

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-160240
(P2019-160240A)

(43) 公開日 令和1年9月19日(2019.9.19)

(51) Int.Cl.	F I	テーマコード (参考)
G06T 7/11 (2017.01)	G06T 7/11	5L096
G06T 7/00 (2017.01)	G06T 7/00 350C	
G10L 21/034 (2013.01)	G10L 21/034	
G10L 21/0208 (2013.01)	G10L 21/0208 100B	

審査請求 未請求 請求項の数 9 O L (全 15 頁)

(21) 出願番号 特願2018-50181 (P2018-50181)
(22) 出願日 平成30年3月16日 (2018.3.16)

(71) 出願人 000004226
日本電信電話株式会社
東京都千代田区大手町一丁目5番1号
(74) 代理人 110002147
特許業務法人酒井国際特許事務所
(72) 発明者 佐々木 琢
東京都千代田区大手町一丁目5番1号 日
本電信電話株式会社内
(72) 発明者 三上 啓太
東京都千代田区大手町一丁目5番1号 日
本電信電話株式会社内
(72) 発明者 森賀 邦広
東京都千代田区大手町一丁目5番1号 日
本電信電話株式会社内
Fターム(参考) 5L096 DA01 GA19 GA51 HA11

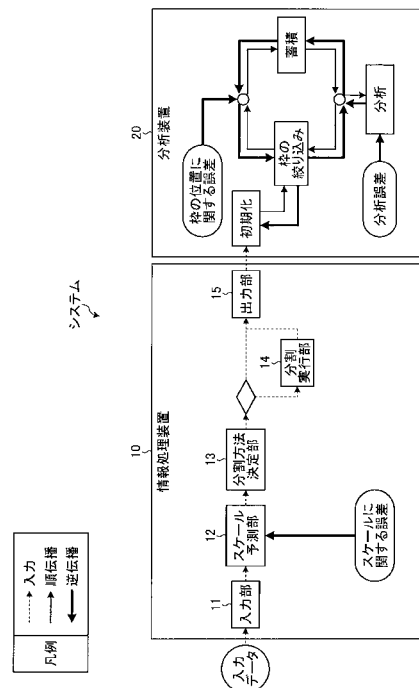
(54) 【発明の名称】 情報処理装置および情報処理方法

(57) 【要約】

【課題】入力データがマルチスケール性を有する場合であっても、入力データにおける特徴部を精度よく分析する。

【解決手段】情報処理装置10は、入力データに対する特徴部の占める割合(スケール)を予測し、予測したスケールが所定値以下の場合、当該入力データを分割して、分析装置20へ出力する。また、情報処理装置10は、予測したスケールが所定値を大きく下回る場合、スケールが小さいほど当該入力データを細かく分割して、分析装置20へ出力する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

データの特徴部を抽出し、分析を行う分析装置で用いられるデータの前処理を行う情報処理装置であって、

前記データの入力を受け付ける入力部と、

前記データに対する前記特徴部の占める割合を予測する予測部と、

前記予測された割合に応じて、前記データに対する分割の方法を決定する分割方法決定部と、

前記決定した分割の方法に基づき、前記データに対して分割を実行する分割実行部と、

を備えることを特徴とする情報処理装置。

10

【請求項 2】

前記予測部は、

データごとに、前記データにおける特徴部の占める割合を示した訓練データを用いた機械学習により、前記データに対する前記特徴部の占める割合を予測する

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 3】

前記分割方法決定部は、

前記データに対する前記特徴部の占める割合が所定値以下である場合、前記データに対して分割を行うと判定する

ことを特徴とする請求項 1 に記載の情報処理装置。

20

【請求項 4】

前記分割方法決定部は、

前記データに対する前記特徴部の占める割合が小さいほど、前記データを細かく分割すると判定する

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 5】

前記データは、画像データまたは映像データであり、

前記特徴部は、前記画像データまたは前記映像データに写る被写体の部分である

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 6】

前記データは、テキストデータであり、

前記特徴部は、前記テキストデータに含まれる所定のキーワードである

ことを特徴とする請求項 1 に記載の情報処理装置。

30

【請求項 7】

前記データは、音声データであり、

前記特徴部は、前記音声データに含まれる、人間の音声、所定の人物の音声、所定の単語を示す音声、および、所定の周波数帯域の音声のいずれかまたはこれらの組み合わせである

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 8】

前記データは、時系列のセンサデータであり、

前記特徴部は、前記時系列のセンサデータに含まれる所定のセンサ値のパターンであることを特徴とする請求項 1 に記載の情報処理装置。

40

【請求項 9】

データの特徴部を抽出し、分析を行う分析装置で用いられるデータの前処理を行う情報処理装置により実行される情報処理方法であって、

前記データの入力を受け付けるステップと、

前記データに対する前記特徴部の占める割合を予測するステップと、

前記予測された割合に応じて、前記データに対する分割の方法を決定するステップと、

前記決定した分割の方法に基づき、前記データに対して分割を実行するステップと

50

を含んだことを特徴とする情報処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置および情報処理方法に関する。

【背景技術】

【0002】

従来、入力されたデータの中の重要な部分（特徴部）と重要でない部分（背景）とに分ける技術がある。例えば、深層学習を用いた技術によれば、画像データの背景を無視し、特徴部だけを抽出したうえで分析を行うことができる。この技術は次の2つの効果を持つ。

