

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6810283号
(P6810283)

(45) 発行日 令和3年1月6日(2021.1.6)

(24) 登録日 令和2年12月14日(2020.12.14)

(51) Int.Cl.

F I

G 1 0 L 15/10 (2006.01)

G 1 0 L 15/10 5 0 0 Z

G 1 0 L 15/183 (2013.01)

G 1 0 L 15/183

G 0 6 T 1/00 (2006.01)

G 0 6 T 1/00 5 0 0 A

G 0 6 F 3/16 (2006.01)

G 0 6 F 3/16 6 5 0

G 0 6 F 3/0484 (2013.01)

G 0 6 F 3/0484 1 5 0

請求項の数 23 (全 55 頁)

(21) 出願番号 特願2019-556201 (P2019-556201)
 (86) (22) 出願日 平成30年9月29日(2018.9.29)
 (65) 公表番号 特表2020-519923 (P2020-519923A)
 (43) 公表日 令和2年7月2日(2020.7.2)
 (86) 国際出願番号 PCT/CN2018/108696
 (87) 国際公開番号 W02019/062931
 (87) 国際公開日 平成31年4月4日(2019.4.4)
 審査請求日 令和1年10月15日(2019.10.15)
 (31) 優先権主張番号 201710913131.2
 (32) 優先日 平成29年9月29日(2017.9.29)
 (33) 優先権主張国・地域又は機関
 中国 (CN)
 (31) 優先権主張番号 201710913272.4
 (32) 優先日 平成29年9月29日(2017.9.29)
 (33) 優先権主張国・地域又は機関
 中国 (CN)

(73) 特許権者 519323517
 シャンハイ カンブリコン インフォメー
 ション テクノロジー カンパニー リミ
 テッド
 SHANGHAI CAMBRICON
 INFORMATION TECHNOLO
 GY CO., LTD.
 中華人民共和国 201306 シャンハ
 イ プドン・ニュー・エリア トンフェイ
 ・ロード 168 ブロック・ビー 6階
 (74) 代理人 100121728
 弁理士 井関 勝守
 (74) 代理人 100165803
 弁理士 金子 修平

最終頁に続く

(54) 【発明の名称】 画像処理装置及び方法

(57) 【特許請求の範囲】

【請求項 1】

音声信号と処理対象画像を受信するための受信モジュールと、
 目標音声命令変換モデルに従って前記音声信号を画像処理命令と前記処理対象画像の処
 理領域である目標領域に変換するための変換モジュールと、
 前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するための処理モ
 ジュールと、
テキスト情報又は前記画像処理命令又は前記目標領域を記憶するための記憶モジュール
と、を備え、

前記変換モジュールは、具体的に、
 音声認識技術、語義理解技術及び前記目標音声命令変換モデルによって前記音声信号を
 前記画像処理命令に変換し、
 前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像
 を領域分割し、前記目標領域を取得するために用いられることを特徴とする画像処理装置
 。

【請求項 2】

前記変換モジュールは、具体的に、
 音声認識技術によって前記音声信号を前記テキスト情報に変換し、
 自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像
 処理命令に変換し、

10

20

前記画像処理命令における前記語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられることを特徴とする請求項 1 に記載の画像処理装置。

【請求項 3】

前記処理モジュールは、具体的に、

予め設定された時間ウィンドウ内で前記記憶モジュールから M (M は 1 より大きい整数) 個の画像処理命令を取得し、

前記 M 個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するために用いられることを特徴とする請求項 1 に記載の画像処理装置。

【請求項 4】

前記処理モジュールは、さらに具体的に、

前記 M 個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、 N 個の画像処理命令を取得し、

前記 N 個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するために用いられることを特徴とする請求項 3 に記載の画像処理装置。

【請求項 5】

前記変換モジュールは、音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するために用いられることを特徴とする請求項 1 に記載の画像処理装置。

【請求項 6】

前記変換モジュールは、

前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、

前記予測命令及びその対応する命令セットの相関係数を確定し、

前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するためにも用いられることを特徴とする請求項 5 に記載の画像処理装置。

【請求項 7】

前記画像処理装置は、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するための訓練モジュールをさらに備えることを特徴とする請求項 1 に記載の画像処理装置。

【請求項 8】

前記処理モジュールは、画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するために用いられることを特徴とする請求項 1 に記載の画像処理装置。

【請求項 9】

前記処理モジュールは、

前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、

前記予測画像及びその対応する目標画像の相関係数を確定し、

前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するためにも用いられることを特徴とする請求項 8 に記載の画像処理装置。

【請求項 10】

前記画像処理装置は訓練モジュールをさらに備え、前記訓練モジュールは、

前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、

前記予測画像及びその対応する目標画像の相関係数を確定し、

前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するためにも用いられることを特徴とする請求項 1 に記載の画像処理装置。

10

20

30

40

50

【請求項 1 1】

前記音声信号と処理対象画像を受信する前に、前記画像処理装置は、
音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するためにも用いられることを特徴とする請求項 1 に記載の画像処理装置。

【請求項 1 2】

前記音声命令変換モデルに対する自己適応訓練はオフライン又はオンラインで行われることを特徴とする請求項 1 1 に記載の画像処理装置。

【請求項 1 3】

前記画像処理装置は、具体的に、
前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、
前記予測命令及びその対応する命令セットの相関係数を確定しと、
前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するために用いられることを特徴とする請求項 1 1 に記載の画像処理装置。

【請求項 1 4】

前記画像処理装置は、画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するために用いられることを特徴とする請求項 1 に記載の画像処理装置。

【請求項 1 5】

前記画像処理装置は、具体的に、
前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、
前記予測画像及びその対応する目標画像の相関係数を確定し、
前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するために用いられることを特徴とする請求項 1 4 に記載の画像処理装置。

【請求項 1 6】

音声信号及び処理対象画像を受信するステップと、
目標音声命令変換モデルに従って前記音声信号を画像処理命令と前記処理対象画像の処理領域である目標領域に変換するステップと、
前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するステップと、
テキスト情報又は前記画像処理命令又は前記目標領域を記憶するステップと、を含み、
前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、
音声認識技術、語義理解技術及び前記音声命令変換モデルによって前記音声信号を前記画像処理命令に変換するステップと、
前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含むことを特徴とする画像処理方法。

【請求項 1 7】

前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、
音声認識技術によって前記音声信号を前記テキスト情報に変換するステップと、
自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像処理命令に変換するステップと、
前記画像処理命令における前記語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、をさらに含むことを特徴とする請求項 1 6 に記載の画像処理方法。

【請求項 1 8】

前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するステップは、
予め設定された時間ウィンドウ内で記憶モジュールから M (M は 1 より大きい整数) 個の画像処理命令を取得するステップと、

前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含むことを特徴とする請求項16に記載の画像処理方法。

【請求項19】

前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップは、

前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N（Nは前記Mよりも小さい整数）個の画像処理命令を取得するステップと、

前記N個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含むことを特徴とする請求項18に記載の画像処理方法。

【請求項20】

前記音声信号と処理対象画像を受信するステップの前に、前記方法は、

音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するステップをさらに含むことを特徴とする請求項16に記載の画像処理方法。

【請求項21】

前記音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するステップは、

前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換するステップと、

前記予測命令及びその対応する命令セットの相関係数を確定するステップと、

前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するステップと、を含むことを特徴とする請求項20に記載の画像処理方法。

【請求項22】

前記音声信号と処理対象画像を受信するステップの前に、前記方法は、

画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するステップをさらに含むことを特徴とする請求項16に記載の画像処理方法。

【請求項23】

前記画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するステップは、

前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得するステップと、

前記予測画像及びその対応する目標画像の相関係数を確定するステップと、

前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するステップと、を含むことを特徴とする請求項22に記載の画像処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像処理分野に関し、特に、画像処理装置及び方法に関する。

【背景技術】

【0002】

ユーザーは、写真を撮った後、より良い画像効果を示すために、コンピュータにおけるPSソフトウェア又は携帯電話におけるレタッチソフトウェアによって画像処理を行うことがある。

【0003】

しかし、コンピュータにおけるPSソフトウェアや携帯電話におけるレタッチソフトウェアを利用して画像処理を行う前に、ユーザーはソフトウェアの使用方法を習得する必要があり、且つソフトウェアの使用方法を習得した後に、命令を手動で入力してコンピュータや携帯電話を制御してレタッチ操作を行う必要がある。ユーザーにとって、このような方式は時間がかかるだけでなく、ユーザーの経験も悪かった。

【発明の概要】

【発明が解決しようとする課題】**【0004】**

本発明の実施例は、音声を入力すると画像を処理できる機能を実現し、画像処理前にユーザーが画像処理ソフトウェアを学習する時間を省き、ユーザーの経験を向上させる画像処理装置及び方法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】**【0005】**

第1の態様として、本発明の実施例は、

音声信号と処理対象画像を受信するための受信モジュールと、

目標音声命令変換モデルに従って前記音声信号を画像処理命令と前記処理対象画像の処理領域である目標領域に変換するための変換モジュールと、

前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するための処理モジュールと、を備える画像処理装置を提供する。

10

【0006】

また、好ましい一実施例において、前記変換モジュールは、具体的に、

音声認識技術によって前記音声信号をテキスト情報に変換し、

自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像処理命令に変換し、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

20

【0007】

また、好ましい一実施例において、前記変換モジュールは、さらに具体的に、

音声認識技術、語義理解技術及び前記目標音声命令変換モデルによって前記音声信号を前記画像処理命令に変換し、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【0008】

また、好ましい一実施例において、前記画像処理装置は、

前記テキスト情報又は前記画像処理命令又は前記目標領域を記憶するための記憶モジュールをさらに備える。

30

【0009】

また、好ましい一実施例において、前記処理モジュールは、具体的に、

予め設定された時間ウィンドウ内で前記記憶モジュールからM(Mは1より大きい整数)個の画像処理命令を取得し、

前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するために用いられる。

【0010】

また、好ましい一実施例において、前記処理モジュールは、さらに具体的に、

前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N(Nは前記Mよりも小さい整数)個の画像処理命令を取得し、

前記N個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するために用いられる。

40

【0011】

また、好ましい一実施例において、前記変換モジュールは、

音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するために用いられる。

【0012】

また、好ましい一実施例において、前記変換モジュールによる前記音声命令変換モデルの自己適応訓練はオフライン又はオンラインで行われる。

【0013】

50

また、好ましい一実施例において、前記変換モジュールによる前記音声命令変換モデルの自己適応訓練はオフライン又はオンラインで行われる。

【0014】

また、好ましい一実施例において、前記変換モジュールによる前記音声命令変換モジュールの自己適応訓練は監督されるか、又は監督されない。

【0015】

また、好ましい一実施例において、前記変換モジュールは、
前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、
前記予測命令及びその対応する命令セットの相関係数を確定し、
前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するためにも用いられる。 10

【0016】

また、好ましい一実施例において、前記画像処理装置は、
音声命令変換モデルに基づいて上記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するための訓練モジュールをさらに備える。

【0017】

また、好ましい一実施例において、前記処理モジュールは、
画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するために用いられる。 20

【0018】

また、好ましい一実施例において、前記処理モジュールによる前記画像処理モデルの自己適応訓練はオフライン又はオンラインで行われる。

【0019】

また、好ましい一実施例において、前記処理モジュールによる前記画像処理モデルの自己適応訓練は監督されるか、又は監督されない。

【0020】

また、好ましい一実施例において、前記処理モジュールは、
前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、
前記予測画像及びその対応する目標画像の相関係数を確定し、
前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するためにも用いられる。 30

【0021】

また、好ましい一実施例において、前記訓練モジュールは、
前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、
前記予測画像及びその対応する目標画像の相関係数を確定し、
前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するためにも用いられる。 40

【0022】

第2の態様として、本発明の実施例は、
音声信号及び処理対象画像を受信するステップと、
目標音声命令変換モデルに従って前記音声信号を画像処理命令と前記処理対象画像の処理領域である目標領域に変換するステップと、
前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するステップと、
を含む画像処理方法を提供する。

【0023】

また、好ましい一実施例において、前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、

音声認識技術によって前記音声信号をテキスト情報に変換するステップと、 50

自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像処理命令に変換するステップと、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0024】

また、好ましい一実施例において、前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、

音声認識技術、語義理解技術及び前記音声命令変換モデルによって前記音声信号を前記画像処理命令に変換するステップと、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

10

【0025】

また、好ましい一実施例において、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記処理対象画像の処理対象領域を取得するステップは、

前記画像処理命令における語義領域の粒度に基づいて、前記処理対象画像を、粒度の同じM個の領域に分割するステップと、

前記画像処理命令に応じて、前記M個の領域から前記目標領域を取得するステップと、を含む。

【0026】

20

また、好ましい一実施例において、前記方法は、

前記テキスト情報又は前記画像処理命令又は前記目標領域を記憶するステップをさらに含む。

【0027】

また、好ましい一実施例において、前記画像処理命令に応じて前記目標領域を処理するステップは、

予め設定された時間ウィンドウ内で前記記憶モジュールからM（Mは1より大きい整数）個の画像処理命令を取得するステップと、

前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含む。

30

【0028】

また、好ましい一実施例において、前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップは、

前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N（Nは前記Mよりも小さい整数）個の画像処理命令を取得するステップと、

前記N個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含む。

【0029】

また、好ましい一実施例において、前記音声信号と処理対象画像を受信するステップの前に、前記方法は、

40

音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するステップをさらに含む。

【0030】

また、好ましい一実施例において、前記音声命令変換モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【0031】

また、好ましい一実施例において、前記音声命令変換モジュールに対する自己適応訓練は監督されるか、又は監督されない。

【0032】

また、好ましい一実施例において、前記音声命令変換モデルに対して自己適応訓練を行

50

って、前記目標音声命令変換モデルを取得するステップは、

前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換するステップと、

前記予測命令及びその対応する命令セットの相関係数を確定するステップと、

前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するステップと、を含む。

【0033】

また、好ましい一実施例において、前記音声信号と処理対象画像を受信するステップの前に、

画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するステップをさらに含む。

【0034】

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【0035】

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練は監督されるか、又は監督されない。

【0036】

また、好ましい一実施例において、前記画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するステップは、

前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得するステップと、

前記予測画像及びその対応する目標画像の相関係数を確定するステップと、

前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するステップと、を含む。

【0037】

本発明の実施例の技術手段において、画像処理装置の受信モジュールは音声信号と処理対象画像を取得し、画像処理装置の変換モジュールは目標音声命令変換モデルに従って前記音声信号を画像処理命令と前記処理対象画像の処理領域である目標領域に変換し、画像処理装置の処理モジュールは前記画像処理命令と前記目標音声命令変換モデルに従って前記目標領域を処理することができることが分かる。従来の画像処理技術と比べると、本発明は、音声によって画像処理を行うので、画像処理前にユーザーが画像処理ソフトウェアを学習する時間を省き、ユーザーの経験を向上させている。

【0038】

第3の態様として、本発明の実施例は、

ユーザーの入力した音声信号を収集するための音声収集装置と、

目標音声命令変換モデルに従って前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換するための命令変換装置と、

前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するための画像処理機と、を備える画像処理装置を提供する。

【0039】

また、好ましい一実施例において、前記命令変換装置は、

前記音声認識技術によって音声信号をテキスト情報に変換するための第1の音声認識装置と、

自然言語処理技術と前記目標音声命令変換モデルによってテキスト情報を前記画像処理命令に変換するための音声テキスト変換装置と、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するための第1の画像認識装置と、を備える。

【0040】

また、好ましい一実施例において、前記命令変換装置は、

前記音声認識技術、語義理解技術及び前記目標音声命令変換モデルによって前記音声信

10

20

30

40

50

号を前記画像処理命令に変換するための第2の音声認識装置と、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するための第2の画像認識装置と、を備える。

【0041】

また、好ましい一実施例において、前記画像処理装置は、

前記テキスト情報又は前記画像処理命令又は前記目標領域を記憶するためのメモリをさらに備える。

【0042】

また、好ましい一実施例において、その特徴として、前記画像処理機は、

予め設定された時間ウィンドウ内で前記メモリからM(Mは1より大きい整数)個の画像処理命令を取得するための命令取得モジュールと、

前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するための処理モジュールと、を備える。

【0043】

また、好ましい一実施例において、前記処理モジュールは、

前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N(Nは前記Mよりも小さい整数)個の画像処理命令を取得し、

前記N個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理する。

【0044】

また、好ましい一実施例において、前記命令変換装置は、

音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するために用いられる。

【0045】

また、好ましい一実施例において、前記命令変換装置による前記音声命令変換モデルの自己適応訓練はオフライン又はオンラインで行われる。

【0046】

また、好ましい一実施例において、前記命令変換装置による前記音声命令変換モデルの自己適応訓練は監督されるか、又は監督されない。

【0047】

また、好ましい一実施例において、前記命令変換装置は、

前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、

前記予測命令及びその対応する命令セットの相関係数を確定し、

前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するためにも用いられる。

【0048】

また、好ましい一実施例において、前記画像処理装置は、

前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するための訓練装置をさらに備える。

【0049】

また、好ましい一実施例において、前記画像処理機は、

画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するために用いられる。

【0050】

また、好ましい一実施例において、前記画像処理機による前記画像処理モデルの自己適応訓練はオフライン又はオンラインで行われる。

【0051】

また、好ましい一実施例において、前記画像処理機による前記画像処理モデルの自己適

10

20

30

40

50

応訓練は監督されるか、又は監督されない。

【 0 0 5 2 】

また、好ましい一実施例において、前記画像処理機は、
前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、
前記予測画像及びその対応する目標画像の相関係数を確定し、
前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するためにも用いられる。

【 0 0 5 3 】

また、好ましい一実施例において、前記訓練装置は、
前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、
前記予測画像及びその対応する目標画像の相関係数を確定し、
前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するためにも用いられる。

10

【 0 0 5 4 】

第4の態様として、本発明の実施例は、
ユーザーの入力した音声信号を収集するステップと、
目標音声命令変換モデルに従って前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換するステップと、
前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するステップと、
を含む画像処理方法を提供する。

20

【 0 0 5 5 】

また、好ましい一実施例において、前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、
音声認識技術によって前記音声信号をテキスト情報に変換するステップと、
自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像処理命令に変換するステップと、
前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【 0 0 5 6 】

また、好ましい一実施例において、前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、
音声認識技術、語義理解技術及び前記音声命令変換モデルによって前記音声信号を前記画像処理命令に変換するステップと、
前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

30

【 0 0 5 7 】

また、好ましい一実施例において、前記方法は、
前記テキスト情報又は前記画像処理命令又は前記目標領域を記憶するステップをさらに含む。

【 0 0 5 8 】

また、好ましい一実施例において、前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するステップは、
予め設定された時間ウィンドウ内で前記メモリからM（Mは1より大きい整数）個の画像処理命令を取得するステップと、
前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含む。

