

(19)日本国特許庁(JP)

(12)特許公報(B1)

(11)特許番号
特許第7265688号
(P7265688)

(45)発行日 令和5年4月26日(2023.4.26)

(24)登録日 令和5年4月18日(2023.4.18)

(51)国際特許分類 F I
G 0 6 T 7/00 (2017.01) G 0 6 T 7/00 3 5 0 B

請求項の数 8 (全17頁)

(21)出願番号	特願2022-573687(P2022-573687)	(73)特許権者	399037405 楽天グループ株式会社 東京都世田谷区玉川一丁目14番1号
(86)(22)出願日	令和3年12月15日(2021.12.15)	(74)代理人	100109380 弁理士 小西 恵
(86)国際出願番号	PCT/JP2021/046260	(74)代理人	100109036 弁理士 永岡 重幸
審査請求日	令和4年11月30日(2022.11.30)	(72)発明者	クエバス ファレス ホスエ 東京都世田谷区玉川一丁目14番1号 楽天グループ株式会社内
早期審査対象出願		(72)発明者	アルサラン ハサン 東京都世田谷区玉川一丁目14番1号 楽天グループ株式会社内
		(72)発明者	サナガヴァラブ ラジャセイカル 東京都世田谷区玉川一丁目14番1号 最終頁に続く

(54)【発明の名称】 情報処理装置、情報処理方法、およびプログラム

(57)【特許請求の範囲】

【請求項1】

オブジェクトを含むオブジェクト画像を取得する取得手段と、
前記取得手段により取得された前記オブジェクト画像を学習モデルに適用することにより、前記オブジェクトに関連する属性である関連属性を予測する予測手段と、を有し、
前記学習モデルは、前記オブジェクトを含む複数の異なるオブジェクトに共通の学習モデルであり、
前記オブジェクトのオブジェクトタイプを示す値と前記複数の異なるオブジェクトに関連する複数の属性について複数の属性値を推定する複数の推定層と、
前記複数の推定層から出力される前記オブジェクトのオブジェクトタイプを示す値と前記複数の属性値を連結して出力する出力層と、を含み、
前記予測手段は、前記複数の属性値から、前記オブジェクトタイプを示す値に応じて設定されている、1つ以上の有効な属性値を取得し、当該1つ以上の有効な属性値に対応する属性を、前記関連属性として予測することを特徴とする情報処理装置。

10

【請求項2】

前記学習モデルは、第1パートと第2パートから構成され、
前記第1パートは、前記オブジェクト画像を入力として、前記オブジェクトの特徴を表す特徴ベクトルを出力し、
前記第2パートは、前記複数の推定層と前記出力層を有する

20

ことを特徴とする請求項 1 に記載の情報処理装置。

【請求項 3】

前記取得手段により取得された前記オブジェクト画像が複数のオブジェクトを含む場合、前記予測手段は、前記複数のオブジェクトのそれぞれ関連する属性を前記関連属性として予測することを特徴とする請求項 1 に記載の情報処理装置。

【請求項 4】

前記オブジェクト画像は、Y C b C r 画像から D C T 変換されたデータのうちの Y 要素、C b 要素、C r 要素から生成された画像データであることを特徴とする請求項 1 から 3 のいずれか 1 項に記載の情報処理装置。

【請求項 5】

前記画像データは、前記 D C T 変換されたデータのうちの Y 要素、C b 要素、C r 要素のサイズ合わせを行って連結したデータであることを特徴とする請求項 4 に記載の情報処理装置。

【請求項 6】

前記予測手段により予測された前記関連属性を出力する出力手段をさらに有することを特徴とする請求項 1 から 5 のいずれか 1 項に記載の情報処理装置。

【請求項 7】

オブジェクトを含むオブジェクト画像を取得する取得工程と、
前記取得工程において取得された前記オブジェクト画像を学習モデルに適用することにより、前記オブジェクトに関連する属性である関連属性を予測する予測工程と、を有し、
前記学習モデルは、前記オブジェクトを含む複数の異なるオブジェクトに共通の学習モデルであり、

前記オブジェクトのオブジェクトタイプを示す値と前記複数の異なるオブジェクトに関連する複数の属性について複数の属性値を推定する複数の推定層と、

前記複数の推定層から出力される前記オブジェクトのオブジェクトタイプを示す値と前記複数の属性値を連結して出力する出力層と、を含み、

前記予測工程では、前記複数の属性値から、前記オブジェクトタイプを示す値に応じて設定されている、1 つ以上の有効な属性値を取得し、当該 1 つ以上の有効な属性値に対応する属性を、前記関連属性として予測する

ことを特徴とする情報処理方法。

【請求項 8】

情報処理をコンピュータに実行させるための情報処理プログラムであって、該プログラムは、前記コンピュータに、

オブジェクトを含むオブジェクト画像を取得する取得処理と、

前記取得処理において取得された前記オブジェクト画像を学習モデルに適用することにより、前記オブジェクトに関連する属性である関連属性を予測する予測処理と、を含む処理を実行させるためのものであり、

前記学習モデルは、前記オブジェクトを含む複数の異なるオブジェクトに共通の学習モデルであり、

前記オブジェクトのオブジェクトタイプを示す値と前記複数の異なるオブジェクトに関連する複数の属性について複数の属性値を推定する複数の推定層と、

前記複数の推定層から出力される前記オブジェクトのオブジェクトタイプを示す値と前記複数の属性値を連結して出力する出力層と、を含み、

前記予測処理は、前記複数の属性値から、前記オブジェクトタイプを示す値に応じて設定されている、1 つ以上の有効な属性値を取得し、当該 1 つ以上の有効な属性値に対応する属性を、前記関連属性として予測することを含む

ことを特徴とする情報処理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

10

20

30

40

50

本発明は、情報処理装置、情報処理方法、およびプログラムに関し、特に、オブジェクトに関連するあらゆる属性を予測するための技術に関する。

【背景技術】

【0002】

近年、インターネットを使って商品の販売を行う電子商取引（E-commerce / eコマース）が盛んに実施されており、そのような電子商取引の実施のためのEC（Electronic Commerce）サイトがウェブ上に多く構築されている。ECサイトは、世界中の各国の言語を用いて構築されることも多く、多くの国に在住するユーザ（消費者）が商品を購入することを可能にしている。ユーザは、PC（Personal Computer）や、スマートフォンといった携帯端末からECサイトにアクセスすることで、実際の店舗に赴くことなく、また時間に関係なく、所望の商品の選択や購入を行うことが可能となっている。