10

【0003】

- ・精度がよい（背景 ノイズに左右されないため）
- ・処理速度が速い（背景の評価をスキップできるため）

【0004】

上記の技術は、例えば、監視カメラの画像または映像に写る被写体、例えば、人物や動物や移動体等の分析等に適用可能である。

【0005】

また、上記のように映像または画像に写る被写体の分析を行う技術として、EDRAM（Enriched Deep Recurrent visual Attention Model）がある。このEDRAMは、入力画像または入力映像に対し、被写体部分をとらえるための枠を移動させ、枠を移動させる都度、枠により切り取られた範囲に対して分析を行う技術である。

20

【0006】

ここで、枠の移動は、画像に対しては縦横の2方向に移動し、映像に対しては縦横に時間軸を加えた3方向に移動することにより行われる。また、画像または映像の被写体を囲うような位置に移動することにより行われる。ここで、枠により切り取られた範囲に対する分析は、例えば、被写体についての以下の分類および照合により行われる。なお、以下は、被写体が人物である場合の分類および照合の例である。

【0007】

- ・分類：人物の属性（例えば、性別、年齢、着ている服等）を推定する
- ・照合：与えられた人物と同一人物か否かを判定する

30

なお、上記の分類には、人物の属性の推定の他、人物の動作等、人物に関する様々な情報、状態を推定することが含まれる。

【0008】

また、EDRAMは、例えば、以下の4つニューラルネットワーク(NN)により構成される。

【0009】

- ・初期化NN：最初の枠を決めるNN
- ・コアNN：過去に枠が何を見てきたかを「記憶」するNN
- ・移動NN：記憶に基づき、枠を最適な位置に移動させるNN
- ・分析NN：記憶に基づき、分析結果を出すNN

40

【0010】

上記の4つのNNの関係を図12に示す。

【0011】

EDRAMは、初期化NNにおいて、例えば、人物を含む画像101を取得すると、画像101の最初の枠を決め、切り取る。そして、コアNNにおいて切り取った枠（例えば、図12に示す1回目の枠）の位置を記憶し、分析NNにおいて1回目の枠内の分析を行い、分析結果を出力する（例えば、30代、女性等）。

【0012】

その後、移動NNにおいて、枠を最適な位置に移動させる。例えば、移動NNは、枠の位置を、図12に示す2回目の枠の位置に移動させる。そして、コアNNにおいて、当該移動に

50

より切り取った枠（例えば、上記の 2 回目の枠）の位置を記憶し、分析NNにおいて 2 回目の枠内の分析を行い、分析結果を出力する。

【 0 0 1 3 】

その後、移動NNにおいて枠をさらに最適な位置に移動させる。例えば、移動NNは、枠の位置を、図 1 2 に示す 3 回目の枠の位置に移動させる。その後、コアNNにおいて、当該移動により切り取った枠（例えば、上記の 3 回目の枠）を記憶し、分析NNにおいて 3 回目の枠内の分析を行い、分析結果を出力する。

【 0 0 1 4 】

EDRAMが、このような処理を繰り返すことで、枠は徐々に絞り込まれていき、やがて枠は画像 1 0 1 内の人物の全身に収束する。よって、EDRAMにおいて、枠を画像内の人物の全身に収束させるためには、初期化NNにより生成される枠が人物を含んでいることが重要である。換言すると、初期化NNにおいて生成される枠（1 回目の枠）が人物を含んでいないと、移動NNにおいて枠を何回絞り込んでも人物は見つかりづらい。

【 0 0 1 5 】

ここで、実験を行ったところ、EDRAMで扱う画像群がマルチスケール性を有する場合、人物を含むような枠の初期化に失敗することが多いという実験結果が得られた。このマルチスケール性とは、画像により写っている人物の大きさ（スケール）が異なる性質である。例えば、図 1 3 に示すように、画像群それぞれの人物の大きさ（スケール）が異なる場合、当該画像群はマルチスケール性を有することになる。

【 0 0 1 6 】

EDRAMで扱う画像群がマルチスケール性を有する場合、人物を含む枠の初期化に失敗し、その結果、画像中の人物の分析精度が低くなってしまいうおそれがある。

【 0 0 1 7 】

このことを、図 1 4 を用いて説明する。例えば、EDRAMで扱う画像群が、すべての画像で人物のスケールがほぼ同じであるデータセット A である場合、何回か訓練すれば、EDRAMにより初期化された 1 回目の枠は、高確率で人物を含むものとなる。つまり、高確率で人物を含むような初期化ができる。一方、EDRAMで扱う画像群が、画像によって人物のスケールが異なるデータセット B である場合、何回訓練しても、EDRAMにより初期化された 1 回目の枠は、高確率で人物を含むようなものにならない。つまり、高確率で人物を含むような初期化ができない。その結果、画像中の人物の分析精度が低くなってしまいうおそれがある。

【 0 0 1 8 】

なお、EDRAMで扱う画像群がマルチスケール性を有する場合に、人物を含むような枠の初期化に失敗してしまうのは、以下の理由によると考えられる。

【 0 0 1 9 】

例えば、図 1 4 のデータセット B の画像 2 0 1 , 2 0 2 , 2 0 3 のように、画像 2 0 1 , 2 0 2 における人物のスケールに対し、画像 2 0 3 における人物のスケールが小さい場合、EDRAMは、画像 2 0 1 , 2 0 2 につられて、画像 2 0 3 についても同じようなスケールの人物を含むような 1 回目の枠を生成してしまう。その結果、EDRAMは、画像 2 0 3 について人物とは異なる場所に 1 回目の枠を生成してしまうからと考えられる（符号 2 0 4 に示す枠参照）。