40

【 0 0 5 9 】

また、好ましい一実施例において、前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップは、
前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N（Nは前

50

記Mよりも小さい整数)個の画像処理命令を取得するステップと、

前記N個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含む。

【0060】

また、好ましい一実施例において、前記音声信号と処理対象画像を受信するステップの前に、前記方法は、

音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するステップをさらに含む。

【0061】

また、好ましい一実施例において、前記音声命令変換モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

10

【0062】

また、好ましい一実施例において、前記音声命令変換モデルに対する自己適応訓練は監督されるか、又は監督されない。

【0063】

また、好ましい一実施例において、前記音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するステップは、

前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換するステップと、

前記予測命令及びその対応する命令セットの相関係数を確定するステップと、

前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するステップと、を含む。

20

【0064】

また、好ましい一実施例において、前記音声信号と処理対象画像を受信するステップの前に、前記方法は、

画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するステップをさらに含む。

【0065】

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【0066】

30

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練は監督されるか、又は監督されない。

【0067】

また、好ましい一実施例において、前記画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するステップは、

前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得するステップと、

前記予測画像及びその対応する目標画像の相関係数を確定するステップと、

前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するステップと、を含む。

40

【0068】

第5の態様として、本発明の実施例は、本発明の実施例の第3の態様の前記画像処理装置を備える画像処理チップも提供する。

【0069】

また、好ましい一実施例において、上記チップは、メインチップと、連携チップとを備え、

上記連携チップは、本発明の実施例の第1の態様に記載の装置を備え、上記メインチップは上記連携チップに起動信号を与え、処理対象画像と画像処理命令が上記連携チップに伝送されるように制御するためのものである。

【0070】

50

第6の態様として、本発明の実施例は、本発明の実施例の第5の態様に記載の画像処理チップを備えるチップパッケージ構造を提供する。

【0071】

第7の態様として、本発明の実施例は、本発明の実施例の第6の態様に記載のチップパッケージ構造を備えるボードカードを提供する。

【0072】

第8の態様として、本発明の実施例は、本発明の実施例の第7の態様に記載のボードカードを備える電子機器を提供する。

【0073】

なお、本発明の実施例の技術手段において、音声収集装置はユーザーの入力した音声信号を収集し、命令変換装置は目標音声命令変換モデルに従って前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換し、画像処理機は前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理することができることが分かる。従来の画像処理技術と比べると、本発明は、音声によって画像処理を行うので、画像処理前にユーザーが画像処理ソフトウェアを学習する時間を省き、ユーザーの経験を向上させている。

10

【0074】

第9の態様として、本発明の実施例は、

音声信号と処理対象画像を入力するための入出力手段と、

前記音声信号と前記処理対象画像を記憶するための記憶手段と、

20

前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換し、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得し、前記処理対象画像を前記記憶手段に記憶するための画像処理手段と、を備え、

前記入出力手段は、前記処理後の画像を出力するためにも用いられる画像処理装置を提供する。

【0075】

また、好ましい一実施例において、前記記憶手段は、ニューロン記憶手段と、重みキャッシュ手段とを備え、前記画像処理手段のニューラルネットワーク演算手段は、ニューラルネットワーク演算サブ手段を備え、

前記ニューロン記憶手段が前記音声信号と前記処理対象画像を記憶するために用いられ、且つ、前記重みキャッシュ手段が目標音声命令変換モデルと目標画像処理モデルを記憶するために用いられる場合、前記ニューラルネットワーク演算サブ手段は前記目標音声命令変換モデルに従って前記音声信号を前記画像処理命令と前記目標領域に変換するために用いられ、

30

前記ニューラルネットワーク演算サブ手段は、前記目標画像処理モデルと前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するためにも用いられ、

前記ニューラルネットワーク演算サブ手段は、前記処理後の画像を前記ニューロン記憶手段に記憶するためにも用いられる。

【0076】

また、好ましい一実施例において、前記記憶手段は汎用データキャッシュ手段を備え、前記画像処理手段のニューラルネットワーク演算手段は汎用演算サブ手段を備え、

40

前記汎用データキャッシュ手段が前記音声信号と前記処理対象画像を受信するために用いられる場合、前記汎用演算サブ手段は前記音声信号を前記画像処理命令と前記目標領域に変換するために用いられ、

前記汎用演算サブ手段は、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するためにも用いられ、

前記汎用演算サブ手段は、前記処理後の画像を前記汎用データ記憶手段に記憶するためにも用いられる。

【0077】

また、好ましい一実施例において、前記ニューラルネットワーク演算サブ手段は、具体

50

的に、

音声認識技術に基づいて前記音声信号をテキスト情報に変換し、

自然言語処理技術と前記目標音声命令変換モデルに基づいて前記テキスト情報を前記画像処理命令に変換し、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【0078】

また、好ましい一実施例において、前記ニューラルネットワーク演算サブ手段は、具体的に、

音声認識技術、語義理解技術及び前記目標音声命令変換モデルに基づいて前記音声信号を前記画像処理命令に変換し、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【0079】

また、好ましい一実施例において、前記汎用演算サブ手段は、具体的に、

音声認識技術に基づいて前記音声信号をテキスト情報に変換し、

自然言語処理技術に基づいて前記テキスト情報を前記画像処理命令に変換し、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【0080】

また、好ましい一実施例において、前記汎用演算サブ手段は、具体的に、

音声認識技術と語義理解技術に基づいて前記音声信号を前記画像処理命令に変換し、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【0081】

また、好ましい一実施例において、前記ニューロン記憶手段は、前記目標領域と前記画像処理命令を記憶するために用いられる。

【0082】

また、好ましい一実施例において、前記汎用データキャッシュ手段は前記目標領域と前記画像処理命令を記憶するために用いられる。

【0083】

また、好ましい一実施例において、前記ニューラルネットワーク演算サブ手段は、

予め設定された時間ウィンドウ内で前記ニューロン記憶手段から M (M は1より大きい整数)個の画像処理命令を取得し、

前記 M 個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、 N (N は前記 M よりも小さい整数)個の画像処理命令を取得し、

前記 N 個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理して、処理後の画像を取得するために用いられる。

【0084】

また、好ましい一実施例において、前記汎用演算サブ手段は、

予め設定された時間ウィンドウ内で前記汎用データキャッシュ手段から M (M は1より大きい整数)個の画像処理命令を取得し、

前記 M 個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、 N (N は前記 M よりも小さい整数)個の画像処理命令を取得し、

前記 N 個の画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するために用いられる。

【0085】

また、好ましい一実施例において、前記ニューラルネットワーク演算サブ手段は、

音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するためにも用いられる。

【 0 0 8 6 】

また、好ましい一実施例において、前記ニューラルネットワーク演算サブ手段は、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するためにも用いられる。

【 0 0 8 7 】

また、好ましい一実施例において、前記ニューラルネットワーク演算サブ手段は、画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するためにも用いられる。

10

【 0 0 8 8 】

また、好ましい一実施例において、前記ニューラルネットワーク演算サブ手段は、前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、前記予測画像及びその対応する目標画像の相関係数を確定し、前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するためにも用いられる。

【 0 0 8 9 】

また、好ましい一実施例において、前記画像処理装置の画像処理手段は、ニューラルネットワーク演算命令と汎用演算命令を含む実行されるべき命令を記憶するための命令キャッシュ手段と、

20

前記ニューラルネットワーク演算命令を前記ニューラルネットワーク演算サブ手段に伝送し、前記汎用演算命令を前記汎用演算サブ手段に伝送するための命令処理手段と、をさらに備える。

【 0 0 9 0 】

第 10 の態様として、本発明の実施例は、音声信号と処理対象画像を入力するステップと、前記音声信号と前記処理対象画像を記憶するステップと、前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換し、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得し、前記処理対象画像を前記記憶手段に記憶するステップと、

30

前記処理後の画像を出力するステップと、を含む画像処理方法を提供する。

【 0 0 9 1 】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップは、

音声認識技術によって前記音声信号をテキスト情報に変換するステップと、自然言語処理技術と目標音声命令変換モデルに基づいて前記テキスト情報を前記画像処理命令に変換するステップと、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

40

【 0 0 9 2 】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップは、

音声認識技術、語義理解技術及び目標音声命令変換モデルに基づいて前記音声信号を前記画像処理命令に変換するステップと、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【 0 0 9 3 】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップは、

音声認識技術に基づいて前記音声信号をテキスト情報に変換するステップと、

50

自然言語処理技術に基づいて前記テキスト情報を前記画像処理命令に変換するステップと、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0094】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップは、

音声認識技術と語義理解技術に基づいて前記音声信号を前記画像処理命令に変換するステップと、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

10

【0095】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップの後に、前記方法は、

前記画像処理命令と前記目標領域を記憶するステップをさらに含む。

【0096】

また、好ましい一実施例において、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するステップは、

予め設定された時間ウィンドウ内で前記ニューロン記憶手段からM（Mは1より大きい整数）個の画像処理命令を取得するステップと、

20

前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N（Nは前記Mよりも小さい整数）個の画像処理命令を取得するステップと、

前記N個の画像処理命令と目標画像処理モデルに基づいて前記目標領域を処理して、処理後の画像を取得するステップと、を含む。

【0097】

また、好ましい一実施例において、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するステップは、

予め設定された時間ウィンドウ内で前記汎用データキャッシュ手段からM（Mは1より大きい整数）個の画像処理命令を取得するステップと、

前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N（Nは前記Mよりも小さい整数）個の画像処理命令を取得するステップと、

30

前記N個の画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するステップと、を含む。

【0098】

また、好ましい一実施例において、前記方法は、

音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するステップをさらに含む。

【0099】

また、好ましい一実施例において、前記音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するステップは、

40

前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換するステップと、

前記予測命令及びその対応する命令セットの相関係数を確定するステップと、

前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するステップと、を含む。

【0100】

また、好ましい一実施例において、前記方法は、

画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するステップをさらに含む。

【0101】

また、好ましい一実施例において、前記画像処理モデルに対して自己適応訓練を行うス

50

テップは、

前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得するステップと、

前記予測画像及びその対応する目標画像の相関係数を確定するステップと、

前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するステップと、を含む。

【0102】

第11の態様として、本発明の実施例は、本発明の実施例の第1の態様の前記画像処理装置を備える画像処理チップも提供している。

【0103】

また、好ましい一実施例において、上記チップは、メインチップと、連携チップとを備え、

上記連携チップは、本発明の実施例の第1の態様に記載の装置を備え、上記メインチップは上記連携チップに起動信号を与え、処理対象画像と画像処理命令が上記連携チップに伝送されるように制御するためのものである。

【0104】

第12の態様として、本発明の実施例は、本発明の実施例の第11の態様に記載の画像処理チップを備えるチップパッケージ構造を提供する。

【0105】

第13の態様として、本発明の実施例は、本発明の実施例の第12の態様に記載のチップパッケージ構造を備えるボードカードを提供する。

【0106】

第14の態様として、本発明の実施例は、本発明の実施例の第13の態様に記載のボードカードを備える電子機器を提供する。

【発明の効果】

【0107】

なお、本発明の実施例の技術手段において、入出力手段は音声信号と処理対象画像を入力し、記憶手段は前記音声信号と前記処理対象画像を記憶し、画像処理手段は前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換し、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得し、前記処理対象画像を前記記憶手段に記憶し、前記入出力手段は前記処理後の画像を出力することができることが分かる。従来の画像処理技術と比べると、本発明は、音声によって画像処理を行うので、画像処理前にユーザーが画像処理ソフトウェアを学習する時間を省き、ユーザーの経験を向上させている。

【図面の簡単な説明】

【0108】

以下では、本発明の実施例又は従来技術における技術案をより明らかに説明するために、実施例又は従来技術の記述に使用する必要がある図面について簡単に説明する。なお、以下の説明における図面は単に本発明の一部の実施例に過ぎず、当業者にとって、これらの図面に基づいて創造的な活動をしないう前提で他の図面を得ることができることは言うまでもない。

【図1】本発明の実施例が提供する画像処理装置の構造模式図である。

【図2】本発明の実施例が提供する他の画像処理装置の部分構造模式図である。

【図3】本発明の実施例が提供する他の画像処理装置の部分構造模式図である。

【図4】本発明の実施例が提供する他の画像処理装置の部分構造模式図である。

【図5】本発明の実施例が提供するチップの構造模式図である。

【図6】本発明の実施例が提供する他のチップの構造模式図である。

【図7】本発明の実施例が提供する画像処理方法のフローチャートである。

【図8】本発明の実施例が提供する画像処理装置の構造模式図である。

【図9】本発明の実施例が提供する他の画像処理装置の部分構造模式図である。

10

20

30

40

50

【図 1 0】本発明の実施例が提供する他の画像処理装置の部分構造模式図である。

【図 1 1】本発明の実施例が提供する他の画像処理装置の部分構造模式図である。

【図 1 2】本発明の実施例が提供する電子機器の構造模式図である。

【図 1 3】本発明の実施例が提供する画像処理方法のフローチャートである。

【図 1 4】本発明の実施例が提供する画像処理装置の構造模式図である。

【図 1 5】本発明の実施例が提供する他の画像処理装置の部分構造模式図である。

【図 1 6】本発明の実施例が提供する電子機器の構造模式図である。

【図 1 7】本発明の実施例が提供する画像処理方法のフローチャートである。

【発明を実施するための形態】

【0109】

10

以下に、それぞれを詳細に説明する。

【0110】

本発明の明細書、特許請求の範囲及び前記図面における「第 1」、「第 2」、「第 3」及び「第 4」等の用語は異なる相手を区別するために使用され、特定の順序を説明することを意図しない。また、用語「含む」、「有する」及びそれらのいずれか変更は、非排他的に含むことを意図している。例えば、一連のステップ又は手段を含む過程、方法、システム、製品又は機器は、列挙されたステップ又は手段に限定されず、列挙していないステップ又は手段を選択的に含むか、或いは、これらの過程、方法、製品又は機器に固有の他のステップ又は手段を選択的に含むことができる。

【0111】

20

本明細書で言及している「実施例」は、実施例と併せて説明する特定の特徴、構造又は特性が本発明の少なくとも一つの実施例に含まれ得ることを意味する。明細書の様々な箇所出現する当該短句は、必ずしも同じ実施例を指しているわけではなく、他の実施例と相互排除する独立又は候補の実施例でもない。当業者であれば、本明細書に記載の実施例が他の実施例と組み合わせることができることを明示的及び暗黙的に理解するであろう。

【0112】

第 1 の態様として、本発明の実施例が提供する画像処理装置の構造模式図である図 1 に示すように、当該画像処理装置 100 は、

音声信号と処理対象画像を受信するための受信モジュール 101 を備える。

【0113】

30

好ましくは、上記画像処理装置 100 は、前処理モジュールをさらに備え、上記受信モジュール 101 が上記音声信号を受信した後、上記前処理モジュールは当該音声信号に対して騒音低減処理を実行する。

【0114】

好ましくは、受信モジュール 101 は、オーディオ収集手段を備え、当該オーディオ収集手段は、音声センサ、マイクロフォン、ピックアップや、他のオーディオ収集機器であってもよい。

【0115】

具体的に、上記受信装置 101 は、上記音声信号を受信する時、環境声音信号も受信する。上記前処理モジュールは、上記環境声音信号に応じて上記音声信号に対して騒音低減処理を実行する。当該環境声音信号は上記音声信号にとって騒音である。

40

【0116】

さらに、上記オーディオ収集手段は、上記音声信号と上記環境声音信号の収集に用いるだけでなく、騒音低減処理も実現するペアマイクロフォンアレイを含んでもよい。

【0117】

好ましくは、1つの可能な実施例において、上記画像処理装置は、第 1 の記憶モジュールと第 2 の記憶モジュールとをさらに備える。上記受信装置が上記音声信号と上記処理対象画像を受信した後、上記画像処理装置は上記音声信号と上記処理対象画像を第 1 の記憶モジュールと第 2 の記憶モジュールにそれぞれ記憶する。

【0118】

50

好ましくは、上記第1の記憶モジュールと上記第2の記憶モジュールは、同一の記憶モジュールであってもよく、異なる記憶モジュールであってもよい。

【0119】

変換モジュール102は、目標音声命令変換モデルに基づいて前記音声信号を画像処理命令と前記処理対象画像の処理領域である目標領域に変換するために用いられる。

【0120】

好ましくは、上記変換モジュール102は、音声認識技術、自然言語処理技術及び画像認識技術に基づいて前記音声信号を画像処理命令と目標領域に変換する前に、上記変換モジュールは上記第1の記憶モジュールから上記音声処理信号を取得する。

【0121】

そのうち、前記変換モジュール102は、

音声認識技術によって前記音声信号をテキスト情報に変換するための第1の音声認識手段1021と、

自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像処理命令に変換するための語義理解手段1022と、

前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するための第1の画像認識手段1023と、を備える。

【0122】

さらに、上記変換モジュール102は、上記画像処理命令における語義領域の粒度を取得するための取得手段1026をさらに備える。

【0123】

例を挙げて上記語義領域を説明すると、上記画像処理装置100が音声信号に応じて上記目標領域が顔領域であると確定した場合、上記語義領域は上記処理対象画像における顔領域であり、上記画像処理装置は顔を粒度として、上記処理対象画像における複数の顔領域を取得する。上記目標領域が背景である場合、上記画像処理装置は上記処理対象画像を背景領域と非背景領域に分割し、上記目標領域が赤色領域である場合、上記画像処理装置は上記処理対象画像を色に従って異なる色の領域に分割する。

【0124】

具体的に、本発明で使用する音声認識技術は、人工ニューラルネットワーク(Artificial Neural Network; ANN)、隠れマルコフモデル(Hidden Markov Model; HMM)等のモデルを含むが、これらに限定されない。上記第1の音声認識手段は、上記音声認識技術に基づいて上記音声信号を処理してもよい。上記自然言語処理技術は、統計的機械学習や、ANN等の方法を利用することを含むが、これらに限定されない。上記語義理解手段は、上記自然言語処理技術に基づいて語義情報を抽出することができる。上記画像認識技術は、エッジ検出に基づく方法、閾値分割方法、領域成長及び分水嶺(watershed)アルゴリズム、階調積分投影(Gray-level Integration Projection)曲線分析、テンプレートマッチング、変形可能テンプレート、ハフ(Hough)変換、スネーク(Snake)演算子、Gaborウェーブレット変換に基づく弾性マップマッチング技術や、アクティブ形状モデル及びアクティブ外観モデル等の方法等のアルゴリズムを含むが、これらに限定されない。上記画像認識手段は、上記画像認識技術に基づいて上記処理対象画像を異なる領域に分割することができる。