10

【0003】

ECサイトでは、ユーザによる購買意欲を増進させることを目的に、過去にユーザが購入した商品の属性（商品に特有な情報）と同様の属性を有する商品を、レコメンデーションする商品として、ユーザが閲覧している画面において合わせて表示することがある。また、ユーザも、所望の商品を購入する場合に、購入しようとする商品の属性から検索する場合がある。このようなことから、電子商取引では、商品の属性を識別することが、サイト運営側や商品提供側にとって共通の課題となっている。

特許文献1には、予め設定された商品の属性を当該商品の画像と対応付けて記憶する技術が開示されている。

20

【先行技術文献】

【特許文献】

【0004】

【文献】特開2016-139189号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

商品の属性は、当該商品によって異なるものであり、また、ECサイトで扱う商品の数は膨大である。特許文献1には、予め設定された商品の属性を当該商品の画像と対応付けて記憶する技術が開示されているが、商品の数が膨大であると、設定する属性の数も膨大になり、処理負荷が増大する。このような処理負荷の増大を回避するための1つの有効な手法は、各商品に関連するあらゆる属性を効率的に予測することである。

30

【0006】

本発明は上記課題に鑑みてなされたものであり、商品といったオブジェクトに関連するあらゆる属性を効率的に予測するための技術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】

【0007】

上記課題を解決するために、本発明による情報処理装置の一態様は、オブジェクトを含むオブジェクト画像を取得する取得手段と、前記取得手段により取得された前記オブジェクト画像を学習モデルに適用することにより、前記オブジェクトに関連するあらゆる属性を予測する予測手段と、を有し、前記学習モデルは、前記オブジェクトを含む複数の異なるオブジェクトに共通の学習モデルであり、前記複数の異なるオブジェクトに関連する複数の属性について複数の属性値を推定する複数の推定層と、前記複数の推定層から出力される複数の属性値を連結して出力する出力層と、を含む。

40

【0008】

前記情報処理装置において、前記学習モデルは、第1パートと第2パートから構成され、前記第1パートは、前記オブジェクト画像を入力として、前記オブジェクトの特徴を表す特徴ベクトルを出力し、前記第2パートは、前記複数の推定層と前記出力層を有し、前記複数の推定層は、前記特徴ベクトルを入力として、前記オブジェクトのオブジェクトタ

50

イブを示す値と前記複数の属性値を出力し、前記出力層は、前記複数の推定層から出力される前記オブジェクトのオブジェクトタイプを示す値と前記複数の属性値を連結して出力しうる。

【0009】

前記情報処理装置において、前記予測手段は、前記学習モデルの第2パートから出力された前記複数の属性値から、前記あらゆる属性を予測しうる。

【0010】

前記情報処理装置において、前記オブジェクトタイプが示す値に応じて、前記複数の属性値のうち1つ以上の有効な属性値が予め設定されており、前記予測手段は、前記複数の属性値から、前記オブジェクトタイプが示す値に応じて前記1つ以上の有効な属性値を取得し、当該1つ以上の有効な属性値に対応する属性を、前記あらゆる属性として予測しうる。

10

【0011】

前記情報処理装置において、前記取得手段により取得された前記オブジェクト画像が複数のオブジェクトを含む場合、前記予想手段は、前記複数のオブジェクトのそれぞれ関連するあらゆる属性を予測しうる。

【0012】

前記情報処理装置において、前記オブジェクト画像は、YCbCr画像からDCI変換されたデータのうちのY要素、Cb要素、Cr要素から生成された画像データでありうる。

また、前記画像データは、前記DCI変換されたデータのうちのY要素、Cb要素、Cr要素のサイズ合わせを行って連結したデータでありうる。

20

【0013】

前記情報処理装置において、前記予測手段により予測された前記あらゆる属性を出力する出力手段をさらに有しうる。

【0014】

上記課題を解決するために、本発明による情報処理方法の一態様は、オブジェクトを含むオブジェクト画像を取得する取得工程と、前記取得手段により取得された前記オブジェクト画像を学習モデルに適用することにより、前記オブジェクトに関連するあらゆる属性を予測する予測工程と、を有し、前記学習モデルは、前記オブジェクトを含む複数の異なるオブジェクトに共通の学習モデルであり、前記複数の異なるオブジェクトに関連する複数の属性について複数の属性値を推定する複数の推定層と、前記複数の推定層から出力される複数の属性値を連結して出力する出力層と、を含む。

30

【0015】

上記課題を解決するために、本発明による情報処理プログラムの一態様は、情報処理をコンピュータに実行させるための情報処理プログラムであって、該プログラムは、前記コンピュータに、オブジェクトを含むオブジェクト画像を取得する取得処理と、前記取得手段により取得された前記オブジェクト画像を学習モデルに適用することにより、前記オブジェクトに関連するあらゆる属性を予測する予測処理と、を含む処理を実行させるためのものであり、前記学習モデルは、前記オブジェクトを含む複数の異なるオブジェクトに共通の学習モデルであり、前記複数の異なるオブジェクトに関連する複数の属性について複数の属性値を推定する複数の推定層と、前記複数の推定層から出力される複数の属性値を連結して出力する出力層と、を含む。

40

【発明の効果】

【0016】

本発明によれば、商品といったオブジェクトに関連するあらゆる属性を効率的に予測することが可能となる。

上記した本発明の目的、態様及び効果並びに上記されなかった本発明の目的、態様及び効果は、当業者であれば添付図面及び請求の範囲の記載を参照することにより下記の発明を実施するための形態から理解できるであろう。

【図面の簡単な説明】

50

【 0 0 1 7 】

【 図 1 】 図 1 は、本発明の実施形態による情報処理装置の機能構成の一例を示すブロックである。

【 図 2 】 図 2 (a) は、学習用データの構成の概念図を示し、図 2 (b) は、第 1 属性 ~ 第 N 属性の具体例を示す。

【 図 3 】 図 3 (a) は、学習用データの構成の一例を示し、図 3 (b) は、属性値の一例を示す。

【 図 4 A 】 図 4 A は、属性予測モデルの第 1 パートのアーキテクチャの一例を示す。

【 図 4 B 】 図 4 B は、図 2 A の入力レイヤにおける処理の流れの概念図を示す。

【 図 5 A 】 図 5 A は、属性予測モデルの第 2 パートのアーキテクチャの一例を示す。

10

【 図 5 B 】 図 5 B は、図 5 A に示す複合属性ベクトルと別の構成例を示す。

【 図 6 】 図 6 は、本発明の実施形態による情報処理装置のハードウェア構成の一例を示すブロックである。

【 図 7 】 図 7 は、本発明の実施形態による情報処理装置により実行される処理を示すフローチャートである。

【 図 8 】 属性予測結果の一例を示す図である。

【 図 9 】 属性予測結果の別の例を説明する図である。

【 発明を実施するための形態 】

【 0 0 1 8 】

以下、添付図面を参照して、本発明を実施するための実施形態について詳細に説明する。以下に開示される構成要素のうち、同一機能を有するものには同一の符号を付し、その説明を省略する。なお、以下に開示される実施形態は、本発明の実現手段としての一例であり、本発明が適用される装置の構成や各種条件によって適宜修正または変更されるべきものであり、本発明は以下の実施形態に限定されるものではない。また、本実施形態で説明されている特徴の組み合わせの全てが本発明の解決手段に必須のものとは限らない。

