実施の形態において、種類又はタグを追加するためのしきい値を、少なくとも50%のようなタグ又は種類を含む結果の最大割合として設定してもよい。

【0109】

本発明の一部の実施の形態によれば、生成したメタデータの種類欄を、クエリー結果に少なくとも2回出現する種類を用いて又はクエリー結果のほとんどに出現する種類に基づいて満たしてもよい。表2に示すように、種類“car”は、結果の三つに出現し、種類“cap”は、結果の一つに出現する。したがって、生成したメタデータは、種類“car”を有するが種類“cap”を有しない。

10

【0110】

本発明の一部の実施の形態によれば、種類をメディアドキュメントに割り当てるためにクエリー結果の種類を用いるのものではなく、クエリーのメディアドキュメントの特徴ベクトルが、一つ以上の分類（例えば、分類656又は分類776）を生成するために分類器（例えば、データベース内のエントリーについてトレーニングされた分類器）に供給される。分類を、複数の値として表すことができ、この場合、各値は、メディアドキュメントが特定の種類の属するという信頼性を表す。例えば、図1に示すクエリーの分類は、クエリーがcarを表現することについての高い信頼性、クエリーがcapを表現することについての低い信頼性及びクエリーがdogを表現することについてのゼロの信頼性を表すことができる。

20

【0111】

動作570において、プロセッサは、メディアドキュメントに対する生成したメタデータを出力する。これは、例えば、メディアドキュメントに関連して表示装置にメタデータを表示することを含んでもよい。ユーザは、自動的に生成したメタデータを視察し、生成したメタデータが結果に対して妥当であること（例えば、種類、タグ、名前及びテキスト記述がメディアドキュメントの性質を正確に反映していること）が正しいことを確認し、必要であれば生成したメタデータを編集してもよい。

【0112】

上述したように、一部の状況において、クエリーが一部のメタデータに既に関連していてもよい。しかしながら、クエリーのメディアドキュメントの特徴ベクトルとデータベースのエントリーの特徴ベクトルとの類似に基づいて他のメタデータを識別する処理を通じて、追加の関連のメタデータを自動的に生成してもよい。さらに、クエリーのメタデータの誤りを、自動的に生成したメタデータに基づいて訂正してもよい。

30

【0113】

生成したメタデータを、他の用途に用いるためにメディアドキュメントに関連して保存してもよい。

【0114】

ポリシーによって制限される検索

本発明の実施の形態の一態様は、アクセスポリシーに従ってデータベースの検索を制限することに関する。例えば、データベースに対してクエリーを行う種々のユーザは、ポリシーに基づいて、ユーザがアクセスすることができるデータベースの一部のみを検索することができる。例えば、5人のユーザ<a1, a2, a3, b1, b2>が存在し、ユーザ<a1, a2, a3>がグループAに属し、ユーザ<b1, b2>がグループBに属し、三つのセットのエントリー（又は三つのセットのメディアドキュメント）<dbA, dbB, dbC>によってデータベースが構成され、dbAは、グループAのメンバーによって見ることができ、dbBは、グループBのメンバーによって見ることができ、dbCは、グループAとグループBの両方のメンバーによって見ることができ、ユーザb1は、dbB及びdbCのエントリーのみの検索が許可（許容）され、dbAのエントリーへのアクセスが許可されない（拒否される）（したがって、検索は、dbAのエントリー

40

50

ーに対して実行されない。)。したがって、ユーザ b 1 によって行われる検索は、d b B 及び d b C のエントリーに限定される。これらのポリシーを、メタデータの異なる欄に異なるように適用することもできる。例えば、ユーザ b 1 は、タグ及びタイトル欄を追加するためにデータベース全体において検索することが許可されるが、テキスト欄を追加するために d b B 及び d b C しか検索することが許可されない。これらのポリシーの適用は、例えば、テキスト及び視覚情報、分類された情報、機密情報及び成人向け又は安全でない資料検索制限のデジタル著作権の行使を含む。

