

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4934220号
(P4934220)

(45) 発行日 平成24年5月16日(2012.5.16)

(24) 登録日 平成24年2月24日(2012.2.24)

(51) Int.Cl.

G 0 6 T 7/20 (2006.01)

F 1

G O 6 T 7/20 3 0 0 A

請求項の数 20 (全 22 頁)

(21) 出願番号 特願2010-531098 (P2010-531098)
 (86) (22) 出願日 平成20年9月4日 (2008.9.4)
 (65) 公表番号 特表2011-501316 (P2011-501316A)
 (43) 公表日 平成23年1月6日 (2011.1.6)
 (86) 國際出願番号 PCT/US2008/075276
 (87) 國際公開番号 WO2009/055148
 (87) 國際公開日 平成21年4月30日 (2009.4.30)
 審査請求日 平成23年8月18日 (2011.8.18)
 (31) 優先権主張番号 11/925,557
 (32) 優先日 平成19年10月26日 (2007.10.26)
 (33) 優先権主張国 米国(US)

早期審査対象出願

(73) 特許権者 000005326
 本田技研工業株式会社
 東京都港区南青山二丁目1番1号
 (73) 特許権者 504325287
 ザ オハイオ ステート ユニバーシティ
 リサーチ ファウンデーション
 アメリカ合衆国 オハイオ 43210,
 コロンバス, ケニー ロード 196
 O
 (74) 代理人 100064414
 弁理士 磯野 道造
 (74) 代理人 100111545
 弁理士 多田 悅夫

最終頁に続く

(54) 【発明の名称】ラベル割当を用いた手サイン認識

(57) 【特許請求の範囲】

【請求項1】

手話を認識するコンピュータを用いる方法であって、
 前記コンピュータは、
 1つ以上のカメラを用いて、対象の部位と前記1つ以上のカメラとの距離を表す画像画
 素を含んだ前記対象の奥行画像を受信し、

前記奥行画像における前記対象の第1の手と第2の手とが重なり合うか隣接していると
 決定する場合に、前記奥行画像から前記第1の手および前記第2の手の手のひらを検出し
 、前記一組の互いにリンクするセグメントを含むリンク構造を生成し、前記一組の互いに
 リンクするセグメント各々を前記第1の部位又は前記第2の部位に割り当てることで、前記
 奥行画像を、前記対象の第1の手を表す第1の部位と前記対象の第2の手を表す第2の部
 位とに分類し、

前記第1の部位と前記第2の部位との形状を、記憶している手形状にマッチングさせて
 、前記奥行画像において前記第1の手と前記第2の手とで表されるサインを出力する
 ことを特徴とする方法。

【請求項2】

前記奥行画像を分類することは、

前記2つ以上のセグメントを前記第1の手と前記第2の手とに分類するためのコストを
 示す、複数の隣接するセグメントが複数の異なる手に分類される場合、増加する項を備え
 るコスト関数を生成し、

前記コスト関数を最小化し、さらに、手の特性を表す制約条件を満たし、前記制約条件が、近接するセグメントどうしは、同じ手に属するセグメントになる可能性が高い第1条件を備えるような、前記第1の手と前記第2の手とについての手形状を決定することを含むことを特徴とする請求項1に記載の方法。

