

(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2017-194806

(P2017-194806A)

(43) 公開日 平成29年10月26日(2017.10.26)

(51) Int.Cl.	F I	テーマコード (参考)
G06K 9/03 (2006.01)	G06K 9/03 C	5B064
G06K 9/66 (2006.01)	G06K 9/66	5B091
G06K 9/72 (2006.01)	G06K 9/72 Z	5B109
G06F 17/27 (2006.01)	G06F 17/27 655	
G06F 17/22 (2006.01)	G06F 17/22 617	

審査請求 有 請求項の数 9 O L (全 18 頁) 最終頁に続く

(21) 出願番号 特願2016-84081 (P2016-84081)  
 (22) 出願日 平成28年4月19日 (2016.4.19)  
 (11) 特許番号 特許第6057112号 (P6057112)  
 (45) 特許公報発行日 平成29年1月11日 (2017.1.11)

(71) 出願人 516092005  
 A I i n s i d e 株式会社  
 東京都渋谷区渋谷三丁目6番18号  
 (74) 代理人 100167667  
 弁理士 安高 史朗  
 (74) 代理人 100174436  
 弁理士 熊谷 大介  
 (72) 発明者 渡久地 択  
 東京都渋谷区渋谷三丁目6番18号 A I  
 i n s i d e 株式会社内  
 Fターム(参考) 5B064 AB03 DA10 DA27 EA08 EA18  
 EA39  
 5B091 CA02 CA05 CB09  
 5B109 LA01

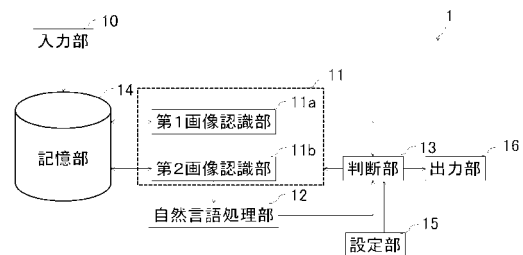
(54) 【発明の名称】 文字認識装置、方法およびプログラム

(57) 【要約】

【課題】 様々な手書き文字が含まれている書類について、高精度に文字認識を行うことができる文字認識装置、方法およびプログラムを提供すること。

【解決手段】 多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第1画像認識部11と、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第2画像認識部12と、第2画像認識部12により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理部13と、第1画像認識部11の出力と自然言語処理部13の出力とを対比する判断部14とを備える。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第 1 画像認識部と、

多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第 2 画像認識部と、

前記第 2 画像認識部により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理部と、

前記第 1 画像認識部により出力されたテキスト候補と、前記自然言語処理部により出力されたテキスト候補とを対比する判断部とを備え、

前記判断部は、所定以上の尤度のテキストを出力する文字認識装置。

**【請求項 2】**

前記判断部は、前記第 1 画像認識部により出力された最も尤度の高いテキスト候補と、前記自然言語処理部により出力された最も尤度の高いテキスト候補とを対比し、所定以上の尤もらしさを得られなかった場合、所定以上の尤もらしさが得られるまで、尤度の高い順に他の候補同士の対比を行う請求項 1 記載の文字認識装置。

**【請求項 3】**

前記判断部は、前記自然言語処理部により出力されたテキスト候補の中で尤度の高いテキスト候補を、前記第 1 画像認識部により出力されたテキスト候補よりも優先的に扱う請求項 1 または 2 記載の文字認識装置。

**【請求項 4】**

前記判断部は、対比した結果、2 つのテキスト候補の差分が所定の閾値を超えない場合に、所定以上の尤度のテキストであると判断して出力する請求項 1 から 3 のいずれか一項に記載の文字認識装置。

**【請求項 5】**

処理にかかる時間を設定する設定部を備え、

前記判断部は、前記第 1 画像認識部により出力された最も尤度の高いテキスト候補と、前記自然言語処理部により出力された最も尤度の高いテキスト候補とを対比し、所定以上の尤もらしさを得られなかった場合、前記設定部によって設定された時間以内において、所定以上の尤もらしさが得られるまで、尤度の高い順に他の候補同士の対比を行う請求項 1 から 4 のいずれか一項に記載の文字認識装置。

**【請求項 6】**

前記判断部は、所定以上の尤もらしさが得られなかったテキストを伏字にして出力する請求項 1 から 5 のいずれか一項に記載の文字認識装置。

**【請求項 7】**

前記第 1 画像認識部は、

項目ごとに適した多層のニューラルネットワークを機械学習により有しており、

前記画像データに含まれる項目を探索し、当該項目に適した多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力し、

前記第 2 画像認識部は、

項目ごとに適した多層のニューラルネットワークを機械学習により有しており、

前記画像データに含まれる項目を探索し、当該項目に適した多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する請求項 1 記載の文字認識装置。

**【請求項 8】**

多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキス

10

20

30

40

50

ト候補と尤度を出力する第1画像認識工程と、

多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第2画像認識工程と、

前記第2画像認識工程により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理工程と、

前記第1画像認識工程により出力されたテキスト候補と、前記自然言語処理工程により出力されたテキスト候補とを対比する判断工程とを備え、

10

前記判断工程は、所定以上の尤度のテキストを出力する文字認識方法。

【請求項9】

多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第1画像認識工程と、

多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第2画像認識工程と、

前記第2画像認識工程により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理工程と、

20

前記第1画像認識工程により出力されたテキスト候補と、前記自然言語処理工程により出力されたテキスト候補とを対比する判断工程と、をコンピュータによって実現するための文字認識プログラムであって、