【0125】

また、好ましい一実施例において、上記音声認識手段1021は上記音声認識技術によって上記音声信号をテキスト情報に変換し、当該テキスト情報を上記第1の記憶モジュールに保存する。上記語義理解手段1022は上記第1の記憶モジュールから上記テキスト情報を取得し、自然言語処理技術と上記目標音声命令変換モデルによって当該テキスト情報を画像処理命令に変換し、画像処理命令を上記第1の記憶モジュールに保存する。上記第1の画像認識手段1023は上記画像処理命令における語義領域の粒度と画像認識技術

10

20

30

40

50

に基づいて上記処理対象画像を領域分割し、上記目標領域を取得し、上記分割結果と上記目標領域を上記第２の記憶モジュールに記憶する。

【０１２６】

また、好ましい一実施例において、上記変換モジュール１０２は、

音声認識技術、自然言語処理技術と前記目標音声命令変換モデルに基づいて上記音声信号を直接に上記画像処理命令に変換し、当該画像処理命令を第１の記憶モジュールに保存するための音声認識手段１０２５と、

上記画像処理命令における上記処理対象画像を操作する語義領域の粒度に基づいて、当該処理対象画像を当該語義領域の粒度に従って分割し、上記処理対象画像を処理する領域である目標領域を取得し、分割後の結果及び上記目標領域を第２の記憶モジュールに記憶する第２の画像認識手段１０２６と、を備えてもよい。

10

【０１２７】

好ましくは、上記受信モジュール１０１が上記音声信号と上記処理対象画像を受信する前に、上記変換モジュール１０２が音声命令変換モデルに対して自己適応訓練を行って、上記目標音声命令変換モデルを取得する。

【０１２８】

また、上記音声命令変換モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【０１２９】

具体的に、上記音声命令変換モデルに対する自己適応訓練がオフラインで行われることは、具体的に、上記変換モジュール１０２がそのハードウェアを基で上記音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得することである。上記音声命令変換モデルに対する自己適応訓練がオンラインで行われることは、具体的に、上記変換モジュールと異なるクラウドサーバが上記音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得することである。上記変換モジュール１０２が上記目標音声命令変換モデルを使用する必要がある場合、当該変換モジュール１０２は上記クラウドサーバから当該目標音声命令変換モデルを取得する。

20

【０１３０】

好ましくは、上記音声命令変換モデルに対する自己適応訓練は監督されるか、又は監督されない。

30

【０１３１】

具体的に、上記音声命令変換モデルに対する自己適応訓練が監督されることは、具体的に、上記変換モジュール１０２が音声命令変換モデルに基づいて上記音声信号を予測命令に変換し、その後、上記予測命令及びその対応する命令セットの相関係数を確定し、当該命令セットは手動で音声信号に従って取得される命令のセットであり、上記変換モジュール１０２が前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得することである。

【０１３２】

また、好ましい一実施例において、上記画像処理装置１００は、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するための訓練モジュール１０５をさらに備える。

40

【０１３３】

一例として、上記音声命令変換モデルに対する自己適応訓練が監督されることは、具体的に、上記変換モジュール１０２又は訓練モジュール１０５が、画像の色の変更や、画像の回転等の関連命令を含む音声信号を受信することである。命令毎に１つの命令セットが対応する。自己適応訓練用の入力した音声信号にとって、対応する命令セットは既知のものであり、上記変換モジュール１０２又は訓練モジュール１０５はこれらの音声信号を音声命令変換モデルの入力データとして、出力後の予測命令を取得する。上記変換モジュール

50

ル 1 0 2 又は訓練モジュール 1 0 5 は、上記予測命令及びその対応する命令セットの相関係数を算出し、上記音声命令変換モデルにおけるパラメータ（重みや、オフセット等）を適応的に更新して、上記音声命令変換モデルの性能を向上し、さらに上記目標音声命令変換モデルを取得する。

【 0 1 3 4 】

前記画像処理装置 1 0 0 は、前記テキスト情報又は前記画像処理命令又は前記目標領域を記憶するための記憶モジュール 1 0 4 をさらに備える。

【 0 1 3 5 】

また、好ましい一実施例において、上記記憶モジュール 1 0 4 は、上記第 1 の記憶モジュール及び第 2 の記憶モジュールと同一の記憶モジュールであってもよく、異なる記憶モジュールであってもよい。

10

【 0 1 3 6 】

処理モジュール 1 0 3 は、前記画像処理命令と目標画像処理モデルに基づいて前記処理対象画像を処理するためのものである。

【 0 1 3 7 】

また、前記処理モジュール 1 0 3 は、予め設定された時間ウィンドウ内で前記記憶モジュールから M（M は 1 より大きい整数）個の画像処理命令を取得するための取得手段 1 0 3 1 と、前記 M 個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するための処理手段 1 0 3 2 と、をさらに備える。

【 0 1 3 8 】

20

好ましくは、前記処理手段 1 0 3 2 は、前記 M 個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N（N は前記 M よりも小さい整数）個の画像処理命令を取得し、前記 N 個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するために用いられる。

【 0 1 3 9 】

具体的に、上記予め設定された時間ウィンドウは、予め設定された時間長であると理解してもよい。予め設定された時間長内で上記取得手段 1 0 3 1 は上記記憶モジュール 1 0 4 から M 個の画像処理命令を取得した後、上記処理手段 1 0 3 2 は上記 M 個の画像処理命令について 2 つずつ比較し、当該 M 個の画像処理命令のうち同じ機能を有する命令を削除して、N 個の画像処理命令を取得する。上記処理手段 1 0 3 2 は上記 N 個の処理命令と上記目標画像処理モデルに基づいて上記処理対象画像を処理する。

30

【 0 1 4 0 】

一例として、上記処理手段 1 0 3 2 は上記 M 個の画像処理命令について 2 つずつ比較する。画像処理命令 A と画像処理命令 B とが同じである場合、上記処理手段 1 0 3 2 は上記画像処理命令 A 及び B のうち最大のオーバーヘッドを削除し、画像処理命令 A と画像処理命令 B とが異なる場合、上記処理手段 1 0 3 2 は上記画像処理命令 A と上記画像処理命令 B の類似度係数を取得する。当該類似度係数が類似度閾値よりも大きい時、上記画像処理命令 A と上記画像処理命令 B の機能が同じであることを確定し、上記処理手段 1 0 3 2 は上記画像処理命令 A 及び B のうち最大のオーバーヘッドを削除し、上記類似度係数が上記類似度閾値よりも小さい時、上記処理手段 1 0 3 2 は上記画像処理命令 A 及び B の機能が異なることを確定する。当該画像処理命令 A 及び B は上記 M 個の処理命令のうちの任意の 2 つである。

40

【 0 1 4 1 】

具体的に、上記処理モジュール 1 0 3 に対し、その入力と出力はいずれも画像である。上記処理モジュール 1 0 3 の A N N や従来のコンピュータビジョン方法に限られない方法によって上記処理対象画像に対して行う処理は、美体（例えば美脚、豊胸）、顔の変更、顔の美化、オブジェクトの変更（猫と犬の交換、シマウマの馬への変更、リンゴとオレンジの交換等）、背景の変更（後の森を畑に置き換え）、遮蔽物除去（例えば片目が顔に遮られると、再度に目を再構成する）、スタイルの変換（1 秒でゴッホ画風へ変更）、位置姿勢の変換（例えば立っている状態から座る状態に変更、顔の正面から側面への変更）、

50

油絵以外のものが油絵に変更、画像背景の色の変更や、画像中のオブジェクト場所の季節の背景の変更等を含むが、これらに限定されない。

【 0 1 4 2 】

好ましくは、上記受信モジュール 1 0 1 が上記音声信号と上記処理対象画像を受信する前に、上記処理モジュール 1 0 3 は画像処理モデルに対して自己適応訓練を行って、上記目標画像処理モデルを取得する。

【 0 1 4 3 】

また、上記画像処理モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【 0 1 4 4 】

具体的に、上記画像処理モデルに対する自己適応訓練がオフラインで行われることは、具体的に、上記処理モジュール 1 0 3 がそのハードウェアを基で上記画像処理モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得することである。上記画像処理モデルに対する自己適応訓練がオンラインで行われることは、具体的に、上記変換モジュールと異なるクラウドサーバが上記画像処理モデルに対して自己適応訓練を行って、目標画像処理モデルを取得することである。上記処理モジュール 1 0 3 が上記目標画像処理モデルを使用する必要がある場合、当該処理モジュール 1 0 3 は上記クラウドサーバから当該目標画像処理モデルを取得する。

【 0 1 4 5 】

好ましくは、上記画像処理モデルに対する自己適応訓練は監督されるか、又は監督されない。

【 0 1 4 6 】

具体的に、上記画像処理モデルに対する自己適応訓練が監督されることは、具体的に、上記処理モジュール 1 0 3 が画像処理モデルに基づいて上記音声信号を予測画像に変換し、その後に、上記予測画像及びその対応する目標画像の相関係数を確定し、当該目標は手動で音声信号に従って処理対象画像を処理することで得られる画像であり、上記処理モジュール 1 0 3 が前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得することである。

【 0 1 4 7 】

また、好ましい一実施例において、上記画像処理装置 1 0 0 は、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するための訓練モジュール 1 0 5 をさらに備える。

【 0 1 4 8 】

一例として、上記画像処理モデルに対する自己適応訓練が監督されることは、具体的に、上記処理モジュール 1 0 3 又は訓練モジュール 1 0 5 が、画像の色の変更や、画像の回転等の関連命令を含む音声信号を受信することである。命令毎にいつの目標画像が対応する。自己適応訓練用の入力した音声信号にとって、対応する目標画像は既知のものであり、上記処理モジュール 1 0 3 又は訓練モジュール 1 0 5 はこれらの音声信号を画像処理モデルの入力データとして、出力後の予測画像を取得する。上記処理モジュール 1 0 3 又は訓練モジュール 1 0 5 は、上記予測画像及びその対応する目標画像の相関係数を算出し、上記画像処理モデルにおけるパラメータ（重みや、オフセット等）を適応的に更新して、上記画像処理モデルの性能を向上し、さらに上記目標画像処理モデルを取得する。

【 0 1 4 9 】

また、好ましい一実施例において、前記画像処理装置 1 0 0 の変換モジュール 1 0 2 は、変換モジュール 1 0 2 における音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得するために用いることができる。また、前記画像処理装置 1 0 0 の処理モジュール 1 0 3 は、処理モジュール 1 0 3 における画像処理モデルに対して自己適応訓練を行って、目標画像処理モデルを取得するために用いることができる。

10

20

30

40

50

【0150】

また、好ましい一実施例において、前記画像処理装置100は、変換モジュール102における音声命令変換モデルと処理モジュール103における画像処理モデルに対してそれぞれ自己適応訓練を行って、目標音声命令変換モデルと目標画像処理モデルを取得するための訓練モジュール105をさらに備える。

【0151】

上記訓練モジュール105は、監督される方法又は監督されない方法によって上記音声命令変換モデル又は上記画像処理モデルにおける構造とパラメータを調整して、当該音声命令変換モデル又は画像処理モデルの性能を向上し、最終に目標音声命令変換モデル又は目標画像処理モデルを取得することができる。

10

【0152】

本実施例において、画像処理装置100はモジュール化されている。本明細書における「モジュール」は、特定用途向け集積回路(application-specific integrated circuit; ASIC)、1つ又は複数のソフトウェア又はファームウェアプログラムを実行するプロセッサ及びメモリ、集積論理回路、及び/又は他の上記機能を提供することができるデバイスを指してもよい。また、以上の受信モジュール101、変換モジュール102、処理モジュール103、記憶モジュール104、及び訓練モジュール105は、図5、図6、図7に示す人工ニューラルネットワークチップによって実現することができる。

【0153】

20

好ましくは、上記画像処理装置100の変換モジュール102又は処理モジュール103の処理手段1032は人工ニューラルネットワークチップである。即ち、上記変換モジュール102と上記処理モジュール103の処理手段1032は独立した2枚の人工ニューラルネットワークチップであり、その構造はそれぞれ図5及び図6に示す通りである。

【0154】

本装置において、変換モジュール102と処理モジュール103は、直列に実行されてもよく、ソフトウェアパイプラインの形態で実行されてもよい。即ち、処理モジュール103が前の画像を処理する時、変換モジュール102は次の画像を処理することができる。そうすると、ハードウェアのスループットを向上し、画像処理効率を向上することができる。

30

【0155】

図5を参照すると、図5は人工ニューラルネットワークチップの構造ブロック模式図である。図5に示すように、当該チップは、制御手段510と、記憶手段520と、入出力手段530とを備える。

【0156】

また、上記制御手段510は、ニューラルネットワーク演算命令と汎用演算命令を含む実行されるべき命令を記憶するための命令キャッシュ手段511を備える。

【0157】

一実施形態において、命令キャッシュ手段511は並べ替えキャッシュであってもよい。

40

【0158】

命令処理手段512は、命令キャッシュ手段からニューラルネットワーク演算命令又は汎用演算命令を取得し、当該命令を処理して上記ニューラルネットワーク演算手段519に供給するためのものである。上記命令処理手段512は、命令キャッシュ手段から命令を取得するための命令取得モジュール513と、取得した命令を復号するための復号モジュール514と、復号後の命令について順に記憶するための命令キューモジュール515と、を備える。

【0159】

スカラレジスタモジュール516は、上記命令に対応するコマンドコードとオペランドを記憶するためのものである。ニューラルネットワーク演算命令に対応するニューラルネ

50

ットワーク演算コマンドコードとオペランド、及び汎用演算命令に対応する汎用演算コマンドコードとオペランドを含んでいる。

【 0 1 6 0 】

処理依存関係モジュール 5 1 7 は、上記命令処理手段 5 1 2 によって送信された命令及びその対応するコマンドコードとオペランドを判断するためのものであり、当該命令と前の命令とが同じデータにアクセスするかどうかを判断し、同じデータにアクセスすると、当該命令を記憶キュー手段 5 1 8 に記憶し、前の命令の実行が完了した後に、記憶キュー手段における当該命令を上記ニューラルネットワーク演算手段 5 1 9 に与える。同じデータにアクセスしないと、当該命令を直接に上記ニューラルネットワーク演算手段 5 1 9 に与える。

10

【 0 1 6 1 】

記憶キュー手段 5 1 8 は、命令が記憶手段にアクセスする時、同じ記憶スペースにアクセスする連続する 2 つの命令を記憶するためのものである。

【 0 1 6 2 】

具体的に、上記連続する 2 つの命令の実行結果の正確性を保証するために、現在の命令が前の命令のデータに依存することが検出された場合、当該連続する 2 つの命令は依存がなくなるまで上記記憶キュー手段 5 1 8 内で待たなければならない。そうしないと、当該連続する 2 つの命令を上記ニューラルネットワーク演算手段に与えることができない。

【 0 1 6 3 】

ニューラルネットワーク演算手段 5 1 9 は、命令処理手段又は記憶キュー手段から伝送される命令を処理するためのものである。

20

【 0 1 6 4 】

記憶手段 5 2 0 は、ニューロンキャッシュ手段 5 2 1 と重みキャッシュ手段 5 2 2 とを備え、ニューラルネットワークデータモデルは上記ニューロンキャッシュ手段 5 2 1 と重みキャッシュ手段 5 2 2 に記憶されている。

【 0 1 6 5 】

入出力手段 5 3 0 は、音声信号を入力し、画像処理命令を出力するためのものである。

【 0 1 6 6 】

一実施形態において、記憶手段 5 2 0 はスクラッチパッドメモリであってもよく、入出力手段 5 3 0 は I O 直接メモリアクセスモジュールであってもよい。

30

【 0 1 6 7 】

具体的に、上記チップ 5 0 0、即ち上記処理モジュール 1 0 2 が音声信号を画像処理命令に変換することは、具体的に以下のステップを含む。

【 0 1 6 8 】

ステップ 5 0 1 では、命令取得モジュール 5 1 3 が命令キャッシュ手段 5 1 1 から 1 つの音声認識用の演算命令を取り出し、演算命令を復号モジュール 5 1 4 に送る。

【 0 1 6 9 】

ステップ 5 0 2 では、復号モジュール 5 1 4 が演算命令を復号し、復号後の命令を命令キュー手段 5 1 5 に送る。

【 0 1 7 0 】

40

ステップ 5 0 3 では、スカラレジスタモジュール 5 1 6 から前記命令に対応するニューラルネットワーク演算コマンドコードとニューラルネットワーク演算オペランドを取得する。

【 0 1 7 1 】

ステップ 5 0 4 では、命令が処理依存関係モジュール 5 1 7 に送られ、処理依存関係モジュール 5 1 7 は命令に対応するコマンドコードとオペランドに対して判断を行い、命令と前の実行完了ではない命令とがデータ上で依存関係が存在するかどうかを判断し、存在しないと、前記命令を直接にニューラルネットワーク演算手段 5 1 9 に送り、存在すると、命令は前の実行完了ではない命令とデータ上でもはや依存関係が存在しないまで記憶キュー手段 5 1 8 で待ちする必要がある、その後前記命令をニューラルネットワーク演算

50

手段 5 1 9 に送信する。

【 0 1 7 2 】

ステップ 5 0 5 では、ニューラルネットワーク演算手段 5 1 9 が命令に対応するコマンドコードとオペランドに従って必要なデータのアドレスとサイズを確定し、記憶手段 5 2 0 から音声命令変換モデルデータ等を含む必要なデータを取り出す。

【 0 1 7 3 】

ステップ 5 0 6 では、ニューラルネットワーク演算手段 5 1 9 が前記命令に対応するニューラルネットワーク演算を実行し、対応する処理を終了し、画像処理命令を取得し、画像処理命令を記憶手段 5 2 0 に書き戻しする。

【 0 1 7 4 】

図 6 を参照すると、図 6 は他の人工ニューラルネットワークチップの構造ブロック模式図である。図 6 に示すように、当該チップは、制御手段 6 1 0 と、記憶手段 6 2 0 と、入出力手段 6 3 0 とを備える。

【 0 1 7 5 】

また、上記制御手段 6 1 0 は、ニューラルネットワーク演算命令と汎用演算命令を含む実行されるべき命令を記憶するための命令キャッシュ手段 6 1 1 を備える。

【 0 1 7 6 】

一実施形態において、命令キャッシュ手段 6 1 1 は並べ替えキャッシュであってもよい。

【 0 1 7 7 】

命令処理手段 6 1 2 は、命令キャッシュ手段からニューラルネットワーク演算命令又は汎用演算命令を取得し、当該命令を処理して上記ニューラルネットワーク演算手段 6 1 9 に与えるためのものである。また、上記命令処理手段 6 1 2 は、命令キャッシュ手段から命令を取得するための命令取得モジュール 6 1 3 と、取得した命令を復号するための復号モジュール 6 1 4 と、復号後の命令について順に記憶するための命令キューモジュール 6 1 5 と、を備える。