【 先行技術文献 】

【 非特許文献 】

【 0 0 2 0 】

【 非特許文献 1 】 Artsiom Ablavatski, Shijian Lu, Jianfei Cai, “ Enriched Deep Recurrent Visual Attention Model for Multiple Object Recognition ”, IEEE WACV 2017, 12 Jun 2017

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 2 1 】

10

20

30

40

50

上記のEDRAMのみならず、入力データから特徴部を抽出し、分析する分析装置において、入力データがマルチスケール性を有する場合、初期化された1回目の枠は特徴部を含まない可能性がある。そのため、入力データを精度よく分析することができない場合がある。そこで、本発明は、前記した問題を解決し、入力データがマルチスケール性を有する場合であっても、入力データの特徴部を精度よく分析することを目的とする。

【課題を解決するための手段】

【0022】

前記した課題を解決するため、本発明は、データの特徴部を抽出し、分析を行う分析装置で用いられるデータの前処理を行う情報処理装置であって、前記データの入力を受け付ける入力部と、前記データに対する前記特徴部の占める割合を予測する予測部と、前記予測された割合に応じて、前記データに対する分割の方法を決定する分割方法決定部と、前記決定した分割の方法に基づき、前記データに対して分割を実行する分割実行部と、を備えることを特徴とする。

10

【発明の効果】

【0023】

本発明によれば、入力データがマルチスケール性を有する場合であっても、入力データの特徴部を精度よく分析することができる。

【図面の簡単な説明】

【0024】

【図1】図1は、システムの構成例を示す図である。

20

【図2】図2は、訓練データの一例を示す図である。

【図3】図3は、画像データの一例を示す図である。

【図4】図4は、画像データの分割の一例を説明するための図である。

【図5】図5は、システムの処理手順の一例を示すフローチャートである。

【図6】図6は、画像データの分割の一例を説明するための図である。

【図7】図7は、ウィンドウスライディング法における人物部分の検出を説明するための図である。

【図8】図8は、YOLO (You Only Look Once) における人物部分の枠決めを説明するための図である。

【図9】図9は、入力データが音声データである場合の特徴部およびスケールを説明するための図である。

30

【図10】図10は、入力データが時系列のセンサデータである場合の特徴部およびスケールを説明するための図である。

【図11】図11は、情報処理プログラムを実行するコンピュータの一例を示す図である。

【図12】図12は、EDRAMによる処理の一例を説明するための図である。

【図13】図13は、マルチスケール性を有する画像群の一例を示す図である。

【図14】図14は、EDRAMにおける人物を含む枠の初期化を説明するための図である。

【発明を実施するための形態】

【0025】

40

[概要]

以下、図面を参照しながら、本発明の実施形態を説明する。まず、図1を用いて、本実施形態の情報処理装置を含むシステムの概要を説明する。

【0026】

システムは、情報処理装置10と、分析装置20とを備える。情報処理装置10は、分析装置20の扱うデータ(入力データ)の前処理を行う。分析装置20は、情報処理装置10により前処理が行われた入力データの分析を行う。例えば、分析装置20は、情報処理装置10により前処理が行われた入力データの特徴部を抽出し、抽出した特徴部に対する分析を行う。

【0027】

50

例えば、入力データが、画像データである場合、入力データの特徴部は、例えば、画像データの人物部分である。この場合、分析装置 20 は、情報処理装置 10 により前処理が行われた画像データから人物部分を抽出し、抽出した人物部分の分析（例えば、人物部分の人物の性別、年齢等の推測）を行う。この分析装置 20 は、例えば、前記したEDRAM等を用いて分析を行う。なお、入力データが、画像データである場合、入力データの特徴部は、人物部分以外であってもよく、例えば、動物や移動体等であってもよい。

【0028】

なお、入力データは、画像データ以外にも、映像データ、テキストデータ、音声データ、時系列のセンサデータであってもよい。なお、以下の説明では、入力データが画像データである場合について説明する。

10

【0029】

分析装置 20 は、例えば、上記のEDRAMにより、情報処理装置 10 により前処理が行われた入力データに基づく枠の初期化、それまでの枠の記憶としての蓄積、前記記憶に基づく枠の絞り込みおよび分析、枠の位置および分析に関する誤差に基づく各NNのパラメータの更新等を行う。各処理には、NNが用いられ、各NNによる処理結果は、例えば図 1 に示すように、順伝搬および逆伝搬する。

【0030】

なお、分析装置 20 は、上記のEDRAM以外にも、スライディングウィンドウ法（後記）や、YOLO（You Only Look Once、後記）等により入力データから特徴部を抽出し、分析を行ってもよい。

20

【0031】

ここで、情報処理装置 10 は、入力データに対する特徴部の割合を占める割合（スケール）の予測結果に基づき入力データの分割を行う。

【0032】

例えば、情報処理装置 10 は、入力データに対する特徴部の割合（スケール）を予測し、予測したスケールが所定値以下の場合（例えば、画像データに対し特徴部となる人物部分が小さい場合）、入力データに対し所定の分割を行う。そして、情報処理装置 10 は、分割した入力データを分析装置 20 へ出力する。一方、予測したスケールが所定値以下の場合（例えば、画像データに対し特徴部となる人物部分が小さい場合）、情報収集装置 10 は、入力データに対する分割は行わずに分析装置 20 へ出力する。

30

【0033】

これにより、分析装置 20 に入力されるデータそれぞれのスケールのばらつきをできるだけ低減できるので、分析装置 20 は入力されるデータの特徴部を精度よく分析することができる。

【0034】

[構成]

引き続き、図 1 を用いて、情報処理装置 10 の構成を説明する。情報処理装置 10 は、入力部 11 と、スケール予測部（予測部）12 と、分割方法決定部 13 と、分割実行部 14 と、出力部 15 とを備える。