前記判断工程は、所定以上の尤度のテキストを出力する文字認識プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像データの文字認識を行う文字認識装置、方法およびプログラムに関する。

30

【背景技術】

【0002】

手書きで文字が記入された書類をイメージスキャナなどで読み取り、OCR(Optical Character Recognition)処理を行うことにより、入力情報を所定の文字コードに変換したデジタルデータを生成する技術が普及している。

【0003】

例えば、特許文献1によれば、機械学習により文字認識を行う文字識別システムが開示されている。文字認識システムは、見本文字画像の入力を受け付ける文字画像入力受付部と、見本文字画像に基づいて文字部品を抽出する文字部品抽出部と、文字部品に基づいて擬似文字モデルを生成する擬似文字モデル生成部と、擬似文字モデルに基づいて文字識別パターンを生成して識別辞書を生成する識別辞書生成部と、により構成されている。

40

【0004】

また、特許文献2によれば、文字認識後に形態素解析をして、品詞尤度と文字類似度の両方が含まれる特徴量ベクトルを用いて、尤もらしさを判定する情報処理装置が開示されている。情報処理装置は、文字認識結果に対して形態素解析を行う形態素解析手段と、文字認識結果の文字について、形態素解析手段による形態素解析結果である対象の文字が属する単語の品詞らしさから作成するP種の品詞それぞれの品詞尤度と、該文字認識結果の各文字の文字類似度によって構成されるP+1個の要素の特徴量ベクトルを作成する特徴量ベクトル作成手段と、特徴量ベクトル作成手段によって作成された特徴量ベクトルから、文字認識結果の各文字の確信度を算出する確信度算出手段とにより構成されている。

50

## 【先行技術文献】

## 【特許文献】

## 【0005】

【特許文献1】特開2015-069256号公報

【特許文献2】特開2014-120059号公報

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0006】

上述した特許文献1, 2によっても、様々な手書き文字（例えば、達筆な手書き文字や薄くて雑な手書き文字など）の認識を高精度に行うことは困難であり、さらに高精度に文字認識を行いたい要望がある。

10

## 【0007】

本発明では、様々な手書き文字が含まれている書類について、高精度に文字認識を行うことができる文字認識装置、方法およびプログラム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】

## 【0008】

上記目的を達成するために、本発明の一態様における文字認識装置は、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第1画像認識部と、多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第2画像認識部と、前記第2画像認識部により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理部と、前記第1画像認識部により出力されたテキスト候補と、前記自然言語処理部により出力されたテキスト候補とを対比する判断部とを備え、前記判断部は、所定以上の尤度のテキストを出力する。

20

## 【0009】

また、本発明の一態様における文字認識装置では、前記第1画像認識部により出力された最も尤度の高いテキスト候補と、前記自然言語処理部により出力された最も尤度の高いテキスト候補とを対比し、所定以上の尤もらしさを得られなかった場合、所定以上の尤もらしさが得られるまで、尤度の高い順に他の候補同士の対比を行う構成でもよい。

30

## 【0010】

また、本発明の一態様における文字認識装置では、前記判断部は、前記自然言語処理部により出力されたテキスト候補の中で尤度の高いテキスト候補を、前記第1画像認識部により出力されたテキスト候補よりも優先的に扱う構成でもよい。

## 【0011】

また、本発明の一態様における文字認識装置では、前記判断部は、対比した結果、2つのテキスト候補の差分が所定の閾値を超えない場合に、所定以上の尤度のテキストであると判断して出力する構成でもよい。

## 【0012】

また、本発明の一態様における文字認識装置では、処理にかかる時間を設定する設定部を備え、前記判断部は、前記第1画像認識部により出力された最も尤度の高いテキスト候補と、前記自然言語処理部により出力された最も尤度の高いテキスト候補とを対比し、所定以上の尤もらしさを得られなかった場合、前記設定部によって設定された時間以内において、所定以上の尤もらしさが得られるまで、尤度の高い順に他の候補同士の対比を行う構成でもよい。

40

## 【0013】

また、本発明の一態様における文字認識装置では、前記判断部は、所定以上の尤もらしさが得られなかったテキストを伏字にして出力する構成でもよい。

## 【0014】

50

また、本発明の一態様における文字認識装置では、前記第1画像認識部は、項目ごとに適した多層のニューラルネットワークを機械学習により有しており、前記画像データに含まれる項目を探索し、当該項目に適した多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力し、前記第2画像認識部は、項目ごとに適した多層のニューラルネットワークを機械学習により有しており、前記画像データに含まれる項目を探索し、当該項目に適した多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する構成でもよい。

【0015】

上記目的を達成するために、本発明の一態様における文字認識方法は、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第1画像認識工程と、多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第2画像認識工程と、前記第2画像認識工程により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理工程と、前記第1画像認識工程により出力されたテキスト候補と、前記自然言語処理工程により出力されたテキスト候補とを対比する判断工程とを備え、前記判断工程は、所定以上の尤度のテキストを出力する。

【0016】

上記目的を達成するために、本発明の一態様における文字認識プログラムは、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第1画像認識工程と、多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第2画像認識工程と、前記第2画像認識工程により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理工程と、前記第1画像認識工程により出力されたテキスト候補と、前記自然言語処理工程により出力されたテキスト候補とを対比する判断工程と、をコンピュータによって実現するための文字認識プログラム

