

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7362076号
(P7362076)

(45)発行日 令和5年10月17日(2023.10.17)

(24)登録日 令和5年10月6日(2023.10.6)

(51)国際特許分類 F I
G 0 6 F 16/532 (2019.01) G 0 6 F 16/532
G 0 6 F 16/583 (2019.01) G 0 6 F 16/583

請求項の数 21 (全25頁)

(21)出願番号	特願2021-87753(P2021-87753)	(73)特許権者	501440684 ソフトバンク株式会社 東京都港区海岸一丁目7番1号
(22)出願日	令和3年5月25日(2021.5.25)	(73)特許権者	500132214 学校法人明星学苑 東京都日野市程久保2丁目1番地1
(65)公開番号	特開2022-180958(P2022-180958 A)	(74)代理人	110002147 弁理士法人酒井国際特許事務所
(43)公開日	令和4年12月7日(2022.12.7)	(72)発明者	堀 隆之 東京都港区海岸一丁目7番1号 ソフト バンク株式会社内
審査請求日	令和4年3月16日(2022.3.16)	(72)発明者	金 容範 東京都港区海岸一丁目7番1号 ソフト バンク株式会社内
		(72)発明者	鈴木 裕真

最終頁に続く

(54)【発明の名称】 情報処理装置、情報処理方法及び情報処理プログラム

(57)【特許請求の範囲】

【請求項1】

セグメンテーションの技術を用いて領域分割された画像のうち、構造を有する物体を含む分割領域である物体領域に関する領域情報、および、姿勢推定の技術を用いて推定された前記物体の構造に関する構造情報に基づいて、前記物体の属性に関する属性情報を抽出する属性情報抽出部と、

前記属性情報抽出部によって抽出された属性情報に基づいて生成された文章であって、前記画像の内容を示す文章と前記画像とを対応付けて共通空間に埋め込むように学習されたVSE (Visual-Semantic Embedding) モデルを生成するモデル生成部と、

を備える情報処理装置。

【請求項2】

前記属性情報抽出部は、

前記画像のうち、前記領域情報として、人物である前記物体を含む分割領域である人物領域に関する人物領域情報、および、前記構造情報として、前記姿勢推定の技術を用いて推定された前記人物の骨格に関する骨格情報に基づいて、前記属性情報として、前記人物の属性に関する人物属性情報を抽出し、

前記モデル生成部は、

前記属性情報抽出部によって抽出された人物属性情報に基づいて生成された文章であって、前記画像の内容を示す文章と前記画像とを対応付けて共通空間に埋め込むように学習された前記VSEモデルを生成する、

請求項 1 に記載の情報処理装置。

【請求項 3】

前記属性情報抽出部は、

前記骨格情報として、前記人物の各関節の関節位置情報に基づいて、前記人物属性情報として、前記人物の姿勢に関する姿勢情報および前記人物の動作に関する動作情報を抽出する、

請求項 2 に記載の情報処理装置。

【請求項 4】

前記属性情報抽出部は、

前記人物領域情報として、前記人物が身に着けている各ファッションアイテムを含む分割領域であるアイテム領域に関するアイテム領域情報に基づいて、前記人物属性情報として、前記人物が身に着けているファッションアイテムの属性に関するアイテム属性情報を抽出する、

10

請求項 2 または 3 に記載の情報処理装置。

【請求項 5】

前記属性情報抽出部は、

前記人物領域情報として、前記人物の各身体部位を含む分割領域である身体部位領域に関する身体部位領域情報に基づいて、前記人物属性情報として、前記人物の身体部位の属性に関する身体部位属性情報を抽出する、

請求項 2 ~ 4 のいずれか 1 つに記載の情報処理装置。

20

【請求項 6】

前記属性情報抽出部は、

前記身体部位領域情報として、前記人物の頭髮の領域に関する情報に基づいて、前記身体部位属性情報として、前記人物の髪型に関する情報を抽出する、

請求項 5 に記載の情報処理装置。

【請求項 7】

前記属性情報抽出部は、

前記身体部位領域情報として、前記人物の顔の領域に関する情報に基づいて、前記身体部位属性情報として、前記人物の表情に関する情報を抽出する、

請求項 5 または 6 に記載の情報処理装置。

30

【請求項 8】

利用者によって入力された第 1 クエリ文を受け付ける受付部と、

前記モデル生成部によって生成された V S E モデルを用いて、前記第 1 クエリ文に関する第 1 画像を検索する検索部と、

前記第 1 画像に含まれる検出対象である第 1 コンセプトを抽出する抽出部と、

をさらに備え、

前記検索部は、

前記 V S E モデルを用いて、前記抽出部によって抽出された第 1 コンセプトに基づく第 2 クエリ文に関する第 2 画像を再検索する、

請求項 1 ~ 7 のいずれか 1 つに記載の情報処理装置。

40

【請求項 9】

前記検出対象は、物体、人物、場面および動作の少なくともいずれか一つを含む

請求項 8 に記載の情報処理装置。

【請求項 10】

前記抽出部によって抽出された第 1 コンセプトに基づいて、前記第 2 クエリ文を生成する生成部をさらに備え、

前記検索部は、

前記 V S E モデルを用いて、前記生成部によって生成された前記第 2 クエリ文に関する前記第 2 画像を再検索する、

請求項 8 に記載の情報処理装置。

50

【請求項 1 1】

前記検索部による検索結果を出力する出力部、をさらに備え、
前記出力部は、
前記抽出部によって抽出された第1コンセプトに関する第1コンセプト情報を出力し、
前記受付部は、
前記出力部によって出力された第1コンセプト情報に基づく前記第2クエリ文を前記利用者から受け付け、
前記検索部は、
前記VSEモデルを用いて、前記受付部によって受け付けられた前記第2クエリ文に関する前記第2画像を再検索する、
請求項8に記載の情報処理装置。

10

【請求項 1 2】

複数の映像または前記複数の映像それぞれに含まれる各シーンである画像を取得する取得部をさらに備え、
前記検索部は、
前記取得部によって取得された画像と前記受付部によって受け付けられた第1クエリ文の組を前記VSEモデルに入力して、前記画像と前記第1クエリ文との第1類似度を前記VSEモデルから出力し、当該出力された前記第1類似度が第1閾値を超える前記第1画像を検索する、
請求項8～11のいずれか1つに記載の情報処理装置。

20

【請求項 1 3】

前記検索部は、
前記取得部によって取得された画像と前記抽出部によって抽出された第1コンセプトに基づく第2クエリ文の組を前記VSEモデルに入力して、前記画像と前記第2クエリ文との第2類似度を前記VSEモデルから出力し、当該出力された前記第2類似度が第2閾値を超える前記第2画像を再検索する、
請求項12に記載の情報処理装置。

【請求項 1 4】

前記抽出部は、
画像が入力された場合に、前記画像に含まれる検出対象であるコンセプトと前記画像とのコンセプト類似度を出力するよう学習された学習済みの機械学習モデルであるコンセプト識別器を用いて、前記第1画像から前記第1コンセプトを抽出する、
請求項8～13のいずれか1つに記載の情報処理装置。

30

【請求項 1 5】

前記抽出部は、
前記検索部によって検索された第1画像を前記コンセプト識別器に入力して、前記第1画像に含まれる検出対象である第1コンセプトと前記第1画像とのコンセプト類似度を前記コンセプト識別器から出力し、当該出力されたコンセプト類似度がコンセプト閾値を超える前記第1コンセプトを抽出する、
請求項14に記載の情報処理装置。

40

【請求項 1 6】

前記抽出部は、
前記出力されたコンセプト類似度がコンセプト閾値を超える前記第1コンセプトの中から、前記第1クエリ文に含まれない文字列に対応する前記第1コンセプトである隠れコンセプトを抽出する、
請求項15に記載の情報処理装置。

【請求項 1 7】

前記受付部は、
前記利用者によって入力されたクエリ画像を受け付け、
前記検索部は、

50

前記 V S E モデルを用いて、前記受付部によって受け付けられたクエリ画像に関する文字列を検索し、当該検索した文字列に基づく第 3 クエリ文に関する第 3 画像を検索する、請求項 8 ~ 16 のいずれか 1 つに記載の情報処理装置。

【請求項 18】

複数の文章または前記複数の文章それぞれに含まれる各テキストである文字列を取得する取得部をさらに備え、

前記検索部は、

前記取得部によって取得された文字列と前記受付部によって受け付けられたクエリ画像の組を前記 V S E モデルに入力して、前記文字列と前記クエリ画像との第 3 類似度を前記 V S E モデルから出力し、当該出力された前記第 3 類似度が第 3 閾値を超える文字列を検索し、当該検索した文字列に基づく前記第 3 クエリ文に関する前記第 3 画像を検索する、請求項 17 に記載の情報処理装置。

10

【請求項 19】

前記抽出部は、

前記第 3 画像に含まれる検出対象である第 3 コンセプトを抽出し、

前記検索部は、

前記 V S E モデルを用いて、前記抽出部によって抽出された第 3 コンセプトに基づく第 4 クエリ文に関する第 4 画像を再検索する、

請求項 18 に記載の情報処理装置。

【請求項 20】

情報処理装置が実行するプログラムにより実現される情報処理方法であって、

セグメンテーションの技術を用いて領域分割された画像のうち、構造を有する物体を含む分割領域である物体領域に関する領域情報、および、姿勢推定の技術を用いて推定された前記物体の構造に関する構造情報に基づいて、前記物体の属性に関する属性情報を抽出する属性情報抽出工程と、

前記属性情報抽出工程によって抽出された属性情報に基づいて生成された文章であって、前記画像の内容を示す文章と前記画像とを対応付けて共通空間に埋め込むように学習された V S E (Visual-Semantic Embedding) モデルを生成するモデル生成工程と、

を含む情報処理方法。

20

【請求項 21】

セグメンテーションの技術を用いて領域分割された画像のうち、構造を有する物体を含む分割領域である物体領域に関する領域情報、および、姿勢推定の技術を用いて推定された前記物体の構造に関する構造情報に基づいて、前記物体の属性に関する属性情報を抽出する属性情報抽出手順と、