【0035】

40

入力部 11 は、入力データの入力を受け付ける。スケール予測部 12 は、入力部 11 により受け付けた入力データに対する特徴部の占める割合（スケール）を予測する。例えば、スケール予測部 12 は、入力データ（画像データ）に人物が写っているとすれば、どの程度のスケールで写っているかを予測する。ここでのスケールの予測には、例えば、機械学習を用いる。この機械学習には、例えば、NNを用いる。NNは、入力データとそのスケールのペアと学習することで、未知の入力データに対してもより正確にスケールを予測することができるようになる。

【0036】

ここで、図 2 を参照しながら、前記NNの学習に用いる訓練データの一例を説明する。例えば、訓練データとして、図 2 に示すように、入力データ（画像データ）と、その画像デ

50

ータにおける特徴部（人物部分）のスケールとを対応付けたデータセットを用意しておく。

【0037】

ここでは、画像データにおける特徴部（人物部分）の占める割合（スケール、R）を、R [15, 30]（カテゴリー1：スケール「大」）、R [10, 15]（カテゴリー2：スケール「中」）、R [5, 10]（カテゴリー3：スケール「小」）の3つのカテゴリーに分けたデータセットを用意した場合の例を示している。そして、スケール予測部12は、このデータセットにフィットするようにNNのパラメータを更新し、予測対象の入力データ（画像データ）が上記のスケール「大」、スケール「中」、スケール「小」のいずれに属するかを判定することによりスケールを予測する。

10

【0038】

例えば、入力データが、図3の符号301に示す画像データと、符号302に示す画像データとである場合を考える。この場合、スケール予測部12は、上記の機械学習の結果を用いて、符号301のように人物が小さく写っている画像データについては「スケールが小さい」と予測し、符号302のように人物が大きく写っている画像データについては「スケールが大きい」と予測する。

【0039】

なお、スケール予測部12は、入力データのスケール（R）を大、中、小等にカテゴライズせず、直接、スケール（R）の値を予測してもよい。

【0040】

なお、入力データが背景を含む画像データである場合、スケール予測部12を実装するNNは、画像データの特徴部の背景となる建物等の大きさ等の大きさから、入力データ（画像データ）が広角で撮られたものか、望遠で撮られたものかを判定し、その結果を正確なスケールの予測に役立てていると考えられる。

20

【0041】

図1の分割方法決定部13は、入力データの分割の方法（分割方法）、つまり、入力データを分割するか否か、また、入力データを分割する場合、当該入力データをいくつに分割するか、どのように分割するか等を決定する。例えば、分割方法決定部13は、スケール予測部12により予測された入力データのスケールに応じて、入力データの分割が必要か否かを判定し、必要と判定した場合はさらに当該入力データをいくつに分割するか、どのように分割するか等を決定する。そして、分割方法決定部13は、当該入力データとその分割方法とを分割実行部14に出力する。一方、分割方法決定部13は、入力データに対する分割は不要と判定した場合、当該入力データを出力部15へ出力する。

30

【0042】

例えば、分割方法決定部13は、図4に示すように、特徴部（人物部分）のスケールが所定値以下である画像データ402について、符号403に示すように4つに分割すると判定する。なお、分割方法決定部13は、入力データのスケールが小さいほど、入力データを細かく分割すると判定してもよい。例えば、スケール予測部12により予測された入力データのスケールが上記の所定値を大きく下回る場合、そのスケールの小ささに応じて、入力データをより細かく分割すると判定してもよい。そして、分割方法決定部13は、画像データ402と画像データ402の分割数の判定結果とを分割実行部14に出力する。

40

【0043】

一方、分割方法決定部13は、図4に示すように、特徴部（人物部分）のスケールが所定値を超える画像データ401については、分割は行わないと判定する。そして、分割方法決定部13は、画像データ401を出力部15に出力する。

【0044】

なお、スケール予測部12は、NNにより実現してもよい。この場合、スケール予測部12は、スケール予測部12が予測したスケールと実際のスケールとの誤差を受け取る。そして、スケール予測部12は、上記の誤差に基づき、スケール予測用いるパラメータの調

50

整を行う。このような処理を繰り返すことで、スケール予測部 12 は、入力データのスケールをより正確に予測できるようになる。

【0045】

図1の分割実行部14は、分割方法決定部13により決定された分割方法に基づき、入力データの分割を行う。そして、分割実行部14は、分割を行った入力データを出力部15へ出力する。例えば、分割実行部14は、図4の画像データ402を符号403に示すように4つに分割し、分割したすべての部分画像を出力部15へ出力する。

【0046】

出力部15は、分割実行部14および分割方法決定部13から出力された入力データを、分析装置20へ出力する。例えば、出力部15は、分割実行部14により4つに分割された画像データ402（図4の符号403参照）と、分割方法決定部13から出力された画像データ401とを分析装置20へ出力する。

【0047】

[処理手順]

次に、図5を用いてシステムの処理手順を説明する。まず、情報処理装置10の入力部11は、入力データを受け付ける（S1）。次に、スケール予測部12は、入力データのスケールを予測する（S2）。そして、分割方法決定部13は、S2で予測された入力データのスケールに基づき、当該入力データを分割するか否か、また分割する場合、どの程度細かく分割するかを決定する（S3：分割方法決定）。