【 0 1 7 8 】

スカラレジスタモジュール 6 1 6 は、上記命令に対応するコマンドコードとオペランドを記憶するためのものである。ニューラルネットワーク演算命令に対応するニューラルネットワーク演算コマンドコードとオペランド、及び汎用演算命令に対応する汎用演算コマンドコードとオペランドを含んでいる。

【 0 1 7 9 】

処理依存関係モジュール 6 1 7 は、上記命令処理手段 6 1 2 によって送信された命令及びその対応するコマンドコードとオペランドを判断するためのものであり、当該命令と前の命令とが同じデータにアクセスするかどうかを判断し、同じデータにアクセスすると、当該命令を記憶キュー手段 6 1 8 に記憶し、前の命令の実行が完了した後に、記憶キュー手段における当該命令を上記ニューラルネットワーク演算手段 6 1 9 に与える。同じデータにアクセスしないと、当該命令を直接に上記ニューラルネットワーク演算手段 6 1 9 に与える。

【 0 1 8 0 】

記憶キュー手段 6 1 8 は、命令が記憶手段にアクセスする時、同じ記憶スペースにアクセスする連続する 2 つの命令を記憶するためのものである。

【 0 1 8 1 】

具体的に、上記連続する 2 つの命令の実行結果の正確性を保証するために、現在の命令が前の命令のデータに依存することが検出された場合、当該連続する 2 つの命令は依存がなくなるまで上記記憶キュー手段 6 1 8 内で待たなければならない。そうしないと、当該連続する 2 つの命令を上記ニューラルネットワーク演算手段に与えることができない。

【 0 1 8 2 】

ニューラルネットワーク演算手段 6 1 9 は、命令処理手段又は記憶キュー手段から伝送される命令を処理するためのものである。

10

20

30

40

50

【 0 1 8 3 】

記憶手段 6 2 0 は、ニューロンキャッシュ手段 6 2 1 と重みキャッシュ手段 6 2 2 とを備え、ニューラルネットワークデータモデルは上記ニューロンキャッシュ手段 6 2 1 と重みキャッシュ手段 6 2 2 に記憶されている。

【 0 1 8 4 】

入出力手段 6 3 0 は、画像処理命令と処理対象画像を入力し、処理後の画像を出力するためのものである。

【 0 1 8 5 】

一実施形態において、記憶手段 6 2 0 はスクラッチパッドメモリであってもよく、入出力手段 6 3 0 は I O 直接メモリアクセスモジュールであってもよい。

10

【 0 1 8 6 】

上記チップ、即ち上記処理モジュール 1 0 3 の処理手段 1 0 3 2 が行う画像処理は、具体的に以下のようなステップを含む。

【 0 1 8 7 】

ステップ 6 0 1 では、命令取得モジュール 6 1 3 が命令キャッシュ手段 6 1 1 から 1 つの変換モジュールの生成した画像処理の命令を取り出し、演算命令を復号モジュール 6 1 4 に送る。

【 0 1 8 8 】

ステップ 6 0 2 では、復号モジュール 6 1 4 が演算命令を復号し、復号後の命令を命令キュー手段 8 1 5 に送信する。

20

【 0 1 8 9 】

ステップ 6 0 3 では、スカラレジスタモジュール 6 1 6 から前記命令に対応するニューラルネットワーク演算コマンドコードとニューラルネットワーク演算オペランドを取得する。

【 0 1 9 0 】

ステップ 6 0 4 では、命令が処理依存関係モジュール 6 1 7 に送られ、処理依存関係モジュール 6 1 7 は命令に対応するコマンドコードとオペランドに対して判断を行い、命令と前の実行完了ではない命令とがデータ上で依存関係が存在するかどうかを判断し、存在しないと、前記命令を直接にニューラルネットワーク演算手段 6 1 9 に送り、存在すると、命令は前の実行完了ではない命令とデータ上でもはや依存関係が存在しないまで記憶キュー手段 6 1 8 で待ちする必要があるし、その後前記命令に対応するマイクロ命令をニューラルネットワーク演算手段 6 1 9 に送信する。

30

【 0 1 9 1 】

ステップ 6 0 5 では、ニューラルネットワーク演算手段 6 1 9 が前記命令に対応するコマンドコードとオペランドに従って必要なデータのアドレスとサイズを確定し、記憶手段 6 2 0 から処理対象画像や、画像処理モデルデータ等を含む必要なデータを取り出す。

【 0 1 9 2 】

ステップ 6 0 6 では、ニューラルネットワーク演算手段 6 1 9 が前記命令に対応するニューラルネットワーク演算を実行し、対応する処理を終了し、処理結果を記憶手段 6 2 0 に書き戻しする。

40

【 0 1 9 3 】

なお、上記変換モジュール 1 0 2 と上記処理モジュール 1 0 3 の処理手段 1 0 3 2 はいずれも人工ニューラルネットワークチップ又は汎用処理チップであってもよい。或いは、上記変換モジュール 1 0 2 と上記処理モジュール 1 0 3 の処理手段 1 0 3 2 のうち、一方が人工ニューラルネットワークチップであり、他方が汎用処理チップであってもよい。

【 0 1 9 4 】

好ましくは、上記画像処理装置は、データ処理装置、ロボット、コンピュータ、タブレット、スマート端末、携帯電話、クラウドサーバ、カメラ、ビデオカメラ、プロジェクタ、腕時計、イヤホン、携帯メモリ又はウェアラブル装置であってもよい。

【 0 1 9 5 】

50

なお、本発明の実施例の技術案において、受信モジュールが音声信号と処理対象画像を受信して取得し、変換モジュールが目標音声命令変換モデルに従って前記音声信号を画像処理命令と前記処理対象画像に対する処理対象領域に変換し、処理モジュールが前記画像処理命令と目標画像処理モデルに従って前記処理対象画像の処理領域を処理することができる。従来の画像処理技術と比べると、本発明は、音声によって画像処理を行うので、画像処理前にユーザーが画像処理ソフトウェアを学習する時間を省き、ユーザーの経験を向上させている。

【0196】

第2の態様として、本発明の実施例は、音声信号と処理対象画像を受信するための受信モジュールと、目標音声命令変換モデルに従って前記音声信号を画像処理命令と前記処理対象画像の処理領域である目標領域に変換するための変換モジュールと、前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するための処理モジュールと、を備える画像処理装置をさらに提供する。

10

【0197】

また、前記変換モジュールは、具体的に、音声認識技術によって前記音声信号をテキスト情報に変換し、自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像処理命令に変換し、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【0198】

20

また、前記変換モジュールは、さらに具体的に、音声認識技術、語義理解技術及び前記目標音声命令変換モデルによって前記音声信号を前記画像処理命令に変換し、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【0199】

また、前記画像処理装置は、前記テキスト情報又は前記画像処理命令又は前記目標領域を記憶するための記憶モジュールをさらに備える。

【0200】

また、前記処理モジュールは、具体的に、予め設定された時間ウィンドウ内で前記記憶モジュールからM(Mは1より大きい整数)個の画像処理命令を取得し、前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するために用いられる。

30

【0201】

また、前記処理モジュールは、さらに具体的に、前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N個の画像処理命令を取得し、前記N個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するために用いられる。

【0202】

また、前記変換モジュールは、音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得する。

【0203】

40

また、前記変換モジュールによる前記音声命令変換モデルの自己適応訓練はオフライン又はオンラインで行われる。

【0204】

また、前記変換モジュールによる前記音声命令変換モデルの自己適応訓練は監督されるか、又は監督されない。

【0205】

また、前記変換モジュールは、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するためにも用いられる。

50

【0206】

また、前記画像処理装置は、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するための訓練モジュールをさらに備える。

【0207】

また、前記処理モジュールは、画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得する。

【0208】

また、前記処理モジュールによる前記画像処理モデルの自己適応訓練はオフライン又はオンラインで行われる。

10

【0209】

また、前記処理モジュールによる前記画像処理モデルの自己適応訓練は監督されるか、又は監督されない。

【0210】

また、前記処理モジュールは、前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、前記予測画像及びその対応する目標画像の相関係数を確定し、前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するためにも用いられる。

20

【0211】

また、前記訓練モジュールは、前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得し、前記予測画像及びその対応する目標画像の相関係数を確定し、前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するためにも用いられる。

【0212】

なお、上記画像処理装置の実現過程は、図1に示す実施例の具体的な実現過程を参照することができ、ここではさらに詳しく説明しない。

【0213】

第3の態様において、本発明の実施例は画像処理方法を提供する。図7に示すように、当該方法は、画像処理装置が音声信号と処理対象画像を受信するステップS701と、前記画像処理装置が目標音声命令変換モデルに従って前記音声信号を画像処理命令と前記処理対象画像の処理領域である目標領域に変換するステップS702と、を含む。

30

【0214】

また、好ましい一実施例において、前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、音声認識技術によって前記音声信号をテキスト情報に変換するステップと、自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像処理命令に変換するステップと、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0215】

40

また、好ましい一実施例において、前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、音声認識技術、語義理解技術及び前記音声命令変換モデルによって前記音声信号を前記画像処理命令に変換するステップと、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0216】

また、好ましい一実施例において、前記方法は、前記テキスト情報又は前記画像処理命令又は前記目標領域を記憶するステップをさらに含む。

【0217】

ステップS703は、前記画像処理装置が前記画像処理命令と目標画像処理モデルに従

50

って前記目標領域を処理するステップである。

【0218】

また、好ましい一実施例において、前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するステップは、予め設定された時間ウィンドウ内で前記記憶モジュールから M (M は1より大きい整数)個の画像処理命令を取得するステップと、前記 M 個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含む。

【0219】

また、好ましい一実施例において、前記 M 個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップは、前記 M 個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、 N (N は前記 M よりも小さい整数)個の画像処理命令を取得するステップと、前記 N 個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含む。

10

【0220】

また、好ましい一実施例において、前記音声信号と処理対象画像を受信するステップの前に、前記方法は、音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得するステップをさらに含む。

【0221】

また、好ましい一実施例において、前記音声命令変換モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

20

【0222】

また、好ましい一実施例において、前記音声命令変換モジュールに対する自己適応訓練は監督されるか、又は監督されない。

【0223】

また、好ましい一実施例において、前記音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得するステップは、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換するステップと、前記予測命令及びその対応する命令セットの相関係数を確定するステップと、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するステップと、を含む。

30

【0224】

また、好ましい一実施例において、前記音声信号と処理対象画像を受信するステップの前に、前記方法は、画像処理モデルに対して自己適応訓練を行って、目標画像処理モデルを取得するステップをさらに含む。

【0225】

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【0226】

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練は監督されるか、又は監督されない。

40

【0227】

また、好ましい一実施例において、前記画像処理モデルに対して自己適応訓練を行って、目標画像処理モデルを取得するステップは、前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得するステップと、前記予測画像及びその対応する目標画像の相関係数を確定するステップと、前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するステップと、を含む。

【0228】

なお、図7に示す方法の各ステップの具体的な実現過程は、上記画像処理装置の具体的な実現過程を参照すればよい。よって、ここではさらに説明しない。

50

【 0 2 2 9 】

本発明の実施例はコンピュータ記憶媒体も提供している。当該コンピュータ記憶媒体はプログラムを記憶することができる。当該プログラムは、実行時に上記した方法実施例に記載の上記のいずれか画像処理方法の一部又は全部のステップを含む。

【 0 2 3 0 】

図 8 を参照すると、図 8 は本発明の実施例が提供する画像処理装置の構造模式図である。図 8 に示すように、当該画像処理装置 8 0 0 は、ユーザーが入力した音声信号を収集するための音声収集装置 8 0 1 を備える。

【 0 2 3 1 】

好ましくは、上記画像処理装置 8 0 0 は、騒音フィルタをさらに備える。上記音声収集装置 8 0 1 が上記音声信号を収集した後、上記騒音フィルタは当該音声信号に対して騒音低減処理を行う。

【 0 2 3 2 】

好ましくは、当該音声収集装置は、音声センサ、マイクロフォン、ピックアップや、他のオーディオ収集装置であってもよい。

【 0 2 3 3 】

具体的に、上記音声収集装置 8 0 1 は、上記音声信号を受信する時、環境音声信号も受信する。上記騒音フィルタは、上記環境音声信号に基づいて上記音声信号に騒音低減処理を行う。当該環境音声信号は、上記音声信号にとっては騒音である。

【 0 2 3 4 】

さらに、上記オーディオ収集装置 8 0 1 は、上記音声信号と上記環境音声信号の収集に用いるだけでなく、騒音低減処理も実現するペアマイクロフォンアレイを含んでもよい。

【 0 2 3 5 】

好ましくは、1つの可能な実施例において、上記画像処理装置は第1のメモリをさらに備える。上記音声収集装置が上記音声信号を収集した後、上記画像処理装置は上記音声信号を第1のメモリに記憶する。

【 0 2 3 6 】

命令変換装置 8 0 2 は、目標音声命令変換モデルに基づいて前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換するために用いられる。

【 0 2 3 7 】

好ましくは、上記命令変換装置 8 0 2 は、音声認識技術、自然言語処理技術及び画像認識技術に基づいて前記音声信号を画像処理命令と目標領域に変換する前に、上記命令変換装置 8 0 2 は上記第1のメモリから上記音声信号を取得する。

【 0 2 3 8 】

また、前記命令変換装置 8 0 2 は、音声認識技術によって前記音声信号をテキスト情報に変換するための第1の音声認識装置 8 0 2 1 と、自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像処理命令に変換するための音声テキスト変換装置 8 0 2 2 と、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するための第1の画像認識装置 8 0 2 3 と、を備える。

【 0 2 3 9 】

さらに、上記命令変換装置 8 0 2 は、上記画像処理命令における語義領域の粒度を取得するための取得モジュール 8 0 2 6 をさらに備える。

【 0 2 4 0 】

例を挙げて上記語義領域を説明すると、上記画像処理装置 8 0 0 が音声信号に応じて上記目標領域が顔領域であると確定した場合、上記語義領域は上記処理対象画像における顔領域であり、上記画像処理装置は顔を粒度として、上記処理対象画像における複数の顔領域を取得する。上記目標領域が背景である場合、上記画像処理装置は上記処理対象画像を背景領域と非背景領域に分割し、上記目標領域が赤色領域である場合、上記画像処理装置は上記処理対象画像を色に従って異なる色の領域に分割する。

10

20

30

40

50

【0241】

具体的に、本発明で使用される音声認識技術は、人工ニューラルネットワーク（Artificial Neural Network；ANN）、隠れマルコフモデル（Hidden Markov Model；HMM）等のモデルを含むが、これらに限定されない。上記第1の音声認識手段は、上記音声認識技術に基づいて上記音声信号を処理してもよい。上記自然言語処理技術は、統計的機械学習や、ANN等の方法を利用することを含むが、これらに限定されない。上記語義理解手段は、上記自然言語処理技術に基づいて語義情報を抽出することができる。上記画像認識技術は、エッジ検出に基づく方法、閾値分割方法、領域成長及び分水嶺（watershed）アルゴリズム、階調積分投影（Gray-level Integration Projection）曲線分析、テンプレートマッチング、変形可能テンプレート、ハフ（Hough）変換、スネーク（Snake）演算子、Gaborウェーブレット変換に基づく弾性マップマッチング技術や、アクティブ形状モデル及びアクティブ外観モデル等の方法等のアルゴリズムを含むが、これらに限定されない。上記画像認識手段は、上記画像認識技術に基づいて上記処理対象画像を異なる領域に分割することができる。

10

【0242】

また、好ましい一実施例において、上記第1の音声認識装置8021は上記音声認識技術によって上記音声信号をテキスト情報に変換し、当該テキスト情報を上記第1のメモリに保存する。上記音声テキスト変換装置8022は上記第1のメモリから上記テキスト情報を取得し、自然言語処理技術と上記目標音声命令変換モデルによって当該テキスト情報を画像処理命令に変換し、上記画像処理命令を上記第1のメモリに保存する。上記第1の画像認識装置8023は上記画像処理命令における語義領域の粒度と画像認識技術に基づいて上記処理対象画像を領域分割し、上記目標領域を取得し、上記分割結果と上記目標領域を上記第2のメモリに記憶する。

20

【0243】

また、好ましい一実施例において、上記命令変換装置802は、音声認識技術、自然言語処理技術及び上記目標音声命令変換モデルに基づいて上記音声信号を直接に上記画像処理命令に変換し、当該画像処理命令を第1のメモリに保存するための第2の音声認識装置8025と、上記画像処理命令における上記処理対象画像を操作する語義領域の粒度に基づいて、当該処理対象画像を当該語義領域の粒度に従って分割し、上記処理対象画像を処理する領域である目標領域を取得し、分割後の結果及び上記目標領域を第2のメモリに記憶するための第2の画像認識装置8026と、を備えてもよい。

30

【0244】

好ましくは、上記音声収集装置801が上記音声信号と上記処理対象画像を受信する前に、上記命令変換装置802が音声命令変換モデルに対して自己適応訓練を行って、上記目標音声命令変換モデルを取得する。

【0245】

また、上記音声命令変換モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【0246】

具体的に、上記音声命令変換モデルに対する自己適応訓練がオフラインで行われることは、具体的に、上記命令変換装置802がそのハードウェアを基で上記音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得することである。上記音声命令変換モデルに対する自己適応訓練がオンラインで行われることは、具体的に、上記命令変換装置802と異なるクラウドサーバが上記音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得することである。上記命令変換装置802が上記目標音声命令変換モデルを使用する必要がある場合、当該命令変換装置802は上記クラウドサーバから当該目標音声命令変換モデルを取得する。

40

【0247】

好ましくは、上記音声命令変換モデルに対する自己適応訓練は監督されるか、又は監督

50

されない。

【0248】

具体的に、上記音声命令変換モデルに対する自己適応訓練が監督されることは、具体的に、上記命令変換装置802が音声命令変換モデルに基づいて上記音声信号を予測命令に変換し、その後、上記予測命令及びその対応する命令セットの相関係数を確定し、当該命令セットは手動で音声信号に従って取得される命令のセットであり、上記命令変換装置802が前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得することである。

【0249】

また、好ましい一実施例において、上記画像処理装置800は、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するための訓練装置805をさらに備える。

10

【0250】

一例として、上記音声命令変換モデルに対する自己適応訓練が監督されることは、具体的に、上記命令変換装置802又は訓練装置805が、画像の色の変更や、画像の回転等の関連命令を含む音声信号を受信することである。命令毎に1つの命令セットが対応する。自己適応訓練用の入力した音声信号にとって、対応する命令セットは既知のものであり、上記命令変換装置802又は訓練装置805はこれらの音声信号を音声命令変換モデルの入力データとして、出力後の予測命令を取得する。上記命令変換装置802又は訓練装置805は、上記予測命令及びその対応する命令セットの相関係数を算出し、上記音声命令変換モデルにおけるパラメータ（重みや、オフセット等）を適応的に更新して、上記音声命令変換モデルの性能を向上し、さらに上記目標音声命令変換モデルを取得する。