前記属性情報抽出手順によって抽出された属性情報に基づいて生成された文章であって、前記画像の内容を示す文章と前記画像とを対応付けて共通空間に埋め込むように学習された V S E (Visual-Semantic Embedding) モデルを生成するモデル生成手順と、

をコンピュータに実行させる情報処理プログラム。

30

【発明の詳細な説明】

【技術分野】

40

【0001】

本発明は、情報処理装置、情報処理方法及び情報処理プログラムに関する。

【背景技術】

【0002】

従来、文字情報から画像を検索可能とするため、画像の内容を示す文字情報（キャプションやタグなど）を生成する技術が知られている。例えば、画像に含まれる人物を検索可能とするため、画像において人物を含む人物領域を特定し、人物領域を複数の部分領域に分割する。そして、複数の部分領域それぞれにおいてクエリ要素を生成し、複数の部分領域のクエリ要素を組み合わせて人物を検索するための検索クエリを生成する技術が知られている。

50

【先行技術文献】

【特許文献】

【0003】

【文献】特開2016-162414号公報

特開2019-219988号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

画像の検索精度を向上させる技術が求められている。

【課題を解決するための手段】

【0005】

本願に係る情報処理装置は、セグメンテーションの技術を用いて領域分割された画像のうち、構造を有する物体を含む分割領域である物体領域に関する領域情報、および、姿勢推定の技術を用いて推定された前記物体の構造に関する構造情報に基づいて、前記物体の属性に関する属性情報を抽出する属性情報抽出部と、前記属性情報抽出部によって抽出された属性情報に基づいて生成された文章であって、前記画像の内容を示す文章と前記画像とを対応付けて共通空間に埋め込むように学習されたVSE (Visual-Semantic Embedding) モデルを生成するモデル生成部と、を備える。

【図面の簡単な説明】

【0006】

【図1】図1は、実施形態に係る情報処理の概要について説明するための図である。

【図2】図2は、実施形態に係るVSE (Visual-Semantic Embedding) モデルとコンセプト識別器について説明するための図である。

【図3】図3は、実施形態に係る情報処理装置の構成例を示す図である。

【図4】図4は、実施形態に係る属性情報の抽出処理手順とVSEモデルの生成処理手順を示す図である。

【図5】図5は、実施形態に係る情報処理手順を示す図である。

【図6】図6は、変形例に係る情報処理手順を示す図である。

【図7】図7は、情報処理装置の機能を実現するコンピュータの一例を示すハードウェア構成図である。

【発明を実施するための形態】

【0007】

以下に、本願に係る情報処理装置、情報処理方法及び情報処理プログラムを実施するための形態（以下、「実施形態」と呼ぶ）について図面を参照しつつ詳細に説明する。なお、この実施形態により本願に係る情報処理装置、情報処理方法及び情報処理プログラムが限定されるものではない。また、以下の各実施形態において同一の部位には同一の符号を付し、重複する説明は省略される。

【0008】

(実施形態)

〔1.はじめに〕

従来、画像に含まれる物体（例えば、人物）の属性に関する属性情報を画像から抽出する技術が知られている。例えば、画像に含まれる人物の姿勢に関する姿勢情報を抽出する姿勢推定 (Pose Estimation) の技術（以下、姿勢推定技術ともいう）が知られている。また、画像に含まれる人物の領域や服装の領域に関する情報を抽出するセグメンテーション (Segmentation) の技術が知られている。

【0009】

しかしながら、セグメンテーションの技術は、画像に含まれる人物の領域や服装の領域を精度よく抽出することができるものの、画像に含まれる人物の姿勢を抽出することはできない。また、姿勢推定の技術は、画像に含まれる人物の姿勢を精度よく抽出することができるものの、画像に含まれる人物の領域や服装の領域を精度よく抽出することは困難で

10

20

30

40

50

ある。

【 0 0 1 0 】

これに対して、一実施形態に係る情報処理装置 1 0 0 は、セグメンテーションの技術と姿勢推定の技術を組み合わせることで、画像に含まれる人物の領域や服装の領域、および画像に含まれる人物の姿勢を精度よく抽出することができる。図 1 は、一実施形態に係る情報処理の概要について説明するための図である。図 1 に示す情報処理は、後述する情報処理装置 1 0 0 (図 3 参照) によって行われる。

【 0 0 1 1 】

図 1 に示すように、一実施形態に係る情報処理装置 1 0 0 は、セグメンテーションの技術を用いて画像を領域分割し、領域分割された画像のうち人物を含む分割領域である人物領域に関する人物領域情報を抽出する。ここで、情報処理装置 1 0 0 は、セグメンテーションの技術を用いて画像を領域分割することで、人物の頭の先から足の先までの全体を含む人物領域を人物ごとに抽出してよい。また、情報処理装置 1 0 0 は、セグメンテーションの技術を用いて画像を領域分割することで、人物の各身体部位 (例えば、頭髮、顔、および人物が身に付けている各ファッションアイテムなど) を含む分割領域である身体部位領域に関する身体部位領域情報を身体部位ごとに抽出してよい。

【 0 0 1 2 】

ここで、セグメンテーションの技術とは、画像を入力として、ピクセルレベルで領域を分割しラベルを付けていく技術である。セグメンテーションの技術は、そのラベリングの意味合いから、3 種類に大別される。セマンティックセグメンテーション (Semantic Segmentation) は、画像上の全ピクセルをクラスに分類する技術である。マンティックセグメンテーションは、物体の種類ごとに画像を領域分割する。また、インスタンスセグメンテーション (Instance Segmentation) は、物体ごとの領域を分割し、かつ物体の種類を認識する技術である。インスタンスセグメンテーションは、物体ごとに画像を領域分割する。また、パノプティックセグメンテーション (Panoptic Segmentation) は、セマンティックセグメンテーションとインスタンスセグメンテーションを組み合わせた技術である。パノプティックセグメンテーションは、人物や動物、自動車などの物体 (数えられるクラス、Thing クラスともいう) に対してインスタンスセグメンテーションを行い、空や道路、芝生などの背景 (数えられないクラス、Stuff クラスともいう) に対してセマンティックセグメンテーションを行う技術である。情報処理装置 1 0 0 は、セマンティックセグメンテーション、インスタンスセグメンテーション、またはパノプティックセグメンテーションの技術を用いて画像を領域分割してよい。また、情報処理装置 1 0 0 は、セマンティックセグメンテーション、インスタンスセグメンテーションおよびパノプティックセグメンテーションの技術を組み合わせて画像を領域分割してよい。

【 0 0 1 3 】

また、情報処理装置 1 0 0 は、姿勢推定の技術を用いて画像に含まれる人物の骨格に関する骨格情報を抽出する。具体的には、情報処理装置 1 0 0 は、姿勢推定に関するあらゆる公知技術を用いて画像から骨格情報を抽出してよい。例えば、情報処理装置 1 0 0 は、姿勢推定モデルと呼ばれる深層学習モデルを用いて、動画や静止画から人物や動物の姿勢 (骨格) を推定する姿勢推定技術を用いて、骨格情報を抽出してよい。なお、情報処理装置 1 0 0 は、1 枚の画像に複数の人物が写っている場合には、姿勢推定処理により、複数の人物について特徴点を検出して、複数の人物の骨格に関する骨格情報を推定してもよい。情報処理装置 1 0 0 は、姿勢推定技術を用いることで、画像中の人物の身体の部位を精緻に推定することができる。

【 0 0 1 4 】

例えば、姿勢推定モデルの一例として、OpenPose (“OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, Zhe Cao et al., 2018) が知られている。OpenPose は、画像に含まれる人物の身体の特徴を示す特徴点 (キーポイントともいう) を検出し、特徴点を結んだ人物の姿勢を推定する姿勢推定モデルである。例えば、OpenPose は、画像に含まれる人物の身体の特徴

10

20

30

40

50

点として、人物の身体の関節の位置を示す関節点を推定し、関節点を連結して生成される人物の身体の骨格を示す骨格モデルを人物の姿勢として検出する。また、例えば、特徴点を検出するタイプの姿勢推定モデルの中には、画像中の人物の身体の30種類の部位を推定可能なものがある。具体的には、特徴点を検出するタイプの姿勢推定モデルを用いると、画像中の人物の身体の部位として、頭、目（右、左）、耳（右、左）、鼻、首、肩（右、中央、左）、肘（右、左）、背骨、手首（右、左）、手（右、左）、親指（右、左）、手先（右、左）、腰（右、中央、左）、膝（右、左）、足首（右、左）、足（右、左）を特定することができる。

【0015】

また、姿勢推定モデルの一例として、DensePose（参考URL：http://openaccess.thecvf.com/content_cvpr_2018/html/Guler_DensePose_Dense_Human_CVPR_2018_paper.html）が知られている。DensePoseは、2次元の画像中の人物の人物領域を検出し、検出した人物領域に対応する3次元身体表面モデルを生成する姿勢推定モデルである。より具体的には、DensePoseは、RGB画像を入力として、RGB画像中の人物の3次元表面のUV座標を推定する。DensePoseを用いると、2次元の画像に写る人物領域から3次元身体表面のUV座標を推定することができるので、2次元の画像に写る人物領域の各人体パーツ（人物の身体の部位）を精緻に推定することができる。DensePoseを用いると、画像中の人物の身体の24種類の部位を推定することができる。具体的には、DensePoseを用いると、RGB画像から、画像中の人物の身体の部位として、頭（左、右）、首、胴体、腕（左、右/上、前/前、後）、脚（左、右/太もも、ふくらはぎ/前、後）、手（左、右）、足（左、右）を特定することができる。