【0048】

S3における分割方法決定の結果、S1で受け付けた入力データを分割しないと判定された場合（S4で「分割なし」）、分割方法決定部13は、当該入力データを出力部15経由で分析装置20に出力する（S6：データ出力）。一方、S3における分割判定の結果、S1で受け付けた入力データを分割すると判定された場合（S4で「分割あり」）、分割実行部14は、分割方法決定部13による判定結果に基づき、入力データに対し所定の分割を行う（S5）。そして、分割実行部14は分割後の入力データを出力部15へ出力する。その後、出力部15は、分割後の入力データを分析装置20へ出力する（S6：データ出力）。S6の後、分析装置20は、情報処理装置10から出力されたデータに対し分析を行う（S7）。

【0049】

このような情報処理装置10によれば、入力データのスケールが所定値以下の場合、スケールに応じて分割を行った上で、分析装置20に出力することもできる。これにより、入力データ群がマルチスケール性を持つ場合であっても、分析装置20に入力されるデータ群のスケールをできるだけ同じにすることができる。その結果、分析装置20は入力データにおける特徴部の分析精度を向上させることができる。

【0050】

[その他の実施形態]

なお、入力データが、図6のように奥行感のある画像データである場合、分割方法決定部13は、遠景の部分は遠景の部分として、近景の部分は近景の部分として分割するような分割の方法を定めてもよい。例えば、分割方法決定部13は、図6に示す画像の奥の方は細かく（小さく）分割し、手前の方は粗く（大きく）分割するような分割の方法を定めてもよい。このようにすることで、入力データに奥行感のある画像データが含まれる場合でも、分析装置20に入力されるデータのスケールをできるだけ同じにすることができる。

【0051】

また、分析装置20は、入力データから特徴部を抽出し、分析を行う装置であれば、前記したEDRAMを用いる装置に限定されない。例えば、分析装置20は、前記したスライディングウィンドウ法や、YOLO等により入力データから特徴部を抽出し、分析を行う装置であってもよい。

【0052】

10

20

30

40

50

例えば、分析装置 20 が、スライディングウィンドウ法により、入力データ（例えば、画像データ）から特徴部（人物部分）を抽出する装置である場合、分析装置 20 は、以下のようにして、画像データから人物部分を抽出し、分析を行う。

【0053】

すなわち、スライディングウィンドウ法を用いる分析装置 20 は、何種類かの大きさの枠（ウィンドウ）を用意しておき、この枠を画像データ上でスライドし、フルスキャンを行うことで、人物部分を検出し、抽出する。これにより、分析装置 20 は、例えば、図 7 に示す画像データ上から 1 人目、2 人目および 3 人目の人物部分を検出し、抽出する。そして、分析装置 20 は、抽出した人物部分の分析を行う。

【0054】

このスライディングウィンドウ法では、枠のサイズを調整する処理を行わないので、画像上に大きく写った人物は大きい枠でないと検出できない、また、画像上に小さく写った人物は小さい枠でないと検出できない。そして、人物部分の検出がうまくできないと、人物部分の分析精度も低下することになる。

【0055】

そこで、スライディングウィンドウ法を用いる分析装置 20 が、上記の情報処理装置 10 からデータそれぞれのスケールをできるだけ同じにしたデータ（画像データ）を受け取ることで、画像データに対し、適切な大きさの枠を用意しやすくなる。その結果、分析装置 20 は、画像データから人物部分を検出しやすくなるので、画像データにおける人物部分の分析精度を向上させることができる。また、分析装置 20 は、画像データに対し様々な大きさの枠を用意する必要がなくなるので、画像データから人物部分を検出する際に要する処理負荷を低減することができる。

【0056】

また、例えば、分析装置 20 が、YOLOにより、入力データ（例えば、画像データ）から特徴部となる人物部分を抽出し、分析を行う装置である場合、分析装置 20 は、以下のようにして、画像データから特徴部となる人物部分を抽出し、分析を行う。

【0057】

すなわち、YOLOを用いる分析装置 20 は、例えば、図 8 に示すように、画像データをグリッドに切り、グリッドごとに人物部分を探しに行く。そして、分析装置 20 は、人物部分を見つけると、その人物部分に枠を合わせる。ここで、YOLOを用いる分析装置 20 は、画像データから人物部分を見つけ、その人物部分に枠を合わせる処理がうまくいかないと、人物部分の検出がうまくいかず、その結果、人物部分の分析精度も低下することになる。

【0058】

そこで、YOLOを用いる分析装置 20 が、上記の情報処理装置 10 からデータそれぞれのスケールをできるだけ同じにしたデータ（画像データ）を受け取ることで、画像データから人物部分を検出しやすくなり、その結果、画像データにおける人物部分の分析精度を向上させることができる。

【0059】

また、前記したとおり、システムの扱う入力データは、画像データ以外にも、映像データ、テキストデータ、音声データ、時系列のセンサデータであってもよい。

【0060】

例えば、入力データがテキストデータである場合、特徴部は、例えば、当該テキストデータにおける、特定の語句、言い回し、表現等である。したがって、情報処理装置 10 は、入力データがテキストデータである場合、入力データのスケールとして、例えば、テキストデータ全体の文字数に対する、上記の特徴部の文字数の占める割合を用いる。

【0061】

そして、情報処理装置 10 は、テキストデータ全体の文字数に対する、上記の特徴部の文字数の占める割合（スケール）ができるだけ同じになるよう、必要に応じてテキストデータの分割を行い、分析装置 20 へ出力する。

【0062】

10

20

30

40

50

このようにすることで、分析装置 20 が、テキストデータにおける、特定の語句、言い回し、表現等の分析を行う分析装置である場合に、分析精度を向上させることができる。