【発明の効果】

【0017】

本発明によれば、高精度に文字認識を行うことができる。

【図面の簡単な説明】

【0018】

【図1】文字認識装置の構成を示すブロック図である。

【図2】画像認識部による特徴抽出とベクトル変換の様子を模式的に示す図である。

【図3】画像認識部による字種の判定の様子を模式的に示す図である。

【図4】第1画像認識部と第2画像認識部による画像認識についての説明に供する図である。

【図5】文字認識装置の動作の流れについての説明に供するフローチャートである。

【発明を実施するための形態】

【0019】

以下、本発明の実施形態に係る文字認識装置、方法およびプログラムについて図面を参照しながら説明する。なお、実施形態を説明する全図において、共通の構成要素には同一の符号を付し、繰り返しの説明を省略する。

【0020】

以下では、一例として、手書き文字が含まれた帳票やアンケート用紙などの書類をスキ

10

20

30

40

50

ャナ等で画像化し、画像化した画像データの文字を認識する文字認識装置の構成と動作について説明する。なお、手書き文字が含まれていない書類、いわゆる、書体データを利用してプリンタによって文字が印刷された書類であっても、印刷されている文字が掠れていたり、または、滲んでいたりすると、文字の認識率が低下する。本実施形態にかかる文字認識装置はこのような書類をスキャナ等で画像化し、画像化した画像データの文字の認識に適用されてもよい。

【0021】

文字認識装置1は、概念的には、バックプロパゲーション（誤差逆伝播法）によって、入力層、一または複数の中間層、および出力層から構成される多層のニューラルネットワークを学習させるアルゴリズムを利用して、文字認識を行う。

10

【0022】

具体的には、文字認識装置1は、図1に示すように、画像データを入力する入力部10と、画像データを認識してテキストを生成する画像認識部11と、画像認識部11により生成されたテキストを自然言語処理する自然言語処理部12と、画像認識部11により生成されたテキストと自然言語処理部12により自然言語処理されたテキストを対比する判断部13とを備える。また、文字認識装置1は、判断部13から出力されるテキストを出力する出力部16を備える。出力部16は、テキストを項目に分けてCSVデータとして出力してもよい。また、文字認識装置1は、画像認識部11、自然言語処理部12および判断部13とを独立に機能させ、それぞれの出力結果が相互に出力結果に影響を与えることにより、所定以上の尤度のテキストを出力する。

20

【0023】

入力部10は、例えば、スキャナ装置によって構成されており、書類を画像化して画像データを生成し、生成した画像データを記憶部14に入力する。

【0024】

ここで、画像認識部11の動作について説明する。画像認識部11は、画像データに基づいて、例えば、罫線抽出、枠構造解析、読取対象枠の位置推定などの文書構造解析を行う。次に、画像認識部11は、文書構造解析の結果を受けて、読取対象である文字行を抽出する。次に、画像認識部11は、文字行画像から文字パターン候補の切出しと、各文字パターンの文字識別を行う。

30

【0025】

つぎに、文字識別の手順について説明する。画像認識部11は、図2に示すように、切出した1個の文字パターンの画像データに対して特徴抽出を行う。そして、画像認識部11は、文字のストロークの方向成分などを抽出して、画像データを1つのベクトルに変換する。図2に示す例では、画像データXが多層のニューラルネットワークに入力され、方向や位置等の特徴を捉えて特徴抽出をされている様子を模式的に示している。また、図2に示す例では、ベクトル $X_1$ と、ベクトル $X_2$ と、ベクトル $X_3$ とに変換された様子を模式的に示している。

【0026】

画像認識部11は、図3に示すように、変換されたベクトルに基づいて、字種が何であるかを判定する。画像認識部11は、当該判定において、事前に大量のパターンを使った分布の様子から、どの字種が特徴空間上のどの辺に分布しているかを保持している辞書データを参照し、未知の入力パターンである画像データの候補を決定する。図3に示す例では、辞書データにおいて、字種「中」、字種「申」および字種「十」の情報が記憶されている様子を概念的に示している。

40

【0027】

画像認識部11は、以上のプロセスにより、複数のテキスト候補（例えば、中、申、十）と、各テキスト候補の尤度を取得する。なお、各テキスト候補の尤度は、特徴空間内における各候補の中心と、未知の入力パターンである画像データとの距離で算出することができる。

【0028】

50

画像認識部 1 1 は、図 1 に示すように、第 1 画像認識部 1 1 a と、第 2 画像認識部 1 1 b とから構成されている。

【 0 0 2 9 】

第 1 画像認識部 1 1 a は、記憶部 1 5 から画像データを読み出し、多層のニューラルネットワークを用いて当該画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する。

【 0 0 3 0 】

第 2 画像認識部 1 1 b は、記憶部 1 5 から画像データを読み出し、多層のニューラルネットワークを用いて当該画像データから特徴点を抽出し、テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する。

10

【 0 0 3 1 】

第 1 画像認識部 1 1 a と第 2 画像認識部 1 1 b との主な違いは、画像認識の機械学習を行う際に利用する学習データの違いである。第 1 画像認識部 1 1 a は 1 文字を 1 文字として出力するように学習データが用意されているのに対し、第 2 画像認識部 1 1 b は 1 文字をより小さい因子である素因子テキストに分離して出力するように学習データが用意されている。