20

【0251】

前記画像処理装置800は、前記テキスト情報又は前記画像処理命令又は前記目標領域を記憶するためのメモリ804をさらに備える。

【0252】

また、好ましい一実施例において、上記メモリ804は、上記第1の記憶モジュール及び第2の記憶モジュールと同一の記憶モジュールであってもよく、異なる記憶モジュールであってもよい。

30

【0253】

画像処理機803は、前記画像処理命令と目標画像処理モデルに基づいて前記処理対象画像を処理するためのものである。

【0254】

また、前記画像処理機803は、予め設定された時間ウィンドウ内で前記記憶モジュールからM（Mは1より大きい整数）個の画像処理命令を取得するための命令取得モジュール8031と、前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するための処理モジュール8032と、を備える。

【0255】

好ましくは、前記処理モジュール8032は、前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N（Nは前記Mよりも小さい整数）個の画像処理命令を取得し、前記N個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するために用いられる。

40

【0256】

具体的に、上記予め設定された時間ウィンドウは、予め設定された時間長であると理解してもよい。予め設定された時間長内で上記取得手段8031は上記記憶モジュール804からM個の画像処理命令を取得した後、上記処理モジュール8032は上記M個の画像処理命令について2つずつ比較し、当該M個の画像処理命令のうち同じ機能を有する命令を削除して、N個の画像処理命令を取得する。上記処理モジュール8032は上記N個の

50

処理命令と上記目標画像処理モデルに基づいて上記処理対象画像を処理する。

【0257】

一例として、上記処理モジュール8032は上記M個の画像処理命令について2つずつ比較する。画像処理命令Aと画像処理命令Bとが同じである場合、上記処理モジュール8032は上記画像処理命令A及びBのうち最大のオーバーヘッドを削除し、画像処理命令Aと画像処理命令Bとが異なる場合、上記処理モジュール8032は上記画像処理命令Aと上記画像処理命令Bの類似度係数を取得する。当該類似度係数が類似度閾値よりも大きい時、上記画像処理命令Aと上記画像処理命令Bの機能が同じであることを確定し、上記処理モジュール8032は上記画像処理命令A及びBのうち最大のオーバーヘッドを削除し、上記類似度係数が上記類似度閾値よりも小さい時、上記処理モジュール8032は上記画像処理命令A及びBの機能が異なることを確定する。当該画像処理命令A及びBは上記N個の処理命令のうちの任意の2つである。

10

【0258】

具体的に、上記画像処理機803に対し、その入力と出力はいずれも画像である。上記画像処理機803のANNや従来のコンピュータビジョン方法に限られない方法によって上記処理対象画像に対して行う処理は、美体（例えば美脚、豊胸）、顔の変更、顔の美化、オブジェクトの変更（猫と犬の交換、シマウマの馬への変更、リンゴとオレンジの交換等）、背景の変更（後の森を畑に置き換え）、遮蔽物除去（例えば片目が顔に遮られると、再度に目を再構成する）、スタイルの変換（1秒でゴッホ画風へ変更）、位置姿勢の変換（例えば立っている状態から座る状態に変更、顔の正面から側面への変更）、油絵以外のものが油絵に変更、画像背景の色の変更や、画像中のオブジェクト場所の季節の背景の変更等を含むが、これらに限定されない。

20

【0259】

好ましくは、上記音声収集装置801が上記音声信号を受信する前に、上記画像処理機803は画像処理モデルに対して自己適応訓練を行って、上記目標画像処理モデルを取得する。

【0260】

また、上記画像処理モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【0261】

具体的に、上記画像処理モデルに対する自己適応訓練がオフラインで行われることは、具体的に、上記画像処理機803がそのハードウェアを基で上記画像処理モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得することである。上記画像処理モデルに対する自己適応訓練はオンラインで行われることは、具体的に、上記画像処理機803と異なるクラウドサーバが上記画像処理モデルに対して自己適応訓練を行って、目標画像処理モデルを取得することである。上記画像処理機803が上記目標画像処理モデルを使用する必要がある場合、当該画像処理機803は上記クラウドサーバから当該目標画像処理モデルを取得する。

30

【0262】

好ましくは、上記画像処理モデルに対する自己適応訓練は監督されるか、又は監督されない。

40

【0263】

具体的に、上記画像処理モデルに対する自己適応訓練が監督されることは、具体的に、上記画像処理機803が画像処理モデルに基づいて上記音声信号を予測画像に変換し、その後、上記予測画像及びその対応する目標画像の相関係数を確定し、当該目標は手動で音声信号に従って処理対象画像を処理することで得られる画像であり、上記画像処理機803が前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得することである。

【0264】

また、好ましい一実施例において、上記画像処理装置800は、前記音声命令変換モデ

50

ルに基づいて前記音声信号を予測命令に変換し、前記予測命令及びその対応する命令セットの相関係数を確定し、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するための訓練装置 805 をさらに備える。

【0265】

一例として、上記画像処理モデルに対する自己適応訓練が監督されることは、具体的に、上記画像処理機 803 又は訓練装置 805 が、画像の色の変更や、画像の回転等の関連命令を含む音声信号を受信することである。命令毎に 1 枚の目標画像が対応する。自己適応訓練用の入力した音声信号にとって、対応する目標画像は既知のものであり、上記画像処理機 803 又は訓練装置 805 はこれらの音声信号を画像処理モデルの入力データとして、出力後の予測画像を取得する。上記画像処理機 803 又は訓練装置 805 は、上記予測画像及びその対応する目標画像の相関係数を算出し、上記画像処理モデルにおけるパラメータ（重みや、オフセット等）を適応的に更新して、上記画像処理モデルの性能を向上し、さらに上記目標画像処理モデルを取得する。

10

【0266】

また、好ましい一実施例において、前記画像処理装置 800 の命令変換装置 802 は、命令変換装置 802 における音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得するために用いることができる。前記画像処理装置 800 の画像処理機 803 は、画像処理機 803 における画像処理モデルに対して自己適応訓練を行って、目標画像処理モデルを取得するために用いることができる。

20

【0267】

また、好ましい一実施例において、前記画像処理装置 800 は、命令変換装置 802 における音声命令変換モデルと画像処理機 803 における画像処理モデルに対してそれぞれ自己適応訓練を行って、目標音声命令変換モデルと目標画像処理モデルを取得するための訓練装置 805 をさらに備える。

【0268】

上記訓練装置 805 は、監督される方法又は監督されない方法によって上記音声命令変換モデル又は上記画像処理モデルにおける構造とパラメータを調整して、当該音声命令変換モデル又は画像処理モデルの性能を向上し、最終に目標音声命令変換モデル又は目標画像処理モデルを取得することができる。

30

【0269】

本実施例において、画像処理装置 800 はモジュール化されている。本明細書における「モジュール」は、特定用途向け集積回路（application-specific integrated circuit; ASIC）、1 つ又は複数のソフトウェア又はファームウェアプログラムを実行するプロセッサ及びメモリ、集積論理回路、及び/又は他の上記機能を提供することができるデバイスを指してもよい。また、以上の音声収集装置 801、命令変換装置 802、画像処理機 803、記憶モジュール 804、及び訓練装置 805 は、図 5、図 6、図 12、図 8 に示す人工ニューラルネットワークチップによって実現することができる。

【0270】

好ましくは、上記画像処理装置 800 の命令変換装置 802 又は画像処理機 803 の処理モジュール 8032 は、人工ニューラルネットワークチップである。即ち、上記命令変換装置 802 と上記画像処理機 803 の処理モジュール 8032 は独立した 2 枚の人工ニューラルネットワークチップであり、その構造はそれぞれ図 5 及び図 6 に示す通りである。

40

【0271】

本装置において、命令変換装置 802 と画像処理機 803 は、直列に実行されてもよく、ソフトウェアパイプラインの形態で実行されてもよい。即ち、画像処理機 803 が前の画像を処理する時、命令変換装置 802 は次の画像を処理することができる。そうすると、ハードウェアのスループットを向上し、画像処理効率を向上することができる。

50

【 0 2 7 2 】

図 5 を参照すると、図 5 は人工ニューラルネットワークチップの構造ブロック模式図である。図 5 に示すように、当該チップは、制御手段 5 1 0 と、記憶手段 5 2 0 と、入出力手段 5 3 0 とを備える。

【 0 2 7 3 】

また、上記制御手段 5 1 0 は、ニューラルネットワーク演算命令と汎用演算命令を含む実行されるべき命令を記憶するための命令キャッシュ手段 5 1 1 を備える。

【 0 2 7 4 】

－実施形態において、命令キャッシュ手段 5 1 1 は並べ替えキャッシュであってもよい。

10

【 0 2 7 5 】

命令処理手段 5 1 2 は、命令キャッシュ手段からニューラルネットワーク演算命令又は汎用演算命令を取得し、当該命令を処理して上記ニューラルネットワーク演算手段 5 1 9 に供給するためのものである。上記命令処理手段 5 1 2 は、命令キャッシュ手段から命令を取得するための命令取得モジュール 5 1 3 と、取得した命令を復号するための復号モジュール 5 1 4 と、復号後の命令について順に記憶するための命令キューモジュール 5 1 5 と、を備える。

【 0 2 7 6 】

スカラーレジスタモジュール 5 1 6 は、上記命令に対応するコマンドコードとオペランドを記憶するためのものである。ニューラルネットワーク演算命令に対応するニューラルネットワーク演算コマンドコードとオペランド、及び汎用演算命令に対応する汎用演算コマンドコードとオペランドを含んでいる。

20

【 0 2 7 7 】

処理依存関係モジュール 5 1 7 は、上記命令処理手段 5 1 2 によって送信された命令及びその対応するコマンドコードとオペランドを判断するためのものであり、当該命令と前の命令とが同じデータにアクセスするかどうかを判断し、同じデータにアクセスすると、当該命令を記憶キュー手段 5 1 8 に記憶し、前の命令の実行が完了した後に、記憶キュー手段における当該命令を上記ニューラルネットワーク演算手段 5 1 9 に与える。同じデータにアクセスしないと、当該命令を直接に上記ニューラルネットワーク演算手段 5 1 9 に与える。

30

【 0 2 7 8 】

記憶キュー手段 5 1 8 は、命令が記憶手段にアクセスする時、同じ記憶スペースにアクセスする連続する 2 つの命令を記憶するためのものである。

【 0 2 7 9 】

具体的に、上記連続する 2 つの命令の実行結果の正確性を保証するために、現在の命令が前の命令のデータに依存することが検出された場合、当該連続する 2 つの命令は依存がなくなるまで上記記憶キュー手段 5 1 8 内で待たなければならない。そうしないと、当該連続する 2 つの命令を上記ニューラルネットワーク演算手段に与えることができない。

【 0 2 8 0 】

ニューラルネットワーク演算手段 5 1 9 は、命令処理手段又は記憶キュー手段から伝送される命令を処理するためのものである。

40

【 0 2 8 1 】

記憶手段 5 2 0 は、ニューロンキャッシュ手段 5 2 1 と重みキャッシュ手段 5 2 2 とを備え、ニューラルネットワークデータモデルは上記ニューロンキャッシュ手段 5 2 1 と重みキャッシュ手段 5 2 2 に記憶されている。

【 0 2 8 2 】

入出力手段 5 3 0 は、音声信号を入力し、画像処理命令を出力するためのものである。

【 0 2 8 3 】

－実施形態において、記憶手段 5 2 0 はスクラッチパッドメモリであってもよく、入出力手段 5 3 0 は I O 直接メモリアクセスモジュールであってもよい。

50

【0284】

具体的に、上記チップ500、即ち上記命令変換装置102が音声信号を画像処理命令に変換することは、具体的に以下のステップを含む。

【0285】

ステップ501では、命令取得モジュール513が命令キャッシュ手段511から1つの音声認識用の演算命令を取り出し、演算命令を復号モジュール514に送る。

【0286】

ステップ502では、復号モジュール514が演算命令を復号し、復号後の命令を命令キュー手段515に送る。

【0287】

ステップ503では、スカラレジスタモジュール516から前記命令に対応するニューラルネットワーク演算コマンドコードとニューラルネットワーク演算オペランドを取得する。

【0288】

ステップ504では、命令が処理依存関係モジュール517に送られ、処理依存関係モジュール517は命令に対応するコマンドコードとオペランドに対して判断を行い、命令と前の実行完了ではない命令とがデータ上で依存関係が存在するかどうかを判断し、存在しないと、前記命令を直接にニューラルネットワーク演算手段519に送り、存在すると、命令は前の実行完了ではない命令とデータ上でもはや依存関係が存在しないまで記憶キュー手段518で待ちする必要がある、その後に前記命令をニューラルネットワーク演算手段519に送信する。

【0289】

ステップ505では、ニューラルネットワーク演算手段519が命令に対応するコマンドコードとオペランドに従って必要なデータのアドレスとサイズを確定し、記憶手段520から音声命令変換モデルデータ等を含む必要なデータを取り出す。

【0290】

ステップ506では、ニューラルネットワーク演算手段519が前記命令に対応するニューラルネットワーク演算を実行し、対応する処理を終了し、画像処理命令を取得し、画像処理命令を記憶手段520に書き戻しする。

【0291】

なお、上記記憶手段520は図5に示すチップのキャッシュ・オン・チップ手段である。

【0292】

図6を参照すると、図6は他の人工ニューラルネットワークチップの構造ブロック模式図である。図6に示すように、当該チップは、制御手段610と、記憶手段620と、入出力手段630とを備える。

【0293】

また、上記制御手段610は、ニューラルネットワーク演算命令と汎用演算命令を含む実行されるべき命令を記憶するための命令キャッシュ手段611を備える。

【0294】

一実施形態において、命令キャッシュ手段611は並べ替えキャッシュであってもよい。

【0295】

命令処理手段612は、命令キャッシュ手段からニューラルネットワーク演算命令又は汎用演算命令を取得し、当該命令を処理して上記ニューラルネットワーク演算手段619に与えるためのものである。また、上記命令処理手段612は、命令キャッシュ手段から命令を取得するための命令取得モジュール613と、取得した命令を復号するための復号モジュール614と、復号後の命令について順に記憶するための命令キューモジュール615と、を備える。

【0296】

スカラレジスタモジュール 6 1 6 は、上記命令に対応するコマンドコードとオペランドを記憶するためのものである。ニューラルネットワーク演算命令に対応するニューラルネットワーク演算コマンドコードとオペランド、及び汎用演算命令に対応する汎用演算コマンドコードとオペランドを含んでいる。

【 0 2 9 7 】

処理依存関係モジュール 6 1 7 は、上記命令処理手段 6 1 2 によって送信された命令及びその対応するコマンドコードとオペランドを判断するためのものであり、当該命令と前の命令とが同じデータにアクセスするかどうかを判断し、同じデータにアクセスすると、当該命令を記憶キュー手段 6 1 8 に記憶し、前の命令の実行が完了した後に、記憶キュー手段における当該命令を上記ニューラルネットワーク演算手段 6 1 9 に与える。同じデータにアクセスしないと、当該命令を直接に上記ニューラルネットワーク演算手段 6 1 9 に与える。

10

【 0 2 9 8 】

記憶キュー手段 6 1 8 は、命令が記憶手段にアクセスする時、同じ記憶スペースにアクセスする連続する 2 つの命令を記憶するためのものである。

【 0 2 9 9 】

具体的に、上記連続する 2 つの命令の実行結果の正確性を保証するために、現在の命令が前の命令のデータに依存することが検出された場合、当該連続する 2 つの命令は依存がなくなるまで上記記憶キュー手段 6 1 8 内で待たなければならない。そうしないと、当該連続する 2 つの命令を上記ニューラルネットワーク演算手段に与えることができない。

20

【 0 3 0 0 】

ニューラルネットワーク演算手段 6 1 9 は、命令処理手段又は記憶キュー手段から伝送される命令を処理するためのものである。

【 0 3 0 1 】

記憶手段 6 2 0 は、ニューロンキャッシュ手段 6 2 1 と重みキャッシュ手段 6 2 2 とを備え、ニューラルネットワークデータモデルは上記ニューロンキャッシュ手段 6 2 1 と重みキャッシュ手段 6 2 2 に記憶されている。

【 0 3 0 2 】

入出力手段 6 3 0 は、画像処理命令と処理対象画像を入力し、処理後の画像を出力するためのものである。

30

【 0 3 0 3 】

一実施形態において、記憶手段 6 2 0 はスクラッチパッドメモリであってもよく、入出力手段 6 3 0 は I/O 直接メモリアクセスモジュールであってもよい。

【 0 3 0 4 】

上記チップ、即ち上記画像処理機 1 0 3 の処理モジュール 1 0 3 2 が行う画像処理は、具体的に以下のようなステップを含む。

【 0 3 0 5 】

ステップ 6 0 1 では、命令取得モジュール 6 1 3 が命令キャッシュ手段 6 1 1 から 1 つの命令変換装置の生成した画像処理の命令を取り出し、演算命令を復号モジュール 6 1 4 に送る。

40

【 0 3 0 6 】

ステップ 6 0 2 では、復号モジュール 6 1 4 が演算命令を復号し、復号後の命令を命令キュー手段 8 1 5 に送信する。

【 0 3 0 7 】

ステップ 6 0 3 では、スカラレジスタモジュール 6 1 6 から前記命令に対応するニューラルネットワーク演算コマンドコードとニューラルネットワーク演算オペランドを取得する。

【 0 3 0 8 】

ステップ 6 0 4 では、命令が処理依存関係モジュール 6 1 7 に送られ、処理依存関係モジュール 6 1 7 は命令に対応するコマンドコードとオペランドに対して判断を行い、命令

50

と前の実行完了ではない命令とがデータ上で依存関係が存在するかどうかを判断し、存在しないと、前記命令を直接にニューラルネットワーク演算手段 619 に送り、存在すると、命令は前の実行完了ではない命令とデータ上でもはや依存関係が存在しないまで記憶キュー手段 618 で待ちする必要がある、その後前記命令に対応するマイクロ命令をニューラルネットワーク演算手段 619 に送信する。

【0309】

ステップ 605 では、ニューラルネットワーク演算手段 619 が前記命令に対応するコマンドコードとオペランドに従って必要なデータのアドレスとサイズを確定し、記憶手段 620 から処理対象画像や、画像処理モデルデータ等を含む必要なデータを取り出す。

【0310】

ステップ 606 では、ニューラルネットワーク演算手段 619 が前記命令に対応するニューラルネットワーク演算を実行し、対応する処理を終了し、処理結果を記憶手段 620 に書き戻しする。