20

【 0 0 1 9 】

[情報処理装置 1 0 0 の機能構成]

本実施形態による情報処理装置 1 0 0 は、商品が含まれる画像（商品画像）を取得し、当該商品に固有に関連するあらゆる属性を予測する。当該属性とは、ユーザが商品を購入する際の指標であり得、例えば、商品の視覚的な特徴や対象となる性別である。なお、本実施形態では、商品に関連する（商品に対する）属性を予測する例を説明するが、商品に限らず、任意のオブジェクトが含まれる画像（オブジェクト画像）から当該オブジェクトに関連する属性（オブジェクトに対する属性）を予測する場合も、本実施形態を適用可能である。

30

【 0 0 2 0 】

図 1 は、本実施形態による情報処理装置 1 の機能構成の一例を示す。

図 1 に示す情報処理装置 1 は、取得部 1 0 1、属性予測部 1 0 2、学習部 1 0 3、出力部 1 0 4、および学習モデル記憶部 1 0 5 を備える。学習モデル記憶部 1 0 5 は、属性予測モデル 1 0 6 を記憶している。

【 0 0 2 1 】

40

取得部 1 0 1 は、商品画像を取得する。取得部 1 0 1 は、商品画像を、ユーザ（オペレータ）による入力部 6 0 5（図 6）を介した入力操作によって取得してもよいし、ユーザの操作により記憶部（図 6 の ROM 6 0 2 や RAM 6 0 3）から取得してもよい。また、取得部 1 0 1 は、通信 I / F 6 0 7（図 6）を介して外部装置から受信した商品画像を取得してもよい。

【 0 0 2 2 】

商品画像は、赤（R）、緑（G）、青（B）の 3 色で色を表現した画像でありうる。また、商品画像は、明るさを表す輝度（Y（Luma））と色の成分（Cb、Cr（Chroma））で表現した画像（RGB 画像から YCbCr 変換された画像（JPEG 画像 / YCbCr 画像））であってもよい。また、商品画像は、情報処理装置 1 0 0 に備えられ

50

た符号化部（不図示）によりYCbCr画像からDCT（Discrete Cosine Transform）変換（圧縮）されたデータ（係数値）であってもよい。また、情報処理装置100以外の装置により（YCbCr変換および）DCT変換された商品画像としてのデータを取得部101が取得するように構成されてもよい。

取得部101は、取得した商品画像を、属性予測部102に出力する。

【0023】

属性予測部102は、取得部101により取得された商品画像を属性予測モデル（ニューラルネットワークモデル）106に適用し、当該商品画像に含まれる商品に関連するあらゆる属性を予測する（後述する属性予測モデル106を用いて予測可能な複数の属性のうち、当該商品に関連する全ての属性を予測する）。属性予測部102による処理については、後述する。

10

【0024】

学習部103は、属性予測モデル106を学習させ、学習済みの属性予測モデル106を学習モデル記憶部105に格納する。学習部103が属性予測モデル106の学習のために用いる学習用（教師用）データの構成について、図2と図3を参照して説明する。

【0025】

図2（a）は学習用データの構成の概念図を示す。図2（a）に示すように、本実施形態では、学習用データは、商品画像と、当該商品画像に含まれる商品の商品タイプ（オブジェクトタイプ）と、当該商品に関連する複数の属性（第1属性～第N属性（ $N > 1$ ））で構成される。

20

ここで、各属性は、当該商品の視覚的な特徴（以下、「属性カテゴリー」と称する）に対応し、各属性の属性値は、当該属性カテゴリーについてのタイプ（以下、「属性タイプ」と称する）を示す。すなわち、商品に関連する複数の属性値は、複数の属性カテゴリーそれぞれについての属性タイプを示す値を示しうる。

【0026】

本実施形態では、電子商取引のサイト（ECサイトといったウェブサイト）で扱われる、衣類、下着、靴、バッグといった、人が身に着けるグッズを、「商品」として扱う例を説明する。よって、例えば、商品タイプとしては、「シャツ」、「スカート」、「ジャケット」、「キャップ」、「スニーカー」、「ブーツ」等が挙げられるが、これらに限定されない。また、本実施形態では、「商品」は人が身に着けるグッズであるとするが、「商品」はこれに限定されない。また、「商品」は電子商取引のサイトで扱われるグッズ/商品に限定されない。

30

【0027】

図2（b）に、属性カテゴリーとしての第1属性～第N属性の具体例を示す。第1属性～第N属性としては、例えば、「丈の長さ」、「首回りタイプ」が挙げられ、それぞれ、第1属性と第2属性に対応する。なお、属性カテゴリーは、例えば、「袖の長さ」、「袖のタイプ」、「（靴の）ヒールの高さ」、「（商品が対象とする）性別」、「色」等も含まれ得る。このように、何番目の属性が、どの属性カテゴリーに対応するかは、あらかじめ所定のテーブル等でマッピングされているものとする。

【0028】

属性カテゴリーは、商品タイプによって有効か無効であり得、その関係はあらかじめ設定されているものとする。例えば、商品タイプが「シャツ」の場合に、属性カテゴリー「丈の長さ」は有効であるが、「ヒールの高さ」は無効である。このとき、属性カテゴリー「丈の長さ」に対応する属性値には有効な属性値（有効なラベル）が設定されるが、属性カテゴリー「ヒールの高さ」には無効な属性値（ダミー値/無効なラベル）が設定される。このように、本実施形態では、商品タイプに応じて、第1属性値～第N属性値のどれに有効な属性値が示され、どれに無効な属性値が示されるかがあらかじめ設定されているものとする。

40

本実施形態では、有効な属性値は、属性カテゴリーについての属性タイプを示す値が設定され（図3（a）と図3（b）を参照）、無効な属性値として「-1」が設定されるも

50

のとする。

【 0 0 2 9 】

図 3 (a) に、学習用データの構成の一例を示す。ここで、商品画像を識別するために、各商品画像に商品画像番号が付されているものとする。前述のように、学習用データは、商品画像に対して、当該商品画像に含まれる商品の商品タイプを示す値と、当該商品の属性カテゴリーについての属性タイプを示す値（第 1 属性値～第 N 属性値）から構成される。