【 0 1 1 5 】

データベースへのエントリーの自動挿入

本発明の実施の形態の一態様によれば、クエリーのメディアドキュメント及びその自動的に生成したメタデータは、新たなエントリーとしてのデータベースへの挿入の候補を形成し、したがって、別のクエリーのメタデータの検索及び生成に更に用いることができるエントリーとなる。

10

【 0 1 1 6 】

しかしながら、検索後に自動的に追加したクエリーの欄が、データベースにエントリーを追加するユーザによって確認される場合、これは、供給されたメタデータ欄の品質がデータベースのエントリーとして安全に用いるのに十分高いことを保証するのに十分でないことがある。例えば、メタデータの誤りによって、新たに不正確にタグが付けられたエントリーによって追加のエントリーに不正確にタグが付けられるというカスケード効果が生じることがあり、これによって、後のクエリーに不正確にタグが付けられる可能性が高くなる。これは、これらのエントリーが分類器のトレーニングに用いられる場合に特に当てはまる。

20

【 0 1 1 7 】

したがって、高品質のメタデータ欄によって特徴付けられたクエリーのみがデータベースのエントリーとして組み込まれる可能性を高くするために、本発明の一実施の形態によれば、新たなエントリーに審査期間 (p r o b a t i o n t i m e) 及びデモクラティックバリデーション (d e m o c r a t i c v a l i d a t i o n) が課される。そのような実施の形態において、クエリーは、一時的にエントリーとしてデータベースに追加される。この一時的なエントリーが異なるユーザからの次のクエリーの消失メタデータの追加に用いられるとともにユーザが自動的に満たされたメタデータ欄の正当性を確認する場合、このエントリーは、有効であるとみなされ、データベースの正規のエントリーのプールに追加される。他の実施の形態において、エントリーのプールに追加する準備ができていないメディアドキュメントは、それが検索結果に出現することができる前にデータベースのアドミニストレータによって正当性が確認される。

30

【 0 1 1 8 】

自動的なメタデータの生成を伴うモデルキャプチャの例

明瞭のために、本発明の一実施の形態による 3 次元 (3 D) モデルをキャプチャするとともにモデルのメタデータを自動的に生成する処理の限定されない例を、図 8 に関連して以下で説明する。

【 0 1 1 9 】

図 8 は、本発明の一実施の形態によるモデルをキャプチャするとともにモデルのメタデータを自動的に生成する工程を示すフローチャートである。

40

【 0 1 2 0 】

便宜上、以下の例は、2017年6月22日に米国特許庁に出願された米国特許出願公開第 15 / 630 , 715 号明細書に記載されたタイプのスキャニングシステムのような 3 D スキャナの使用を仮定し、その開示の全体を参照によりここに組み込む。そのような 3 D スキャナは、3 次元物体の形状及び文字情報を取得することができる。

【 0 1 2 1 】

動作 810 において、3 D スキャナは、ランプのような特定の物理的な 3 次元物体の 3 次元的な形状及び文字を取得する。3 D モデルを生成するためにキャプチャした未加工デ

50

ータ（例えば、2次元画像）の3D形状及び文字を算出するための計算の一部を3Dスキャナで行うことができ、それに対し、残りの部分を、ローカルサーバ又はリモートサーバで行うことができる。この動作の出力は、特定の物のテクスチャード加工の3次元モデル（3Dモデル）（例えば、ランプの3Dモデル）である。

【0122】

動作820において、3Dモデルのデータベースは、キャプチャした3Dモデルを用いることによってクエリーが行われる。例えば、本発明の一実施の形態によれば、キャプチャした3Dモデル（例えば、キャプチャしたランプのモデル）は、ローカルサーバ又はリモートサーバにアップロードされ、ローカルサーバ又はリモートサーバは、この3Dモデルの検索を、分類及びラベル付けされた3Dモデルのデータベースに対して行う。便宜上、データベースのモデルの全てのメタデータ欄（例えば、名前、種類、タグ及びテキスト記述）の全てが追加されるとともにクエリーモデルのメタデータ欄のいずれも追加されないと仮定する。

