(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2019年3月7日(07.03.2019)

WIPO | PCT

(10) 国際公開番号
WO 2019/045101 A1

(51) 国際特許分類:
G06T 7/00 (2017.01) *G06T 7/11* (2017.01)

(21) 国際出願番号： PCT/JP2018/032635

(22) 国際出願日： 2018年9月3日(03.09.2018)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(30) 優先権データ：
特願 2017-169632 2017年9月4日(04.09.2017) JP

(71) 出願人: 国立大学法人 東京大学 (**THE UNIVERSITY OF TOKYO**) [JP/JP]; 〒1138654 東京都文京区本郷七丁目３番１号 Tokyo (JP).

(72) 発明者: 相澤 清晴 (**AIZAWA Kayoharu**); 〒1138654 東京都文京区本郷七丁目３番１号 国立大学法人東京大学内 Tokyo (JP).

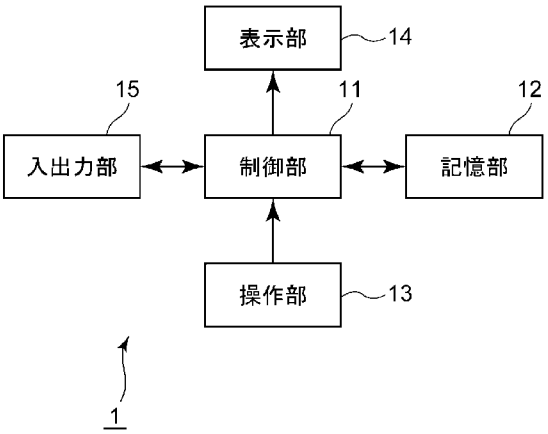
小川 徹 (**OGAWA Toru**); 〒1138654 東京都文京区本郷七丁目３番１号 国立大学法人東京大学内 Tokyo (JP).

(74) 代理人: 竹居 信利 (**TAKEI Nobutoshi**); 〒1600022 東京都新宿区新宿６丁目７－１ エルプリメント新宿３０８ Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** IMAGE PROCESSING DEVICE AND PROGRAM

(54) 発明の名称： 画像処理装置及びプログラム



11 Control unit
12 Storage unit
13 Operation unit
14 Display unit
15 Input/output unit

(57) **Abstract:** An image processing device for accepting comics image data and generating, on the basis of the comics image data, information that specifies a frame portion, information that specifies a face portion, information that specifies a body portion, and information that specifies a text portion through use of a result obtained by machine learning so as to specify each of the frame portion, the face position, the body portion, and the text portion, respectively, the information being supplied to prescribed information processing.

(57) 要約：漫画画像データを受け入れ、当該漫画画像データに基づいて、コマ部分、顔部分、身体部分、文字部分のそれぞれを特定するよう機械学習した結果を用い、コマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とをそれぞれ生成して、当該情報が所定の情報処理に供される画像処理装置である。

WO 2019/045101 A1

［続葉有］

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告(条約第21条(3))

明　細　書

発明の名称 ： 画像処理装置及びプログラム

技術分野

[0001] 本発明は、画像処理装置及びプログラムに関する。

背景技術

[0002] 近年では、漫画の画像データを加工して、セリフ部分を抽出して他国語用に翻訳する技術や、コマごとに配列を変更して、スマートフォン等の画面に適した状態とする技術が考えられている。

[0003] このような処理を行うにあたり、従来から、色や文字認識処理の結果等を用いて、描かれた人物の部分やセリフ部分を特定する処理等が考えられている（非特許文献１）。

先行技術文献

非特許文献

[0004] 非特許文献1：Christophe Rigaud, et. al., Speech ballon and speaker association for comics and manga understanding., Proceedings of the 13th International Conference on Document Analysis and Recognition, pp. 351-355, IEEE, 2015

発明の概要

発明が解決しようとする課題

[0005] 一方、近年では機械学習により画像中から物体を検出する技術が開発され、広く研究されている。しかしながら、従来の一般物体検出の処理では、画像中の特定の部分には一つの物体が含まれるとの前提で検出が行われるため、多数の検出対象が互いに重なりあっている場合については考慮されていない。

[0006] ところが漫画画像データにおいては、コマの内側に（ないしは複数のコマにまたがって）人物の身体や顔が描画され、また、これらの各部に重なり合わせてセリフの文字が配置されることが一般的である。従って、機械学習に

よる物体検出処理をそのまま適用したのでは、コマ、登場人物の身体、顔、セリフといった部分がそれぞれ十分な精度で検出できない。

[0007] 本発明は上記実情に鑑みて為されたもので、漫画画像データの処理において、機械学習処理を用いて、コマ、顔部分、身体部分、及び文字部分の認識精度を、従来のものに比べて向上できる画像処理装置及びプログラムを提供することを、その目的の一つとする。

課題を解決するための手段

[0008] 上記従来例の問題点を解決するための本発明は、画像処理装置であって、漫画画像データを受け入れる受入手段と、画像データから、当該画像データ内に描画された漫画のコマ部分を検出するよう機械学習された状態にあるフレーム検出手段と、画像データから、当該画像データ内に描画された顔部分を検出するよう機械学習された状態にある顔検出手段と、画像データから、当該画像データ内に描画された身体部分を検出するよう機械学習された状態にある身体検出手段と、画像データから、当該画像データ内に含まれる文字部分を検出するよう機械学習された状態にある文字検出手段と、前記受け入れた漫画画像データに基づいて、前記フレーム検出手段が検出したコマ部分を特定する情報と、前記顔検出手段が検出した、顔部分を特定する情報と、前記身体検出手段が検出した、身体部分を特定する情報と、前記文字検出手段が検出した、文字部分を特定する情報と、を生成する検出情報生成手段と、を含み、前記生成されたコマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とが所定の情報処理に供されることとしたものである。

発明の効果

[0009] 本発明によれば、機械学習処理を用いて、漫画画像データのうちからコマ、顔部分、身体部分、及び文字部分を認識する際の認識精度を、従来のものに比べて向上できる。

図面の簡単な説明

[0010] [図1]本発明の実施の形態に係る画像処理装置の構成例を表すブロック図であ

る。

[図2]本発明の実施の形態に係る画像処理装置の例を表す機能ブロック図である。

[図3]本発明の実施の形態に係る画像処理装置が処理の対象とする漫画画像データの概要例を表す説明図である。

[図4]本発明の実施の形態に係る画像処理装置の検出処理部の概要例を表す内部機能ブロック図である。

[図5]本発明の実施の形態に係る画像処理装置の検出処理部のもう一つの例を表す内部機能ブロック図である。

[図6]本発明の実施の形態に係る画像処理装置の検出処理部のもう一つの例による処理の概要例を表す説明図である。

[図7]本発明の実施の形態に係る画像処理装置の検出処理部の構成例を表す機能ブロック図である。

発明を実施するための形態

[0011] 本発明の実施の形態について図面を参照しながら説明する。本発明の実施の形態に係る画像処理装置１は、図１に例示するように、制御部１１と、記憶部１２と、操作部１３と、表示部１４と、入出力部１５とを含んで構成されている。

[0012] 制御部１１は、ＣＰＵ等のプログラム制御デバイスであり、記憶部１２に格納されたプログラムを実行して、漫画画像データを受け入れ、当該受け入れた漫画画像データに基づいて、コマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報と、を生成する。本実施の形態の制御部１１は、これらの各部分を特定する処理において、画像データから、当該画像データ内に描画された漫画のコマ部分を検出するよう機械学習された状態にあるフレーム検出器と、画像データから、当該画像データ内に描画された顔部分を検出するよう機械学習された状態にある顔検出器と、画像データから、当該画像データ内に描画された身体部分を検出するよう機械学習された状態にある身体検出器と、画像データから、