【 0 0 3 2 】

ここで、第 1 画像認識部 1 1 a による画像認識と第 2 画像認識部 1 1 b による画像認識の具体例について説明する。以下では、第 1 画像認識部 1 1 a および第 2 画像認識部 1 1 b、例えば、図 4 ( a ) に示すように、画像データ A 1 について画像認識を行う場合について説明する。

20

【 0 0 3 3 】

第 1 画像認識部 1 1 a は、画像データ A 1 を分割する処理を行う。本実施例では、分割する処理により、画像データ A 1 は、図 4 ( b ) に示すように、4 つの画像データ a 1 , a 2 , a 3 , a 4 に分割できたものとする。

【 0 0 3 4 】

第 1 画像認識部 1 1 a は、多層のニューラルネットワークを用いて画像データ a 1 から特徴点を抽出し、複数の候補 (例えば、「高」, 「喬」, 「富」, 「畜」等) を生成し、各候補の尤度を計算する。

【 0 0 3 5 】

第 1 画像認識部 1 1 a は、多層のニューラルネットワークを用いて画像データ a 2 から特徴点を抽出し、複数の候補 (例えば、「校」, 「核」, 「梓」, 「検」等) を生成し、各候補の尤度を計算する。

30

【 0 0 3 6 】

第 1 画像認識部 1 1 a は、多層のニューラルネットワークを用いて画像データ a 3 から特徴点を抽出し、複数の候補 (例えば、「時」, 「暁」, 「待」, 「晤」等) を生成し、各候補の尤度を計算する。

【 0 0 3 7 】

第 1 画像認識部 1 1 a は、多層のニューラルネットワークを用いて画像データ a 4 から特徴点を抽出し、複数の候補 (例えば、「代」, 「付」, 「何」, 「仕」等) を生成し、各候補の尤度を計算する。

40

【 0 0 3 8 】

第 1 画像認識部 1 1 a は、各候補の尤度に基づいて、例えば、「高校時代」, 「喬核暁付」等をテキスト候補として判断部 1 3 に出力する。

【 0 0 3 9 】

第 2 画像認識部 1 1 b は、第 1 画像認識部 1 1 a よりも小さい単位である素因子テキストを生成するように画像データ A 1 を分割する処理を行う。素因子テキストとは、1 文字をより小さい因子で表したテキストである。例えば、「校」という画像データから小さい因子で表した「木」と「交」が素因子テキストである。本実施例では、分割する処理により、画像データ A 1 は、図 4 ( c ) に示すように、6 つの画像データ b 1 , b 2 , b 3 ,

50

b 4 , b 5 , b 6 に分割できたものとする。

【 0 0 4 0 】

第 2 画像認識部 1 1 b は、多層のニューラルネットワークを用いて画像データ b 1 から特徴点を抽出し、複数の素因子テキストの候補（例えば、「高」、「喬」、「富」、「畜」等）を生成し、各候補の尤度を計算する。

【 0 0 4 1 】

第 2 画像認識部 1 1 b は、多層のニューラルネットワークを用いて画像データ b 2 から特徴点を抽出し、複数の素因子テキストの候補（例えば、「木」、「不」、「六」、「禾」等）を生成し、各候補の尤度を計算する。

【 0 0 4 2 】

第 2 画像認識部 1 1 b は、多層のニューラルネットワークを用いて画像データ b 3 から特徴点を抽出し、複数の素因子テキストの候補（例えば、「交」、「定」、「氣」、「充」等）を生成し、各候補の尤度を計算する。

【 0 0 4 3 】

第 2 画像認識部 1 1 b は、多層のニューラルネットワークを用いて画像データ b 4 から特徴点を抽出し、複数の素因子テキストの候補（例えば、「日」、「日」、「月」等）を生成し、各候補の尤度を計算する。

【 0 0 4 4 】

第 2 画像認識部 1 1 b は、多層のニューラルネットワークを用いて画像データ b 5 から特徴点を抽出し、複数の素因子テキストの候補（例えば、「寺」、「圭」、「茉」、「苦」等）を生成し、各候補の尤度を計算する。

【 0 0 4 5 】

第 2 画像認識部 1 1 b は、多層のニューラルネットワークを用いて画像データ b 6 から特徴点を抽出し、複数の素因子テキストの候補（例えば、「代」、「付」、「何」、「仕」等）を生成し、各候補の尤度を計算する。

【 0 0 4 6 】

第 2 画像認識部 1 1 b は、各素因子テキストの尤度に基づいて、例えば、「高木交日寺代」、「喬不定日圭付」等をテキスト候補として判断部 1 3 に出力する。

【 0 0 4 7 】

自然言語処理部 1 3 は、第 2 画像認識部 1 1 b により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する。

【 0 0 4 8 】

具体的には、自然言語処理部 1 3 は、「高木交日寺代」について、「高木」「交日寺代」に分離したり、「高」「木交」「日寺代」に分離したりして、それぞれに対して形態素解析を行って、それぞれの尤度を計算する。

【 0 0 4 9 】

また、自然言語処理部 1 3 は、「高木」「交日寺代」について、「日」と「寺」を接合して「時」にし、「高木」「交時代」にして、形態素解析を行って、尤度を計算する。

【 0 0 5 0 】

また、自然言語処理部 1 3 は、「高」「木交」「日寺代」について、「木」と「交」を接合して「校」にし、前のテキスト「高」と組み合わせて「高校」にし、「高校」「日寺代」にして、形態素解析を行って、尤度を計算する。