【0311】

なお、上記記憶手段 620 は図 6 に示すチップのキャッシュ・オン・チップ手段である。

【0312】

なお、上記命令変換装置 102 と上記画像処理機 103 の処理モジュール 1032 はいずれも人工ニューラルネットワークチップ又は汎用処理チップであってもよい。或いは、上記命令変換装置 102 と上記画像処理機 103 の処理モジュール 1032 のうち、一方が人工ニューラルネットワークチップであり、他方が汎用処理チップであってもよい。

【0313】

好ましくは、上記画像処理装置は、データ処理装置、ロボット、コンピュータ、タブレット、スマート端末、携帯電話、クラウドサーバ、カメラ、ビデオカメラ、プロジェクタ、腕時計、イヤホン、携帯メモリ又はウェアラブル装置であってもよい。

【0314】

なお、本発明の実施例の技術案において、音声収集装置がユーザーの入力した音声信号を取得し、命令変換装置が目標音声命令変換モデルに従って前記音声信号を画像処理命令と前記処理対象画像に対する処理対象領域に変換し、画像処理機が前記画像処理命令と目標画像処理モデルに従って前記処理対象画像の処理領域を処理することができることが分かる。従来の画像処理技術と比べると、本発明は、音声によって画像処理を行うので、画像処理前にユーザーが画像処理ソフトウェアを学習する時間を省き、ユーザーの経験を向上させている。

【0315】

また、好ましい一実施例において、画像処理チップは上記の図 8 に示す画像処理装置を備える。

【0316】

また、上記チップは、メインチップと連携チップとを備え、上記連携チップは、本発明の実施例の第 1 の態様に記載の装置を備え、上記メインチップは上記連携チップに起動信号を与え、処理対象画像と画像処理命令が上記連携チップに伝送されるように制御するためのものである。

【0317】

また、上記連携チップは、上記の図 5 及び図 6 に示すチップを備える。

【0318】

好ましくは、上記画像処理チップは、ビデオカメラ、携帯電話、コンピュータ、ノートブック、タブレット又は他の画像処理装置に用いることができる。

【0319】

また、好ましい一実施例において、本発明の実施例は、上記の画像処理チップを備えるチップパッケージ構造を提供する。

【0320】

10

20

30

40

50

また、好ましい一実施例において、本発明の実施例は、上記のチップパッケージ構造を備えるボードカードを提供する。

【0321】

また、好ましい一実施例において、本発明の実施例は、上記のボードカードを備える電子機器を提供する。

【0322】

1つの可能な実施例では、本発明の実施例が他の電子機器を提供している。当該電子機器は、上記ボードカード、対話型インタフェース、制御手段、及び音声収集装置を含んでいる。

【0323】

図8に示すように、音声収集装置は音声を受信し、音声と処理対象画像を入力データとしてボードカードの内部のチップに伝えるために用いられる。

【0324】

好ましくは、上記画像処理チップは、人工ニューラルネットワーク処理チップであってもよい。

【0325】

好ましくは、音声収集装置は、マイクロフォン又はマルチアレイマイクロフォンであってもよい。

【0326】

また、ボードカードの内部のチップは図5及び図6に示すものと同じ実施例を含み、対応する出力データ（即ち処理後の画像）を取得し、対話型インタフェースに伝送するためのものである。

【0327】

また、対話型インタフェースは上記チップ（人工ニューラルネットワークプロセッサと見なすことができる）の出力データを受信し、適切な形式のフィードバック情報に変換してユーザーに表示させる。

【0328】

また、制御手段はユーザーの操作又は命令を受信し、画像処理装置全体の動作を制御する。

【0329】

好ましくは、上記電子機器は、データ処理装置、ロボット、コンピュータ、タブレット、スマート端末、携帯電話、クラウドサーバ、カメラ、ビデオカメラ、プロジェクタ、腕時計、イヤホン、携帯メモリ又はウェアラブル装置であってもよい。

【0330】

図13を参照すると、図13は本発明の実施例が提供する画像処理方法のフローチャートである。図13に示すように、当該方法は次のようなステップを含む。

【0331】

S1301では、画像処理装置がユーザーの入力した音声信号を収集する。

【0332】

S1302では、前記画像処理装置が目標音声命令変換モデルに従って前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換する。

【0333】

また、好ましい一実施例において、前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、音声認識技術によって前記音声信号をテキスト情報に変換するステップと、自然言語処理技術と前記目標音声命令変換モデルによって前記テキスト情報を前記画像処理命令に変換するステップと、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0334】

また、好ましい一実施例において、前記目標音声命令変換モデルに従って前記音声信号

10

20

30

40

50

を画像処理命令と目標領域に変換するステップは、音声認識技術、語義理解技術及び前記音声命令変換モデルによって前記音声信号を前記画像処理命令に変換するステップと、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0335】

また、好ましい一実施例において、前記方法は、前記テキスト情報又は前記画像処理命令又は前記目標領域を記憶するステップをさらに含む。

【0336】

S1303では、前記画像処理装置が前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理する。

10

【0337】

また、好ましい一実施例において、前記画像処理命令と目標画像処理モデルに従って前記目標領域を処理するステップは、予め設定された時間ウィンドウ内で前記記憶モジュールからM(Mは1より大きい整数)個の画像処理命令を取得するステップと、前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含む。

【0338】

また、好ましい一実施例において、前記M個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップは、前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N(Nは前記Mよりも小さい整数)個の画像処理命令を取得するステップと、前記N個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理するステップと、を含む。

20

【0339】

また、好ましい一実施例において、前記音声信号と処理対象画像を受信するステップの前に、前記方法は、音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得するステップをさらに含む。

【0340】

また、好ましい一実施例において、前記音声命令変換モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【0341】

30

また、好ましい一実施例において、前記音声命令変換モデルに対する自己適応訓練は監督されるか、又は監督されない。

【0342】

また、好ましい一実施例において、前記音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得するステップは、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換するステップと、前記予測命令及びその対応する命令セットの相関係数を確定するステップと、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するステップと、を含む。

【0343】

40

また、好ましい一実施例において、前記音声信号と処理対象画像を受信するステップの前に、前記方法は、画像処理モデルに対して自己適応訓練を行って、目標画像処理モデルを取得するステップをさらに含む。

【0344】

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【0345】

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練は監督されるか、又は監督されない。

【0346】

50

また、好ましい一実施例において、前記画像処理モデルに対して自己適応訓練を行って、目標画像処理モデルを取得するステップは、前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得するステップと、前記予測画像及びその対応する目標画像の相関係数を確定するステップと、前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するステップと、を含む。

【0347】

なお、図13に示す方法の各ステップの具体的実現過程は上記画像処理装置の具体的実現過程を参照してもよい。ここではさらに詳しく説明しない。

【0348】

図14を参照すると、図14は本発明の実施例が提供する画像処理装置の構造模式図である。図14に示すように、当該画像処理装置は、音声信号と処理対象画像を入力するための入出力手段130を備えている。

【0349】

好ましくは、上記画像処理装置は、騒音フィルタをさらに備え、上記入出力手段130が上記音声信号を取得した後、上記騒音フィルタは当該音声信号に対して騒音低減処理を行う。

【0350】

好ましくは、当該入出力手段130は、音声センサ、マイクロフォン、ピックアップや、他のオーディオ収集装置であってもよい。

【0351】

具体的に、上記入出力手段130は、上記音声信号を取得する時、環境声音信号も取得する。上記騒音フィルタは、上記環境声音信号に応じて上記音声信号に対して騒音低減処理を実行する。当該環境声音信号は上記音声信号の騒音であると見なしてもよい。

【0352】

さらに、上記入出力手段130は、上記音声信号と上記環境声音信号の収集に用いるだけでなく、騒音低減処理も実現するペアマイクロフォンアレイを含んでもよい。

【0353】

記憶手段120は、前記音声信号と前記処理対象画像を記憶するためのものである。

【0354】

画像処理手段110は、前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換し、前記画像処理命令に従って前記目標領域を処理して、処理後の画像を取得し、前記処理対象画像を前記記憶手段に記憶するためのものである。

【0355】

好ましくは、前記記憶手段120は、ニューロン記憶手段121と、重みキャッシュ手段122とを備え、前記画像処理手段110のニューラルネットワーク演算手段113は、ニューラルネットワーク演算サブ手段1131を備え、前記ニューロン記憶手段121が前記音声信号と前記処理対象画像を記憶するために用いられ、且つ、前記重みキャッシュ手段122が目標音声命令変換モデルと目標画像処理モデルを記憶するために用いられる場合、前記ニューラルネットワーク演算サブ手段1131は、前記目標音声命令変換モデルに基づいて前記音声信号を前記画像処理命令と前記目標領域に変換するために用いられ、前記ニューラルネットワーク演算サブ手段1131は、前記目標画像処理モデルと前記画像処理命令に従って前記目標領域を処理して、処理後の画像を取得するためにも用いられ、前記処理後の画像を前記ニューロン記憶手段に記憶するためにも用いられる。

【0356】

さらに、前記ニューラルネットワーク演算サブ手段1131は、具体的に、音声認識技術に基づいて前記音声信号をテキスト情報に変換し、自然言語処理技術と前記目標音声命令変換モデルに従って前記テキスト情報を前記画像処理命令に変換し、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

10

20

30

40

50

【 0 3 5 7 】

さらに、前記ニューラルネットワーク演算サブ手段 1 1 3 1 は、具体的に、音声認識技術、語義理解技術及び前記目標音声命令変換モデルに基づいて前記音声信号を前記画像処理命令に変換し、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【 0 3 5 8 】

さらに、前記ニューロン記憶手段 1 2 1 は、前記目標領域と前記画像処理命令を記憶するために用いられる。

【 0 3 5 9 】

具体的に、前記ニューラルネットワーク演算サブ手段 1 1 3 1 は、予め設定された時間ウィンドウ内で前記ニューロン記憶手段から M (M は 1 より大きい整数) 個の画像処理命令を取得し、前記 M 個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N (N は前記 M よりも小さい整数) 個の画像処理命令を取得し、前記 N 個の画像処理命令と前記目標画像処理モデルに基づいて前記目標領域を処理して、処理後の画像を取得するために用いられる。

10

【 0 3 6 0 】

具体的に、上記記憶手段 1 2 0 のニューロン記憶手段 1 2 1 が上記音声信号と上記処理対象画像を記憶し、且つ、上記重みキャッシュ手段 1 2 2 が上記目標音声命令変換モデルを記憶する時、上記ニューラルネットワーク演算サブ手段 1 1 3 1 は、音声認識技術に基づいて前記音声信号をテキスト情報に変換し、自然音声処理技術と上記目標音声命令変換モデルに基づいて上記テキスト情報を画像処理命令に変換し、当該画像処理命令における語義領域の粒度と画像認識技術に基づいて上記処理対象画像を領域分割し、上記目標領域を取得するか、又は、上記ニューラルネットワーク演算サブ手段 1 1 3 1 は、音声認識技術、語義理解技術及び上記目標音声命令変換モデルに基づいて上記音声信号を画像処理命令に変換し、上記画像処理命令における語義領域の粒度と画像認識技術に基づいて上記処理対象画像を領域分割し、上記目標領域を取得する。

20

【 0 3 6 1 】

さらに、上記ニューラルネットワーク演算サブ手段 1 1 3 1 は、上記画像処理命令と上記目標領域を上記ニューロンキャッシュ手段 1 2 1 に記憶する。上記ニューラルネットワーク演算サブ手段 1 1 3 1 は、上記重みキャッシュ手段 1 2 2 から上記目標音声命令変換モデルを取得し、予め設定された時間ウィンドウ内で上記ニューロン記憶手段 1 2 1 から M 個の画像処理命令と目標領域を取得し、上記 M 個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N 個の画像処理命令を取得する。上記ニューラルネットワーク演算サブ手段 1 1 3 1 は、上記 N 個の画像処理命令と上記目標画像処理モデルに基づいて上記目標領域を処理して、処理後の画像を取得する。

30

【 0 3 6 2 】

好ましくは、前記記憶手段は汎用データキャッシュ手段を備え、前記画像処理手段のニューラルネットワーク演算手段は汎用演算サブ手段を備え、前記汎用データキャッシュ手段が前記音声信号と前記処理対象画像を受信するために用いられる場合、前記汎用演算サブ手段は前記音声信号を前記画像処理命令と前記目標領域に変換するために用いられ、前記汎用演算サブ手段は、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するためにも用いられ、前記汎用演算サブ手段は、前記処理後の画像を前記汎用データ記憶手段に記憶するためにも用いられる。

40

【 0 3 6 3 】

さらに、前記汎用演算サブ手段は、具体的に、音声認識技術に基づいて前記音声信号をテキスト情報に変換し、自然言語処理技術に基づいて前記テキスト情報を前記画像処理命令に変換し、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【 0 3 6 4 】

さらに、前記汎用演算サブ手段は、具体的に、音声認識技術と語義理解技術に基づいて

50

前記音声信号を前記画像処理命令に変換し、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するために用いられる。

【0365】

さらに、前記汎用データキャッシュ手段は前記目標領域と前記画像処理命令を記憶するために用いられる。

【0366】

具体的に、前記汎用演算サブ手段は、予め設定された時間ウィンドウ内で前記汎用データキャッシュ手段からM(Mは1より大きい整数)個の画像処理命令を取得し、前記M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N(Nは前記Mよりも小さい整数)個の画像処理命令を取得し、前記N個の画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するために用いられる。

10

【0367】

具体的に、上記記憶手段120の汎用データキャッシュ手段123が上記音声信号と上記処理対象画像を記憶する時、上記汎用演算サブ手段1132は、音声認識技術に基づいて上記音声信号をテキスト情報に変換し、自然言語処理技術に基づいて上記テキスト情報を画像処理命令に変換し、上記画像処理命令の語義領域の粒度と画像認識技術に基づいて上記処理対象画像を領域分割して、上記目標領域を取得するか、又は、上記汎用演算サブ手段1132は、音声認識技術と語義理解技術に基づいて上記音声信号を上記画像処理命令に変換し、当該画像処理命令における語義領域の粒度と画像認識技術に基づいて上記処理対象画像を領域分割して、上記目標領域を取得する。

20

【0368】

さらに、上記汎用演算サブ手段1132は上記画像処理命令と上記目標領域を上記汎用データキャッシュ手段123に記憶する。上記汎用演算サブ手段1132は上記汎用データキャッシュ手段から上記目標領域を取得し、予め設定された時間ウィンドウ内で上記汎用データキャッシュ手段からM個の画像処理命令を取得し、当該M個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、N個の画像処理命令を取得し、当該N個の画像処理命令に応じて上記目標領域を処理して、処理後の画像を取得する。

【0369】

具体的に、上記予め設定された時間ウィンドウは、予め設定された時間長であると理解してもよい。予め設定された時間長内で上記ニューラルネットワーク演算サブ手段1131は上記ニューロン記憶手段121からM個の画像処理命令を取得するか、又は、上記汎用演算サブ手段1132は上記汎用データキャッシュ手段からM個の画像処理命令を取得した後、上記ニューラルネットワーク演算サブ手段1131又は上記汎用演算サブ手段1132は上記M個の画像処理命令について2つずつ比較し、当該M個の画像処理命令のうち同じ機能を有する命令を削除して、N個の画像処理命令を取得する。上記ニューラルネットワーク演算サブ手段1131又は上記汎用演算サブ手段1132は上記N個の処理命令と上記目標画像処理モデルに基づいて上記処理対象画像を処理する。

30

【0370】

一例として、上記ニューラルネットワーク演算サブ手段1131又は上記汎用演算サブ手段1132は上記M個の画像処理命令について2つずつ比較する。画像処理命令Aと画像処理命令Bとが同じである場合、上記ニューラルネットワーク演算サブ手段1131又は上記汎用演算サブ手段1132は上記画像処理命令A及びBのうち最大のオーバーヘッドを削除し、画像処理命令Aと画像処理命令Bとが異なる場合、上記ニューラルネットワーク演算サブ手段1131又は上記汎用演算サブ手段1132は上記画像処理命令Aと上記画像処理命令Bの類似度係数を取得する。当該類似度係数が類似度閾値よりも大きい時、上記画像処理命令Aと上記画像処理命令Bの機能が同じであることを確定し、上記ニューラルネットワーク演算サブ手段1131又は上記汎用演算サブ手段1132は上記画像処理命令A及びBのうち最大のオーバーヘッドを削除し、上記類似度係数が上記類似度閾値よりも小さい時、上記ニューラルネットワーク演算サブ手段1131又は上記汎用演算

40

50

サブ手段 1 1 3 2 は上記画像処理命令 A 及び B の機能が異なることを確定する。当該画像処理命令 A 及び B は上記 N 個の処理命令のうちの任意の 2 つである。

【 0 3 7 1 】

前記入出力手段 1 0 4 は、前記処理後の画像を出力するためにも用いられる。

【 0 3 7 2 】

また、上記画像処理手段は、上記音声信号に応じて上記処理対象画像を処理して、処理後の画像を取得した後、上記入出力手段によって当該処理後の画像を出力する。

【 0 3 7 3 】

例を挙げて上記語義領域を説明すると、上記画像処理装置が音声信号に応じて上記目標領域が顔領域であると確定した場合、上記語義領域は上記処理対象画像における顔領域であり、上記画像処理装置は顔を粒度として、上記処理対象画像における複数の顔領域を取得する。上記目標領域が背景である場合、上記画像処理装置は上記処理対象画像を背景領域と非背景領域に分割し、上記目標領域が赤色領域である場合、上記画像処理装置は上記処理対象画像を色に従って異なる色の領域に分割する。

【 0 3 7 4 】

具体的に、本発明で使用される音声認識技術は、人工ニューラルネットワーク (Artificial Neural Network; ANN)、隠れマルコフモデル (Hidden Markov Model; HMM) 等のモデルを含むが、これらに限定されない。上記第 1 の音声認識手段は、上記音声認識技術に基づいて上記音声信号を処理してもよい。上記自然言語処理技術は、統計的機械学習や、ANN 等の方法を利用することを
20
含むが、これらに限定されない。上記語義理解手段は、上記自然言語処理技術に基づいて語義情報を抽出することができる。上記画像認識技術は、エッジ検出に基づく方法、閾値分割方法、領域成長及び分水嶺 (watershed) アルゴリズム、階調積分投影 (Gray-level Integration Projection) 曲線分析、テンプレートマッチング、変形可能テンプレート、ハフ (Hough) 変換、スネーク (Snake) 演算子、Gabor ウェーブレット変換に基づく弾性マップマッチング技術や、アクティブ形状モデル及びアクティブ外観モデル等の方法等のアルゴリズムを含むが、これらに限定されない。上記画像認識手段は、上記画像認識技術に基づいて上記処理対象画像を異なる領域に分割することができる。