当該画像データ内に含まれる文字部分を検出するよう機械学習された状態にある文字検出器とを用いる。

[0013] またこの制御部１１は、生成されたコマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを用いて所定の情報処理を実行する。この情報処理としては、例えば、各部を表す画像を出力する処理や、文字部分を特定する情報により特定された範囲内の文字列に対する光学文字認識処理や、コマ部分ごとに画像データを分割する分割処理等がある。これらの制御部１１の動作については、後に詳しく述べる。

[0014] 記憶部１２は、メモリデバイス等であり、制御部１１により実行されるプログラムを保持する。このプログラムは、コンピュータ可読かつ、非一時的な記録媒体に格納されて提供され、この記憶部１２に格納されたものであってもよい。また、この記憶部１２は、制御部１１のワークメモリとしても動作する。

[0015] 操作部１３は、マウスやキーボード等であり、利用者の指示操作を受け入れて制御部１１に出力する。表示部１４は、例えばディスプレイ等であり、制御部１１から入力される指示に基づいて情報を表示出力する。

[0016] 入出力部１５は、例えばネットワークインタフェース等であり、外部からデータ（画像データ等）を受信して、制御部１１に出力する。またこの入出力部１５は、制御部１１から入力される指示に従って、データを外部の装置等に送出する。

[0017] 次に制御部１１の動作について説明する。本実施の形態の制御部１１は、記憶部１２に格納されたプログラムを実行することで、機能的には、図２に例示するように、受入部２１と、検出処理部２２と、検出情報生成部２３と、情報処理部２４とを含んで構成される。また検出処理部２２は、フレーム検出部３１と、顔検出部３２と、身体検出部３３と、文字検出部３４とを含む。

[0018] 受入部２１は、漫画画像データを受け入れて検出処理部２２に出力する。

ここで漫画画像データは、一般的には、顔部分（Ｆ）と身体部分（Ｂ）と文字部分（Ｃ）とが互いに重なり合って描画された画像データであり（図３）、少なくとも一つのコマ（Ｍ）を含む。また、この受入部２１は、検出処理部２２におけるニューラルネットワークを利用するため、漫画画像データを拡大または縮小して、ニューラルネットワークの入力に適したサイズにリサイズする。

[0019] 検出処理部２２のフレーム検出部３１は、画像データから、当該画像データ内に描画された漫画のコマ部分を検出するよう機械学習された状態にあるフレーム検出器を有する。具体的に、このフレーム検出部３１が備えるフレーム検出器は、Ｒ－ＣＮＮ（Regions with CNN features）（Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014）や、Fast Ｒ－ＣＮＮ（Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE International Conference on Computer Vision. 2015）、Faster Ｒ－ＣＮＮ（Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015）、YOLO（You Only Look Once）（Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." arXiv preprint arXiv: 1506.02640 (2015)）、あるいは、ＳＳＤ（シングル・ショット・マルチボックス・ディテクタ；Single Shot MultiBox Detector）（Liu, Wei, et al. "SSD: Single Shot MultiBox Detector." arXiv preprint arXiv:1512.02325 (2015)）など、種々の方法で構成されたニューラルネットワークを採用して実現できる。

[0020] 図４にその概略を示すように、ＳＳＤ等のニューラルネットワークを採用した検出器４０は、ベースネットワーク部４１と、分類器４２とを含んで構成される。ここでベースネットワーク部４１は、検出対象の候補が含まれる画像の範囲と、当該範囲内の画像の特徴量とを出力する。また分類器４２は

、出力された画像の範囲に、検出対象（フレーム検出部３１の場合、漫画画像データのコマを区分する枠線）が含まれるか否かを、出力された特徴量に基づいて判断する。

[0021] このようなＳＳＤ等を採用した検出器４０は、検出対象の範囲（フレーム検出部３１の場合、漫画画像データのコマを区分する枠線に外接する形状の範囲）を人為的に指定した画像データのサンプルを用いて機械学習させる。ここで機械学習の具体的方法や、検出器４０の利用方法については、広く知られているので、ここでの詳しい説明を省略する。

[0022] 顔検出部３２は、画像データから、当該画像データ内に描画されたキャラクタの顔部分を検出するよう機械学習された状態にある顔検出器を有する。この顔検出器も、フレーム検出部３１が備えるフレーム検出器と同様、ＳＳＤ等、種々の方法で構成されたニューラルネットワークを採用して実現できる。この顔検出器は、検出対象の範囲である、漫画画像データに含まれるキャラクタの顔に外接する所定形状の範囲を人為的に指定した画像データのサンプルを用いて機械学習させる。

[0023] 身体検出部３３は、画像データから、当該画像データ内に描画されたキャラクタの身体部分を検出するよう機械学習された状態にある身体検出器を有する。この身体検出器も、フレーム検出部３１が備えるフレーム検出器と同様、ＳＳＤ等、種々の方法で構成されたニューラルネットワークを採用して実現できる。この身体検出器は、検出対象の範囲である、漫画画像データに含まれるキャラクタの身体に外接する所定形状の範囲を人為的に指定した画像データのサンプルを用いて機械学習させる。

[0024] 文字検出部３４は、画像データから、当該画像データ内に描画された文字部分を検出するよう機械学習された状態にある文字検出器を有する。この文字検出器も、フレーム検出部３１が備えるフレーム検出器と同様、ＳＳＤ等、種々の方法で構成されたニューラルネットワークを採用して実現できる。この文字検出器は、検出対象の範囲である、漫画画像データに含まれる文字部分に外接する所定形状の範囲を人為的に指定した画像データのサンプルを

用いて機械学習させる。

[0025] 検出情報生成部２３は、受入部２１が受け入れた漫画画像データについて、フレーム検出部３１が検出したコマ部分を特定する情報と、顔検出部３２が検出した、顔部分を特定する情報と、身体検出部３３が検出した、身体部分を特定する情報と、文字検出部３４が検出した、文字部分を特定する情報とを生成する。

[0026] 情報処理部２４は、生成されたコマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを用いて所定の情報処理を実行する。この情報処理としては、例えば、特定された文字部分の画像に対して光学的文字認識（ＯＣＲ）を行い、その結果を出力する処理等がある。また、情報処理部２４は、光学的文字認識の結果、得られた文字列を、機械翻訳処理により他言語に翻訳して出力してもよい。

[0027] 本実施の形態の一例は以上の構成を備え、次のように動作する。なお、以下の説明では、制御部１１によるフレーム検出部３１，顔検出部３２，身体検出部３３，及び文字検出部３４は、ＳＳＤを採用し、それぞれ、予め画像データから、当該画像データ内に描画された漫画のコマ部分、顔部分、身体部分、及び文字部分を検出するよう機械学習した状態にあるものとする。

[0028] 画像処理装置１は、利用者から入力される漫画の画像データ（機械学習のサンプルに含まれないもの）を処理の対象として、当該処理対象の画像データに対して並列的に、フレーム検出器と、顔検出器と、身体検出器と、文字検出器とにより、コマ部分、キャラクタの顔部分、身体部分、及び文字部分をそれぞれ検出して、それぞれ検出した画像の範囲を特定する情報を得る。

[0029] そして画像処理装置１は、コマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを用いて所定の情報処理、例えば特定された文字部分の画像に対して光学的文字認識（ＯＣＲ）を行い、当該光学的文字認識の結果、得られた文字列を、機械翻訳処理により他言語に翻訳して出力する。