【 0 0 5 1 】

また、自然言語処理部 1 3 は、「高校」「日寺代」について、「日」と「寺」を接合して「時」にし、「高校」「時代」にして、形態素解析を行って、尤度を計算する。

【 0 0 5 2 】

このようにして、自然言語処理部 1 3 は、第 2 画像認識部 1 1 b から出力されたテキスト候補を、全てのパターンで接合、切り離しの処理を行う。例えば、「私の高木交日寺代

10

20

30

40

50

は」というテキストについて形態素解析を行うと、「主語」「格助詞」「名詞(苗字)」「名詞」「名詞」「係助詞」となり、自然言語的な観点から不当な並びとなるため、当該テキストに対しては、尤度が低いという評価を行う。一方、「私の高校時代は」というテキストについて形態素解析を行うと、「主語」「格助詞」「名詞」「係助詞」となり、自然言語的な観点から適当な並びとなるため、当該テキストに対しては、尤度が高いという評価を行う。素因子テキストに分離し、全てのパターンで組み合わせたテキスト候補に対して形態素解析による評価を行うことで、画像認識のエラーによる誤りを減らすことができる。

**【0053】**

判断部14は、第1画像認識部11aにより出力されたテキスト候補(例えば、「高校時代」と)と、自然言語処理部13により出力されたテキスト候補(例えば、「高木交日寺代」と)とを対比する。

10

**【0054】**

判断部14は、対比した結果に基づいて、所定以上の尤度のテキスト(例えば、「高校時代」と)を出力する。

**【0055】**

画像認識を重視した第1画像認識部11aと、自然言語を重視した第2画像認識部11bという二つの異なる観点のテキストを対比することで、文字認識装置1は、より高精度に文字認識を行うことができる。

**【0056】**

また、判断部14は、第1画像認識部11aにより出力された最も尤度の高いテキスト候補と、自然言語処理部13により出力された最も尤度の高いテキスト候補とを対比し、所定以上の尤もらしさ(例えば、一致率が98パーセント以上)を得られなかった場合、所定以上の尤もらしさが得られるまで、尤度の高い順に他の候補同士の対比を行う構成でもよい。

20

**【0057】**

判断部14は、第1画像認識部11aと自然言語処理部13で得られたすべての出力パターンの中から、どの回答が尤もらしいかの重みを画像認識部11にフィードバックする。また、判断部14は、出力パターンには存在しない「Xという文字である可能性」をフィードバックする。画像認識部11は、当該フィードバックを受けて、再度重み付けに基づく画像認識を行う。

30

**【0058】**

例えば、判断部14は、第1画像認識部11aにより出力された最も尤度の高いテキスト候補である「喬校時付」と、自然言語処理部13により出力された最も尤度の高いテキスト候補である「高木交時代」とを対比し、差分が大きく、所定以上の尤もらしさを得られなかった場合、第1画像認識部11aにより出力された次に尤度の高いテキスト候補である「高校時代」と、自然言語処理部13により出力された最も尤度の高いテキスト候補である「高校時代」とを対比し、所定以上の尤もらしさが得られた場合、「高校時代」を出力する。

**【0059】**

よって、文字認識装置1は、様々な手書き文字が含まれている書類について、所定以上の尤もらしさが得られるまで処理を繰り返すので、信頼性の高い高精度な文字認識を行うことができる。

40

**【0060】**

また、判断部14は、自然言語処理部13により出力されたテキスト候補の中で、尤度の高いテキスト候補を第1画像認識部11aにより出力されたテキスト候補よりも優先的に扱う構成でもよい。

**【0061】**

当該構成の場合には、文字認識装置1は、より自然言語処理部13による処理を優先するので、自然言語処理に適しているといえ、文章の文字認識に対して強みを発揮する。

50

## 【0062】

また、判断部14は、対比した結果、2つのテキスト候補の差分が所定の閾値を超えない場合に、所定以上の尤度のテキストであると判断して出力する構成でもよい。つまり、判断部14は、自然言語処理部13により出力されたテキスト候補と、第1画像認識部11aにより出力されたテキスト候補とがほぼ一致している場合に、所定以上の尤度のテキストであると判断する。

## 【0063】

当該構成の場合には、文字認識装置1は、自然言語処理部13と第1画像認識部11aの二つの処理の結果を利用して文字認識を行っているので、高精度に文字認識を行うことができる。

10

## 【0064】

文字認識装置1は、図1に示すように、処理にかかる時間を設定する設定部15を備える構成でもよい。設定部15は、例えば、ユーザの指示にしたがって、処理時間を10分などに設定する。

## 【0065】

判断部14は、第1画像認識部11aにより出力された最も尤度の高いテキスト候補と、自然言語処理部13により出力された最も尤度の高いテキスト候補とを対比し、所定以上の尤もらしさを得られなかった場合、設定部15によって設定された時間以内において、所定以上の尤もらしさが得られるまで、尤度の高い順に他の候補同士の対比を行う。

20

## 【0066】

よって、文字認識装置1は、設定された時間内において、所定以上の尤もらしさが得られるまで処理を繰り返すので、認識処理を何回行っても所定以上の尤もらしさが得られないような認識困難な画像データに対して何度も試行を繰り返すことがないメリットがある。