【0063】

また、例えば、入力データが音声データである場合、特徴部は、例えば、背景ノイズのある音声データにおける人間の音声や、背景ノイズがない音声データにおける、特定の単語またはフレーズ、特定の人物の音声、特定の周波数帯域等である。したがって、情報処理装置 10 は、入力データが音声データである場合、入力データのスケールとして、例えば、音声データに対する、上記の人間の音声の SN 比 (Signal-to-Noise ratio)、または、音声データ全体の時間の長さに対する、特定の単語またはフレーズの時間の長さを用いる。また、音声データにおける特定の周波数帯域を用いる場合、情報処理装置 10 は、

10

【0064】

そして、情報処理装置 10 は、音声データ全体に対する、上記の特徴部 (人間の音声の SN 比、特定の単語またはフレーズの時間の長さ、特定の周波数帯域の幅) の占める割合 (スケール) ができるだけ同じになるよう、必要に応じて音声データの分割を行い、分析装置 20 へ出力する。

【0065】

このようにすることで、分析装置 20 が、音声データにおける、人間の音声、特定の単語またはフレーズ、特定の人物の音声、特定の周波数帯域等の分析を行う場合に、分析精度を向上させることができる。

20

【0066】

また、入力データが時系列のセンサデータである場合、特徴部は、例えば、何らかの異常を示すセンサの値のパターン等である。一例を挙げると、センサの値自体は通常時取り入る範囲 (正常範囲) だが、異常時特有のパターンを繰り返す場合がある (図 10 参照)。このような場合、異常を検知し、分析するため、時系列のセンサデータのうち、センサの値自体は正常の範囲だが、異常時特有のパターンを示す部分を特徴部として用いる。

【0067】

したがって、情報処理装置 10 は、入力データが時系列のセンサデータである場合、入力データのスケールとして、例えば、時系列のセンサデータのうち、センサの値自体は正常の範囲だが、異常時特有のパターンを示す部分の波長を用いる (図 10 参照)。そして、情報処理装置 10 は、時系列のセンサデータ全体に対する、上記の特徴部 (センサの値自体は正常範囲だが、異常時特有のパターンを示す部分) の波長の占める割合 (スケール) ができるだけ同じになるよう、必要に応じて時系列のセンサデータの分割を行い、分析装置 20 へ出力する。

30

【0068】

このようにすることで、分析装置 20 が、時系列のセンサデータから、異常の検知および分析を行う場合に、分析精度を向上させることができる。

【0069】

また、入力データは、ビデオ映像 (映像データ) であってもよい。この場合、特徴部は、例えば、ビデオ映像において人物が特定の動作をしているフレームである。そして、情報処理装置 10 は、ビデオ映像全体のフレーム数に対する、上記の特徴部 (ビデオ映像で人物が特定の動作をしているフレーム) の占める割合 (スケール) ができるだけ同じになるよう、必要に応じてビデオ映像のフレームの分割を行い、分析装置 20 へ出力する。

40

【0070】

このようにすることで、分析装置 20 が、ビデオ映像において人物が特定の動作をしているフレームを分析する場合に、分析精度を向上させることができる。

【0071】

[プログラム]

50

また、上記の実施形態で述べた情報処理装置 10 の機能を実現するプログラムを所望の情報処理装置（コンピュータ）にインストールすることによって実装できる。例えば、パッケージソフトウェアやオンラインソフトウェアとして提供される上記のプログラムを情報処理装置に実行させることにより、情報処理装置を情報処理装置 10 として機能させることができる。ここで言う情報処理装置には、デスクトップ型またはノート型のパーソナルコンピュータ、ラック搭載型のサーバコンピュータ等が含まれる。また、その他にも、情報処理装置にはスマートフォン、携帯電話機や PHS（Personal Handyphone System）等の移動体通信端末、さらには、PDA（Personal Digital Assistants）等がその範疇に含まれる。また、情報処理装置 10 を、クラウドサーバに実装してもよい。

【0072】

図 11 を用いて、上記のプログラム（情報処理プログラム）を実行するコンピュータの一例を説明する。図 11 に示すように、コンピュータ 1000 は、例えば、メモリ 1010 と、CPU 1020 と、ハードディスクドライブインタフェース 1030 と、ディスクドライブインタフェース 1040 と、シリアルポートインタフェース 1050 と、ビデオアダプタ 1060 と、ネットワークインタフェース 1070 とを有する。これらの各部は、バス 1080 によって接続される。

【0073】

メモリ 1010 は、ROM（Read Only Memory）1011 および RAM（Random Access Memory）1012 を含む。ROM 1011 は、例えば、BIOS（Basic Input Output System）等のブートプログラムを記憶する。ハードディスクドライブインタフェース 1030 は、ハードディスクドライブ 1090 に接続される。ディスクドライブインタフェース 1040 は、ディスクドライブ 1100 に接続される。ディスクドライブ 1100 には、例えば、磁気ディスクや光ディスク等の着脱可能な記憶媒体が挿入される。シリアルポートインタフェース 1050 には、例えば、マウス 1110 およびキーボード 1120 が接続される。ビデオアダプタ 1060 には、例えば、ディスプレイ 1130 が接続される。

【0074】

ここで、図 11 に示すように、ハードディスクドライブ 1090 は、例えば、OS 1091、アプリケーションプログラム 1092、プログラムモジュール 1093 およびプログラムデータ 1094 を記憶する。前記した実施形態で説明した各種データや情報は、例えばハードディスクドライブ 1090 やメモリ 1010 に記憶される。