10

【0123】

データベースのクエリーを、図5の動作510及び520に示すのと略同一の手法で行ってもよい。動作510において、特徴ベクトルを、トレーニングされた畳み込みニューラルネットワーク（CNN）を用いながらクエリーの3Dモデルに基づいて算出（又は抽出）する（そのようなデータベースでトレーニングされたCNNの重みを再算出してもよく、データベースの3Dモデルの各々の特徴ベクトルを、同一のCNNを用いて再算出してもよく、これらの特徴ベクトルをデータベースに格納してもよい。）。したがって、クエリーの3Dモデルの特徴ベクトルを、分類及び検索のために用いることができる。

20

【0124】

動作830において、クエリーの3Dモデルに対してメタデータ欄を自動的に生成する。上述したように、分類を、クエリー3Dモデルの特徴ベクトルが分類器に供給されるときにクエリーの3Dモデルの種類のセットとしてもよく、検索出力を、（例えば、類似する特徴ベクトルを有する）クエリーモデルに類似する3Dモデルを有するエントリーのセットとしてもよい。このような類似するエントリーモデルのセットを、名前、タグ及びテキスト記述のような種類以外のメタデータ欄の自動的な追加に用いることができる。

【0125】

動作840において、クエリーとして供給された3Dモデルの自動的に生成したメタデータ欄の全てがサーバによって追加されるとともに検証のためにユーザに対して表示される。動作850において、ユーザは、サーバによって提示されたメタデータの自動的に生成した値を検証及び修正する。一旦、メタデータが検証（場合によっては編集）されると、結果的に得られるメタデータを動作860で出力することができ、3Dモデルは、3次元モデルを有する電子商取引リスティングの作成のような他の状況で用いられ、この場合、自動的に生成したメタデータを、電子商取引リスティングの種々の箇所を自動的に満たすために用いることができる。

30

【0126】

電子商取引アプリケーションのメディアドキュメント

上述したように、自動的に生成したメタデータを有する3次元（3D）モデルのようなメディアドキュメントのメタデータを自動的に生成するアプリケーションの一例は、電子商取引アプリケーションにおけるものである。特に、販売される製品の3Dモデルを提供することは、ショッピング体験を向上させることができ、その結果、顧客が3D環境内で仮想的に商品を扱うことができるようにすることによる顧客の契約が増加し、これによって、製品のサイズ及び形状の更に実体験のように感じる理解を提供する。さらに、販売される製品の2D画像は、普通であり、購入者によって典型的に予測される。

40

【0127】

製品についての視覚情報（例えば、画像及び3次元モデル）に加えて、電子商取引アプリケーションにおけるリスティングは、一般的には、名前（又はタイトル）、テキスト記述、タグ（又はキーワード）のような多量の関連のデータを含み、（通常カテゴリーと称

50

される)種類に分かれている。特に、大規模な電子商取引のウェブサイトにおいて、リスティングカテゴリーは、数千以上ほどになることができ、サブカテゴリー関係を作成するツリー構造に分かれており、リストされた製品のそれぞれは一つ以上のカテゴリーに属する。

【0128】

したがって、販売者が販売中の製品をリストすることを所望するとき、販売者は、販売する製品のメタデータ欄を満たす必要がある。一般的には、これは、手動の処理であり、製品リスティングページによって要求される情報を収集するとともに書式を埋めるのにかなりの時間及び労力を必要とすることがある。特に、リスティングがモバイルプラットフォーム(例えば、スマートフォン)から作成されるとき、この動作は、時間がかかるとともに面倒である。