【 0 3 7 5 】

好ましくは、上記入出力手段 1 3 0 が上記音声信号と上記処理対象画像を取得する前に、上記ニューラルネットワーク演算サブ手段 1 1 3 1 は音声命令変換モデルに対して自己適応訓練を行って、上記目標音声命令変換モデルを取得する。

【 0 3 7 6 】

また、音声命令変換モデルに対する上記ニューラルネットワーク演算サブ手段 1 1 3 1 の自己適応訓練はオフライン又はオンラインで行われる。

【 0 3 7 7 】

具体的に、上記音声命令変換モデルに対する自己適応訓練がオフラインで行われることは、具体的に、上記ニューラルネットワーク演算サブ手段 1 1 3 1 がそのハードウェアを基で上記音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデル
40
を取得することである。上記音声命令変換モデルに対する自己適応訓練がオンラインで行われることは、具体的に、ニューラルネットワーク演算サブ手段 1 1 3 1 と異なるクラウドサーバが上記音声命令変換モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得することである。上記ニューラルネットワーク演算サブ手段 1 1 3 1 が上記目標音声命令変換モデルを使用する必要がある場合、当該ニューラルネットワーク演算サブ手段 1 1 3 1 は上記クラウドサーバから当該目標音声命令変換モデルを取得する。

【 0 3 7 8 】

好ましくは、音声命令変換モデルに対する上記ニューラルネットワーク演算サブ手段 1 1 3 1 の自己適応訓練は監督されるか、又は監督されない。

【 0 3 7 9 】

10

20

30

40

50

具体的に、上記音声命令変換モデルに対する自己適応訓練が監督されることは、具体的に、上記ニューラルネットワーク演算サブ手段 1 1 3 1 が音声命令変換モデルに基づいて上記音声信号を予測命令に変換し、その後、上記予測命令及びその対応する命令セットの相関係数を確定し、当該命令セットは手動で音声信号に従って取得される命令のセットであり、上記ニューラルネットワーク演算サブ手段 1 1 3 1 が前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得することである。

【 0 3 8 0 】

一例として、上記音声命令変換モデルに対する自己適応訓練が監督されることは、具体的に、上記ニューラルネットワーク演算サブ手段 1 1 3 1 が、画像の色の変更や、画像の回転等の関連命令を含む音声信号を取得することである。命令毎に 1 つの命令セットが対応する。自己適応訓練用の入力した音声信号にとって、対応する命令セットは既知のものであり、上記ニューラルネットワーク演算サブ手段 1 1 3 1 はこれらの音声信号を音声命令変換モデルの入力データとして、出力後の予測命令を取得する。上記ニューラルネットワーク演算サブ手段 1 1 3 1 は、上記予測命令及びその対応する命令セットの相関係数を算出し、上記音声命令変換モデルにおけるパラメータ（重みや、オフセット等）を適応的に更新して、上記音声命令変換モデルの性能を向上し、さらに上記目標音声命令変換モデルを取得する。

【 0 3 8 1 】

具体的に、上記画像処理手段 1 1 0 に対し、その入力と出力はいずれも画像である。上記画像 1 0 3 の A N N や従来のコンピュータビジョン方法に限られない方法によって上記処理対象画像に対して行う処理は、美体（例えば美脚、豊胸）、顔の変更、顔の美化、オブジェクトの変更（猫と犬の交換、シマウマの馬への変更、リンゴとオレンジの交換等）、背景の変更（後の森を畑に置き換え）、遮蔽物除去（例えば片目が顔に遮られると、再度に目を再構成する）、スタイルの変換（1 秒でゴッホ画風へ変更）、位置姿勢の変換（例えば立っている状態から座る状態に変更、顔の正面から側面への変更）、油絵以外のものが油絵に変更、画像背景の色の変更や、画像中のオブジェクト場所の季節の背景の変更等を含むが、これらに限定されない。

【 0 3 8 2 】

好ましくは、上記ニューラルネットワーク演算サブ手段 1 1 3 1 が上記音声信号を受信する前に、当該ニューラルネットワーク演算サブ手段 1 1 3 1 は画像処理モデルに対して自己適応訓練を行って、上記目標画像処理モデルを取得する。

【 0 3 8 3 】

また、画像処理モデルに対する上記ニューラルネットワーク演算サブ手段 1 1 3 1 の自己適応訓練はオフライン又はオンラインで行われる。

【 0 3 8 4 】

具体的に、上記画像処理モデルに対する自己適応訓練がオフラインで行われることは、具体的に、上記ニューラルネットワーク演算サブ手段 1 1 3 1 がそのハードウェアを基で上記画像処理モデルに対して自己適応訓練を行って、目標音声命令変換モデルを取得することである。上記画像処理モデルに対する自己適応訓練がオンラインで行われることは、具体的に、上記ニューラルネットワーク演算サブ手段 1 1 3 1 と異なるクラウドサーバが上記画像処理モデルに対して自己適応訓練を行って、目標画像処理モデルを取得することである。上記ニューラルネットワーク演算サブ手段 1 1 3 1 が上記目標画像処理モデルを使用する必要がある場合、当該ニューラルネットワーク演算サブ手段 1 1 3 1 は上記クラウドサーバから当該目標画像処理モデルを取得する。

【 0 3 8 5 】

好ましくは、画像処理モデルに対する上記ニューラルネットワーク演算サブ手段 1 1 3 1 の自己適応訓練は監督されるか、又は監督されない。

【 0 3 8 6 】

具体的に、上記画像処理モデルに対する上記ニューラルネットワーク演算サブ手段 1 1

10

20

30

40

50

31の自己適応訓練が監督されることは、具体的に、上記ニューラルネットワーク演算サブ手段1131が画像処理モデルに基づいて上記音声信号を予測画像に変換し、その後、上記予測画像及びその対応する目標画像の相関係数を確定し、当該目標は手動で音声信号に従って処理対象画像を処理することで得られる画像であり、上記ニューラルネットワーク演算サブ手段1131が前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化し、前記目標画像処理モデルを取得することである。

【0387】

一例として、上記画像処理モデルに対する自己適応訓練が監督されることは、具体的に、上記ニューラルネットワーク演算サブ手段1131が、画像の色の変更や、画像の回転等の関連命令を含む音声信号を取得することである。命令毎に1枚の目標画像が対応する。自己適応訓練用の入力した音声信号にとって、対応する目標画像は既知のものであり、上記ニューラルネットワーク演算サブ手段1131はこれらの音声信号を画像処理モデルの入力データとして、出力後の予測画像を取得する。上記ニューラルネットワーク演算サブ手段1131は、上記予測画像及びその対応する目標画像の相関係数を算出し、上記画像処理モデルにおけるパラメータ（重みや、オフセット等）を適応的に更新して、上記画像処理モデルの性能を向上し、さらに上記目標画像処理モデルを取得する。

【0388】

また、上記画像処理装置の画像処理手段110は、ニューラルネットワーク演算命令と汎用演算命令を含む実行されるべき命令を記憶するための命令キャッシュ手段111と、前記ニューラルネットワーク演算命令を前記ニューラルネットワーク演算サブ手段に伝送し、前記汎用演算命令を前記汎用演算サブ手段に伝送するための命令処理手段112と、をさらに備える。

【0389】

なお、上記画像処理装置の画像処理手段113におけるニューラルネットワーク演算サブ手段1131が、画像処理操作や、上記画像処理モデルと上記音声命令変換モデルに対して自己適応訓練を行う過程において、上記命令処理手段112は上記命令キャッシュ手段111からニューラルネットワーク演算命令を取得し、上記ニューラルネットワーク演算サブ手段1131に伝送して、当該ニューラルネットワーク演算サブ手段1131を駆動する。上記汎用演算サブ手段1132が画像処理操作を行う過程において、上記命令処理手段112は上記命令キャッシュ手段111から汎用演算命令を取得して上記汎用演算サブ手段1132に伝送して、当該汎用演算サブ手段1132を駆動する。

【0390】

本実施例において、上記画像処理装置は手段の形式で現れる。本明細書における「手段」は、特定用途向け集積回路(application-specific integrated circuit; ASIC)、1つ又は複数のソフトウェア又はファームウェアプログラムを実行するプロセッサ及びメモリ、集積論理回路、及び/又は他の上記機能を提供することができるデバイスを指してもよい。

【0391】

なお、本発明の実施例の技術案において、入出力手段は音声信号と処理対象画像を入力し、記憶手段は前記音声信号と前記処理対象画像を記憶し、画像処理手段は前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換し、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得し、前記処理対象画像を前記記憶手段に記憶し、前記入出力手段は前記処理後の画像を出力することができる。従来の画像処理技術と比べると、本発明は、音声によって画像処理を行うので、画像処理前にユーザーが画像処理ソフトウェアを学習する時間を省き、ユーザーの経験を向上させている。

【0392】

図15を参照すると、図15は本発明の実施例が提供する他の画像処理装置の構造ブロック模式図である。図15に示すように、当該チップは、画像処理手段210と、記憶手段220と、入出力手段230とを備えている。

【0393】

また、上記画像処理手段210は、ニューラルネットワーク演算命令と汎用演算命令を含む実行されるべき命令を記憶するための命令キャッシュ手段211を備える。

【0394】

一実施形態において、上記命令キャッシュ手段211は並べ替えキャッシュであってもよい。

【0395】

命令処理手段212は、命令キャッシュ手段からニューラルネットワーク演算命令又は汎用演算命令を取得し、当該命令を処理して上記ニューラルネットワーク演算手段213に供給するためのものである。上記命令処理手段212は、命令キャッシュ手段から命令

10

【0396】

スカラーレジスタモジュール217は、ニューラルネットワーク演算命令に対応するニューラルネットワーク演算コマンドコードとオペランド、及び汎用演算命令に対応する汎用演算コマンドコードとオペランドを含んでいる。

【0397】

処理依存関係モジュール218は、上記命令処理手段212によって送信された命令及びその対応するコマンドコードとオペランドを判断するためのものであり、当該命令と前

20

【0398】

記憶キュー手段219は、命令が記憶手段にアクセスする時、同じ記憶スペースにアクセスする連続する2つの命令を記憶するためのものである。

【0399】

具体的に、上記連続する2つの命令の実行結果の正確性を保証するために、現在の命令が前の命令のデータに依存することが検出された場合、当該連続する2つの命令は依存がなくなるまで上記記憶キュー手段219内で待たなければならない。そうしないと、当該連続する2つの命令を上記ニューラルネットワーク演算手段に与えることができない。

30

【0400】

ニューラルネットワーク演算手段213は、命令処理モジュール又は記憶キュー手段から伝送される命令を処理するためのものである。

【0401】

記憶手段220は、ニューロンキャッシュ手段521と重みキャッシュ手段522とを備え、ニューラルネットワークデータモデルは上記ニューロンキャッシュ手段221と重みキャッシュ手段222に記憶されている。

40

【0402】

入出力手段230は、音声信号を入力し、画像処理命令を出力するためのものである。

【0403】

一実施形態において、記憶手段220はスクラッチパッドメモリであってもよく、入出力手段230はI/O直接メモリアクセスモジュールであってもよい。

【0404】

具体的に、上記画像処理装置のニューラルネットワーク演算サブ手段が音声信号を画像処理命令に変換することは、具体的に次のようなステップを含む。

【0405】

ステップAでは、命令取得モジュール214が命令キャッシュ手段211から1つの音

50

声認識用のニューラルネットワーク演算命令を取り出し、演算命令を復号モジュール 2 1 5 に送る。

【 0 4 0 6 】

ステップ B では、復号モジュール 2 1 5 が演算命令を復号し、復号後の命令を命令キュー手段 2 1 6 に送る。

【 0 4 0 7 】

ステップ C では、スカラレジスタモジュール 2 1 7 から前記命令に対応するニューラルネットワーク演算コマンドコードとニューラルネットワーク演算オペランドを取得する。

【 0 4 0 8 】

ステップ D では、命令が処理依存関係モジュール 2 1 8 に送られ、当該処理依存関係モジュール 2 1 8 は命令に対応するコマンドコードとオペランドに対して判断を行い、命令と前の実行完了ではない命令とがデータ上で依存関係が存在するかどうかを判断し、存在しないと、前記命令を直接にニューラルネットワーク演算手段 2 1 3 に送り、存在すると、命令は前の実行完了ではない命令とデータ上でもはや依存関係が存在しないまで記憶キュー手段 2 1 9 で待ちする必要がある、その後前記命令をニューラルネットワーク演算手段 2 1 3 に送信する。

10

【 0 4 0 9 】

ステップ E では、ニューラルネットワーク演算サブ手段 2 1 3 1 が命令に対応するコマンドコードとオペランドに従って必要なデータのアドレスとサイズを確定し、記憶手段 2 2 0 から音声命令変換モデルデータ等を含む必要なデータを取り出す。

20

【 0 4 1 0 】

ステップ F では、ニューラルネットワーク演算サブ手段 2 1 3 1 が前記命令に対応するニューラルネットワーク演算を実行し、対応する処理を終了し、画像処理命令を取得し、画像処理命令を記憶手段 2 2 0 のニューロン記憶手段 2 2 1 に書き戻しする。

【 0 4 1 1 】

具体的に、上記画像処理装置の汎用演算サブ手段が音声信号を画像処理命令に変換することは、具体的に次のようなステップを含む。

【 0 4 1 2 】

ステップ A ' では、命令取得モジュール 2 1 4 が命令キャッシュ手段 2 1 1 から 1 つの音声認識用の汎用演算命令を取り出し、演算命令を復号モジュール 2 1 5 に送る。

30

【 0 4 1 3 】

ステップ B ' では、復号モジュール 2 1 5 が演算命令を復号し、復号後の命令を命令キュー手段 2 1 6 に送る。

【 0 4 1 4 】

ステップ C ' では、スカラレジスタモジュール 2 1 7 から前記命令に対応する汎用演算コマンドコードと汎用演算オペランドを取得する。

【 0 4 1 5 】

ステップ D ' では、命令が処理依存関係モジュール 2 1 8 に送られ、当該処理依存関係モジュール 2 1 8 は命令に対応するコマンドコードとオペランドに対して判断を行い、命令と前の実行完了ではない命令とがデータ上で依存関係が存在するかどうかを判断し、存在しないと、前記命令を直接にニューラルネットワーク演算手段 2 1 3 に送り、存在すると、命令は前の実行完了ではない命令とデータ上でもはや依存関係が存在しないまで記憶キュー手段 2 1 9 で待ちする必要がある、その後前記命令をニューラルネットワーク演算手段 2 1 3 に送信する。

40

【 0 4 1 6 】

ステップ E ' では、汎用演算サブ手段 2 1 3 2 が命令に対応するコマンドコードとオペランドに従って必要なデータのアドレスとサイズを確定し、記憶手段 2 2 0 から音声命令変換モデルデータ等を含む必要なデータを取り出す。

【 0 4 1 7 】

ステップ F ' では、汎用演算サブ手段 2 1 3 2 が前記命令に対応する汎用演算を実行し

50

、対応する処理を終了し、画像処理命令を取得し、画像処理命令を記憶手段 220 の汎用データキャッシュ手段 223 に書き戻しする。

【0418】

なお、画像処理過程において、上記ニューラルネットワーク演算サブ手段 213 のニューラルネットワーク演算サブ手段 2131 と汎用演算サブ手段 2132、上記記憶手段 220 のニューロン記憶手段 221、重みキャッシュ手段 222 と汎用データキャッシュ手段 223 と上記入出力手段 230 の具体的な操作過程は、図 14 に示す実施例の関連説明を参照することができ、ここではさらに説明しない。

【0419】

なお、上記記憶手段 220 は図 15 に示す画像処理装置のキャッシュ・オン・チップ手
段である。 10

【0420】

好ましくは、上記画像処理装置は、データ処理装置、ロボット、コンピュータ、タブレット、スマート端末、携帯電話、クラウドサーバ、カメラ、ビデオカメラ、プロジェクタ、腕時計、イヤホン、携帯メモリ又はウェアラブル装置であってもよい。

【0421】

また、好ましい一実施例において、画像処理チップは上記した図 14 に示す画像処理装置を備えている。

【0422】

また、上記チップは、メインチップと、連携チップとを備え、上記連携チップは、本発明の実施例の第 1 の態様に記載の装置を備え、上記メインチップは、上記連携チップに起動信号を与え、処理対象画像と画像処理命令が上記連携チップに伝送されるように制御するためのものである。 20

【0423】

好ましくは、上記画像処理チップは、ビデオカメラ、携帯電話、コンピュータ、ノートブック、タブレット又は他の画像処理装置に用いることができる。

【0424】

また、好ましい一実施例において、本発明の実施例はチップパッケージ構造を提供しており、当該チップパッケージ構造は上記画像処理チップを備えている。

【0425】

また、好ましい一実施例において、本発明の実施例はボードカードを提供しており、当該ボードカードは上記チップパッケージ構造を備えている。 30

【0426】

また、好ましい一実施例において、本発明の実施例は電子機器を提供しており、当該電子機器は上記ボードカードを備えている。

【0427】

また、好ましい一実施例において、本発明の実施例は他の電子機器を提供しており、当該電子機器は上記ボードカード、対話型インタフェース、制御手段と音声収集装置を備えている。

【0428】

図 16 に示すように、上記音声収集装置は音声を受信し、音声と処理対象画像を入力データとしてボードカードの内部の画像処理チップに伝えるために用いられる。 40

【0429】

好ましくは、上記画像処理チップは人工ニューラルネットワーク処理チップであってもよい。

【0430】

好ましくは、音声収集装置はマイクロフォン又はマルチアレイマイクロフォンであってもよい。

【0431】

また、ボードカードの内部のチップは図 14 及び図 15 に示すものと同じ実施例を含み 50

、対応する出力データ（即ち処理後の画像）を取得し、対話型インタフェースに伝送するためのものである。

【0432】

また、対話型インタフェースは上記チップ（人工ニューラルネットワークプロセッサと見なすことができる）の出力データを受信し、適切な形式のフィードバック情報に変換してユーザーに表示させる。

【0433】

また、画像処理手段はユーザーの操作又は命令を受信し、画像処理装置全体の動作を制御する。

【0434】

好ましくは、上記電子機器は、データ処理装置、ロボット、コンピュータ、タブレット、スマート端末、携帯電話、クラウドサーバ、カメラ、ビデオカメラ、プロジェクタ、腕時計、イヤホン、携帯メモリ又はウェアラブル装置であってもよい。

【0435】

図17を参照すると、図17は本発明の実施例が提供する画像処理方法のフローチャートである。図17に示すように、当該方法は次のようなステップを含む。

【0436】

S1701では、画像処理装置が音声信号と処理対象画像を入力する。

【0437】

S1702では、画像処理装置が前記音声信号と前記処理対象画像を記憶する。

【0438】

S1703では、画像処理装置が前記音声信号を画像処理命令と処理対象画像の処理領域である目標領域に変換し、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得し、前記処理対象画像を前記記憶手段に記憶する。