## 【0067】

また、判断部14は、所定以上の尤もらしさが得られなかったテキストを伏字にして出力する構成でもよい。

## 【0068】

伏字とは、文字認識ができなかった箇所を示すものであり、例えば、「    」や「    」などである。具体的には、「高校時代」の「時」の箇所に対して所定以上の尤もらしさが得られなかった場合には、判断部14は、「高校      代」を出力する。

30

## 【0069】

よって、文字認識装置1は、文字識別できた箇所と文字識別ができなかった箇所を明示して出力することができる。なお、伏字にした箇所について、正しい文字（本実施例では、「    」の箇所は「時」である）を文字認識装置1にフィードバックしてもよい。当該フィードバックにより、文字認識装置1は、次回の文字認識において、前回「    」となった画像データを正しいテキスト「校」として出力することができる。

## 【0070】

また、第2画像認識部11bは、項目ごとに適した多層のニューラルネットワークを機械学習により有しており、画像データに含まれる項目を探索し、当該項目に適した多層のニューラルネットワークを用いて画像データから特徴点を抽出し、テキストよりも小さい因子である素因子テキストに分離されたテキスト候補を出力する構成でもよい。

40

## 【0071】

例えば、「申込日」のような項目には、「2015年5月1日」等の申込日に関する情報が入力されることが予想できる。つまり、「申込日」のような項目は、数字「0～9」と、漢字「年」、「月」、「日」が入力され、他の文字は入力されない。よって、第2画像認識部11bは、認識する項目が「申込日」のような場合には、数字「0～9」と、漢字「年」、「月」、「日」とを出力するようなニューラルネットワークを用いて画像データから特徴点を抽出し、テキストよりも小さい因子である素因子テキストに分離されたテキスト候補を出力する。

50

## 【 0 0 7 2 】

また、「氏名\_\_フリガナ」のような項目には、「トッキョタロウ」等の氏名のカタカナに関する情報が入力されることが予想できる。つまり、「氏名\_\_フリガナ」のような項目は、カタカナ「ア～ン」等が入力され、他の文字（漢字、数字等）は入力されない。よって、第2画像認識部11bは、認識する項目が「氏名\_\_フリガナ」のような場合には、カタカナを出力するようなニューラルネットワークを用いて画像データから特徴点を抽出し、テキストよりも小さい因子である素因子テキストに分離されたテキスト候補を出力する。

## 【 0 0 7 3 】

また、「電話番号（TEL）」のような項目には、「03-3581-1111」等の電話番号に関する情報が入力されることが予想できる。つまり、「電話番号（TEL）」のような項目は、数字「0～9」とハイフン「-」が入力され、他の文字（漢字、ひらがな等）は入力されない。よって、第2画像認識部11bは、認識する項目が「電話番号（TEL）」のような場合には、数字「0～9」とハイフン「-」を出力するようなニューラルネットワークを用いて画像データから特徴点を抽出し、テキストよりも小さい因子である素因子テキストに分離されたテキスト候補を出力する。

10

## 【 0 0 7 4 】

よって、文字認識装置1は、項目ごとに適した多層のニューラルネットワークを利用して画像データからテキスト候補を出力するので、効率的に高精度に文字認識を行うことができる。

20

## 【 0 0 7 5 】

ここで、文字認識装置1の動作の流れについて、図5に示すフローチャートを参照しながら説明する。

## 【 0 0 7 6 】

ステップS1において、第1画像認識部11aは、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する。

## 【 0 0 7 7 】

ステップS2において、第2画像認識部11bは、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、テキストよりも小さい因子である素因子テキストに分離されたテキスト候補を出力する。

30

## 【 0 0 7 8 】

ステップS3において、自然言語処理部12は、ステップS2の工程により出力されたテキスト候補に対して、隣接する素因子テキストを接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する。

## 【 0 0 7 9 】

ステップS4において、判断部13は、ステップS1の工程により出力されたテキスト候補と、ステップS3の工程により出力されたテキスト候補とを対比する。本工程において、判断部13は、所定以上の尤度（例えば、一致率が98パーセント以上）のテキストを出力する。

40

## 【 0 0 8 0 】

よって、文字認識装置1は、様々な手書き文字が含まれている書類について、高精度に文字認識を行うことができる。

## 【 0 0 8 1 】

また、本実施例では、主に、様々な手書き文字が含まれている書類について、高精度に文字認識を行うことができる文字認識装置1の構成と動作について説明したが、これに限られず、各構成要素を備え、様々な手書き文字が含まれている書類について、高精度に文字認識を行うための方法、およびプログラムとして構成されてもよい。

## 【 0 0 8 2 】

さらに、文字認識装置1を構成する各機能を実現するためのプログラムをコンピュータ

50

で読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、指示することによって実現してもよい。

【0083】

具体的には、当該プログラムは、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第1画像認識工程と、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第2画像認識工程と、第2画像認識工程により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストに形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理工程と、第1画像認識工程により出力されたテキスト候補と、自然言語処理工程により出力されたテキスト候補とを対比する判断工程と、をコンピュータによって実現するためのプログラムである。また、判断工程は、所定以上の尤度のテキストを出力する。

10

【0084】

さらに、ここでいう「コンピュータシステム」とは、OSや周辺機器等のハードウェアを含むものとする。また、「コンピュータで読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。

【0085】

さらに「コンピュータで読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時刻の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時刻プログラムを保持しているものも含んでもよい。また、上記プログラムは、前述した機能の一部を実現するためのものであってもよく、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであってもよい。