10

【0129】

したがって、本発明の実施の形態の態様による自動的なメタデータ生成によって、電子商取引ポスティング(e-commerce postings)のメタデータ欄を、キャプチャしたメタデータ(例えば、製品の写真又は物の3Dスキャン)を用いて自動的に追加することができる。本発明の一実施の形態を2次元画像及び3次元モデル本発明の実施の形態は、それに限定されず、ビデオのような他の形態のメディアに適用されてもよい(例えば、特徴ベクトルを、2D画像又は3Dモデルではなくビデオから抽出してもよく、結果的に得られる特徴ベクトルを、類似するエントリーからメタデータを自動的に検索するために画像に表された物を分類するとともに類似する物の画像及び/又はビデオを識別するのに用いてもよい。)

20

【0130】

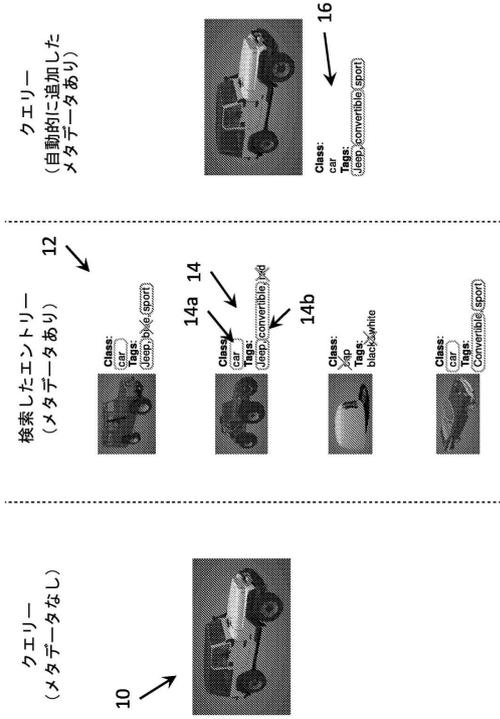
このような種々のメタデータ欄(タイトル、種類、タグ、テキスト記述)の自動的な追加を、リスティングの自己更新構造化データベース(self-updating organized database)によって取得し、これによって、効率的な分類動作及び検索動作を可能にする。

【0131】

本発明を所定の例示的な実施の形態に関連して説明したが、本発明が開示した実施の形態に限定されず、添付した特許請求の範囲の精神及び範囲並びにその等価物に含まれる種々の変形及び等価の配置をカバーすることを意図することを理解すべきである。