[0030] ［ベースネットワークを共用する例］

またここまでの説明では、フレーム検出部３１，顔検出部３２，身体検出部３３，及び文字検出部３４は、それぞれ独立したベースネットワークと、検出器を備えるものとしたが、本実施の形態はこの例に限られない。例えば一つのベースネットワークをフレーム検出部３１，顔検出部３２，身体検出部３３，文字検出部３４が共用してもよい。

[0031] すなわちこの例では、フレーム検出部３１，顔検出部３２，身体検出部３３，及び文字検出部３４は、図５に例示するように、それぞれに共通して、検出対象の候補となる画像の範囲と、当該範囲内の画像の特徴量とを機械学習した状態にあり、処理対象となった画像データに基づき、検出対象の候補となる画像の範囲と、当該範囲内の画像の特徴量とを出力するベースネットワーク部４１′と、フレーム検出部３１，顔検出部３２，身体検出部３３，及び文字検出部３４のそれぞれに対応して、独立して設けられる分類器４２ａ，４２ｂ，４２ｃ，４２ｄとを備える。

[0032] なお、この例でも、ベースネットワーク部４１′及び分類器４２ａ，ｂ，ｃ，ｄは、ＳＳＤに基づくニューラルネットワークとしてよいが、次の点でＳＳＤを変形して用いる。すなわち一般的なＳＳＤの出力段では、物体を検出する領域の候補（アンカーボックス）が予め複数定められており（複数のアンカーボックスの集合をアンカーセットと呼ぶ）、当該複数の領域の候補のうちから、対象となる物体が含まれる領域を特定する。

[0033] 本実施の形態のここでの例では、出力段より前のネットワーク（ベースネットワーク部４１′）は１つとするが、出力段において、アンカーセット（各アンカーセットには、例えば８７３２個のアンカーボックスが含まれる）をフレーム検出部３１，顔検出部３２，身体検出部３３，及び文字検出部３４のそれぞれに対応して、４つ複製して、各分類器４２ａ，ｂ，ｃ，ｄとして用いる。すなわちこの例におけるＳＳＤでは、各アンカーボックスについて、当該アンカーボックス内で検出した物体の領域との位置特定誤差（左上座標の情報と幅及び高さの情報とからなる４次元の情報）と、物体が含まれ得るとされる確信度（ここではシグモイド関数により正規化しておく）との

合計５次元の情報を出力するが、アンカーボックスが複製した４つの同じアンカーセット中のアンカーボックスのうち、第１のアンカーセットＡ１中のアンカーボックスについては画像データのうちコマ部分を機械学習させた状態とする。また第２のアンカーセットＡ２中のアンカーボックスについては、画像データのうち顔部分を機械学習させた状態とし、第３のアンカーセットＡ３中のアンカーボックスについては、画像データのうち身体部分を機械学習させた状態とし、第４のアンカーセットＡ４中のアンカーボックスについては、画像データのうち文字部分を機械学習させた状態とする（図６）。

[0034] 具体的には、出力段が出力する第１のアンカーセット中のアンカーボックスの情報については、学習用のサンプルを入力したときに、コマの枠線に外接する矩形が推定されることとなるよう、出力段から誤差を逆伝播して、分類器４２ａ及びベースネットワーク部４１′のパラメータを更新する。

[0035] 同様に、出力段が出力する第２のアンカーセット中のアンカーボックスの情報については、学習用のサンプルを入力したときに、キャラクタの顔部分に外接する矩形が推定されることとなるよう、出力段から誤差を逆伝播して、分類器４２ｂ及びベースネットワーク部４１′のパラメータを更新する。また出力段が出力する第３のアンカーセット中のアンカーボックスの情報については、学習用のサンプルを入力したときに、キャラクタの身体に外接する矩形が推定されることとなるよう、出力段から誤差を逆伝播して、分類器４２ｃ及びベースネットワーク部４１′のパラメータを更新する。さらに出力段が出力する第４のアンカーセット中のアンカーボックスの情報については、学習用のサンプルを入力したときに、文字に外接する矩形が推定されることとなるよう、出力段から誤差を逆伝播して、分類器４２ｄ及びベースネットワーク部４１′のパラメータを更新する。

[0036] なお、第ｉ（ｉ＝１，２，３，４）のアンカーセット中のａ番目のアンカーボックス（ａ＝１，２，…８７３２）に対する、ｍ番目のサンプル（ミニバッチ学習を行うこととして、ミニバッチサイズをＭとすると、ｍ＝１，２，…，Ｍ）の割り当てｓ（ｍ，ｉ，ａ）とその重なりＪ（ｍ，ｉ，ａ）とを

次のように定義する。

[数1]

$$s(m,i,a) = \mathrm{argmax}_{g,c=t(m,g)} Jaccard(B_a, B_{m,g})$$

$$J(m,i,a) = Jaccard\left(B_a, B_{m,s(m,i,a)}\right)$$

ここで、ｇは１以上、Ｇ（ｍ）以下の整数であり、Ｇ（ｍ）は、上記ｍ番目のサンプルに含まれる正解の個数であり、ｔ（ｍ，ｇ）、及びＢ（ｍ，ｇ）は上記ｍ番目のサンプルのｇ番目の正解のクラス（コマ、顔、身体、文字のいずれであるかを表す情報）と、外接矩形とを表す。

[0037] そして損失関数（Loss関数）Ｌ（ｚ）を、位置特定誤差Ｌloc（ｍ，ｚ）と、確信度Ｌconf（ｍ，ｚ）との和として次のように設定する。

[数2]

$$L(z) = \frac{1}{\sum_{m=1,\cdots,M} |A(m,pos)|} \sum_{m=1,\cdots,M} \left(L_{loc}(m,z) + L_{conf}(m,z)\right)$$

ここで、ｚはニューラルネットワークの出力を表し、Ａ（ｍ，ｐｏｓ）は、ｍ番目のサンプルについてオブジェクトが割り当てられたアンカーボックスの添字集合であり、具体的には、

[数3]

$$A(m,pos) = \{(c,a) | J(m,i,a) \geq 0.5\}$$

などとしておく。

[0038] なお、Ｌloc（ｍ，ｚ）及び、Ｌconf（ｍ，ｚ）は、次のように定義しておく。

[数4]

$$L_{loc}(m,z)=\sum_{A(m,pos)} huber\left(z(i,a,2\cdots5),\log\Delta B\right)$$

ここで、$\Delta B=B(m,s(m,i,a))-B(a)$

$$L_{conf}(m,z)=\sum_{A(m,pos)\cup A(m,neg)} l(m,i,a,z)$$

ここで、$l(m,i,a,z)=sigmoid(z(i,a,1),J(m,i,a)\geq 0.5)$

なお、Ａ（ｍ，ｎｅｇ）は、ハードネガティブ（hard negative）の集合であって、オブジェクトに割り当てられていないアンカーボックスのうち、ｌ（ｍ，ｉ，ａ，ｚ）が大きい順に上位ｋ｜Ａ（ｍ，ｐｏｓ）｜個を選択して得られる（ハードネガティブマイニングと呼ばれる方法であるので、ここでの詳細な説明を省略する）。また、huber()は、ヒューバー関数（huber関数）である。この関数についても広く知られているのでここでの詳細な説明を省略する。