20

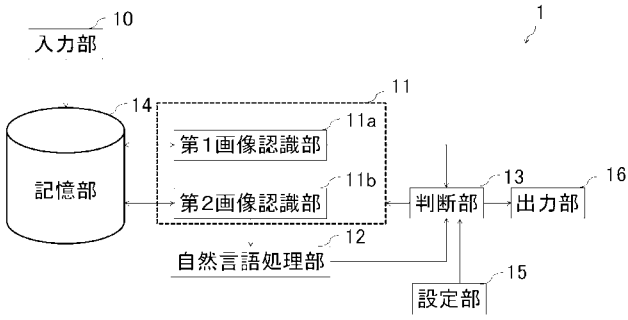
【符号の説明】

【0086】

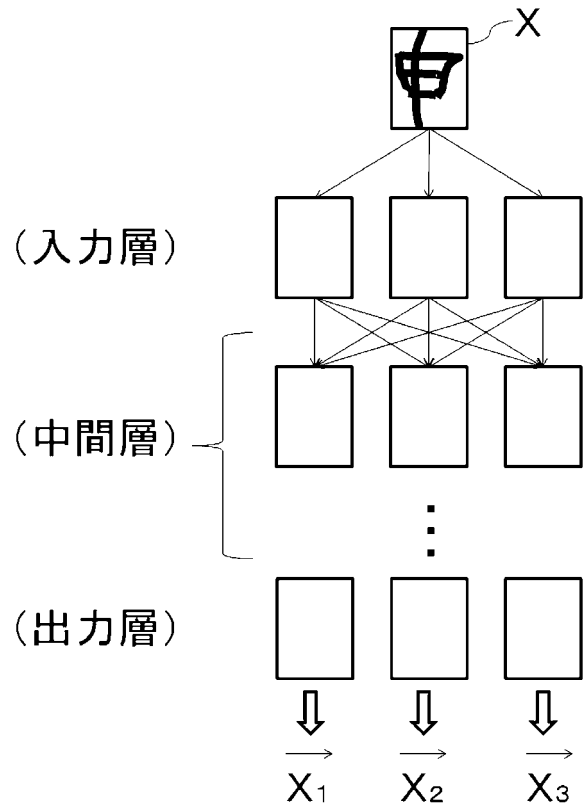
1 文字認識装置、10 入力部、11 画像認識部、11a 第1画像認識部、11b 第2画像認識部、12 自然言語処理部、13 判断部、14 記憶部、15 設定部、16 出力部。

30

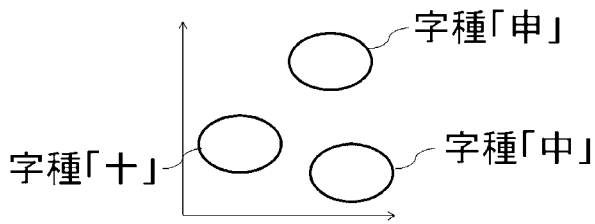
【 図 1 】



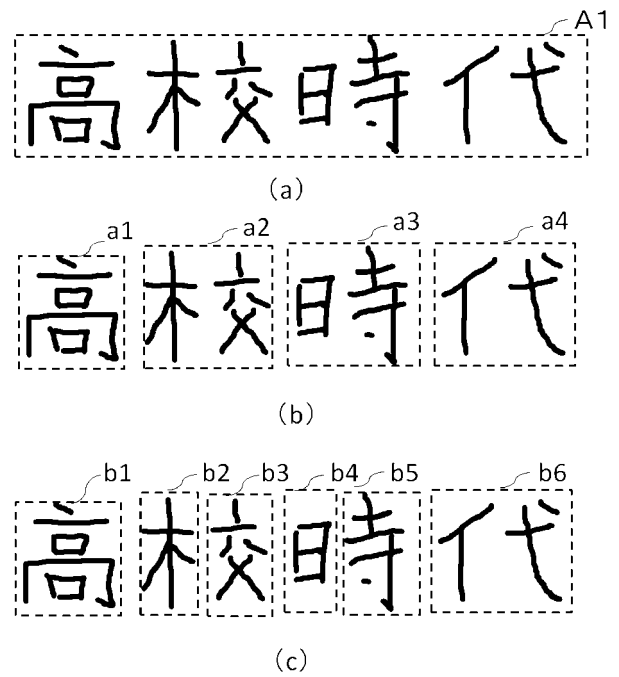
【 図 2 】



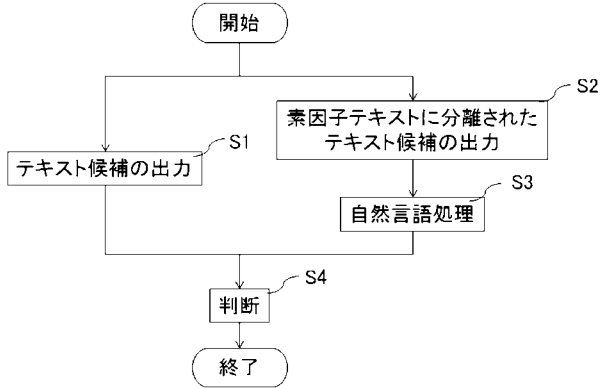
【 図 3 】



【 図 4 】



【図 5】



## 【手続補正書】

【提出日】平成28年9月15日(2016.9.15)

## 【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

## 【特許請求の範囲】

## 【請求項 1】

多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第 1 画像認識部と、

多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第 2 画像認識部と、

前記第 2 画像認識部により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストにそれぞれ形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理部と、