【0016】

続いて、情報処理装置100は、抽出した人物領域情報および骨格情報に基づいて、画像に含まれる人物の属性に関する人物属性情報を抽出する。例えば、情報処理装置100は、人物属性情報の一例として、例えば、抽出された骨格情報に基づいて、人物の姿勢に関する姿勢情報を抽出する。また、情報処理装置100は、人物属性情報の一例として、例えば、抽出された骨格情報の時間変化のパターンに基づいて、人物の動作に関する動作情報を抽出する。例えば、情報処理装置100は、画像に含まれる人物が立っている状態から座っている状態に姿勢を変化させたという情報を抽出する。また、情報処理装置100は、人物属性情報の一例として、例えば、抽出した各身体部位の身体部位領域の情報に基づいて、人物が身に着けている服装、髪型、および表情に関する情報を抽出する。例えば、情報処理装置100は、画像に含まれる人物の服装が青い服であるという情報を抽出する。続いて、情報処理装置100は、抽出した人物属性情報に基づいて、画像の内容を示す文章を生成する。例えば、情報処理装置100は、画像に含まれる人物が立っている状態から座っている状態に姿勢を変化させたという情報と人物の服装が青い服であるという情報とに基づいて、画像の内容を示す文章の一例として、「青い服を来た人がレストラン内で赤いソファに座った。」という文章を生成してよい。

【0017】

上述したように、一実施形態に係る情報処理装置100は、セグメンテーションの技術を用いて領域分割された画像のうち、人物を含む分割領域である人物領域に関する人物領域情報、および、姿勢推定の技術を用いて推定された人物の骨格に関する骨格情報に基づいて、人物の属性に関する人物属性情報を抽出する。また、一実施形態に係る情報処理装置100は、抽出された人物属性情報に基づいて生成された文章であって、画像の内容を示す文章と画像とを対応付けて共通空間に埋め込むように学習されたVSEモデルを生成する。

【0018】

このように、一実施形態に係る情報処理装置100は、セグメンテーションの技術と姿勢推定の技術を組み合わせることで、画像に含まれる人物の領域や服装の領域、および画像に含まれる人物の姿勢を精度よく抽出することができる。これにより、情報処理装置1

10

20

30

40

50

00は、画像に含まれる人物領域および人物領域の構成要素（例えば、各関節の関節位置情報、身体部位領域、およびファッションアイテムの領域など）を階層的に分解することができる。また、情報処理装置100は、分解された構成要素の階層的な関係性に基づいて、人物領域や骨格情報から、人物の属性に関する人物属性情報を適切に抽出することができる。また、情報処理装置100は、画像から適切に抽出された人物属性情報に基づいて、画像の内容を示す文章を適切に生成することができる。したがって、一実施形態に係る情報処理装置100は、画像の検索精度を向上させることができる。

【0019】

なお、上述した例では、情報処理装置100が、セグメンテーションの技術を用いて画像から人物領域情報を抽出し、抽出した人物領域情報に基づいて、人物属性情報を抽出する場合について説明したが、これに限られない。例えば、情報処理装置100は、セグメンテーションの技術を用いて、画像から人物領域を抽出する。続いて、情報処理装置100は、姿勢推定の技術を用いて、人物領域から人物領域に含まれる人物の骨格に関する骨格情報を抽出してよい。このように、情報処理装置100は、姿勢推定の技術を用いて骨格情報を推定するために、セグメンテーションの技術を用いて抽出された人物領域の情報をを用いることで、より精度よく姿勢情報を推定することができる。また、情報処理装置100は、人物属性情報を推定するために、画像情報、人物領域情報、および骨格情報を組み合わせた情報を用いてよい。これにより、情報処理装置100は、より精度よく人物属性情報を推定することができる。

【0020】

また、近年、膨大な数の映像コンテンツを効率よく整理、管理するために、画像の内容を自動で分析し、画像の内容を示す文字情報（キャプションやタグなど）を自動で生成・付与する技術が知られている。これにより、文字情報から画像を検索可能とすることができる。ここで、本願明細書における「画像」とは、映像などの動画であってもよいし、映像に含まれる各シーン（静止画）であってもよい。

【0021】

このような背景の下、文字情報から画像を検索する代表的な2つの手法を比較した研究が知られている。具体的には、(1)画像に含まれる物体、人物、場面および動作等の検出対象（以下、コンセプトともいう）をあらかじめ学習した学習済みの機械学習モデル（以下、コンセプト識別器ともいう）を用いてクエリ文から画像を検索する手法と、(2)画像の特徴を示す画像特徴量と、画像の内容を言語で表現した言語表現の特徴を示す言語特徴量とが対応付けられて埋め込まれた共通空間を用いて、クエリ文とマッチする画像を検索する手法とを比較した研究が知られている（参考文献；「Comparison and Evaluation of Video Retrieval Approaches Using Query Sentences」、IMIP 2020: Proceedings of the 2020 2nd International Conference on Intelligent Medicine and Image Processing, April 2020, Pages 103-107、<https://doi.org/10.1145/3399637.3399657>）。

【0022】

上記の研究によると、(1)に示すコンセプト識別器を用いる手法と(2)に示すVSEを用いる手法は、相補的であることが示されている。そこで、本願発明の発明者は、(1)に示すコンセプト識別器を用いる手法と(2)に示すVSEに基づく手法を統合することで、画像の検索精度を向上させる技術を提案する。具体的には、一実施形態に係る情報処理装置100は、コンセプト識別器を用いて画像から適切なコンセプトを抽出し、VSEモデルを用いて抽出されたコンセプトと類似する画像を再検索する。例えば、情報処理装置100は、視覚グラフ情報と文字グラフ情報とを対応付けて共通空間に埋め込むように学習されたVSEモデル（以下、VSEモデルと記載する場合がある）を用いて抽出されたコンセプトと類似する画像を再検索する。これにより、一実施形態に係る情報処理装置100は、例えば、利用者から受け付けたクエリ文に明示されていないコンセプト（例えば、暗示的なコンセプト）の中から、適切なコンセプトを抽出して、抽出したコンセプトと類似する画像を再検索することができる。したがって、本願発明の一実施形態によ

10

20

30

40

50

れば、画像の検索精度を向上させることができる。

【0023】

ここから、図2を用いて、実施形態に係るVSEモデルとコンセプト識別器について説明する。図2は、実施形態に係るVSEモデルとコンセプト識別器について説明するための図である。

【0024】

図2の左側は、実施形態に係るVSEモデルの一例を示す。例えば、図2の左側に示すVSEモデルは、画像から文を検索する場合、画像をVSEモデルに入力して、画像の特徴を示す画像特徴量に対応する特徴ベクトル（以下、画像の特徴ベクトルともいう）を生成する。続いて、VSEモデルは、生成した画像の特徴ベクトルを文の特徴を示す文特徴量に対応する特徴ベクトル（以下、文の特徴ベクトルともいう）との共通の空間にマッピングしたのち、画像の特徴ベクトルと類似する文の特徴ベクトルに対応する文を検索結果とする。例えば、VSEモデルは、画像の特徴ベクトルと文の特徴ベクトルとの類似度が所定の閾値を超えるような文の特徴ベクトルに対応する文を検索結果として出力する。

10

【0025】

また、VSEモデルは、画像から画像特徴量を抽出することができる。VSEモデルは、画像から画像特徴量を抽出することができる任意の公知技術により実現されてよい。例えば、VSEモデルは、畳み込みニューラルネットワーク（CNN：Convolutional Neural Network）を備えてよい。そして、VSEモデルは、CNNを用いて、画像から画像特徴量を抽出してよい。また、例えば、VSEモデルは、物体認識用に開発されたResNet（Residual Network）（Kaiming He et al., 2015）、AlexNet（Krizhevsky et al., 2012）、VGGNet（Simonyan et al., 2014）、GoogLeNet（Szegedy et al., 2014）、SENet（Squeeze-and-Excitation Networks）（Jie Hu et al., 2018）、EfficientNet（Tan et al., 2019）、またはZFNet（Matthew et al., 2013）を備えてよい。そして、VSEモデルは、ResNet、AlexNet、VGGNet、GoogLeNet、SENet、EfficientNet、またはZFNetを用いて、画像から画像特徴量を抽出してよい。また、例えば、VSEモデルは、物体検出用に開発されたFaster R-CNN（Shaoqing Ren et al., 2015）、YOLO（You Look Only Once）（Joseph Redmon et al., 2015）、またはSSD（Single Shot MultiBox Detector）（Wei Liu., 2015）を備えてよい。そして、VSEモデルは、Faster R-CNN、YOLO、またはSSDを用いて、画像から画像特徴量を抽出してよい。

20

30

【0026】

また、図2の左側に示すVSEモデルは、文から画像を検索する場合、文をVSEモデルに入力して、文の特徴ベクトルを生成する。続いて、VSEモデルは、生成した文の特徴ベクトルを画像の特徴ベクトルとの共通の空間にマッピングしたのち、文の特徴ベクトルと類似する画像の特徴ベクトルに対応する画像を検索結果とする。例えば、VSEモデルは、文の特徴ベクトルと画像の特徴ベクトルとの類似度が所定の閾値を超えるような画像の特徴ベクトルに対応する画像を検索結果として出力する。

【0027】

また、VSEモデルは、言語表現（文章、フレーズ、又はキーワードなど）から言語表現の特徴を示す言語特徴量を抽出する。VSEモデルは、言語特徴量を抽出することができる任意の公知技術により実現されてよい。例えば、VSEモデルは、再帰型ニューラルネットワーク（RNN：Recurrent Neural Network）を備えてよい。そして、VSEモデルは、RNNを用いて、言語表現から言語特徴量を抽出してよい。また、例えば、VSEモデルは、GRU（Gated Recurrent Unit）またはLSTM（Long Short Term Memory）を備えてよい。そして、VSEモデルは、GRUまたはLSTMを用いて、言語表現から言語特徴量を抽出してよい。また、例えば、VSEモデルは、Transformer（Ashish Vaswani et al., 2017）、TransformerをベースとしたBERT（Bidirectional Encoder Representations from Transformers）

40

50

、GPT-3 (Generative Pre-Training3) またはT5 (Text-to-Text Transfer Transformer) を備えてよい。そして、VSEモデルは、Transformer、BERT、GPT-3 またはT5 を用いて、言語表現から言語特徴量を抽出してよい。