[0039] 以上のようにフレーム検出部３１，顔検出部３２，身体検出部３３，及び文字検出部３４が構成された本実施の形態の画像処理装置１においても、フレーム検出部３１，顔検出部３２，身体検出部３３，及び文字検出部３４のそれぞれに対応する分類器４２ａ，ｂ，ｃ，ｄが、それぞれコマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを推定するので、これらを用いて所定の情報処理を実行する。

[0040] ［画像データを分割する例］

また本実施の形態の制御部１１は、受入部２１の動作として、入力された漫画画像データの全体を、拡大または縮小して、ニューラルネットワークの入力に適したサイズにリサイズするのではなく、入力された漫画画像データを、所定の条件に基づいて複数の分割部分に分割し、当該分割して得られた分割部分（部分的な漫画画像データ、以下、部分画像データと呼ぶ）を、ニューラルネットワークの入力に適したサイズにリサイズして、検出処理部２

２に出力してもよい。

[0041] ここで上記所定の条件は、例えば、元の漫画画像データ（幅ｗ，高さｈ）を、２×２個に分割（それぞれが幅ｗ／２，高さｈ／２となるような、重なり合わない４つの領域に分割）するとの条件であってもよい。またこの条件は、漫画画像データの内容に基づき、例えば、白色（背景色）が連続する部分で分割するとの条件であってもよい。さらに、この所定の条件は、コマごとに分割するとの条件であってもよい。

[0042] 本実施の形態のこの例では、検出処理部２２のフレーム検出部３１，顔検出部３２，身体検出部３３，及び文字検出部３４は、部分画像データのそれぞれからコマ部分（コマごとに分割しない場合）、顔部分、身体部分及び文字部分を検出する。なお、この例では、機械学習の処理も、分割して得られた部分画像データを用いて行うこととしてもよい。

[0043] そして検出情報生成部２３は、部分画像データごとにフレーム検出部３１が検出したコマ部分を特定する情報と、顔検出部３２が検出した、顔部分を特定する情報と、身体検出部３３が検出した、身体部分を特定する情報と、文字検出部３４が検出した、文字部分を特定する情報とを生成し、これらをまとめて元の漫画画像データにおける、コマ部分、顔部分、身体部分、及び文字部分のそれぞれを特定する情報を生成する。

[0044] 情報処理部２４は、検出情報生成部２３により生成されたコマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを用いて所定の情報処理を実行する。

[0045] また、上述の通り、部分画像データに分割する所定の条件として、コマごとに分割するとの条件であってもよい。この場合、制御部１１は、検出処理部２２のフレーム検出部３１としての動作によりコマ部分を検出し、当該検出したコマ部分ごとに分割して部分画像データを生成することとしてもよい。

[0046] すなわち本実施の形態のこの例では、図７に例示するように、フレーム検出部３１が出力する、コマ部分（コマを区分する枠線）に外接する多角形（

または円弧等の曲線を少なくとも一部に含んでもよい）を特定する情報を、顔検出部３２，身体検出部３３，文字検出部３４に出力する。そして、顔検出部３２，身体検出部３３，文字検出部３４のそれぞれが、フレーム検出部３１が出力する情報で特定されるコマ部分ごとに、各コマ部分をそれぞれ部分画像データとして、部分画像データのそれぞれから顔部分、身体部分及び文字部分を検出する。なお、この例でも、顔検出部３２，身体検出部３３，文字検出部３４に係る機械学習の処理は、分割して得られた部分画像データを用いて行うこととしてもよい。

[0047] そしてこの例でも、検出情報生成部２３は、フレーム検出部３１が検出したコマ部分ごとに、顔検出部３２が検出した、顔部分を特定する情報と、身体検出部３３が検出した、身体部分を特定する情報と、文字検出部３４が検出した、文字部分を特定する情報とを生成し、これらをまとめて元の漫画画像データにおける、コマ部分、顔部分、身体部分、及び文字部分のそれぞれを特定する情報を生成する。

[0048] 情報処理部２４は、検出情報生成部２３により生成されたコマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを用いて所定の情報処理を実行する。

[0049] ［検出結果の合成］

また処理の対象とする画像データを分割する場合、制御部１１は、分割前の画像データについても、ニューラルネットワークの入力に適したサイズにリサイズして、検出処理部２２としての動作を行ってもよい。すなわち、制御部１１は、フレーム検出部３１，顔検出部３２，身体検出部３３，及び文字検出部３４の動作として、分割前の画像データのそれぞれからコマ部分、顔部分、身体部分及び文字部分を検出する。

[0050] そして制御部１１は、ここで検出したコマ部分、顔部分、身体部分及び文字部分を特定する情報を記憶しておき、さらに処理の対象とする画像データを、所定の条件に基づいて複数の分割部分に分割し、当該分割して得られた部分画像データごとに、ニューラルネットワークの入力に適したサイズにリ

サイズして、検出処理部２２としての動作を行う。

[0051] この例によると、分割前の画像データについて検出されたコマ部分、顔部分、身体部分及び文字部分を特定する情報と、分割後に得られた部分画像データごとの顔部分、身体部分及び文字部分を特定する情報とが得られることとなる。

[0052] そして制御部１１は、分割前の画像データから検出されたコマ部分、顔部分、身体部分及び文字部分を特定する情報と、分割後の部分画像データのそれぞれから検出された顔部分、身体部分及び文字部分を特定する情報とを用い、分割前、または分割後のいずれか少なくとも一方から顔部分、身体部分及び文字部分が検出されたならば、検出情報生成部２３は、当該少なくとも一方から検出した顔部分、身体部分及び文字部分を特定する情報を生成して出力する（各部分の検出結果をそれぞれ統合して出力する）。

[0053] この例では、いわば、分割前の画像データから検出した顔部分、身体部分、文字部分のそれぞれと、分割後の画像データから検出した顔部分、身体部分、文字部分のそれぞれとの論理和が、処理対象となった画像データから検出した顔部分、身体部分、文字部分として、当該処理対象となった画像データから検出した顔部分、身体部分、文字部分を特定する情報が出力される。

[0054] なお、コマ部分は、顔部分、身体部分、または文字部分よりも一般的に高い精度で検出できるため、分割前の画像データ（または分割後の画像データであってもよい）のいずれか一方のみから検出すれば十分と考えられるが、制御部１１は、コマ部分についても、分割前の画像データまたは分割後の画像データの少なくともいずれかから検出した場合に、当該コマ部分を特定する情報を出力するようにしてもよい。

[0055] また、このように、いずれかから検出された各部分（コマ部分、顔部分、身体部分、文字部分のそれぞれ）の情報を出力する場合は、重複している部分の情報については、重複を除いて出力する。

[0056] またここでは分割前の画像データと、分割後の画像データとのいずれかから検出されたコマ部分、顔部分、身体部分及び文字部分を特定する情報を出

力することとした。つまり、例えば処理対象の画像データが漫画の１ページ分の画像データである場合、ページ全体で検出したものと、部分ごとに分割した分割部分ごとに検出したものとの「ＯＲ（論理和）」をとることとした。しかしながら、本実施の形態のこの例は、これに限られず、分割前の画像データと、分割後の画像データとの双方から共通して検出されたコマ部分、顔部分、身体部分及び文字部分を特定する情報を出力してもよい（つまり、例えば処理対象の画像データが漫画の１ページ分の画像データである場合、ページ全体で検出したものと、部分ごとに分割した分割部分ごとに検出したものとの「ＡＮＤ（論理積）」をとってもよい）。