30

【 図 1 】



【 図 3 】

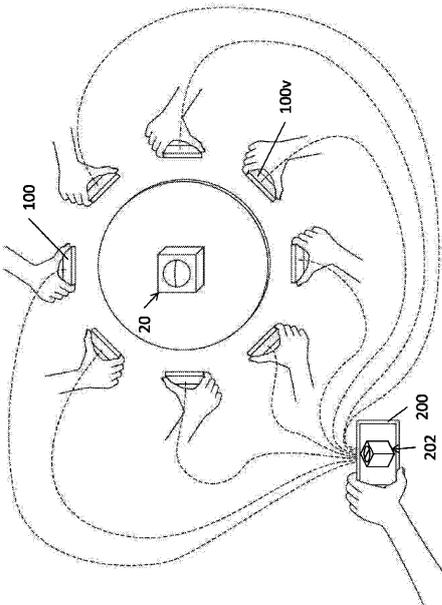


FIG. 3

【 図 2 】

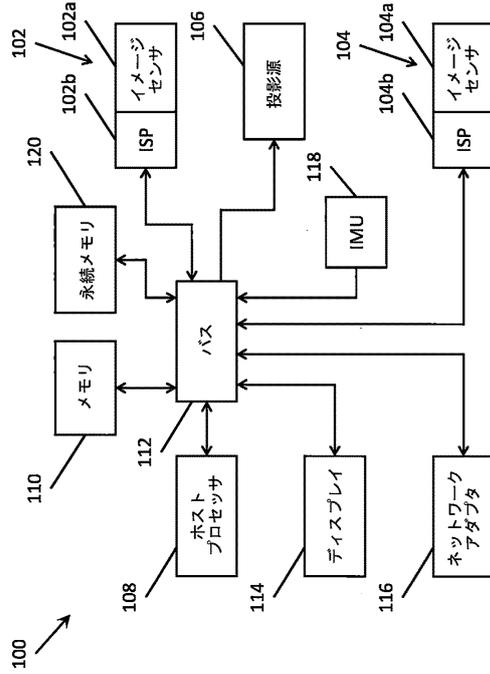
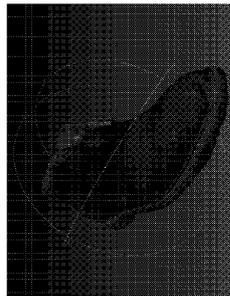