【 0 0 3 0 】

図 3 (a) において、商品画像番号 = 1 の学習用データについて説明する。商品画像番号 = 1 の商品画像に含まれる商品のタイプ（商品タイプ）は「シャツ」であり、「シャツ」は商品タイプ値 = 1 に対応する。また、当該商品画像番号 = 1 の「シャツ」に対して、第 1 属性値 = 1、第 2 属性値 = 2 である。すなわち、「シャツ」に対しては、2 つの有効な値が示されている。図 3 (b) は、属性値の詳細の一例を示す。

図 3 (b) を参照すると、第 1 属性（「丈の長さ」）についての属性値 = 1 は「腰まで」を示し、第 2 属性（首回りタイプ）についての属性値 = 2 は「V 型」を示す。よって、当該商品画像番号 = 1 の「シャツ」は、「腰まで」の長さ（第 1 属性値 = 1）であり、「首回りタイプ」は V 型（第 2 属性値 = 2）であることを表している。

【 0 0 3 1 】

また、図 3 (a) において、商品画像番号 = 2 の学習用データについて説明する。商品画像番号 = 2 の商品画像に含まれる商品のタイプは「スカート」であり、「スカート」は商品タイプ値 = 2 に対応する。また、当該商品画像番号 = 2 の「スカート」に対する、第 1 属性値 = 3、第 2 属性値 = - 1 である。すなわち、「スカート」に対しては、有効な属性値と無効な属性値が示されている。

図 3 (b) を参照すると、第 1 属性（「丈の長さ」）についての属性値 = 2 は「膝まで」を示す。よって、当該商品画像番号 = 1 の「スカート」は、「膝まで」の長さ（第 1 属性値 = 3）を表すことがわかる。一方で、第 2 属性値 = - 1 は、無効な属性値であり、当該商品画像番号 = 1 の「スカート」の場合は、「首回りタイプ」に該当する特徴（属性）を有さないことを意味する。

【 0 0 3 2 】

図 1 の説明に戻り、出力部 1 0 4 は、属性予測部 1 0 2 により予測されたあらゆる属性の情報（属性予測結果）を出力する。出力部 1 0 4 は、例えば、取得部 1 0 2 により取得された商品画像と対応付けて、属性予測結果を出力してもよい。出力部 1 0 4 は、属性予測結果を、表示部 6 0 6（図 6）に表示してもよい。また、属性予測対象の商品画像がユーザ装置といった外部装置から取得された場合は、出力部 1 0 4 は、通信 I / F 6 0 7（図 6）を介して当該外部装置の表示部に表示されるように、当該外部装置へ送信してもよい。

【 0 0 3 3 】

[属性予測部 1 0 2 の処理]

次に、属性予測部 1 0 2 の処理について、属性予測モデル 1 0 6 と併せて説明する。属性予測部 1 0 2 は、取得部 1 0 1 により取得された商品画像を、属性予測モデル 1 0 6 に適用し、教師あり学習を行うことにより、共有（shared）特徴ベクトルを生成（抽出）し、当該ベクトルから、商品に関連するあらゆる属性の全てを予測する。なお、特徴ベクトルとは、特徴を表す値 / 情報を表す。

【 0 0 3 4 】

属性予測モデル 1 0 6 は、商品画像に含まれ得る複数の異なる商品（商品タイプ）に対して共通な学習モデルであり、共有して使用されることを特徴とする。例えば、商品画像に含まれ得る商品タイプが、「シャツ」、「スカート」、「ジャケット」、「キャップ」、「スニーカー」、「ブーツ」である場合、属性予測モデル 1 0 6 は少なくとも、「シャツ」、「スカート」、「ジャケット」、「キャップ」、「スニーカー」、「ブーツ」のそれぞれに関連するあらゆる属性の全てを予測するように構成される。

10

20

30

40

50

【 0 0 3 5 】

属性予測モデル106は、第1パートと第2パートから構成される。第1パートは、商品画像から共有特徴ベクトルを生成（抽出）するためのレイヤ（共有特徴ベクトル抽出レイヤ（Shared feature vector extraction layers））である。また、第2パートは、生成された共有特徴ベクトルから商品タイプを示す値と複数の属性値を予測して出力するための複数のレイヤ（属性特定（推定）レイヤ（Attribute specific layers））である。

【 0 0 3 6 】

図4Aに、属性予測モデル106の第1パートのアーキテクチャの一例を示す。

属性予測モデル106の第1パートは、画像認識モデルを適用した機械学習のための学習モデルである。図4Aに示すように、属性予測モデル106の第1パートは、入力レイヤL41と、複数のレイヤ（第1レイヤL42～第NレイヤL45）で構成される。第1レイヤL42～第NレイヤL45は、複数の畳み込み層を含んで構成される中間層と、クラスを分類/予測する出力層から構成され、入力画像41から共有特徴ベクトル42を生成して出力する。

当該中間層として、例えば、Google ResearchによるEfficientNetが使用される。EfficientNetが使用される場合、各畳み込み層は、MBConv（Mobile Inverted Bottleneck Convolution）が使用される。中間層では特徴マップが抽出され、出力層では、当該当該マップから次元を減らしつつ、最終的な共有特徴ベクトル42を生成するように構成される。なお、畳み込み層の数は、特定の数に限定されない。

【 0 0 3 7 】

図4Bに、図4Aの入力レイヤL41における処理の流れの概念図を示す。本実施形態では、属性予測部102は、RGBで表現された画像からYCbCr変換され、かつ、DCT変換されたデータを入力画像41として用いるように構成される。当該変換処理は、取得部101により行われてもよいし、情報処理装置100以外の装置により行われてもよい。

【 0 0 3 8 】

入力画像41の構成要素である、YCbCr変換され、かつ、DCT変換された画像データ（係数値）を、DCT-Y411、DCT-Cb412、DCT-Cr413と表す。DCT-Y411、DCT-Cb412、DCT-Cr413は、一例として、それぞれ[64, 80, 80]、[64, 40, 40]、[64, 40, 40]の成分を有し、各次元（Dimensionality）は[チャンネルの数（n_channels）、幅（width）、高さ（height）]を表す。

【 0 0 3 9 】

続いて、属性予測部102は、DCT-Cb412とDCT-Cr413に対してアップサンプリング処理を行い、DCT-Cb414とDCT-Cr415を生成する。その後、属性予測部102は、DCT-Y411、DCT-Cb414、DCT-Cr415をチャンネルごとに連結（Concatenate）し、連結DCTデータ416を生成する。連結DCTデータ416は、次の第1レイヤL42に入力される。すなわち、Y要素、Cb要素、Cr要素のサイズ合わせを行い、連結した連結DCTデータ416（画像データ）が生成される。