[0057] さらに、ここでは分割の態様を一種類としたが、複数種類の分割態様で分割して得た複数種類の部分画像データを生成してもよい。例えば、コマごとに分割して得た部分画像データと、２×２の４分割した部分画像データと…といったように複数種類の態様で分割して得られた部分画像データ（さらに分割前の画像データを加えてもよい）のいずれか少なくとも一つから（あるいはそれぞれから共通して）検出されたコマ部分、顔部分、身体部分及び文字部分を特定する情報を出力することとしてもよい。この場合も、重複が生じる場合は、重複を除いて出力する。

[0058] ［実施形態の効果］

このように本実施の形態によれば、漫画画像データ内で互いに重なり合い、または包含関係となるコマ部分、キャラクタの顔部分、身体部分、及び文字部分の各部に対応してそれぞれ独立した検出器（または分類器）により、それぞれ検出を行うので、従来の機械学習を利用した検出に比べ、検出の精度を向上できる。

符号の説明

[0059] １　画像処理装置、１１　制御部、１２　記憶部、１３　操作部、１４　表示部、１５　入出力部、２１　受入部、２２　検出処理部、２３　検出情報生成部、２４　情報処理部、３１　フレーム検出部、３２　顔検出部、３３　身体検出部、３４　文字検出部、４０　検出器、４１，４１′　ベース

ネットワーク部、４２　分類器。

請求の範囲

[請求項1] 漫画画像データを受け入れる受入手段と、

画像データから、当該画像データ内に描画された漫画のコマ部分を検出するよう機械学習された状態にあるフレーム検出手段と、

画像データから、当該画像データ内に描画された顔部分を検出するよう機械学習された状態にある顔検出手段と、

画像データから、当該画像データ内に描画された身体部分を検出するよう機械学習された状態にある身体検出手段と、

画像データから、当該画像データ内に含まれる文字部分を検出するよう機械学習された状態にある文字検出手段と、

前記受け入れた漫画画像データに基づいて、前記フレーム検出手段が検出したコマ部分を特定する情報と、前記顔検出手段が検出した、顔部分を特定する情報と、前記身体検出手段が検出した、身体部分を特定する情報と、前記文字検出手段が検出した、文字部分を特定する情報と、を生成する検出情報生成手段と、

を含み、

前記生成されたコマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とが所定の情報処理に供される画像処理装置。

[請求項2] 請求項１記載の画像処理装置であって、

前記フレーム検出手段と、顔検出手段と、身体検出手段と、文字検出手段とは、それぞれに共通して、検出対象の候補となる画像の範囲と、当該範囲内の画像の特徴量とを機械学習した状態にあり、処理対象となった画像データに基づき、検出対象の候補となる画像の範囲と、当該範囲内の画像の特徴量とを出力するベースネットワーク部と、

前記フレーム検出手段と、顔検出手段と、身体検出手段と、文字検出手段とのそれぞれに対応して設けられる分類器であって、前記画像の特徴量に基づき、対応する画像の範囲内に含まれる画像が、それぞ

れコマ部分、顔部分、身体部分、文字部分であるか否かを分類する分類器とを含む画像処理装置。

[請求項3] 請求項１または２に記載の画像処理装置であって、

前記検出情報生成手段は、前記受け入れた漫画画像データを、所定の条件に基づいて複数の分割部分に分割して得られた分割部分のそれぞれのうちから、前記顔検出手段と、身体検出手段と、文字検出手段とにより、分割部分ごとに、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを生成する画像処理装置。

[請求項4] 請求項３に記載の画像処理装置であって、

前記検出情報生成手段は、フレーム検出手段が検出したコマ部分を特定する情報を用いて、当該特定されたコマ部分のそれぞれを前記分割部分として、当該分割部分のそれぞれのうちから、前記顔検出手段と、身体検出手段と、文字検出手段とにより、分割部分であるコマ部分ごとに、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを生成する画像処理装置。

[請求項5] 請求項３または４に記載の画像処理装置であって、

前記検出情報生成手段は、前記受け入れた漫画画像データを、分割する前の画像データから前記顔検出手段と、身体検出手段と、文字検出手段とにより検出した顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを生成し、

当該生成した情報と、前記分割部分のそれぞれのうちから検出した、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とを統合して出力する画像処理装置。

[請求項6] 請求項１から５のいずれか一項に記載の画像処理装置であって、

前記フレーム検出手段と、顔検出手段と、身体検出手段と、文字検出手段とは、シングル・ショット・マルチボックス・ディテクタ（ＳＳＤ）を用いて構成される画像処理装置。

[請求項7] コンピュータを、

漫画画像データを受け入れる受入手段と、

画像データから、当該画像データ内に描画された漫画のコマ部分を検出するよう機械学習された状態にあるフレーム検出手段と、

画像データから、当該画像データ内に描画された顔部分を検出するよう機械学習された状態にある顔検出手段と、

画像データから、当該画像データ内に描画された身体部分を検出するよう機械学習された状態にある身体検出手段と、

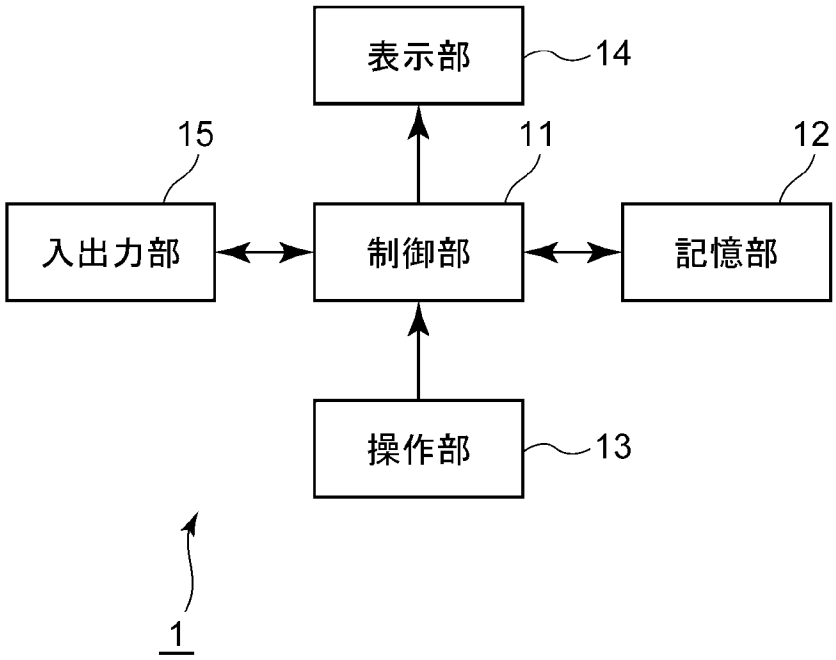
画像データから、当該画像データ内に含まれる文字部分を検出するよう機械学習された状態にある文字検出手段と、

前記受け入れた漫画画像データに基づいて、前記フレーム検出手段が検出したコマ部分を特定する情報と、前記顔検出手段が検出した、顔部分を特定する情報と、前記身体検出手段が検出した、身体部分を特定する情報と、前記文字検出手段が検出した、文字部分を特定する情報と、を生成する検出情報生成手段と、