【0075】

そして、CPU 1020 が、ハードディスクドライブ 1090 に記憶されたプログラムモジュール 1093 やプログラムデータ 1094 を必要に応じて RAM 1012 に読み出して、上述した各手順を実行する。

【0076】

なお、上記の情報処理プログラムに係るプログラムモジュール 1093 やプログラムデータ 1094 は、ハードディスクドライブ 1090 に記憶される場合に限られず、例えば、着脱可能な記憶媒体に記憶されて、ディスクドライブ 1100 等を介して CPU 1020 によって読み出されてもよい。あるいは、上記のプログラムに係るプログラムモジュール 1093 やプログラムデータ 1094 は、LAN や WAN（Wide Area Network）等のネットワークを介して接続された他のコンピュータに記憶され、ネットワークインタフェース 1070 を介して CPU 1020 によって読み出されてもよい。また、コンピュータ 1000 は、CPU 1020 に代えて、GPU（Graphics Processing Unit）を用いて処理を実行してもよい。

【符号の説明】

【0077】

- 10 情報処理装置
- 11 入力部
- 12 スケール予測部

10

20

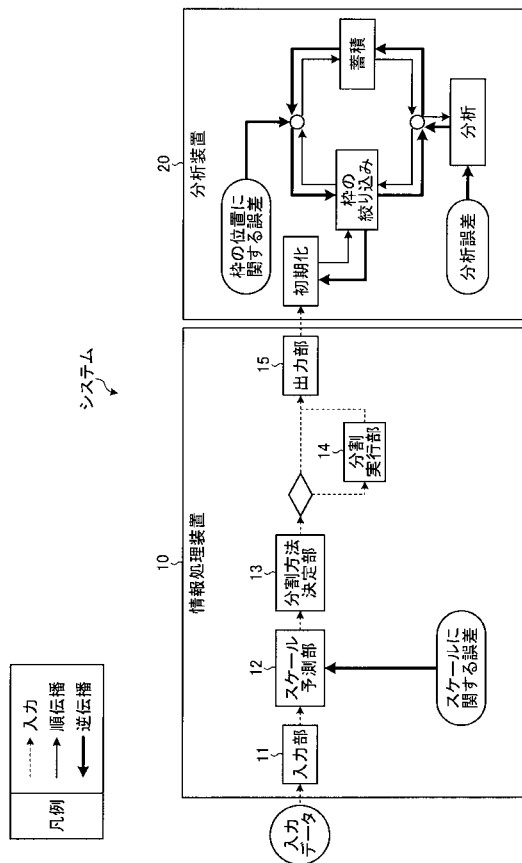
30

40

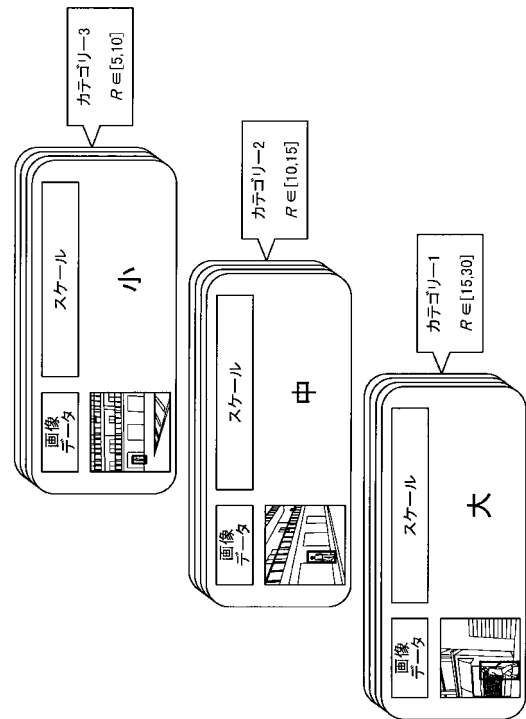
50