【 0 0 4 0 】

本実施形態ではこのように、学習モデル（属性予測モデル106）への入力画像として、RGB領域の画像の代わりに、YCbCr画像からDCT変換された画像データ（係数値）、すなわち、画像の圧縮領域を使用する。これにより、学習モデルの処理に必要なデータ量が大幅に減少し、また、圧縮領域で処理されることから学習モデルを深く構成する必要がないという効果がある。

【 0 0 4 1 】

属性予測部102は、第1レイヤL42に入力された連結DCTデータ416から、図

10

20

30

40

50

4 A に示したように第 1 レイヤ L 4 2 ~ 第 N レイヤ L 4 5 を介して、共有特徴ベクトル 4 2 を生成する。共有特徴ベクトル 4 2 は、属性予測モデル 1 0 6 の第 2 パートで共有される特徴ベクトルである。

【 0 0 4 2 】

次に、属性予測部 1 0 2 は、生成した共有特徴ベクトルを、属性予測モデル 1 0 6 の第 2 パートに入力する。図 5 A に、属性予測モデル 1 0 6 の第 2 パートのアーキテクチャの一例を示す。

属性予測モデル 1 0 6 の第 2 パートは、共有特徴ベクトルを入力として、商品に関連する複数の属性のそれぞれについて属性値を出力する、並列に設けられた複数のレイヤ（複数の推定層）と、当該複数のレイヤ（レイヤブランチ）から出力される複数の属性値を連結して出力する出力層とを含んで構成される。

10

【 0 0 4 3 】

具体的には、当該第 2 パートは、複数のレイヤとして、共有特徴ベクトルから商品タイプ（商品タイプ値 5 1）を予測するサブレイヤ L 5 1 1 ~ L 5 1 4 と、商品に対する複数の属性の属性値（第 1 の属性値 5 2、第 2 属性値 5 3、・・・、第 N の属性値 5 4）を予測するサブレイヤ L 5 2 1 ~ L 5 2 4、L 5 3 1 ~ L 5 3 4、・・・、L 5 4 1 ~ L 5 4 4 と、出力層として、当該複数のレイヤから出力される複数の属性値を連結して出力する連結部 5 5 から構成される。このような構成により、当該第 2 パートは、商品タイプを示す値と複数の属性値のそれぞれの属性値を予測して出力する。

【 0 0 4 4 】

商品タイプ値 5 1 を予測するサブレイヤは、図 5 A に示すように、例えば、畳み込み層（Conv）L 5 1 1、バッチ正規化（Batch Normalization）L 5 1 2、活性化関数層（ReLU）L 5 1 3、および予測（出力）層 L 5 1 4 で構成される。第 1 の属性値 5 2 ~ 第 N の属性値 5 4 を予測するサブレイヤも同様に構成される。

20

【 0 0 4 5 】

属性予測部 1 0 2 は、商品タイプ値 5 1 および第 1 属性値 5 2 ~ 第 N 属性値 5 2 を予測すると、続いて、商品タイプ値 5 1 および第 1 の属性値 3 2 ~ 第 N の属性値 5 4 を連結して共通の特徴空間に埋めこみ、当該特徴空間上で複合特徴ベクトル（複合属性値）5 6 を生成する。複合属性ベクトル 5 6 は、図 5 A に示すように、商品タイプ値 5 1 から始まり、第 1 の属性値 5 2 ~ 第 N の属性値 5 4 が順に続くように連結されている。なお、当該構成は、出力された複数の属性値を表す一形態であり、特定の構成に限定されない。

30

【 0 0 4 6 】

複合属性ベクトル 5 6 を生成後、属性予測部 1 0 2 は、複合属性ベクトル 5 6 の最初に配置されている、商品タイプ値 5 1 を読み出し（解読して）取得する。そして、属性予測部 1 0 2 は、取得した商品タイプ値 5 1 に対応する属性値（すなわち、有効な属性値）のみを、複合属性ベクトル 5 6 から読み出して取得する。

【 0 0 4 7 】

例えば、図 3 (a) を参照し、商品タイプ値 5 1 が「 2 」（商品タイプ = 「スカート」）の場合は、第 1 属性値は有効な属性値であるが、第 2 属性値は無効な属性値である。上述したように、商品タイプに応じて、第 1 属性値 ~ 第 N 属性値のどれに有効な属性値が示され、どれに無効な属性値が示されるかがあらかじめ設定されている。よって、属性予測部 1 0 2 は、複合属性ベクトル 5 6 から読み出した商品タイプ値 5 1 が「 2 」の場合は、第 1 属性値 5 2 を読み出し、第 2 属性値 5 3 は読み出さないという処理を行うことができる。属性予測部 1 0 2 は、第 3 属性値 5 4 以降に対しても同様に、商品タイプ値 5 1 に対応する属性値（有効な属性値）を読み出し、対応しない属性値（無効な属性値）を読み出さずに、読み出し処理を完了することができる。

40

【 0 0 4 8 】

これにより、属性予測部 1 0 2 は、商品タイプ値 5 1 に対して有効に設定されている属性値のみを読み出し、商品タイプ値に対応せず無効な属性値を読み出さないため、読み出し処理がより効率的に行われ、処理負荷も低くなる。

50

【 0 0 4 9 】

属性予測部 1 0 2 は、読み出した 1 つ以上の属性値に対応する属性の情報を属性予測結果として生成する。例えば、属性予測部 1 0 2 は、読み出した 1 つ以上の属性値のそれぞれに対応する属性タイプの名前を、図 3 (b) に示したような対応表から取得し、当該属性の名前を、商品タイプと併せて属性予測結果として生成することができる。

【 0 0 5 0 】

また、商品画像が複数の商品を含む場合は、当該複数の商品それぞれを含む複数の商品領域 (R e g i o n o f I n t e r e s t) を、属性予測モデル 1 0 6 の入力画像 (D C T 変換された画像データ) とすることができる。この場合、属性予測部 1 0 2 は、当該複数の商品領域を属性予測モデル 1 0 6 に適用し、当該複数の商品領域に対して予測された、商品タイプ値および第 1 の属性値 ~ 第 N の属性値を連結して共通の特徴空間に埋めこみ、当該特徴空間上で複合特徴ベクトル (複合属性値) を生成する。このように生成された複合特徴ベクトルの構成例を図 5 B に示す。図 5 B は、商品の数が n の場合に、出力層としての連結部 5 5 から出力された複合特徴ベクトルの構成例を示している。出力層は、複数の商品のそれぞれ関連する複数の属性それぞれについての属性値を連結して出力する。