前記第 1 画像認識部により出力されたテキスト候補と、前記自然言語処理部により出力されたテキスト候補とを対比する判断部とを備え、

前記判断部は、所定以上の尤度のテキストを出力する文字認識装置。

## 【請求項 2】

前記判断部は、前記第 1 画像認識部により出力された最も尤度の高いテキスト候補と、前記自然言語処理部により出力された最も尤度の高いテキスト候補とを対比し、所定以上の尤もらしさを得られなかった場合、所定以上の尤もらしさが得られるまで、尤度の高い

順に他の候補同士の対比を行う請求項 1 記載の文字認識装置。

【請求項 3】

前記判断部は、前記自然言語処理部により出力されたテキスト候補の中で尤度の高いテキスト候補を、前記第 1 画像認識部により出力されたテキスト候補よりも優先的に扱う請求項 1 または 2 記載の文字認識装置。

【請求項 4】

前記判断部は、対比した結果、2 つのテキスト候補の差分が所定の閾値を超えない場合に、所定以上の尤度のテキストであると判断して出力する請求項 1 から 3 のいずれか一項に記載の文字認識装置。

【請求項 5】

処理にかかる時間を設定する設定部を備え、

前記判断部は、前記第 1 画像認識部により出力された最も尤度の高いテキスト候補と、前記自然言語処理部により出力された最も尤度の高いテキスト候補とを対比し、所定以上の尤もらしさを得られなかった場合、前記設定部によって設定された時間以内において、所定以上の尤もらしさが得られるまで、尤度の高い順に他の候補同士の対比を行う請求項 1 から 4 のいずれか一項に記載の文字認識装置。

【請求項 6】

前記判断部は、所定以上の尤もらしさが得られなかったテキストを伏字にして出力する請求項 1 から 5 のいずれか一項に記載の文字認識装置。

【請求項 7】

前記第 1 画像認識部は、

項目ごとに適した多層のニューラルネットワークを機械学習により有しており、

前記画像データに含まれる項目を探索し、当該項目に適した多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力し、

前記第 2 画像認識部は、

項目ごとに適した多層のニューラルネットワークを機械学習により有しており、

前記画像データに含まれる項目を探索し、当該項目に適した多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する請求項 1 記載の文字認識装置。

【請求項 8】

多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第 1 画像認識工程と、

多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第 2 画像認識工程と、

前記第 2 画像認識工程により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストにそれぞれ形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理工程と、

前記第 1 画像認識工程により出力されたテキスト候補と、前記自然言語処理工程により出力されたテキスト候補とを対比する判断工程とを備え、

前記判断工程は、所定以上の尤度のテキストを出力する文字認識方法。

【請求項 9】

多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第 1 画像認識工程と、

多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第 2 画像認識工程と、

前記第 2 画像認識工程により出力されたテキスト候補に対して、隣接する素因子テキス

トの接合および切り離しを行って、組み合わせられる複数のパターンのテキストにそれぞれ形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理工程と、

前記第1画像認識工程により出力されたテキスト候補と、前記自然言語処理工程により出力されたテキスト候補とを対比する判断工程と、をコンピュータによって実現するための文字認識プログラムであって、

前記判断工程は、所定以上の尤度のテキストを出力する文字認識プログラム。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0008

【補正方法】変更

【補正の内容】

【0008】

上記目的を達成するために、本発明の一態様における文字認識装置は、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第1画像認識部と、多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第2画像認識部と、前記第2画像認識部により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストにそれぞれ形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理部と、前記第1画像認識部により出力されたテキスト候補と、前記自然言語処理部により出力されたテキスト候補とを対比する判断部とを備え、前記判断部は、所定以上の尤度のテキストを出力する。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0015

【補正方法】変更

【補正の内容】

【0015】

上記目的を達成するために、本発明の一態様における文字認識方法は、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第1画像認識工程と、多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離されたテキスト候補と尤度を出力する第2画像認識工程と、前記第2画像認識工程により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストにそれぞれ形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理工程と、前記第1画像認識工程により出力されたテキスト候補と、前記自然言語処理工程により出力されたテキスト候補とを対比する判断工程とを備え、前記判断工程は、所定以上の尤度のテキストを出力する。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0016

【補正方法】変更

【補正の内容】

【0016】

上記目的を達成するために、本発明の一態様における文字認識プログラムは、多層のニューラルネットワークを用いて画像データから特徴点を抽出し、複数のテキスト候補と尤度を出力する第1画像認識工程と、多層のニューラルネットワークを用いて前記画像データから特徴点を抽出し、前記テキストよりも小さい因子である素因子テキストに分離され

たテキスト候補と尤度を出力する第2画像認識工程と、前記第2画像認識工程により出力されたテキスト候補に対して、隣接する素因子テキストの接合および切り離しを行って、組み合わせられる複数のパターンのテキストにそれぞれ形態素解析を行い、自然言語的な観点から尤もらしい複数のテキスト候補と尤度を出力する自然言語処理工程と、前記第1画像認識工程により出力されたテキスト候補と、前記自然言語処理工程により出力されたテキスト候補とを対比する判断工程と、をコンピュータによって実現するための文字認識プログラムであって、前記判断工程は、所定以上の尤度のテキストを出力する文字認識プログラムである。

フロントページの続き

(51)Int.Cl.

F I

テーマコード(参考)

G 0 6 F 17/27 6 1 5