【0028】

図2の右側は、実施形態に係るコンセプト識別器の出力結果の一例を示す。コンセプト識別器は、コンセプトを含む画像が入力された場合に、画像に含まれるコンセプトと画像との類似度を示すコンセプト類似度を出力するよう学習された学習済みの機械学習モデルである。ここで、コンセプト識別器が学習するコンセプト(検出対象ともいう)には、画像に含まれる物体や人物等の対象物に限らず、画像の場面(シーン)および画像に含まれる人物や動物等の動作(走っている、座っている等)等の概念が含まれる。例えば、図2の右側に示すコンセプト識別器の出力結果は、バイクの横に男性が立っている画像がコンセプト識別器に入力された場合に、画像に含まれる男性の髪の毛の色、男性が着ている服装、男性の体の部位、バイクの色、背景の山や海、赤い橋といった対象を検出し、対象のクラス(カテゴリ)を出力したものである。なお、図2の右側では図示を省略しているが、コンセプト識別器は、画像に含まれるコンセプトのクラスとともに、画像に含まれるコンセプトが当該コンセプトのクラス(カテゴリ)に該当する確率を出力する。このように、コンセプト識別器は、コンセプトを含む画像が入力された場合に、画像に含まれるコンセプトを検出するとともに、検出されたコンセプトのクラスを推定する。すなわち、コンセプト識別器は、コンセプトを含む画像が入力された場合に、コンセプト類似度として、画像に含まれる各コンセプトが推定された各コンセプトのクラスに該当する確率をそれぞれ出力する。

10

20

【0029】

〔2. 情報処理装置の構成〕

次に、図3を用いて、実施形態に係る情報処理装置の構成について説明する。図3は、実施形態に係る情報処理装置の構成例を示す図である。図3に示すように、情報処理装置100は、通信部110と、記憶部120と、入力部130と、出力部140と、制御部150とを有する。

【0030】

(通信部110)

通信部110は、例えば、NIC (Network Interface Card)、モデムチップ及びアンテナモジュール等によって実現される。また、通信部110は、ネットワークN(図示略)と有線又は無線で接続される。

30

【0031】

(記憶部120)

記憶部120は、例えば、RAM (Random Access Memory)、フラッシュメモリ等の半導体メモリ素子、又は、ハードディスク、光ディスク等の記憶装置によって実現される。例えば、記憶部120は、複数の映像または複数の映像それぞれに含まれる各シーンである画像のデータベースである映像プールを記憶する。また、記憶部120は、複数の文章または複数の文章それぞれに含まれる各テキストである文字列のデータベースであるキャプションプールを記憶する。

40

【0032】

(入力部130)

入力部130は、利用者から各種操作の入力を受け付ける。例えば、入力部130は、タッチパネル機能により表示面(例えば出力部140)を介して利用者からの各種操作を受け付けてもよい。また、入力部130は、情報処理装置100に設けられたボタンや、情報処理装置100に接続されたキーボードやマウスからの各種操作を受け付けてもよい。例えば、入力部130は、利用者からクエリ文の入力を受け付けてよい。また、入力部130は、利用者からクエリ画像の入力を受け付けてよい。

【0033】

(出力部140)

50

出力部 140 は、例えば液晶ディスプレイや有機 E L (Electro-Luminescence) ディスプレイ等によって実現される表示画面であり、各種情報を表示するための表示装置である。出力部 140 は、制御部 150 の制御に従って、各種情報を表示する。なお、情報処理装置 100 にタッチパネルが採用される場合には、入力部 130 と出力部 140 とは一体化される。また、以下の説明では、出力部 140 を画面と記載する場合がある。

【0034】

(制御部 150)

制御部 150 は、コントローラ (Controller) であり、例えば、CPU (Central Processing Unit)、MPU (Micro Processing Unit)、ASIC (Application Specific Integrated Circuit) や FPGA (Field Programmable Gate Array) 等によって、情報処理装置 100 の内部の記憶装置に記憶されている各種プログラム (情報処理プログラムの一例に相当) が RAM 等の記憶領域を作業領域として実行されることにより実現される。図 3 に示す例では、制御部 150 は、属性情報抽出部 151 と、モデル生成部 152 と、受付部 153 と、取得部 154 と、検索部 155 と、抽出部 156 と、生成部 157 を有する。

10

【0035】

(属性情報抽出部 151)

属性情報抽出部 151 は、セグメンテーションの技術を用いて領域分割された画像のうち、構造を有する物体を含む分割領域である物体領域に関する領域情報を抽出する。例えば、属性情報抽出部 151 は、構造を有する物体の一例として、人物を含む分割領域である人物領域に関する人物領域情報を抽出する。例えば、属性情報抽出部 151 は、人物領域情報の一例として、人物が身に着けている各ファッションアイテムを含む分割領域であるアイテム領域に関するアイテム領域情報を抽出してよい。

20

【0036】

また、属性情報抽出部 151 は、人物領域情報の一例として、人物の各身体部位を含む分割領域である身体部位領域に関する身体部位領域情報を抽出してよい。例えば、属性情報抽出部 151 は、身体部位領域情報の一例として、人物の頭髮の領域に関する情報を抽出してよい。また、属性情報抽出部 151 は、身体部位領域情報の一例として、人物の顔の領域に関する情報を抽出してよい。

【0037】

また、属性情報抽出部 151 は、姿勢推定の技術を用いて推定された物体の構造に関する構造情報を抽出する。例えば、属性情報抽出部 151 は、構造情報の一例として、姿勢推定の技術を用いて推定された人物の骨格に関する骨格情報を抽出する。例えば、属性情報抽出部 151 は、骨格情報の一例として、人物の各関節の関節位置情報を抽出する。

30

【0038】

また、属性情報抽出部 151 は、領域情報と構造情報を抽出すると、抽出した領域情報と構造情報に基づいて、物体の属性に関する属性情報を抽出する。例えば、属性情報抽出部 151 は、抽出した人物領域情報と骨格情報に基づいて、画像に含まれる人物の属性に関する人物属性情報を抽出する。

【0039】

例えば、属性情報抽出部 151 は、人物属性情報の一例として、抽出した人物の各関節の関節位置情報に基づいて、人物の姿勢に関する姿勢情報を抽出する。例えば、属性情報抽出部 151 は、人物が立っている状態、座っている状態、または右手を挙げている状態である等の姿勢情報を抽出してよい。

40

【0040】

また、属性情報抽出部 151 は、人物属性情報の一例として、抽出した人物の各関節の関節位置情報の時間変化のパターンに基づいて、人物の動作に関する動作情報を抽出する。例えば、属性情報抽出部 151 は、人物が立っている状態から座っている状態に姿勢を変化させたという動作情報を抽出してよい。また、属性情報抽出部 151 は、人物が走っている、歩いている、または右手を振っている等の動作情報を抽出してよい。

50

【 0 0 4 1 】

また、属性情報抽出部 1 5 1 は、人物属性情報の一例として、抽出したアイテム領域情報に基づいて、人物が身に着けているファッションアイテムの属性に関するアイテム属性情報を抽出する。例えば、属性情報抽出部 1 5 1 は、アイテム属性情報の一例として、人物が身に着けている衣服や靴、帽子、カバンなどのファッションアイテムの種類や色、形状、材質等を示す情報を抽出してよい。

【 0 0 4 2 】

また、属性情報抽出部 1 5 1 は、人物属性情報の一例として、人物の身体部位の属性に関する身体部位属性情報を抽出する。例えば、属性情報抽出部 1 5 1 は、身体部位属性情報の一例として、人物の頭髮の領域に関する情報に基づいて、人物の髪型に関する情報を抽出する。例えば、属性情報抽出部 1 5 1 は、人物の髪型の種類や色、形状、毛質等を示す情報を抽出してよい。また、属性情報抽出部 1 5 1 は、身体部位領域情報の一例として、人物の顔の領域に関する情報に基づいて、人物の表情に関する情報を抽出する。例えば、属性情報抽出部 1 5 1 は、人物の表情の種類（笑っている、怒っている等）や表情の度合い（少し笑っている、とても怒っている等）等を示す情報を抽出してよい。

10

【 0 0 4 3 】

なお、上述した例では、属性情報抽出部 1 5 1 が、構造を有する物体の一例として、画像に含まれる人物の属性に関する人物属性情報を抽出する場合について説明したが、属性情報抽出部 1 5 1 は、構造を有する物体であれば、人物以外のどのような物体の属性情報を抽出してもよい。例えば、属性情報抽出部 1 5 1 は、構造を有する物体の一例として、ドア部分、窓ガラス部分、およびタイヤ部分といったパーツを組み合わせて構成される車両の属性に関する属性情報を抽出してよい。例えば、属性情報抽出部 1 5 1 は、画像のうち、領域情報として、車両である物体を含む分割領域である車両領域に関する車両領域情報、および、構造情報として、姿勢推定の技術を用いて推定された車両の骨格に関する骨格情報に基づいて、属性情報として、車両の属性に関する車両属性情報を抽出する。

20

【 0 0 4 4 】

（モデル生成部 1 5 2）

モデル生成部 1 5 2 は、属性情報抽出部 1 5 1 によって抽出された属性情報に基づいて生成された文章であって、画像の内容を示す文章と画像とを対応付けて共通空間に埋め込むように学習された V S E モデルを生成する。具体的には、モデル生成部 1 5 2 は、属性情報抽出部 1 5 1 によって抽出された人物属性情報に基づいて生成された文章であって、画像の内容を示す文章と画像とを対応付けて共通空間に埋め込むように学習された V S E モデルを生成する。

30

【 0 0 4 5 】

例えば、モデル生成部 1 5 2 は、属性情報抽出部 1 5 1 によって抽出された人物属性情報の特徴を示す人物属性特徴ベクトルを生成する。例えば、モデル生成部 1 5 2 は、セグメンテーションの技術を用いて抽出された人物領域の特徴を示す人物領域特徴ベクトルを生成する。また、モデル生成部 1 5 2 は、姿勢推定の技術を用いて推定された人物の骨格情報の特徴を示す骨格情報特徴ベクトルを生成する。続いて、モデル生成部 1 5 2 は、生成した人物領域特徴ベクトルと骨格情報特徴ベクトルをつなぎ合わせることで、人物属性特徴ベクトルを生成する。また、モデル生成部 1 5 2 は、生成した画像の内容を示す文章の特徴を示す文章特徴ベクトルを生成する。モデル生成部 1 5 2 は、人物属性特徴ベクトルと文章特徴ベクトルを生成すると、生成した人物属性特徴ベクトルと文章特徴ベクトルとが共通空間において類似するように V S E モデルを学習することで、V S E モデルを生成する。