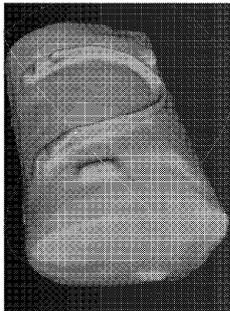
FIG. 1

FIG. 2

【 図 4 A 】



【 図 4 B 】



【図5】

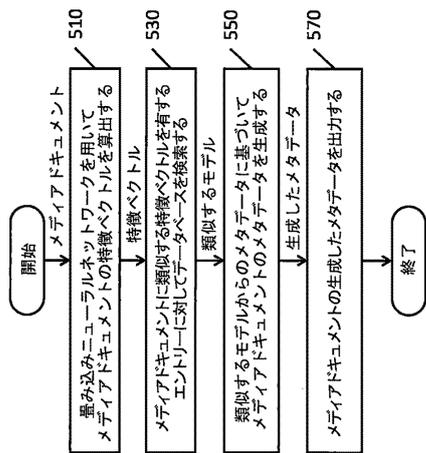


FIG. 5

【図6A】

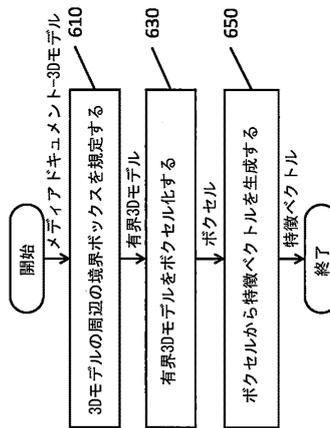


FIG. 6A

【図6B】

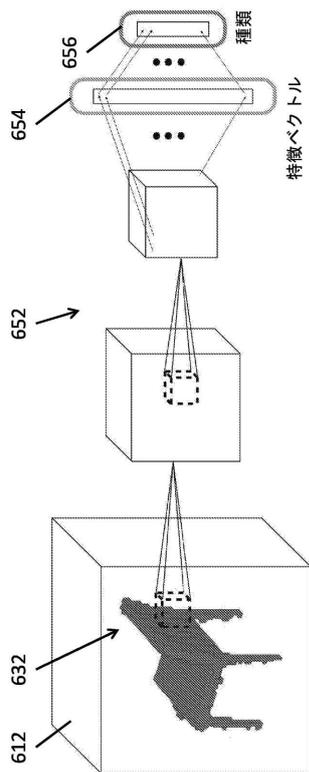


FIG. 6B

【図7A】

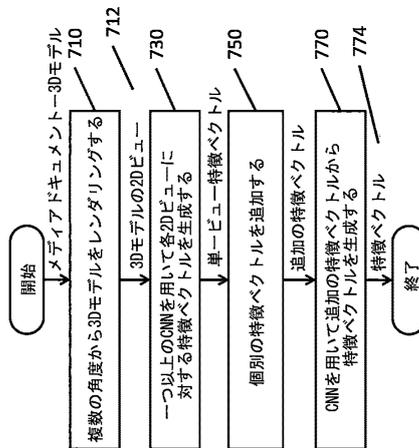


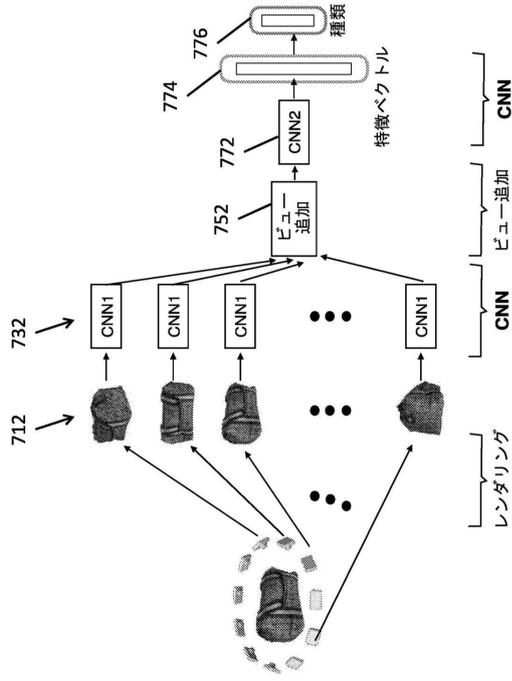
FIG. 7A

500 ↗

510 ↗

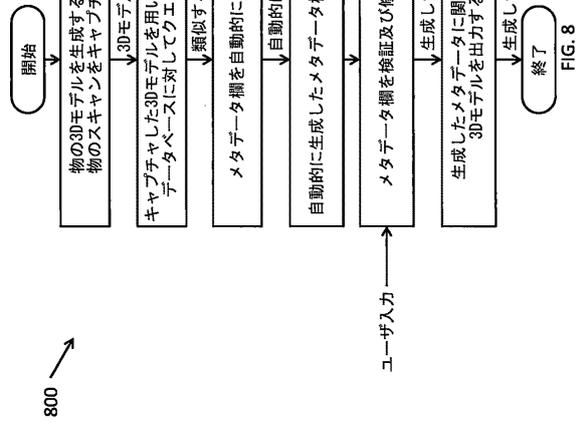
510' ↗

【 図 7 B 】



【 図 8 】

FIG. 7B



フロントページの続き

- (72)発明者 カルロ ダル ムットー
アメリカ合衆国, カリフォルニア 94303, パロ アルト, イースト ベイショア ロード
2225, スイート 110, シーノオー アキフィ, インコーポレイティド
- (72)発明者 アルビーゼ メーモ
アメリカ合衆国, カリフォルニア 94303, パロ アルト, イースト ベイショア ロード
2225, スイート 110, シーノオー アキフィ, インコーポレイティド
- (72)発明者 ダック グエン
アメリカ合衆国, カリフォルニア 94303, パロ アルト, イースト ベイショア ロード
2225, スイート 110, シーノオー アキフィ, インコーポレイティド
- (72)発明者 アップス ラフィー
アメリカ合衆国, カリフォルニア 94303, パロ アルト, イースト ベイショア ロード
2225, スイート 110, シーノオー アキフィ, インコーポレイティド
- (72)発明者 ジェyson トラチュウスキー
アメリカ合衆国, カリフォルニア 94303, パロ アルト, イースト ベイショア ロード
2225, スイート 110, シーノオー アキフィ, インコーポレイティド

審査官 吉田 誠

- (56)参考文献 特開2010-067014(JP, A)
特開2015-176485(JP, A)
国際公開第2014/103061(WO, A1)
鈴木 一史, 同値類に基づく回転不変特徴量を用いた3次元物体モデルの類似検索, 電子情報通
信学会論文誌, 日本, 社団法人電子情報通信学会, 2003年 8月 1日, 第J86-D-II巻第
8号, 1234-1243ページ
- (58)調査した分野(Int.Cl., DB名)
G06F 16/00 - 16/958
G06T 1/00