このような場合も、属性予測部 1 0 2 は、商品タイプ値 (1) ~ (n) のそれぞれに対応する (すなわち、商品タイプ値 (1) の場合は、第 1 属性値 (1) ~ 第 N 属性値 (n))、属性値を読み出し、対応しない属性値を読み出さずに、読み出し処理を完了することができる。これにより、情報処理装置 1 0 0 は、当該複数のそれぞれに関連するあらゆる属性を予測することができる。

【 0 0 5 1 】

[情報処理装置 1 0 0 のハードウェア構成]

図 6 は、本実施形態による情報処理装置 1 0 0 のハードウェア構成の一例を示すブロック図である。

本実施形態による情報処理装置 1 0 0 は、単一または複数の、あらゆるコンピュータ、モバイルデバイス、または他のいかなる処理プラットフォーム上にも実装することができる。

図 6 を参照して、情報処理装置 1 0 0 は、単一のコンピュータに実装される例が示されているが、本実施形態による情報処理装置 1 0 0 は、複数のコンピュータを含むコンピュータシステムに実装されてよい。複数のコンピュータは、有線または無線のネットワークにより相互通信可能に接続されてよい。

【 0 0 5 2 】

図 6 に示すように、情報処理装置 1 0 0 は、CPU 6 0 1 と、ROM 6 0 2 と、RAM 6 0 3 と、HDD 6 0 4 と、入力部 6 0 5 と、表示部 6 0 6 と、通信 I / F 6 0 7 と、システムバス 6 0 8 とを備えてよい。情報処理装置 1 0 0 はまた、外部メモリを備えてよい。

CPU (C e n t r a l P r o c e s s i n g U n i t) 6 0 1 は、情報処理装置 1 0 0 における動作を統括的に制御するものであり、データ伝送路であるシステムバス 6 0 8 を介して、各構成部 (6 0 2 ~ 6 0 7) を制御する。

【 0 0 5 3 】

ROM (R e a d O n l y M e m o r y) 6 0 2 は、CPU 6 0 1 が処理を実行するために必要な制御プログラム等を記憶する不揮発性メモリである。なお、当該プログラムは、HDD (H a r d D i s k D r i v e) 6 0 4、SSD (S o l i d S t a t e D r i v e) 等の不揮発性メモリや着脱可能な記憶媒体 (不図示) 等の外部メモリに記憶されていてもよい。

RAM (R a n d o m A c c e s s M e m o r y) 6 0 3 は、揮発性メモリであり、CPU 6 0 1 の主メモリ、ワークエリア等として機能する。すなわち、CPU 6 0 1 は、処理の実行に際してROM 6 0 2 から必要なプログラム等をRAM 6 0 3 にロードし、当該プログラム等を実行することで各種の機能動作を実現する。

【 0 0 5 4 】

HDD 6 0 4 は、例えば、CPU 6 0 1 がプログラムを用いた処理を行う際に必要な各

10

20

30

40

50

種データや各種情報等を記憶している。また、HDD 604には、例えば、CPU 601がプログラム等を用いた処理を行うことにより得られた各種データや各種情報等が記憶される。

入力部 605は、キーボードやマウス等のポインティングデバイスにより構成される。

表示部 606は、液晶ディスプレイ(LCD)等のモニターにより構成される。表示部 606は、入力部 605と組み合わせて構成されることにより、GUI(Graphical User Interface)として機能してもよい。

【0055】

通信I/F 607は、情報処理装置100と外部装置との通信を制御するインタフェースである。

通信I/F 607は、ネットワークとのインタフェースを提供し、ネットワークを介して、外部装置との通信を実行する。通信I/F 607を介して、外部装置との間で各種データや各種パラメータ等が送受信される。本実施形態では、通信I/F 607は、イーサネット(登録商標)等の通信規格に準拠する有線LAN(Local Area Network)や専用線を介した通信を実行してよい。ただし、本実施形態で利用可能なネットワークはこれに限定されず、無線ネットワークで構成されてもよい。この無線ネットワークは、Bluetooth(登録商標)、ZigBee(登録商標)、UWB(Ultra Wide Band)等の無線PAN(Personal Area Network)を含む。また、Wi-Fi(Wireless Fidelity)(登録商標)等の無線LAN(Local Area Network)や、WiMAX(登録商標)等の無線MAN(Metropolitan Area Network)を含む。さらに、LTE/3G、4G、5G等の無線WAN(Wide Area Network)を含む。なお、ネットワークは、各機器を相互に通信可能に接続し、通信が可能であればよく、通信の規格、規模、構成は上記に限定されない。

【0056】

図6に示す情報処理装置100の各要素のうち少なくとも一部の機能は、CPU 601がプログラムを実行することで実現することができる。ただし、図6に示す情報処理装置100の各要素のうち少なくとも一部の機能が専用のハードウェアとして動作するようにしてもよい。この場合、専用のハードウェアは、CPU 601の制御に基づいて動作する。

【0057】

<処理の流れ>

図7に、本実施形態による情報処理装置100により実行される処理のフローチャートを示す。図6に示す処理は、情報処理装置100のCPU 601がROM 602等に格納されたプログラムをRAM 603にロードして実行することによって実現される。

【0058】

S71では、取得部101は、属性予測対象の商品を含む商品画像を取得する。例えば、取得部101は、情報処理装置100の操作者が、情報処理装置100を操作して、任意の電子商取引のサイトにアクセスした上で任意の商品を含む商品画像を選択することにより、当該商品画像を取得する。また、取得部101は、ユーザ装置といった外部装置から送信された商品画像または商品画像を示すURLを取得することで、商品画像を取得することができる。1つの商品画像において含まれる属性予測対象の商品は1つに限らず、当該商品画像に複数の属性予測対象の商品が含まれてもよい。

【0059】

S72では、属性予測部102は、取得部101により取得された商品画像を属性予測モデル106に入力し、共有特徴ベクトル42を生成する(図4Aを参照)。続いて、S73では、属性予測部102は、S72で生成された共有特徴ベクトル42を入力として、複合属性ベクトル(複合属性値)56を生成する(図5A、図5Bを参照)。S74では、属性予測部102は、S73で生成され複合属性ベクトル56から、商品に関連する複数の属性の全てを予測して取得する。当該予測処理については上述の通りである。