- 1 3 分割方法決定部
- 1 4 分割実行部
- 1 5 出力部
- 2 0 分析装置

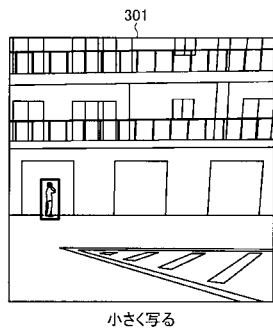
【 図 1 】



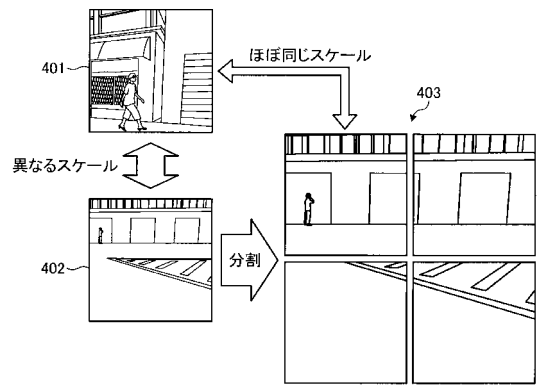
【 図 2 】



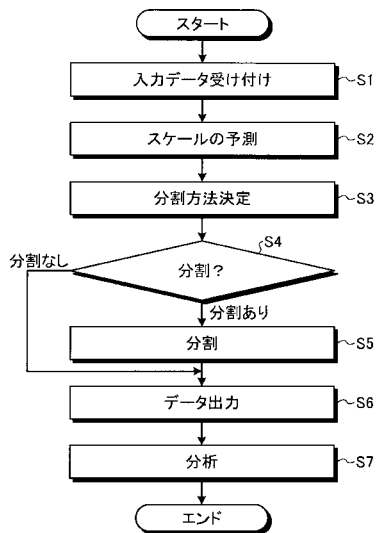
【 図 3 】



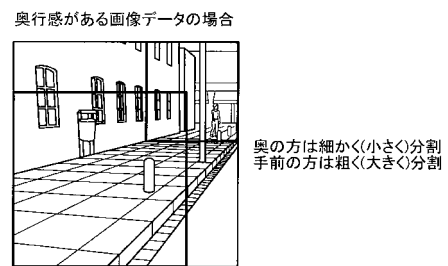
【 図 4 】



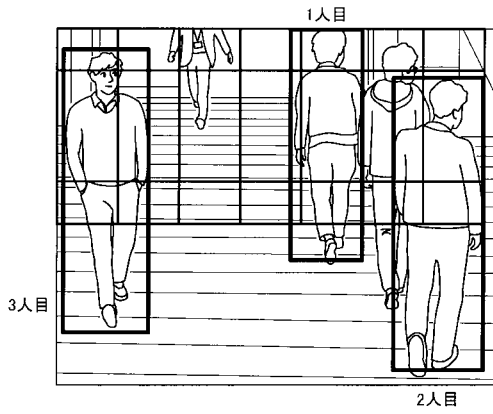
【 図 5 】



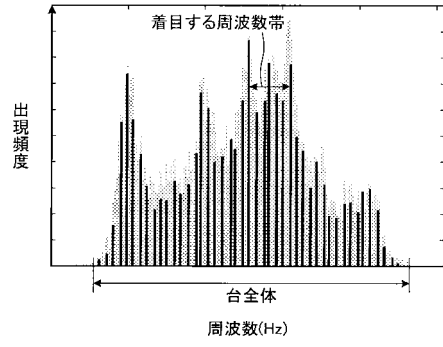
【 図 6 】



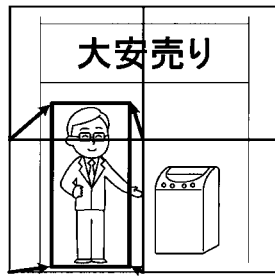
【 図 7 】



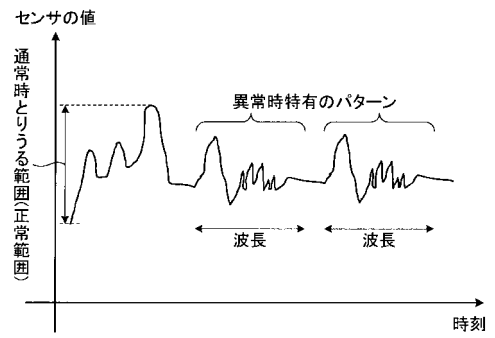
【 図 9 】



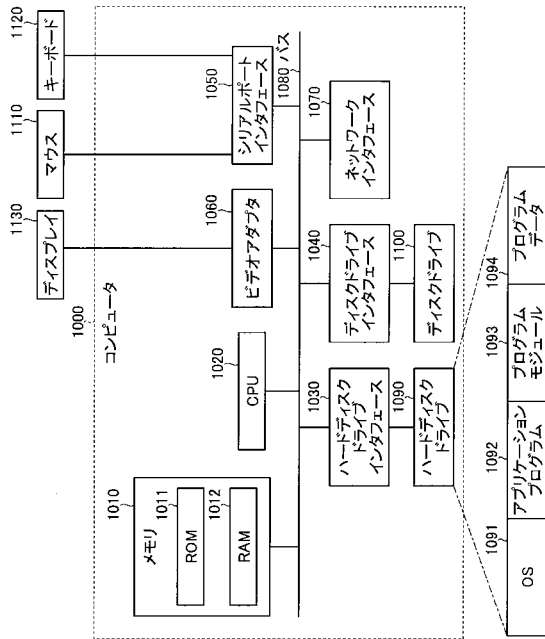
【 図 8 】



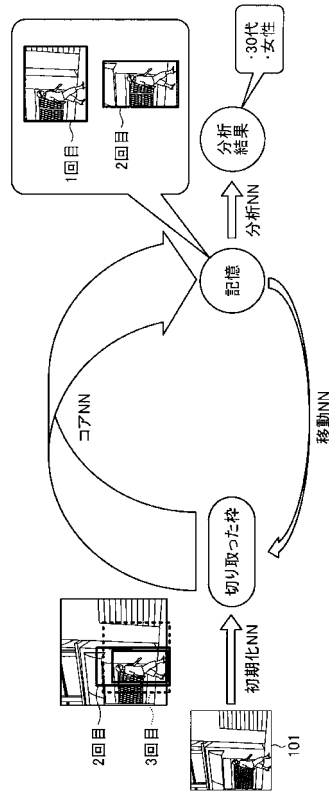
【 図 10 】



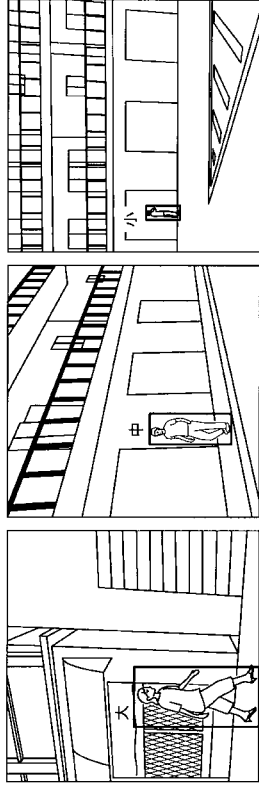
【 図 11 】



【 図 12 】



【 図 1 3 】



【 図 1 4 】