40

【 0 0 4 6 】

（受付部 1 5 3）

受付部 1 5 3 は、利用者によって入力されたクエリ文を受け付ける。例えば、受付部 1 5 3 は、入力部 1 3 0 を介して利用者が入力したクエリ文を受け付ける。ここで、本願明細書におけるクエリ文とは、完全な文章でなくてもよく、例えば、キーワードやフレーズ

50

であってもよい。以下では、受付部 153 が利用者から最初に受け付けたクエリ文を「第 1 クエリ文」と記載する。例えば、受付部 153 は、第 1 クエリ文の一例として、利用者から「person in a car」というフレーズを受け付ける。

【0047】

(取得部 154)

取得部 154 は、映像プールから画像を取得する。例えば、取得部 154 は、受付部 153 が第 1 クエリ文を受け付けると、記憶部 120 を参照して、複数の映像または複数の映像それぞれに含まれる各シーンである画像を映像プールから取得する。例えば、取得部 154 は、N 個 (N は自然数) の画像 # 11 ~ 画像 # 1N を映像プールから取得する。

【0048】

(検索部 155)

検索部 155 は、モデル生成部 152 によって生成された VSE モデルを用いて、受付部 153 によって受け付けられた第 1 クエリ文に関する第 1 画像を検索する。具体的には、検索部 155 は、取得部 154 が画像を取得すると、受付部 153 によって受け付けられた第 1 クエリ文と取得部 154 によって取得された画像の組を VSE モデルに入力する。例えば、検索部 155 は、第 1 クエリ文である「person in a car」と N 個の画像 # 11 ~ 画像 # 1N それぞれとの組を VSE モデルに入力する。

【0049】

続いて、検索部 155 は、第 1 クエリ文と画像との第 1 類似度を VSE モデルから出力する。例えば、検索部 155 は、第 1 クエリ文と N 個の画像 # 11 ~ 画像 # 1N それぞれとの類似度 # 11 ~ 類似度 # 1N それぞれを出力する。続いて、検索部 155 は、出力された第 1 類似度が第 1 閾値を超える第 1 画像を検索する。例えば、類似度 # 11 ~ 類似度 # 13 は第 1 閾値を超えるが、類似度 # 14 ~ 類似度 # 1N は第 1 閾値以下であるとする。このとき、検索部 155 は、N 個の画像 # 11 ~ 画像 # 1N の中から、第 1 クエリ文との第 1 類似度が第 1 閾値を超える画像 # 11 ~ 画像 # 13 を第 1 画像として取得する。

【0050】

なお、検索部 155 は、出力された第 1 類似度が第 1 閾値を超える第 1 画像を検索する代わりに、出力された第 1 類似度が高い方から順にいくつかの第 1 画像を検索してよい。例えば、第 1 クエリ文と N 個の画像 # 11 ~ 画像 # 1N それぞれとの類似度 # 11 ~ 類似度 # 1N のうち、類似度 # 11 の類似度が最も高く、N が大きくなるほど類似度が低いとする。このとき、検索部 155 は、N 個の画像 # 11 ~ 画像 # 1N の中から、第 1 クエリ文との第 1 類似度が高い方から順に、例えば、3 つの画像 # 11 ~ 画像 # 13 を第 1 画像として取得してよい。

【0051】

(抽出部 156)

抽出部 156 は、第 1 画像に関するコンセプトを抽出する。具体的には、抽出部 156 は、コンセプトを含む画像が入力された場合に、画像に含まれるコンセプトと画像とのコンセプト類似度を出力するよう学習された学習済みの機械学習モデルであるコンセプト識別器を用いて、第 1 画像から第 1 画像に関するコンセプトを抽出する。例えば、抽出部 156 は、検索部 155 によって第 1 画像が検索されると、検索部 155 によって検索された第 1 画像をコンセプト識別器に入力する。例えば、抽出部 156 は、検索部 155 によって検索された第 1 画像である画像 # 11 ~ 画像 # 13 それぞれをコンセプト識別器に入力する。なお、以下では、簡単のため、第 1 画像が画像 # 11 のみである場合について説明する。

【0052】

続いて、抽出部 156 は、第 1 画像に含まれるコンセプトと第 1 画像とのコンセプト類似度をコンセプト識別器から出力する。例えば、抽出部 156 は、画像 # 11 に含まれるコンセプトである「car_interior」と画像 # 11 とのコンセプト類似度 # 21 である「90%」をコンセプト識別器から出力する。また、例えば、抽出部 156 は、画像 # 11 に含まれるコンセプトである「自動車」と画像 # 11 とのコンセプト類似度 # 22 である「

10

20

30

40

50

80%」をコンセプト識別器から出力する。また、例えば、抽出部156は、画像#11に含まれるコンセプトである「バイク」と画像#11とのコンセプト類似度#23である「70%」をコンセプト識別器から出力する。

【0053】

続いて、抽出部156は、出力されたコンセプト類似度がコンセプト閾値を超えるコンセプトを抽出する。例えば、コンセプト閾値が「85%」であるとする。このとき、抽出部156は、コンセプト類似度がコンセプト閾値である「85%」を超えるコンセプトである「car_interior」を抽出する。続いて、抽出部156は、出力されたコンセプト類似度がコンセプト閾値を超えるコンセプトの中から、第1クエリ文に含まれないコンセプトである隠れコンセプトを抽出する。例えば、抽出部156は、第1クエリ文に含まれない文字列を含むコンセプトを隠れコンセプトとして抽出する。例えば、抽出部156は、第1クエリ文である「person in a car」に含まれない文字列である「interior」を含むコンセプトである「car_interior」を隠れコンセプトとして抽出する。

10

【0054】

なお、抽出部156は、出力されたコンセプト類似度がコンセプト閾値を超えるコンセプトを抽出する代わりに、出力されたコンセプト類似度が高い方から順にいくつかのコンセプトを抽出してよい。例えば、第1画像とM個（Mは自然数）のコンセプト#11～コンセプト#1Mそれぞれとのコンセプト類似度#11～コンセプト類似度#1Mのうち、コンセプト類似度#11のコンセプト類似度が最も高く、Mが大きくなるほどコンセプト類似度が低いとする。このとき、抽出部156は、M個のコンセプト#11～コンセプト#1Mの中から、第1画像とのコンセプト類似度が高い方から順に、例えば、3つのコンセプト#11～コンセプト#13を抽出してよい。

20

【0055】

（生成部157）

以下では、抽出部156によって抽出された隠れコンセプトに基づいて生成されたクエリ文のことを「第2クエリ文」と記載する。生成部157は、抽出部156によって抽出された隠れコンセプトに基づいて、第2クエリ文を生成する。例えば、生成部157は、抽出部156によって抽出された隠れコンセプトに基づいて、第1クエリ文を更新して、第2クエリ文を生成してよい。例えば、生成部157は、抽出部156によって抽出された隠れコンセプトである「car_interior」を含む第2クエリ文を生成してよい。出力部140は、生成部157によって生成された第2クエリ文を出力する。例えば、出力部140は、生成部157によって生成された第2クエリ文の一例として、「car_interior」を出力する。受付部153は、出力部140によって出力された第2クエリ文を利用者から受け付ける。例えば、受付部153は、出力部140によって出力された第2クエリ文である「car_interior」を利用者から受け付ける。

30

【0056】

なお、生成部157が第2クエリ文を生成する代わりに、出力部140によって出力された隠れコンセプトに基づいて利用者が第2クエリ文を生成してもよい。受付部153は、利用者によって生成された第2クエリ文を利用者から受け付けてもよい。

【0057】

40

また、取得部154は、受付部153が第2クエリ文を受け付けると、記憶部120を参照して、複数の映像または複数の映像それぞれに含まれる各シーンである画像を映像プールから取得する。検索部155は、VSEモデルを用いて、受付部153によって受け付けられた第2クエリ文に関する第2画像を再検索する。例えば、検索部155は、受付部153によって受け付けられた第2クエリ文である「car_interior」に関する第2画像を再検索する。例えば、検索部155は、受付部153によって受け付けられた第2クエリ文と取得部154によって取得された画像の組をVSEモデルに入力する。続いて、検索部155は、画像と第2クエリ文との第2類似度をVSEモデルから出力する。続いて、検索部155は、出力された第2類似度が第2閾値を超える第2画像を再検索する。出力部140は、検索部155によって再検索された第2画像を検索結果として出力する。

50

このようにして、出力部 1 4 0 は、例えば、隠れコンセプトである「car_interior」に基づいて生成された第 2 クエリ文である「car_interior」に関する第 2 画像を検索結果として出力する。

【 0 0 5 8 】

上述した例では、抽出部 1 5 6 が、第 1 クエリ文である「person in a car」に含まれない文字列「interior」を含むコンセプトである「car_interior」を隠れコンセプトとして抽出する場合について説明したが、他の例について説明する。例えば、受付部 1 5 3 は、第 1 クエリ文の一例として、「destroyed old building」というフレーズを利用者から受け付ける。検索部 1 5 5 は、V S E モデルを用いて、受付部 1 5 3 によって受け付けられた第 1 クエリ文である「destroyed old building」に関する第 1 画像を検索する。抽出部 1 5 6 は、第 1 画像をコンセプト識別器に入力して、第 1 画像に含まれるコンセプトである「ruin」を抽出する。続いて、抽出部 1 5 6 は、第 1 クエリ文である「destroyed old building」に含まれない文字列である「ruin」を含むコンセプトである「ruin」を隠れコンセプトとして抽出する。

10

【 0 0 5 9 】

〔 3 . 情報処理のフロー 〕

次に、図 4 を用いて、実施形態に係る情報処理の手順について説明する。図 4 は、実施形態に係る情報処理の一例を示すフローチャートである。図 4 では、属性情報抽出部 1 5 1 が、処理対象となる画像を取得する（ステップ S 1 1 ）。例えば、属性情報抽出部 1 5 1 は、処理対象となる画像を映像プールから取得する。