【0060】

なお、商品画像が複数の商品を含む場合は、取得部 102 は、例えば公知の画像処理技術により、当該複数の商品それぞれを含む複数の商品領域 (Region of Interest) を取得し、属性予測部 102 に出力することができる。そして、属性予測部 102 は、それぞれの商品領域 (部分画像) に対して、S72 ~ S74 の処理を行うことができる。

【0061】

[属性予測結果例]

次に、図8と図9を参照して、本実施形態による情報処理装置100による属性予測結果の例について説明する。

図8は、属性予測結果の一例として、商品画像80に対する属性予測結果81を示す。当該属性予想結果80は、情報処理装置100の表示部606に表示されてもよいし、通信I/F607を介してユーザ装置といった外部装置に送信され、当該外部装置において表示されてもよい。

【0062】

商品画像80は、属性予測対象の商品を含む商品画像であり、例えば、情報処理装置100の操作者が、情報処理装置100を操作して、任意の電子商取引のサイトにアクセスした上で選択された画像である。また、商品画像80は、ユーザ装置といった外部装置から送信された画像であってもよい。

【0063】

図8の例では、商品画像80に対して商品タイプは「ジャケット」と予測され、属性(各属性カテゴリーについての属性タイプ)として、「丈の長さ：尻まで」、「首回りタイプ：ラウンド」、「袖の長さ：手首まで」、「袖のスタイル：一般」、「襟のタイプ：フード」、「閉じ：完全」、「閉じのタイプ：チャック」、「性別：男性」、「色：グレー」が予測されている。

【0064】

このように、図8では、属性予測結果81として、商品画像80に含まれる商品(商品タイプ：「ジャケット」)に関連する属性の全てが表示されている。これにより、当該属性予測結果を受け取った操作者やユーザは、当該商品の特徴をよりよく把握することが可能となる。

【0065】

次に、図9(a)~(c)を参照して、属性予測結果の別の例として、商品画像90に対する属性予測結果について説明する。図9(a)~(c)は、属性予測結果の別の例を説明する図である。図9(a)に示すように、商品画像90は複数の商品(キャップとシャツ)を含み、取得部101が、例えば公知の画像処理技術により、当該複数の商品それぞれを含む複数の商品領域(Region of Interest)91、92を取得する(図9(b))。例えば、取得部101は、YOLO(You only look once)による学習モデルを用いた機械学習により、複数の商品領域91、92を取得するように構成されてもよい。続いて、属性予測部102は、それぞれの商品領域(部分画像)91、92に対して属性予測処理を行う。

【0066】

図9(c)に、商品画像80に含まれる商品領域91、92に対する属性予測結果93を示す。具体的には、「キャップ」については、「キャップタイプ：ベースボール」、「性別：男性」、「色：白」と予測され、「シャツ」については、「丈の長さ：腰まで」、「首回りタイプ：ラウンド」、「袖の長さ：短い」、「性別：ユニセックス」、「色：グレー」と予測されている。図8の例と同様に、当該属性予想結果93は、情報処理装置100の表示部606に表示されてもよいし、通信I/F607を介してユーザ装置といった外部装置に送信され、当該外部装置において表示されてもよい。

【0067】

このように、図9(c)では、属性予測結果93として、商品画像90に含まれる複数の商品(商品タイプ：「キャップ」、「シャツ」)のそれぞれに関連する属性の全てが表

10

20

30

40

50

示されている。これにより、当該属性予測結果を受け取った操作者やユーザは、当該複数の商品の特徴をよりよく把握することが可能となる。

【0068】

このように、本実施形態による情報処理装置100は、商品画像に含まれる1つ以上の商品について、関連する属性の全てを機械学習により予測し、当該機械学習で用いる学習モデルは、商品画像に含まれ得る全ての商品について共通な学習モデルである。また、情報処理装置100は、商品に関連する属性のみを予測し、関連しない属性は予測しないように構成される。これにより、1つの学習モデルを用いて、あらゆる商品それぞれに対して、関連するあらゆる属性の全てを予測することが可能となり、効率的な属性予測が可能となる。

10

【0069】

また、学習モデルに入力される商品画像は、YCbCr画像からDCT変換された画像データ(係数値)である。このような画像データを用いることにより、学習モデルの処理に必要なデータ量が大幅に減少し、また、圧縮領域で処理されることから学習モデルを深く構成する必要がないという効果がある。

【0070】

なお、上記実施形態では、複合特徴ベクトル311は4つの特徴ベクトルから生成される例を説明したが、結合される特徴ベクトルは4つに限定されない。例えば、第2特徴ベクトル302と色特徴ベクトル304から複合特徴ベクトル311が生成され、当該複合特徴ベクトル311から類似画像が検索されてもよい。また、機械学習により生成された他の特徴ベクトルを結合した複合特徴ベクトル311から類似画像が検索されるように構成されてもよい。

20

【0071】

なお、上記実施形態では商品に関連する属性の数は複数であることを想定したが、当然に1つの場合も、本実施形態を適用可能である。

【0072】

なお、上記において特定の実施形態が説明されているが、当該実施形態は単なる例示であり、本発明の範囲を限定する意図はない。本明細書に記載された装置及び方法は上記した以外の形態において具現化することができる。また、本発明の範囲から離れることなく、上記した実施形態に対して適宜、省略、置換及び変更をなすこともできる。かかる省略、置換及び変更をなした形態は、請求の範囲に記載されたもの及びこれらの均等物の範疇に含まれ、本発明の技術的範囲に属する。

30

【符号の説明】

【0073】

10：ユーザ装置、100：情報処理装置、101：取得部、102：属性予測部、103：学習部、104：出力部、105：学習モデル記憶部、106、属性予測モデル106

40

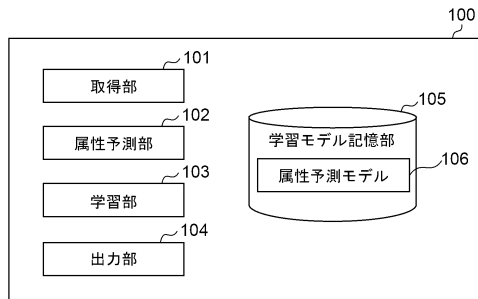
50

【要約】

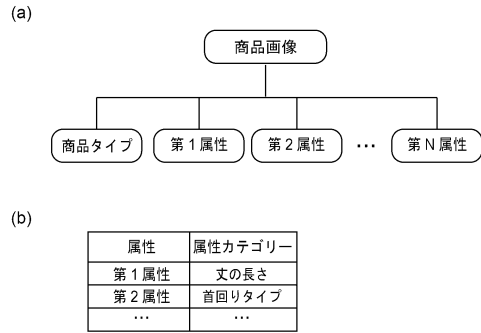
情報処理装置(1)は、オブジェクトを含むオブジェクト画像を取得する取得手段(101)と、前記取得手段により取得された前記オブジェクト画像を学習モデルに適用することにより、前記オブジェクトに関連するあらゆる属性を予測する予測手段(102)と、を有し、前記学習モデルは、前記オブジェクトを含む複数の異なるオブジェクトに共通の学習モデルであり、前記複数の異なるオブジェクトに関連する複数の属性について複数の属性値を推定する複数の推定層と、を含む。