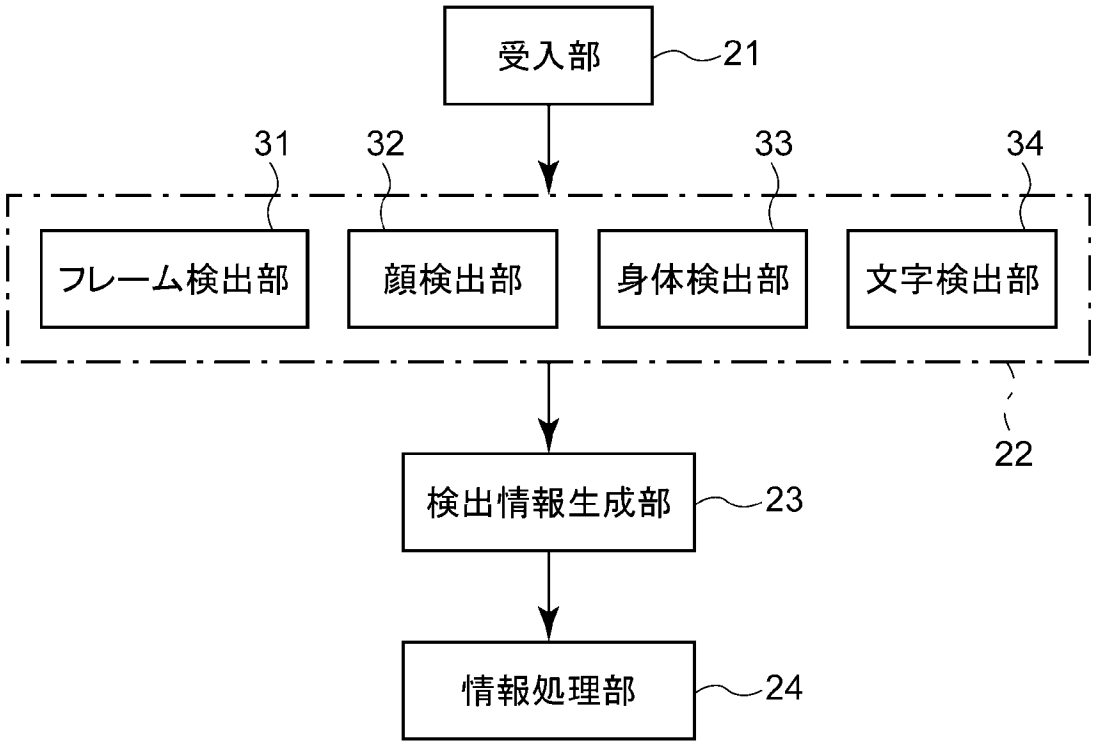
として機能させ、

前記生成されたコマ部分を特定する情報と、顔部分を特定する情報と、身体部分を特定する情報と、文字部分を特定する情報とが所定の情報処理に供されるプログラム。

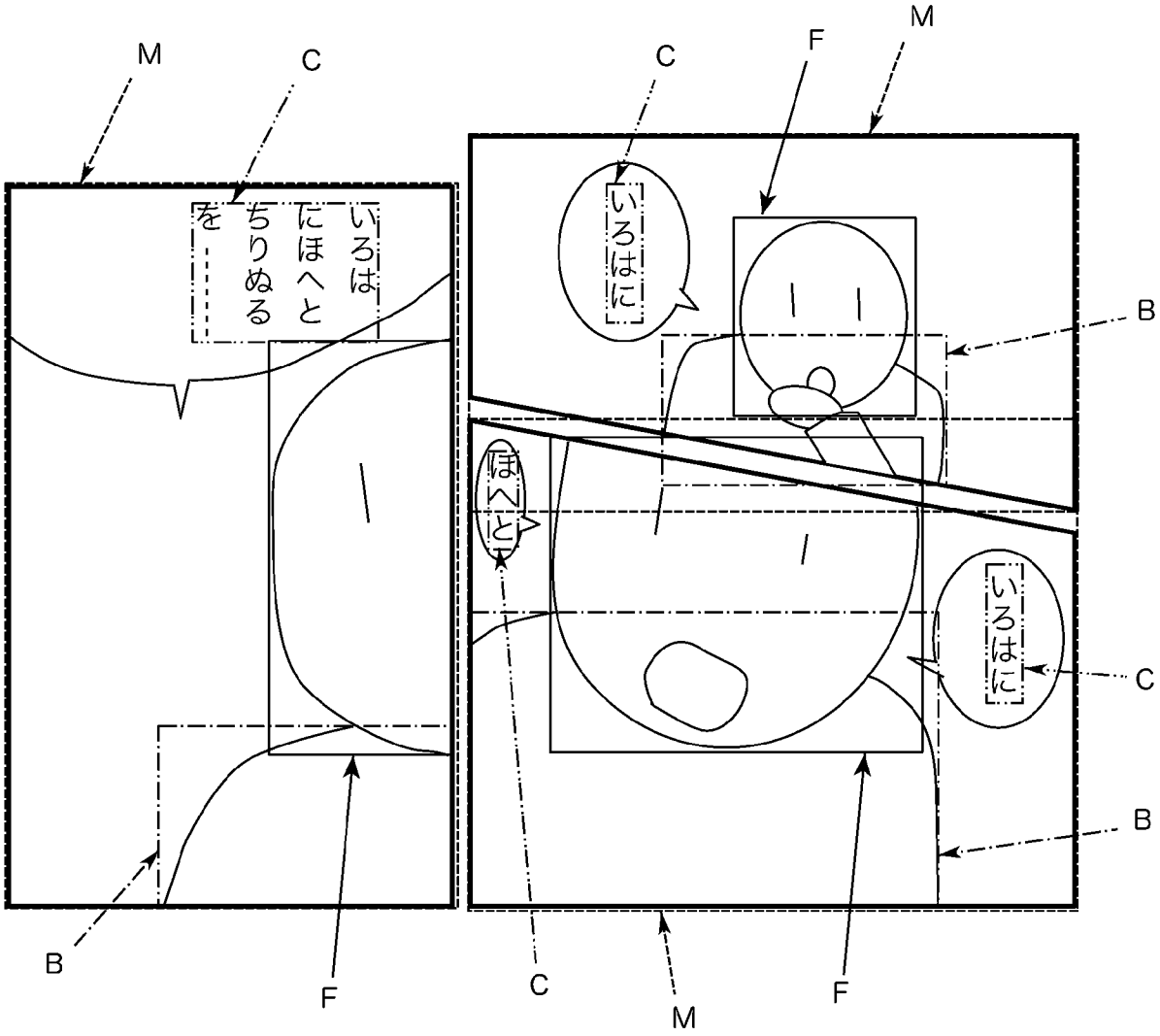
[図1]