20

【 0 0 6 0 】

続いて、属性情報抽出部 1 5 1 は、セグメンテーションの技術を用いて、画像のうち人物を含む人物領域に関する人物領域情報を抽出する。また、属性情報抽出部 1 5 1 は、姿勢推定の技術を用いて、画像に含まれる人物の骨格情報を抽出する（ステップ S 1 2 ）。

【 0 0 6 1 】

続いて、属性情報抽出部 1 5 1 は、人物領域情報および骨格情報に基づいて、画像に含まれる人物の属性に関する人物属性情報を抽出する（ステップ S 1 3 ）。属性情報抽出部 1 5 1 は、人物属性情報を抽出すると、抽出された人物属性情報に基づいて、画像の内容を示す文章を生成する（ステップ S 1 4 ）。

【 0 0 6 2 】

モデル生成部 1 5 2 は、属性情報抽出部 1 5 1 が生成した文章と画像とを対応付けて共通空間に埋め込むように学習された V S E モデルを生成する（ステップ S 1 5 ）。

30

【 0 0 6 3 】

次に、図 5 を用いて、実施形態に係る情報処理の手順について説明する。図 5 は、実施形態に係る情報処理の一例を示すフローチャートである。図 5 では、受付部 1 5 3 が、利用者によって入力された第 1 クエリ文を受け付ける（ステップ S 1 0 1 ）。取得部 1 5 4 は、受付部 1 5 3 が第 1 クエリ文を受け付けると、複数の映像または複数の映像それぞれに含まれる各シーンである画像を映像プールから取得する（ステップ S 1 0 2 ）。

【 0 0 6 4 】

検索部 1 5 5 は、受付部 1 5 3 によって受け付けられた第 1 クエリ文と取得部 1 5 4 によって取得された画像の組を V S E モデルに入力する。続いて、検索部 1 5 5 は、第 1 クエリ文と画像との第 1 類似度を V S E モデルから出力する（ステップ S 1 0 3 ）。続いて、検索部 1 5 5 は、出力された第 1 類似度が第 1 閾値を超える第 1 画像を検索する（ステップ S 1 0 4 ）。

40

【 0 0 6 5 】

抽出部 1 5 6 は、検索部 1 5 5 によって検索された第 1 画像をコンセプト識別器に入力する（ステップ S 1 0 5 ）。続いて、抽出部 1 5 6 は、第 1 画像に含まれるコンセプトと第 1 画像とのコンセプト類似度をコンセプト識別器から出力する（ステップ S 1 0 6 ）。続いて、抽出部 1 5 6 は、出力されたコンセプト類似度がコンセプト閾値を超えるコンセプトを抽出する。続いて、抽出部 1 5 6 は、出力されたコンセプト類似度がコンセプト閾

50

値を超えるコンセプトの中から、第1クエリ文に含まれないコンセプトである隠れコンセプトを抽出する(ステップS107)。

【0066】

出力部140は、抽出部156によって抽出された隠れコンセプトを出力する(ステップS108)。利用者は、出力部140によって出力された隠れコンセプトに基づいて第1クエリ文を更新し、新たな第2クエリ文を生成する(ステップS109)。受付部153は、出力部140によって出力された隠れコンセプトに基づいて更新された第2クエリ文を利用者から受け付ける。例えば、受付部153は、入力部130を介して、利用者によって更新された第2クエリ文を利用者から受け付ける(ステップS110)。取得部154は、受付部153が第2クエリ文を受け付けると、複数の映像または複数の映像それぞれに含まれる各シーンである画像を映像プールから取得する(ステップS111)。

10

【0067】

検索部155は、受付部153によって受け付けられた第2クエリ文と取得部154によって取得された画像の組をVSEモデルに入力する。続いて、検索部155は、画像と第2クエリ文との第2類似度をVSEモデルから出力する(ステップS112)。続いて、検索部155は、出力された第2類似度が第2閾値を超える第2画像を再検索する(ステップS113)。出力部140は、検索部155によって再検索された第2画像を検索結果として出力する(ステップS114)。

【0068】

〔4.変形例〕

次に、図6を用いて、変形例に係る情報処理の手順について説明する。図6は、変形例に係る情報処理の一例を示すフローチャートである。図6では、受付部153が、利用者によって入力されたクエリ画像を受け付ける(ステップS201)。例えば、受付部153は、入力部130を介して利用者が入力したクエリ画像を受け付ける。ここで、本願明細書におけるクエリ画像とは、画像全体でなくともよく、例えば、画像の一部であってもよい。

20

【0069】

取得部154は、受付部153がクエリ画像を受け付けると、記憶部120を参照して、複数の文章または複数の文章それぞれに含まれる各テキストである文字列をキャプションプールから取得する(ステップS202)。

30

【0070】

また、検索部155は、取得部154が文字列を取得すると、受付部153によって受け付けられたクエリ画像と取得部154によって取得された文字列の組をVSEモデルに入力する。続いて、検索部155は、クエリ画像と文字列との第3類似度をVSEモデルから出力する(ステップS203)。続いて、検索部155は、出力された第3類似度が第3閾値を超える文字列を検索する(ステップS204)。

【0071】

生成部157は、検索部155によって検索された文字列に基づいて第3クエリ文を生成する。出力部140は、生成部157によって生成された第3クエリ文を出力する。受付部153は、出力部140によって出力された第3クエリ文を利用者から受け付ける(ステップS205)。取得部154は、受付部153が第3クエリ文を受け付けると、複数の映像または複数の映像それぞれに含まれる各シーンである画像を映像プールから取得する(ステップS206)。

40

【0072】

検索部155は、受付部153によって受け付けられた第3クエリ文と取得部154によって取得された画像の組をVSEモデルに入力する。続いて、検索部155は、画像と第3クエリ文との第1類似度をVSEモデルから出力する(ステップS207)。続いて、検索部155は、出力された第1類似度が第1閾値を超える第3画像を検索する(ステップS208)。

【0073】

50

抽出部 156 は、検索部 155 によって検索された第 3 画像をコンセプト識別器に入力する（ステップ S 209）。続いて、抽出部 156 は、第 3 画像に含まれるコンセプトと第 3 画像とのコンセプト類似度をコンセプト識別器から出力する（ステップ S 210）。続いて、抽出部 156 は、出力されたコンセプト類似度がコンセプト閾値を超えるコンセプトを抽出する。続いて、抽出部 156 は、出力されたコンセプト類似度がコンセプト閾値を超えるコンセプトの中から、第 3 クエリ文に含まれないコンセプトである隠れコンセプトを抽出する（ステップ S 211）。

【0074】

出力部 140 は、抽出部 156 によって抽出された隠れコンセプトを出力する（ステップ S 212）。利用者は、出力部 140 によって出力された隠れコンセプトに基づいて第 3 クエリ文を更新し、新たな第 4 クエリ文を生成する（ステップ S 213）。受付部 153 は、出力部 140 によって出力された隠れコンセプトに基づいて更新された第 4 クエリ文を利用者から受け付ける。例えば、受付部 153 は、入力部 130 を介して、利用者によって更新された第 4 クエリ文を利用者から受け付ける（ステップ S 214）。取得部 154 は、受付部 153 が第 4 クエリ文を受け付けると、複数の映像または複数の映像それぞれに含まれる各シーンである画像を映像プールから取得する（ステップ S 215）。

【0075】

検索部 155 は、受付部 153 によって受け付けられた第 4 クエリ文と取得部 154 によって取得された画像の組を V S E モデルに入力する。続いて、検索部 155 は、画像と第 4 クエリ文との第 2 類似度を V S E モデルから出力する（ステップ S 216）。続いて、検索部 155 は、出力された第 2 類似度が第 2 閾値を超える第 4 画像を再検索する（ステップ S 217）。出力部 140 は、検索部 155 によって再検索された第 4 画像を検索結果として出力する（ステップ S 218）。

【0076】

なお、ステップ S 205 において、利用者は、出力部 140 によって出力された第 3 クエリ文を変更することができる。受付部 153 は、入力部 130 を介して、利用者によって変更された第 3 クエリ文を利用者から受け付ける。

【0077】

〔5. 効果〕

上述してきたように、実施形態に係る情報処理装置 100 は、属性情報抽出部 151 と、モデル生成部 152 を有する。属性情報抽出部 151 は、セグメンテーションの技術を用いて領域分割された画像のうち、構造を有する物体を含む分割領域である物体領域に関する領域情報、および、姿勢推定の技術を用いて推定された物体の構造に関する構造情報に基づいて、物体の属性に関する属性情報を抽出する。モデル生成部 152 は、属性情報抽出部 151 によって抽出された属性情報に基づいて生成された文章であって、画像の内容を示す文章と画像とを対応付けて共通空間に埋め込むように学習された V S E (Visual-Semantic Embedding) モデルを生成する。

【0078】

このように、情報処理装置 100 は、セグメンテーションの技術と姿勢推定の技術を組み合わせることで、画像に含まれる構造を有する物体（例えば、車両や人物）の領域や物体の構成要素（各部位や各パーツ）の領域、および画像に含まれる物体の姿勢を精度よく抽出することができる。これにより、情報処理装置 100 は、画像に含まれる物体領域および物体領域の構成要素（例えば、物体の構造、各部位の領域など）を階層的に分解することができる。また、情報処理装置 100 は、分解された構成要素の階層的な関係性に基づいて、物体領域や構造情報から、物体の属性に関する属性情報を適切に抽出することができる。また、情報処理装置 100 は、画像から適切に抽出された属性情報に基づいて、画像の内容を示す文章を適切に生成することができる。したがって、本願発明に係る情報処理装置 100 は、画像の検索精度を向上させることができる。

【0079】

また、属性情報抽出部 151 は、画像のうち、領域情報として、人物である物体を含む

10

20

30

40

50

分割領域である人物領域に関する人物領域情報、および、構造情報として、姿勢推定の技術を用いて推定された人物の骨格に関する骨格情報に基づいて、属性情報として、人物の属性に関する人物属性情報を抽出する。モデル生成部 152 は、属性情報抽出部 151 によって抽出された人物属性情報に基づいて生成された文章であって、画像の内容を示す文章と画像とを対応付けて共通空間に埋め込むように学習された VSE モデルを生成する。