【図面】

【図1】



【図2】



10

【図3】

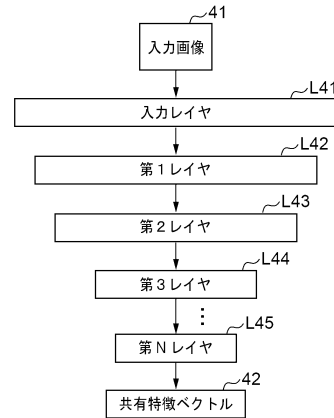
(a)

商品画像番号	商品タイプ	商品タイプ値	第1属性値	第2属性値	...	第N属性値
1	シャツ	1	1	2
2	スカート	2	3	-1
3	シャツ	1	1	1
4	ジャケット	4	2	1
5	キャップ	5	-1	-1
...

(b)

属性	属性カテゴリー	属性値：属性タイプ
第1属性	丈の長さ	1：腰まで
		2：尻まで
		3：膝まで
第2属性	首回りタイプ	1：ラウンド
		2：V型
		...
...

【図4A】



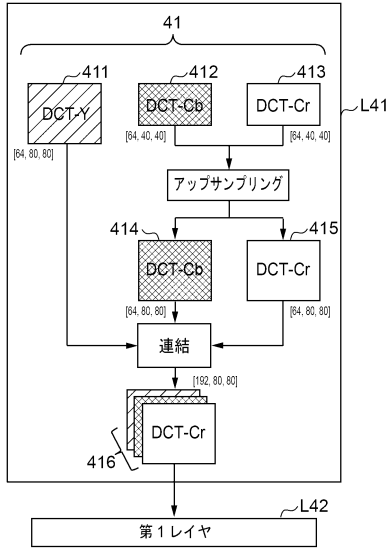
20

30

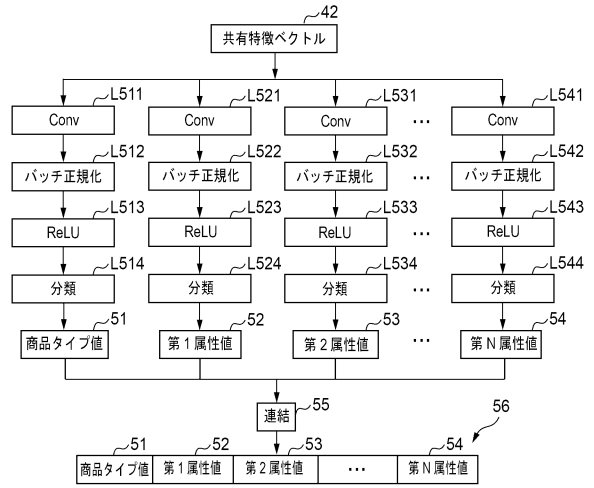
40

50

【図 4 B】

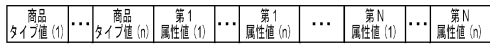


【図 5 A】

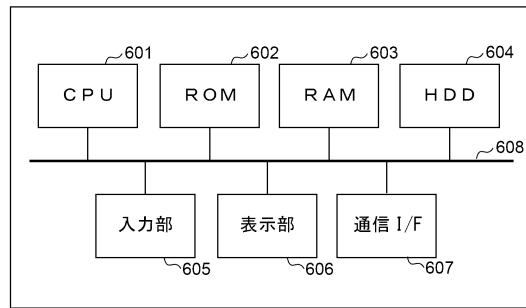


10

【図 5 B】



【図 6】



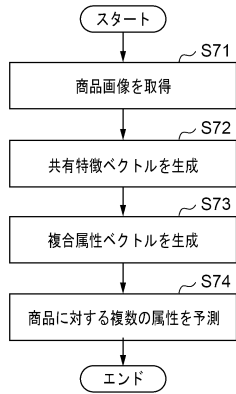
20

30

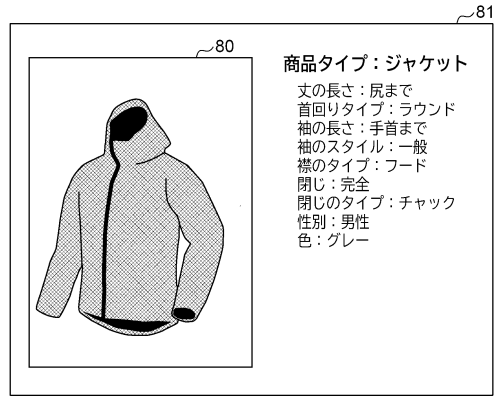
40

50

【 図 7 】

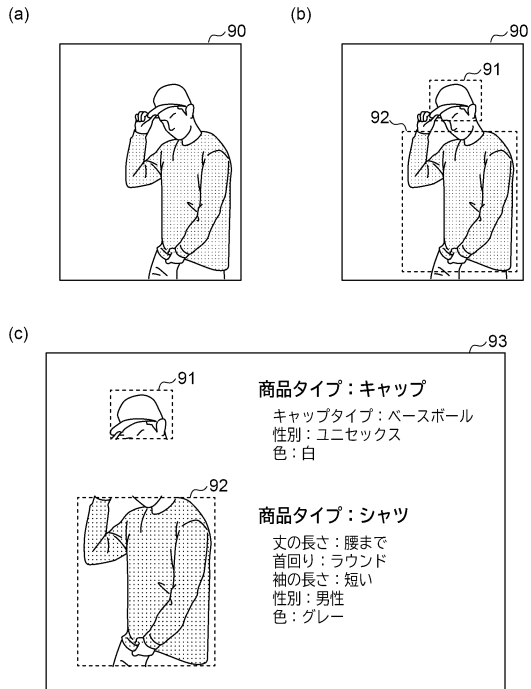


【 図 8 】



10

【 図 9 】



20

30

40

50

フロントページの続き

楽天グループ株式会社内

審査官 千葉 久博

- (56)参考文献 特開2020-71871(JP,A)
特開2018-106284(JP,A)
特開2005-250712(JP,A)
特開2005-101712(JP,A)
米国特許出願公開第2020/0320769(US,A1)
- (58)調査した分野 (Int.Cl., DB名)
G06T 7/00