表示部
14
15
11
12
入出力部
制御部
記憶部
操作部
13
1

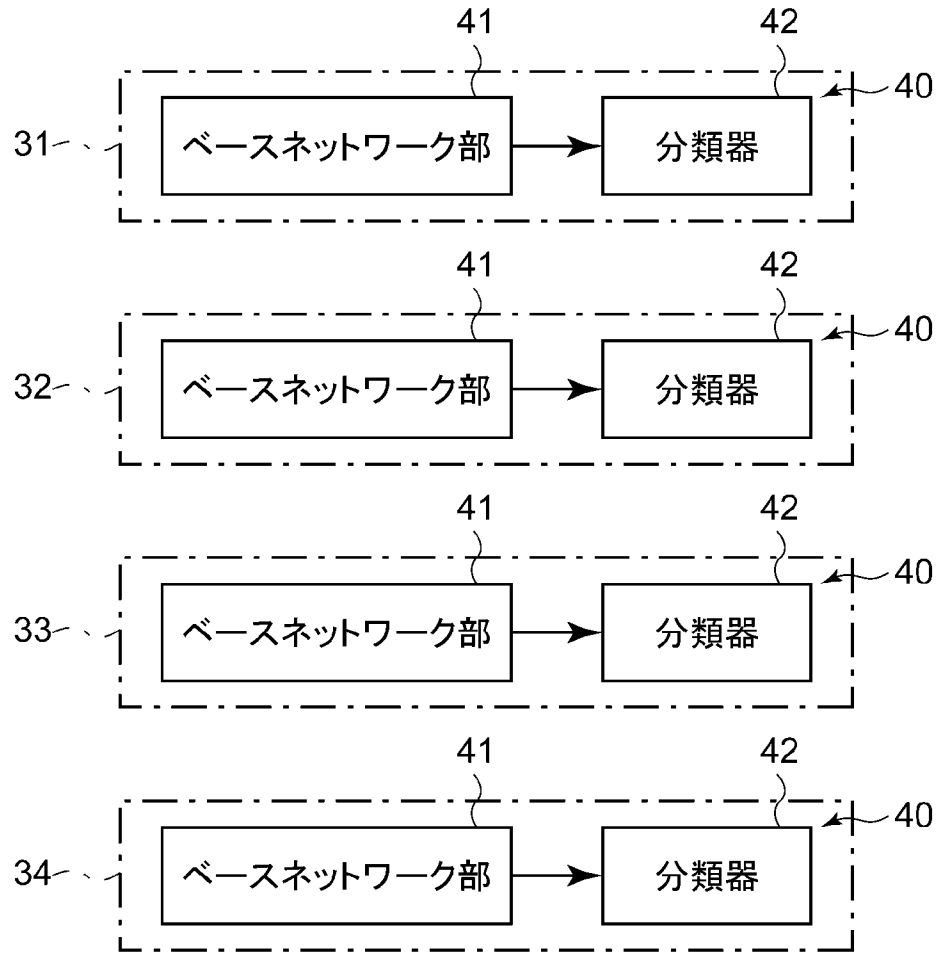
[図2]


受入部
21
31
32
33
34
フレーム検出部
顔検出部
身体検出部
文字検出部
22
検出情報生成部
23
情報処理部
24

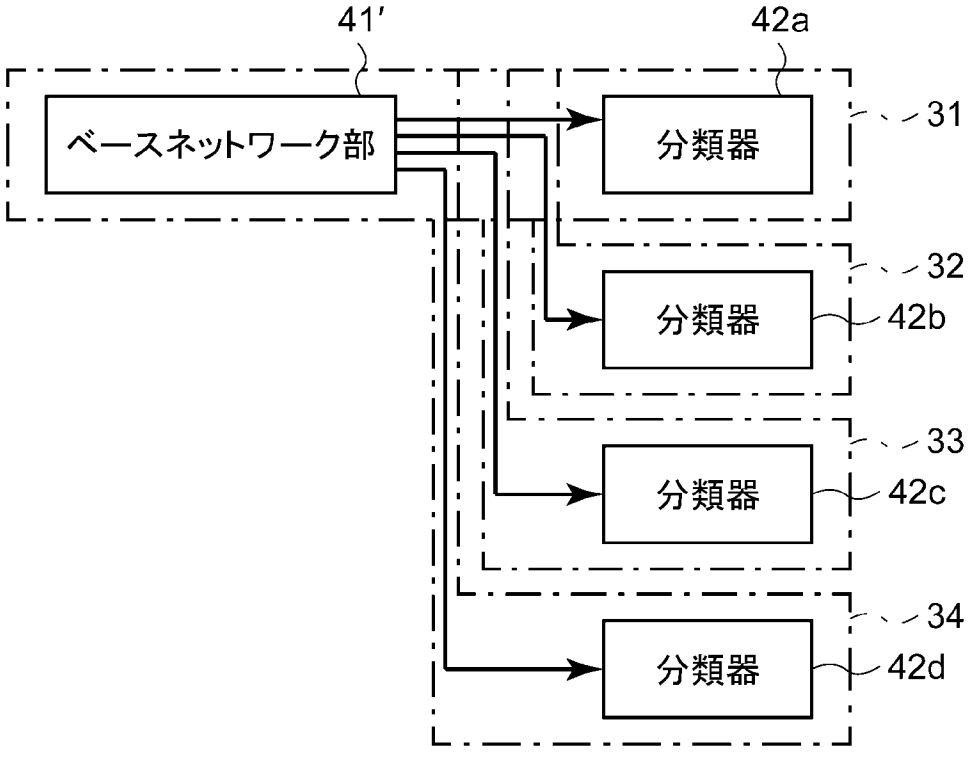
[図3]



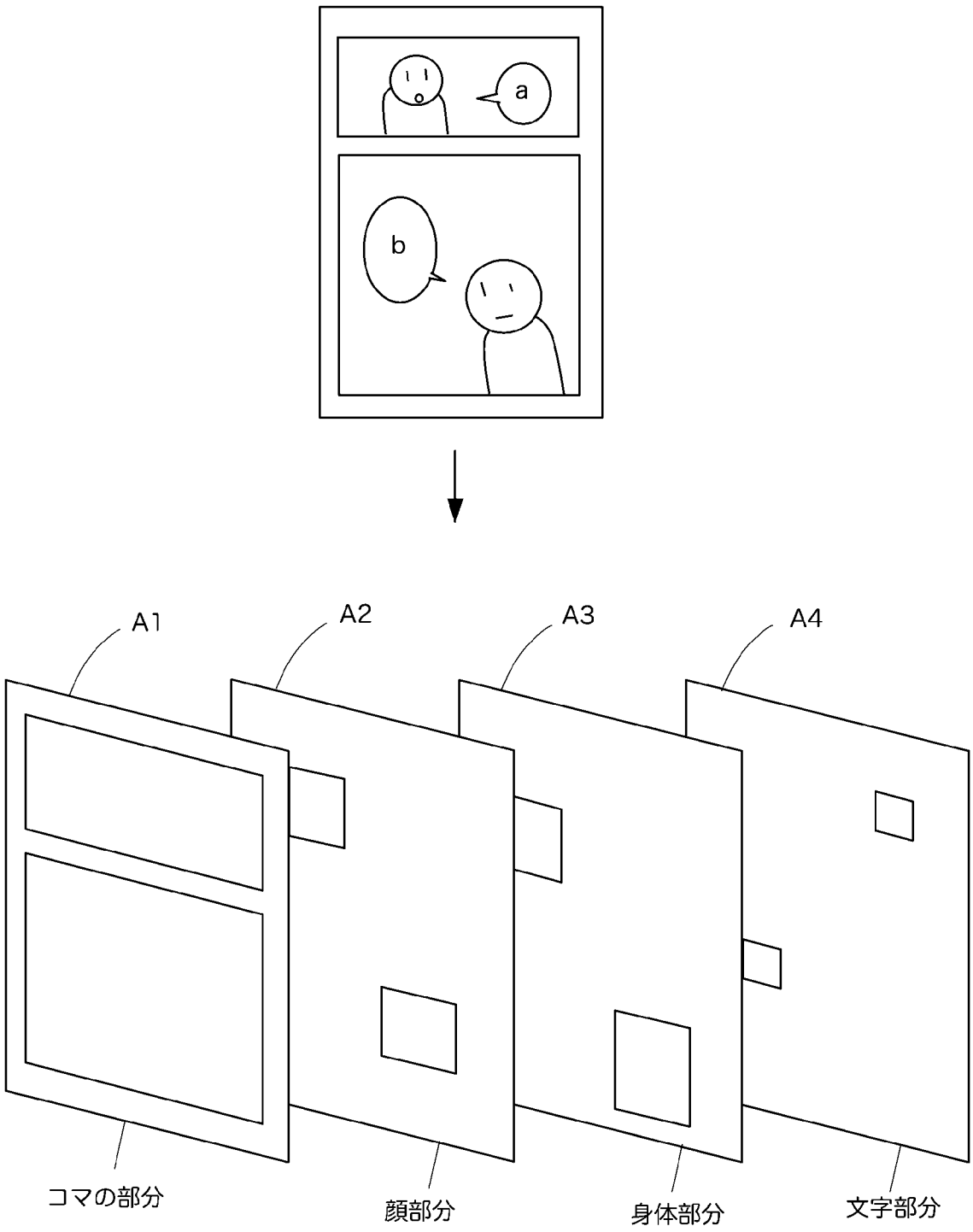
[図4]