【0080】

情報処理装置 100 は、セグメンテーションの技術と姿勢推定の技術を組み合わせることで、画像に含まれる人物の領域や服装の領域、および画像に含まれる人物の姿勢を精度よく抽出することができる。これにより、情報処理装置 100 は、画像に含まれる人物領域および人物領域の構成要素（例えば、各関節の関節位置情報、身体部位領域、およびファッションアイテムの領域など）を階層的に分解することができる。また、情報処理装置 100 は、分解された構成要素の階層的な関係性に基づいて、人物領域や骨格情報から、人物の属性に関する人物属性情報を適切に抽出することができる。また、情報処理装置 100 は、画像から適切に抽出された人物属性情報に基づいて、画像の内容を示す文章を適切に生成することができる。したがって、本願発明に係る情報処理装置 100 は、画像の検索精度を向上させることができる。

【0081】

また、属性情報抽出部 151 は、骨格情報として、人物の各関節の関節位置情報に基づいて、人物属性情報として、人物の姿勢に関する姿勢情報および人物の動作に関する動作情報を抽出する。

【0082】

これにより、情報処理装置 100 は、画像に含まれる人物の姿勢および動作に関する情報を適切に抽出することができるので、画像の内容を示す文章を適切に生成することができる。

【0083】

また、属性情報抽出部 151 は、人物領域情報として、人物が身に着けている各ファッションアイテムを含む分割領域であるアイテム領域に関するアイテム領域情報に基づいて、人物属性情報として、人物が身に着けているファッションアイテムの属性に関するアイテム属性情報を抽出する。

【0084】

これにより、情報処理装置 100 は、画像に含まれる人物のファッションアイテムに関する情報を適切に抽出することができるので、画像の内容を示す文章を適切に生成することができる。

【0085】

また、属性情報抽出部 151 は、人物領域情報として、人物の各身体部位を含む分割領域である身体部位領域に関する身体部位領域情報に基づいて、人物属性情報として、人物の身体部位の属性に関する身体部位属性情報を抽出する。例えば、属性情報抽出部 151 は、身体部位領域情報として、人物の頭髮の領域に関する情報に基づいて、身体部位属性情報として、人物の髪型に関する情報を抽出する。また、属性情報抽出部 151 は、身体部位領域情報として、人物の顔の領域に関する情報に基づいて、身体部位属性情報として、人物の表情に関する情報を抽出する。

【0086】

これにより、情報処理装置 100 は、画像に含まれる人物の髪型や表情に関する情報を適切に抽出することができるので、画像の内容を示す文章を適切に生成することができる。

【0087】

また、情報処理装置 100 は、受付部 153 と、検索部 155 と、抽出部 156 を有する。受付部 153 は、利用者によって入力された第 1 クエリ文を受け付ける。検索部 155 は、モデル生成部 152 によって生成された VSE モデルを用いて、第 1 クエリ文に関する第 1 画像を検索する。抽出部 156 は、第 1 画像に関するコンセプトを抽出する。検索部 155 は、VSE モデルを用いて、抽出部 156 によって抽出されたコンセプトに基

10

20

30

40

50

づく第2クエリ文に関する第2画像を再検索する。

【0088】

これにより、情報処理装置100は、VSEを用いることで、利用者によって入力されたクエリ文に関する画像を適切に検索することができる。また、情報処理装置100は、適切に検索された画像からコンセプトを抽出したうえで、抽出したコンセプトに基づいて画像を再検索することができる。したがって、情報処理装置100は、画像の検索精度を向上させることができる。

【0089】

また、情報処理装置100は、生成部157をさらに備える。生成部157は、抽出部156によって抽出されたコンセプトに基づいて、第2クエリ文を生成する。検索部155は、VSEモデルを用いて、生成部157によって生成された第2クエリ文に関する第2画像を再検索する。

10

【0090】

これにより、情報処理装置100は、適切なコンセプトに基づいて適切なクエリ文を生成することができる。例えば、情報処理装置100は、適切な検索キーワードを追加（または不適切な検索キーワードを排除）することで、検索精度を向上させることを可能にする。したがって、情報処理装置100は、適切なクエリ文に基づいて画像を再検索することができるので、画像の検索精度を向上させることができる。

【0091】

また、情報処理装置100は、検索部155による検索結果を出力する出力部140をさらに備える。出力部140は、抽出部156によって抽出されたコンセプトを出力する。受付部153は、出力部140によって出力されたコンセプトに基づく第2クエリ文を利用者から受け付ける。検索部155は、VSEモデルを用いて、受付部153によって受け付けられた第2クエリ文に関する第2画像を再検索する。

20

【0092】

これにより、情報処理装置100は、利用者が、適切なコンセプトに基づいて適切なクエリ文を生成するのに助けることができる。例えば、情報処理装置100は、利用者が適切な検索キーワードを追加（または不適切な検索キーワードを排除）することで、検索精度を向上させることを可能にする。したがって、情報処理装置100は、適切なクエリ文に基づいて画像を再検索することができるので、画像の検索精度を向上させることができる。

30

【0093】

また、情報処理装置100は、取得部154をさらに備える。取得部154は、複数の映像または複数の映像それぞれに含まれる各シーンである画像を取得する。検索部155は、取得部154によって取得された画像と受付部153によって受け付けられた第1クエリ文の組をVSEモデルに入力して、画像と第1クエリ文との第1類似度をVSEモデルから出力し、出力された第1類似度が第1閾値を超える第1画像を検索する。

【0094】

これにより、情報処理装置100は、VSEに基づく処理により、適切な画像を選択することができる。

40

【0095】

また、検索部155は、取得部154によって取得された画像と抽出部156によって抽出されたコンセプトに基づく第2クエリ文の組をVSEモデルに入力して、画像と第2クエリ文との第2類似度をVSEモデルから出力し、出力された第2類似度が第2閾値を超える第2画像を再検索する。

【0096】

これにより、情報処理装置100は、VSEに基づく処理とコンセプト識別器に基づく処理を回すことで、適切なコンセプトを選択することができる。例えば、情報処理装置100は、利用者が入力したクエリ文に明示されていない内容（例えば、暗示的な内容）に関するコンセプトを抽出することができる。

50

【 0 0 9 7 】

また、抽出部 1 5 6 は、コンセプトを含む画像が入力された場合に、画像に含まれるコンセプトと画像とのコンセプト類似度を出力するよう学習された学習済みの機械学習モデルであるコンセプト識別器を用いて、第 1 画像から第 1 画像に関するコンセプトを抽出する。

【 0 0 9 8 】

これにより、情報処理装置 1 0 0 は、コンセプト識別器を用いることで、適切に検索された画像から適切なコンセプトを抽出することができる。また、情報処理装置 1 0 0 は、適切なコンセプトを抽出したうえで、適切なコンセプトに基づいて画像を再検索することができる。したがって、情報処理装置 1 0 0 は、画像の検索精度を向上させることができる。

10

【 0 0 9 9 】

また、抽出部 1 5 6 は、検索部 1 5 5 によって検索された第 1 画像をコンセプト識別器に入力して、第 1 画像に含まれるコンセプトと第 1 画像とのコンセプト類似度をコンセプト識別器から出力し、出力されたコンセプト類似度がコンセプト閾値を超えるコンセプトを抽出する。

【 0 1 0 0 】

これにより、情報処理装置 1 0 0 は、適切なコンセプトを抽出することができる。

【 0 1 0 1 】

また、抽出部 1 5 6 は、出力されたコンセプト類似度がコンセプト閾値を超えるコンセプトの中から、第 1 クエリ文に含まれないコンセプトである隠れコンセプトを抽出する。

20

【 0 1 0 2 】

これにより、情報処理装置 1 0 0 は、利用者が入力したクエリ文に明示されていない内容（例えば、暗示的な内容）に関するコンセプトを抽出することができる。また、情報処理装置 1 0 0 は、利用者が入力したクエリ文に明示されていないコンセプト（例えば、暗示的なコンセプト）に基づいて画像を再検索することができる。したがって、情報処理装置 1 0 0 は、画像の検索精度を向上させることができる。

また、受付部 1 5 3 は、利用者によって入力されたクエリ画像を受け付ける。検索部 1 5 5 は、V S E モデルを用いて、受付部 1 5 3 によって受け付けられたクエリ画像に関する文字列を検索し、検索した文字列に基づく第 3 クエリ文に関する第 3 画像を検索する。

30

【 0 1 0 3 】

これにより、情報処理装置 1 0 0 は、利用者が入力したクエリ画像に明示されていない内容（例えば、暗示的な内容）に関するコンセプトを抽出することができる。

【 0 1 0 4 】

また、取得部 1 5 4 は、複数の文章または複数の文章それぞれに含まれる各テキストである文字列を取得する。検索部 1 5 5 は、取得部 1 5 4 によって取得された文字列と受付部 1 5 3 によって受け付けられたクエリ画像の組を V S E モデルに入力して、文字列とクエリ画像との第 3 類似度を V S E モデルから出力し、出力された第 3 類似度が第 3 閾値を超える文字列を検索し、検索した文字列に基づく第 3 クエリ文に関する第 3 画像を検索する。

40

【 0 1 0 5 】

これにより、情報処理装置 1 0 0 は、V S E に基づく処理とコンセプト識別器に基づく処理を回すことで、適切なコンセプトを選択することができる。

【 0 1 0 6 】

また、抽出部 1 5 6 は、第 3 画像に関するコンセプトを抽出する。検索部 1 5 5 は、V S E モデルを用いて、抽出部 1 5 6 によって抽出されたコンセプトに基づく第 4 クエリ文に関する第 4 画像を再検索する。

【 0 1 0 7 】

これにより、情報処理装置 1 0 0 は、利用者が入力したクエリ画像に明示されていない内容（例えば、暗示的な内容）に関するコンセプトを抽出することができる。また、情報

50

処理装置 100 は、利用者が入力したクエリ画像に明示されていないコンセプト（例えば、暗示的なコンセプト）に基づいて画像を再検索することができる。したがって、情報処理装置 100 は、画像の検索精度を向上させることができる。