【0439】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップは、音声認識技術に基づいて前記音声信号をテキスト情報に変換するステップと、自然言語処理技術と目標音声命令変換モデルに基づいて前記テキスト情報を前記画像処理命令に変換するステップと、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0440】

また、好ましい一実施例において、前記目標音声命令変換モデルに従って前記音声信号を画像処理命令と目標領域に変換するステップは、音声認識技術、語義理解技術及び前記音声命令変換モデルによって前記音声信号を前記画像処理命令に変換するステップと、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0441】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップは、音声認識技術、語義理解技術及び目標音声命令変換モデルに基づいて前記音声信号を前記画像処理命令に変換するステップと、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0442】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップは、音声認識技術に基づいて前記音声信号をテキスト情報に変換するステップと、自然言語処理技術に基づいて前記テキスト情報を前記画像処理命令に変換するステップと、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0443】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップは、音声認識技術と語義理解技術に基づいて前記音声信号を前記画像処理命令に変換するステップと、前記画像処理命令における語義領域の粒度と画像認識技術に基づいて前記処理対象画像を領域分割し、前記目標領域を取得するステップと、を含む。

【0444】

また、好ましい一実施例において、前記音声信号を画像処理命令と目標領域に変換するステップの後に、前記方法は、前記画像処理命令と前記目標領域を記憶するステップをさらに含む。

【0445】

また、好ましい一実施例において、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するステップは、予め設定された時間ウィンドウ内で前記ニューロン記憶手段から M (M は1より大きい整数)個の画像処理命令を取得するステップと、前記 M 個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、 N (N は前記 M よりも小さい整数)個の画像処理命令を取得するステップと、前記 N 個の画像処理命令と目標画像処理モデルに基づいて前記目標領域を処理して、処理後の画像を取得するステップと、を含む。

10

【0446】

また、好ましい一実施例において、前記画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するステップは、予め設定された時間ウィンドウ内で前記汎用データキャッシュ手段から M (M は1より大きい整数)個の画像処理命令を取得し、前記 M 個の画像処理命令のうち同じ機能を有する画像処理命令を削除して、 N (N は前記 M よりも小さい整数)個の画像処理命令を取得するステップと、前記 N 個の画像処理命令に応じて前記目標領域を処理して、処理後の画像を取得するステップと、を含む。

20

【0447】

S1704では、画像処理装置が前記処理後の画像を出力する。

【0448】

また、好ましい一実施例において、前記方法は、音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するステップをさらに含む。

【0449】

また、好ましい一実施例において、前記音声命令変換モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

30

【0450】

また、好ましい一実施例において、前記音声命令変換モデルに対する自己適応訓練は監督されるか、又は監督されない。

【0451】

また、好ましい一実施例において、前記音声命令変換モデルに対して自己適応訓練を行って、前記目標音声命令変換モデルを取得するステップは、前記音声命令変換モデルに基づいて前記音声信号を予測命令に変換するステップと、前記予測命令及びその対応する命令セットの相関係数を確定するステップと、前記予測命令及びその対応する命令セットの相関係数に従って前記音声命令変換モデルを最適化して、前記目標音声命令変換モデルを取得するステップと、を含む。

40

【0452】

また、好ましい一実施例において、前記方法は、画像処理モデルに対して自己適応訓練を行って、前記目標画像処理モデルを取得するステップをさらに含む。

【0453】

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練はオフライン又はオンラインで行われる。

【0454】

また、好ましい一実施例において、前記画像処理モデルに対する自己適応訓練は監督されるか、又は監督されない。

50

【 0 4 5 5 】

また、好ましい一実施例において、前記画像処理モデルに対して自己適応訓練を行うステップは、前記処理対象画像に対する前記画像処理モデルの処理によって予測画像を取得するステップと、前記予測画像及びその対応する目標画像の相関係数を確定するステップと、前記予測画像及びその対応する目標画像の相関係数に従って前記画像処理モデルを最適化して、前記目標画像処理モデルを取得するステップと、を含む。

【 0 4 5 6 】

なお、図 1 7 に示す方法の各ステップの具体的な実現過程は上記画像処理装置の具体的な実現過程を参照することができ、ここではさらに説明しない。

【 0 4 5 7 】

なお、前記の各方法実施例について、簡単に説明するために、それらを一連の動作の組み合わせとして記述したが、当業者であれば、本発明が説明された動作順序によって限定されないことを理解すべきである。それは、本発明によれば、幾つかのステップは他の順序で、又は同時に実行され得るからである。さらに、当業者であれば、明細書に説明された実施例はいずれも好ましい実施例であり、係る動作及びモジュールは必ずしも本発明にとって必要なものではないことを理解すべきである。

【 0 4 5 8 】

上記実施例において、各実施例に対する説明はいずれもそれぞれの重点があり、ある実施例で詳しく説明しない部分は、他の実施例の関連説明を参照することができる。

【 0 4 5 9 】

本願で提供される幾つかの実施例において、開示された装置は、他の形態によって実現されてもよいことを理解すべきである。例えば、以上に説明された装置実施例は単に例示的なものである。例えば前記手段の分割は、単に論理的な機能分割であり、実際実現する場合に別の分割形態を採用してもよい。例えば複数の手段又はモジュールは組み合わせてもよく、他のシステムに統合してもよく、一部の特徴を無視するか、又は実行しなくてもよい。一方、表れ又は検討している構成要素同士の結合又は直接結合又は通信接続は、幾つかのインターフェース、装置又は手段を介した間接的結合又は通信接続であってもよく、電氣的又は他の形式であってもよい。

【 0 4 6 0 】

前記した別々の構成要素として説明された手段は、物理的に分離されてもされなくてもよい。手段として現れた部材も物理手段でもそうでなくてもよい。即ち、一箇所に位置してもよく、複数のネットワーク手段に分布されてもよい。なお、実際の必要に応じてその中の一部又は全部手段を選んで本実施例の技術案の目的を実現することができる。

【 0 4 6 1 】

また、本発明の各実施例における各機能手段は、1つの処理手段に統合されてもよく、各手段が物理的に独立して存在してもよく、1つの手段に2又は2以上の手段を統合してもよい。上記の統合した手段は、ハードウェアで実現されてもよく、ソフトウェア機能手段で実現されてもよい。

【 0 4 6 2 】

前記統合した手段は、ソフトウェア機能手段で実現され、且つ独立の製品として販売又は使用される場合、コンピュータ読み取り可能なメモリに記憶してもよい。このような理解に基づいて、本発明の技術案は実質的に、又は従来技術に対して貢献した部分又は当該技術案の全部又は一部をソフトウェア製品の形で具体化することができる。当該コンピュータソフトウェア製品はメモリに記憶され、1台のコンピュータデバイス（パーソナルコンピュータ、サーバ又はネットワークデバイス等であり得る）に本発明の各実施例に記載の方法の全部又は一部のステップを実行させるための複数の命令を含まれる。また、前記のメモリは、Uディスク、読み取り専用メモリ（ROM、Read - Only Memory）、ランダムアクセスメモリ（RAM、Random Access Memory）、リムーバブルハードディスク、磁気ディスク又は光ディスク等の様々なプログラムコードを記憶可能な媒体を含む。

10

20

30

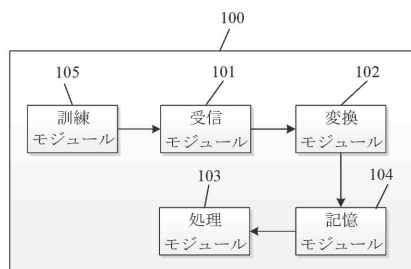
40

50

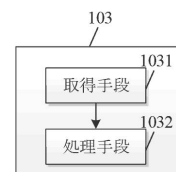
【 0 4 6 3 】

以上より、本発明の実施例を詳細に説明し、具体的な例にて本発明の原理及び実施形態を説明したが、以上の実施例の説明は単に本発明の方法及びその核心構想の理解に寄与するためのものである。同時に、当業者であれば、本発明の構想によって、具体的な実施形態及び応用範囲で変更することができる。よって、本明細書の内容を本発明の制限であると理解してはならない。

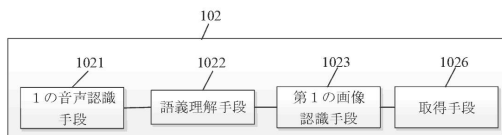
【 図 1 】



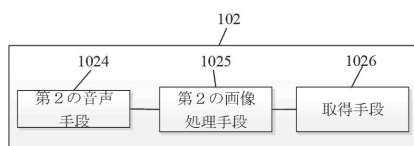
【 図 4 】



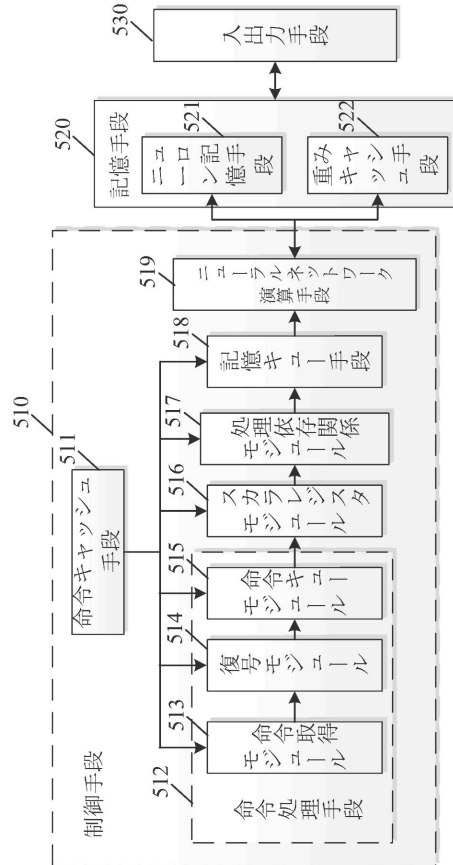
【 図 2 】



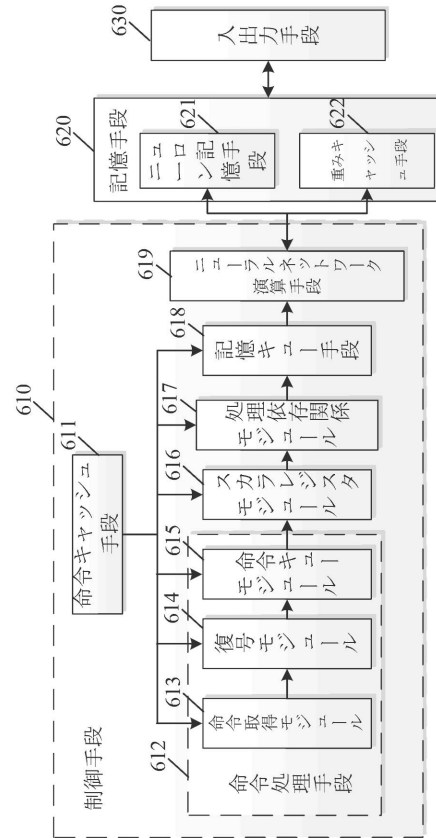
【 図 3 】



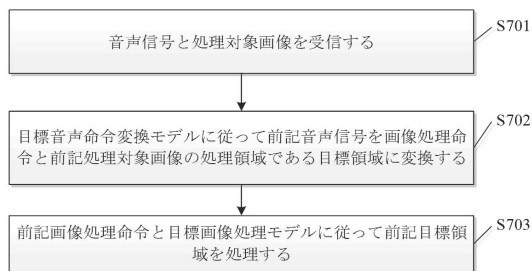
【図 5】



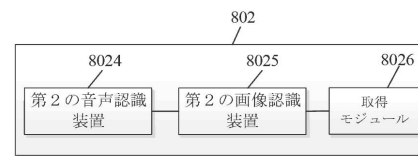
【図 6】



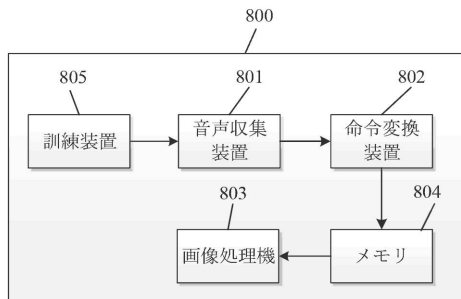
【図 7】



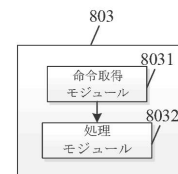
【図 10】



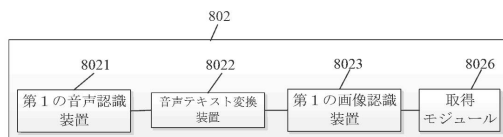
【図 8】



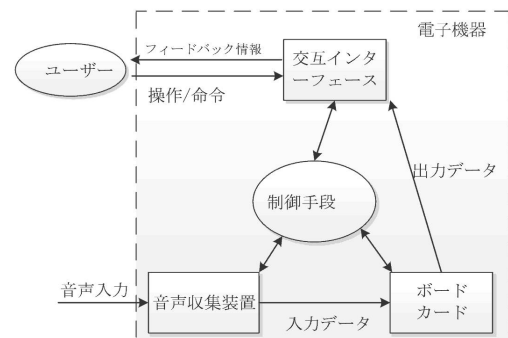
【図 11】



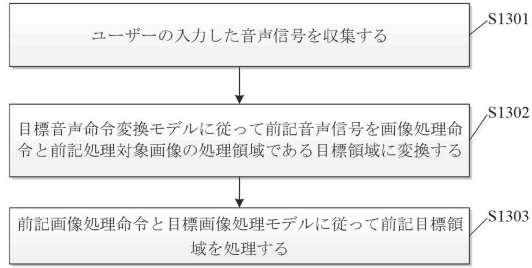
【図 9】



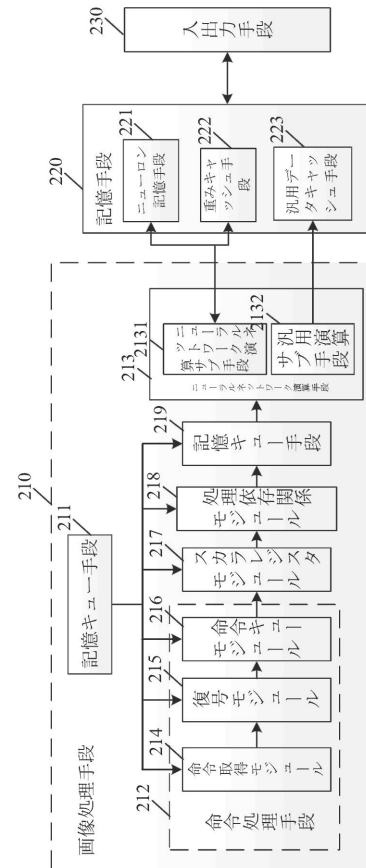
【図 12】



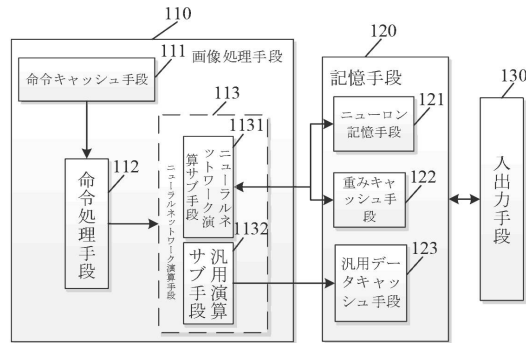
【図 13】



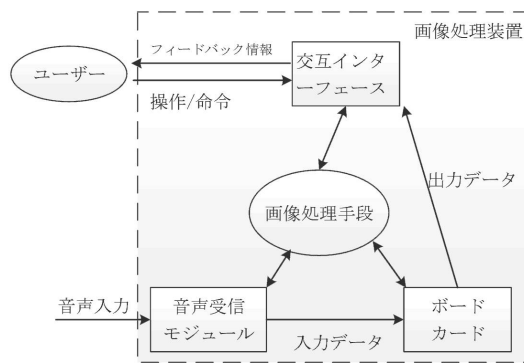
【図 15】



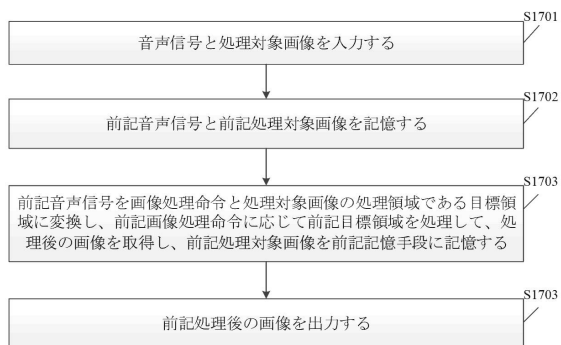
【図 14】



【図 16】



【図 17】



フロントページの続き

- (31)優先権主張番号 201711121244.5
(32)優先日 平成29年11月14日(2017.11.14)
(33)優先権主張国・地域又は機関
中国(CN)

早期審査対象出願

- (74)代理人 100170900
弁理士 大西 渉
(72)発明者 チェン ティエンシ
中華人民共和国 201306 シャンハイ プドン・ニュー・エリア トンフェイ・ロード 1
68 ブロック・ビー 6階
(72)発明者 フ シュアイ
中華人民共和国 201306 シャンハイ プドン・ニュー・エリア トンフェイ・ロード 1
68 ブロック・ビー 6階
(72)発明者 チェン シャオピン
中華人民共和国 201306 シャンハイ プドン・ニュー・エリア トンフェイ・ロード 1
68 ブロック・ビー 6階

審査官 岩田 淳

- (56)参考文献 中国特許出願公開第106156310(CN,A)
米国特許出願公開第2017/0142327(US,A1)
中国特許出願公開第107025629(CN,A)
特開2009-134433(JP,A)
特開平03-075860(JP,A)
中国特許出願公開第106814639(CN,A)
特開2002-175540(JP,A)
特開2005-202567(JP,A)
山下 澄奈, 外3名, DCGANを用いたイラスト画像生成の一手法, 情報処理学会 研究報告 数理モデル化と問題解決(MPS), [online], 日本, 情報処理学会, 2017年 2月27日, 第1-6ページ
品川 政太郎, 外3名, DNNによる画像操作システム, 電子情報通信学会技術研究報告, 日本, 一般社団法人電子情報通信学会 The Institute of Electronics, Information and Communication Engineers, 2017年 6月, 第117巻, 第106号, 第13-16ページ
CELLA LAO ROUSSEAU, Bye-bye basic editing, hello voice-controlled Photoshop!, 2017年 1月12日, [2020年5月26日検索], <URL:https://www.imore.com/bye-bye-basic-editing-hello-voice-controlled-photoshop>

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 17/26
G06F 3/01
G06F 3/048 - 3/0489
G06F 3/16
G06T 1/00 - 19/20