41
42
40
31
ベースネットワーク部
分類器
41
42
40
32
ベースネットワーク部
分類器
41
42
40
33
ベースネットワーク部
分類器
41
42
40
34
ベースネットワーク部
分類器

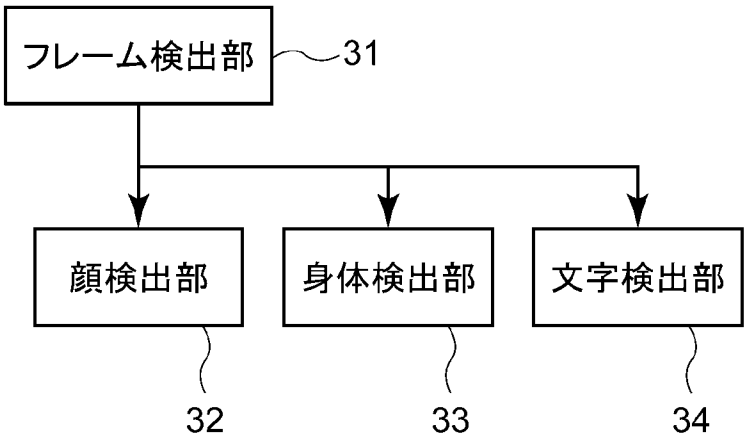
[図5]


41′
42a
ベースネットワーク部
分類器
31
32
分類器
42b
33
分類器
42c
34
分類器
42d

[図6]



[図7]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2018/032635

A. CLASSIFICATION OF SUBJECT MATTER
Int. Cl. G06T7/00(2017.01)i, G06T7/11(2017.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int. Cl. G06T7/00, G06T7/11

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2018
Registered utility model specifications of Japan 1996-2018
Published registered utility model applications of Japan 1994-2018

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	柳澤秀彰,渡辺裕, Faster R-CNN を用いたマンガの構造解析, 第31回 画像符号化シンポジウム 第21回 映像メディア処理シンポジウム, 30 November 2016 (intake date), pp. 80-81 (YANAGISAWA, Hideaki, WATANABE, Hiroshi. Structural Analysis of Comic Images using Faster R-CNN. PCSJ/IMPS 2016)	1-4, 6, 7 5
Y A	藤本東,小川徹,山本和慶,松井勇佑,山崎俊彦,相澤清晴, 学術用アノテーション付き漫画画像データセットの構築と解析, 電子情報通信学会技術研究報告, 15 March 2017 (intake date), vol. 116, no. 64, pp. 35-40 (FUJIMOTO, Azuma, OGAWA, Toru, YAMAMOTO, Kazuyoshi, MATSUI, Yusuke, YAMASAKI, Toshihiko, AIZAWA, Kiyoharu. Creation and Analysis of Academic Manga Image Dataset with Annotations. IEICE Technical Report)	1-4, 6, 7 5

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10.10.2018	23.10.2018
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2018/032635

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	JP 2011-238043 A (KDDI CORP.) 24 November 2011, paragraphs [0031], [0035] (Family: none)	3, 4, 6 1, 2, 5, 7
Y A	JP 2002-229766 A (EASTMAN KODAK CO.) 16 August 2002, paragraph [0023] & US 2002/0065088 A1, paragraph [0031] & EP 1211631 A1 & DE 60126896 T2 & FR 2817429 A1 & AT 355567 T	3, 4, 6 1, 2, 5, 7
Y A	福井宏,山下隆義,山内悠嗣,藤吉弘亘, Deep Learningを用いた歩行者検出の研究動向, 電子情報通信学会技術研究報告, 17 January 2017 (intake date), vol. 116, no. 366, pp. 37-46, non-official translation (FUKUI, Hiroshi, YAMASHITA, Takayoshi, YAMAUCHI, Yuji, FUJIYOSHI, Hironobu. Research Trends in Pedestrian Detection Using Deep Learning. IEICE Technical Report)	6 1-5, 7

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１８／０３２６３５

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G06T7/00(2017.01)i, G06T7/11(2017.01)i

Ｂ．　調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

Int.Cl.　G06T7/00, G06T7/11

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０１８年
日本国実用新案登録公報	１９９６－２０１８年
日本国登録実用新案公報	１９９４－２０１８年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

Ｃ．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
Ｙ Ａ	柳澤　秀彰、渡辺　裕, Ｆａｓｔｅｒ　Ｒ－ＣＮＮ　を用いたマンガの構造解析, 第３１回　画像符号化シンポジウム　第２１回　映像メディア処理シンポジウム，2016.11.30（受入日），p. ８０－８１	１－４，６，７ ５

☑　Ｃ欄の続きにも文献が列挙されている。

☐　パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー

「Ａ」特に関連のある文献ではなく、一般的技術水準を示すもの

「Ｅ」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「Ｌ」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）

「Ｏ」口頭による開示、使用、展示等に言及する文献

「Ｐ」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献

「Ｔ」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「Ｘ」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Ｙ」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「＆」同一パテントファミリー文献

国際調査を完了した日 １０．１０．２０１８	国際調査報告の発送日 ２３．１０．２０１８
国際調査機関の名称及びあて先 日本国特許庁（ＩＳＡ／ＪＰ） 郵便番号１００－８９１５ 東京都千代田区霞が関三丁目４番３号	特許庁審査官（権限のある職員）　５Ｈ　５２８４ 川▲崎▼　博章 電話番号　０３－３５８１－１１０１　内線　３５３１

Ｃ（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
Ｙ Ａ	藤本　東、小川　徹、山本　和慶、松井　勇佑、山崎　俊彦、相澤　清晴， 学術用アノテーション付き漫画画像データセットの構築と解析， 電子情報通信学会技術研究報告，2017.03.15（受入日）， Ｖｏｌ．１１６、Ｎｏ．４６４，ｐ．３５－４０	１－４，６，７ ５
Ｙ Ａ	JP 2011-238043 A（ＫＤＤＩ株式会社）2011.11.24， 段落［００３１］、［００３５］（ファミリーなし）	３，４，６ １，２，５，７
Ｙ Ａ	JP 2002-229766 A（イーストマン　コダック　カンパニー） 2002.08.16，段落［００２３］ & US 2002/0065088 A1，段落［００３１］ & EP 1211631 A1 & DE 60126896 T2 & FR 2817429 A1 & AT 355567 T	３，４，６ １，２，５，７
Ｙ Ａ	福井　宏、山下　隆義、山内　悠嗣、藤吉　弘亘， Ｄｅｅｐ　Ｌｅａｒｎｉｎｇを用いた歩行者検出の研究動向， 電子情報通信学会技術研究報告，2017.01.17（受入日）， Ｖｏｌ．１１６、Ｎｏ．３６６，ｐ．３７－４６	６ １－５，７