【0108】

〔6. ハードウェア構成〕

また、上述してきた実施形態に係る情報処理装置 100 は、例えば図 7 に示すような構成のコンピュータ 1000 によって実現される。図 7 は、情報処理装置 100 の機能を実現するコンピュータの一例を示すハードウェア構成図である。コンピュータ 1000 は、CPU 1100、RAM 1200、ROM 1300、HDD 1400、通信インターフェイス (I/F) 1500、入出力インターフェイス (I/F) 1600、及びメディアインターフェイス (I/F) 1700 を備える。

10

【0109】

CPU 1100 は、ROM 1300 または HDD 1400 に格納されたプログラムに基づいて動作し、各部の制御を行う。ROM 1300 は、コンピュータ 1000 の起動時に CPU 1100 によって実行されるブートプログラムや、コンピュータ 1000 のハードウェアに依存するプログラム等を格納する。

【0110】

HDD 1400 は、CPU 1100 によって実行されるプログラム、及び、かかるプログラムによって使用されるデータ等を格納する。通信インターフェイス 1500 は、所定の通信網を介して他の機器からデータを受信して CPU 1100 へ送り、CPU 1100 が生成したデータを所定の通信網を介して他の機器へ送信する。

20

【0111】

CPU 1100 は、入出力インターフェイス 1600 を介して、ディスプレイやプリンタ等の出力装置、及び、キーボードやマウス等の入力装置を制御する。CPU 1100 は、入出力インターフェイス 1600 を介して、入力装置からデータを取得する。また、CPU 1100 は、生成したデータを入出力インターフェイス 1600 を介して出力装置へ出力する。なお、CPU 1100 の代わりに、MPU (Micro Processing Unit)、また多大な計算パワーを必要とすることから GPU (Graphics Processing Unit) を用いてもよい。

【0112】

メディアインターフェイス 1700 は、記録媒体 1800 に格納されたプログラムまたはデータを読み取り、RAM 1200 を介して CPU 1100 に提供する。CPU 1100 は、かかるプログラムを、メディアインターフェイス 1700 を介して記録媒体 1800 から RAM 1200 上にロードし、ロードしたプログラムを実行する。記録媒体 1800 は、例えば DVD (Digital Versatile Disc)、PD (Phase change rewritable Disk) 等の光学記録媒体、MO (Magneto-Optical disk) 等の光磁気記録媒体、テープ媒体、磁気記録媒体、または半導体メモリ等である。

30

【0113】

例えば、コンピュータ 1000 が情報処理装置 100 として機能する場合、コンピュータ 1000 の CPU 1100 は、RAM 1200 上にロードされたプログラムを実行することにより、制御部 150 の機能を実現する。コンピュータ 1000 の CPU 1100 は、これらのプログラムを記録媒体 1800 から読み取って実行するが、他の例として、他の装置から所定の通信網を介してこれらのプログラムを取得してもよい。

40

【0114】

以上、本願の実施形態のいくつかを図面に基づいて詳細に説明したが、これらは例示であり、発明の開示の欄に記載の態様を始めとして、当業者の知識に基づいて種々の変形、改良を施した他の形態で本発明を実施することが可能である。

【0115】

〔7. その他〕

また、上記実施形態及び変形例において説明した各処理のうち、自動的に行われるもの

50

として説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。この他、上記文書中や図面中で示した処理手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。例えば、各図に示した各種情報は、図示した情報に限られない。

【0116】

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的な形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。

10

【0117】

また、上述してきた実施形態及び変形例は、処理内容を矛盾させない範囲で適宜組み合わせることが可能である。

【0118】

また、上述してきた「部(section、module、unit)」は、「手段」や「回路」などに読み替えることができる。例えば、検索部は、検索手段や検索回路に読み替えることができる。

【符号の説明】

【0119】

- 100 情報処理装置
- 110 通信部
- 120 記憶部
- 130 入力部
- 140 出力部
- 150 制御部
- 151 属性情報抽出部
- 152 モデル生成部
- 153 受付部
- 154 取得部
- 155 検索部
- 156 抽出部
- 157 生成部

20

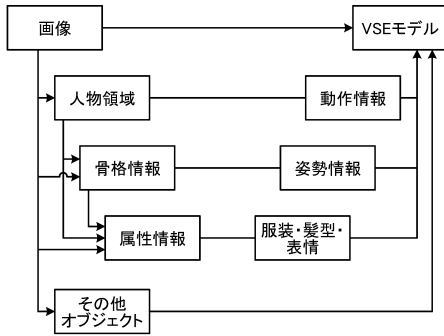
30

40

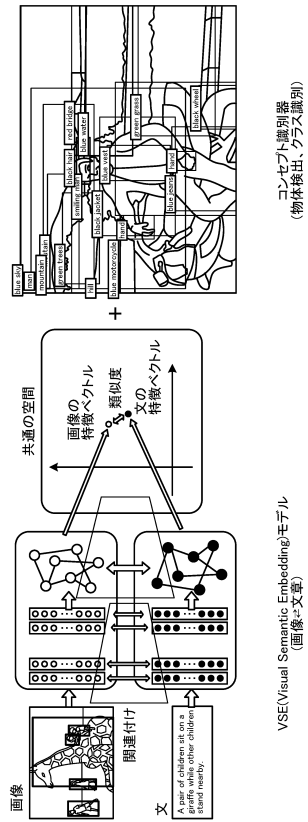
50

【 図 面 】

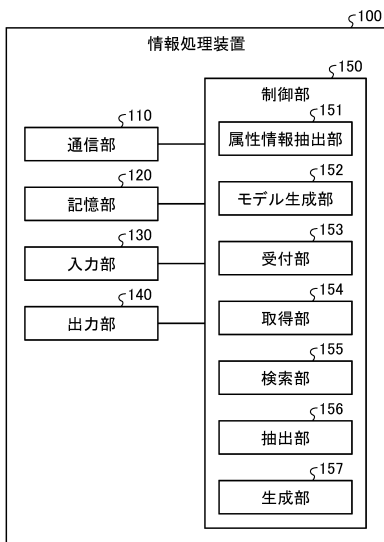
【 図 1 】



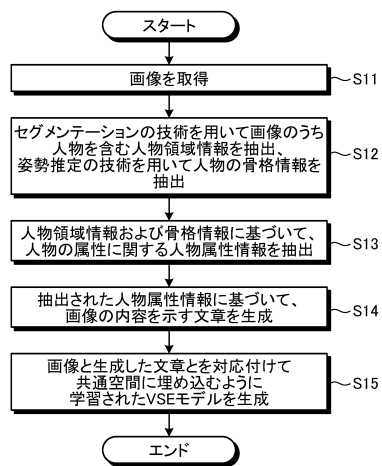
【 図 2 】



【 図 3 】



【 図 4 】



10

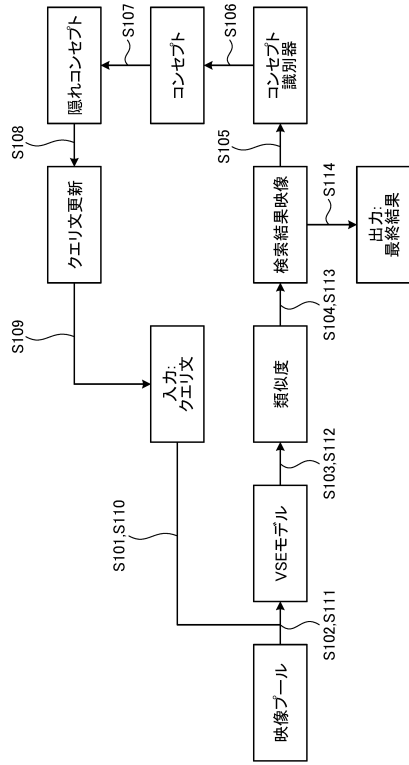
20

30

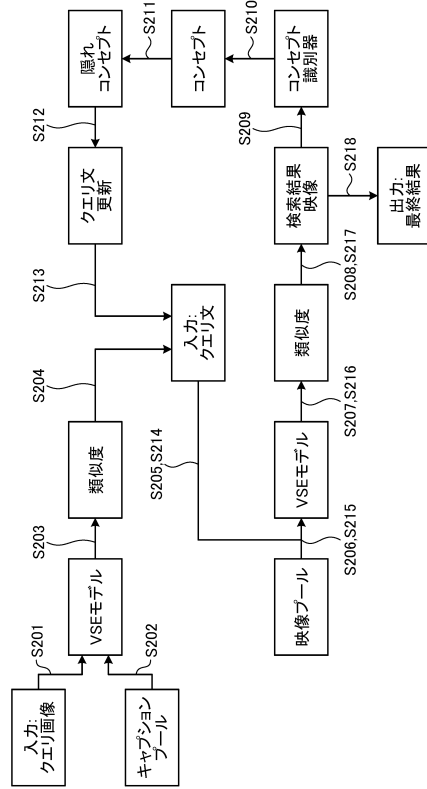
40

50

【図 5】



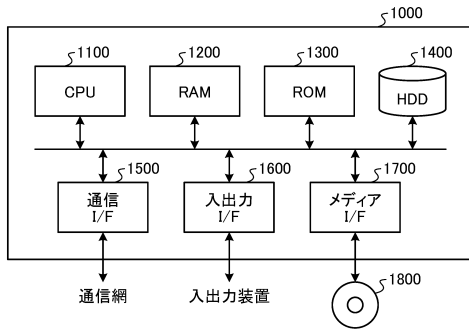
【図 6】



10

20

【図 7】



30

40

50

フロントページの続き

東京都港区海岸一丁目7番1号 ソフトバンク株式会社内

(72)発明者 植木 一也

東京都日野市程久保2丁目1番地1 学校法人明星学苑 明星大学内

審査官 和田 財太

(56)参考文献 国際公開第2021/064963(WO, A1)

特開平11-306322(JP, A)

特開2018-142160(JP, A)

特開2018-073429(JP, A)

(58)調査した分野 (Int.Cl., DB名)

G06F 16/00