【請求項3】

前記制約条件は、

画像画素の第1クラスタから生成した第1セグメントは、前記第1クラスタの第1の厚さが閾値よりも大きい場合には、2本以上の指を含む第2条件と、

画像画素の第2クラスタから生成した第2セグメントは、前記第2クラスタの第2の厚さが前記閾値以下である場合には、1本の指を含む第3条件と、

前記第1の手と前記第2の手とに分類されるセグメントの長さの合計は、最大値を超えない第4条件とを更に含む

ことを特徴とする請求項2に記載の方法。

【請求項4】

前記奥行画像を分類することは、前記リンク構造に基づいて、セグメントを表すノードを含み、かつ前記ノード間の連結間関係を示すような非循環有向グラフの構造を生成し、

前記ノードの第1のセットを前記第1の手にグループ化し、前記ノードの第2のセットを前記第2の手にグループ化するような、ノードの組み合わせを生成し、

両方の手の特徴に基づいて、前記ノードの組み合わせを評価する

ことを特徴とする請求項1に記載の方法。

【請求項5】

前記ノードの組み合わせを評価することは、

同じ手に分類された第1のノードと第2のノードとのリンクを表す第1のスコアを求め、

セグメントが分類された手のひらと前記セグメントとの距離を表す第2のスコアを求め、

前記同じ手に分類された、リンクするセグメントどうしの角度を表す第3のスコアを求め、

分岐ノードと前記分岐ノードが分類された手のひらとの距離を表す第4のスコアを求め、この場合の前記分岐ノードは、3つ以上のノードにリンクするノード、もしくは手のひらと2つ以上のノードにリンクするノードを表し、

前記第1のスコアと、前記第2のスコアと、前記第3のスコアと、前記第4のスコアとの合計値が最も小さい組み合わせを決定する

ことを含むことを特徴とする請求項4に記載の方法。

【請求項6】

前記コンピュータは、

前記第1の手または前記第2の手を表す画像画素に基づいてスケルトン線を生成し、

前記スケルトン線と直交し、かつ前記第1の手と前記第2の手の輪郭となる画像画素のエッジと交差する線の極大長さを求めて、前記第1の手のひらまたは前記第2の手のひらの中央部を含んでいる線の候補を決定し、

前記第1の手または前記第2の手の指を表す画像画素に最も近接する前記線の候補の中央部を、前記手のひらの中央部として決定することによって、

前記第1の手のひらと前記第2の手のひらとを検出することを更に含むことを特徴とする請求項1に記載の方法。

【請求項7】

前記第1の手および前記第2の手を表す前記画像画素を細線化することによって、前記リンク構造を生成する

ことを含むことを特徴とする請求項1に記載の方法。

【請求項8】

プロセッサによって実行可能な指令を記憶する非一時的なコンピュータ読み取り可能媒

10

20

30

40

50

体であって、

前記指令を実行することで、プロセッサは、
1つ以上のカメラを用いて、対象の部位と前記1つ以上のカメラとの距離を表す画像画素
を含んだ前記対象の奥行画像を受信し、

前記奥行画像における前記対象の第1の手と第2の手とが重なり合うか隣接していると
決定する場合に、前記奥行画像から前記第1の手および前記第2の手の手のひらを検出し
、前記一組の互いにリンクするセグメントを含むリンク構造を生成し、前記一組の互いに
リンクするセグメント各々を前記第1の部位又は前記第2の部位に割り当てることで、前記
奥行画像を、前記対象の第1の手を表す第1の部位と前記対象の第2の手を表す第2の部
位とに分類し、

前記第1の部位と前記第2の部位との形状を、記憶している手形状にマッチングさせて
、前記奥行画像において前記第1の手と前記第2の手とで表されるサインを出力する
ことを特徴とする媒体。

【請求項9】

前記奥行画像を分類する指令は、

前記2つ以上のセグメントを前記第1の手と前記第2の手とに分類するためのコストを
示す、複数の隣接するセグメントが複数の異なる手に分類される場合、増加する項を備え
るコスト関数を生成し、

前記コスト関数を最小化し、さらに、手の特性を表す制約条件を満たし、前記制約条件
が、近接するセグメントどうしは、同じ手に属するセグメントになる可能性が高い第1条件
を備えるような、前記第1の手と前記第2の手とについての手形状を決定する指令を含
む

ことを特徴とする請求項8に記載の媒体。

【請求項10】

前記制約条件は、

画像画素の第1クラスタから生成した第1セグメントは、前記第1クラスタの第1の厚
さが閾値よりも大きい場合には、2本以上の指を含む第2条件と、

画像画素の第2クラスタから生成した第2セグメントは、前記第2クラスタの第2の厚
さが前記閾値以下である場合には、1本の指を含む第3条件と、

前記第1の手と前記第2の手とに分類されるセグメントの長さの合計は、最大値を超
えない第4条件とを更に含む

ことを特徴とする請求項8に記載の媒体。

【請求項11】

前記デジタル画像を分類する指令は、

前記リンク構造に基づいて、セグメントを表すノードを含み、かつ前記ノード間の連結
間関係を示すような非循環有向グラフの構造を生成し、

前記ノードの第1のセットを前記第1の手にグループ化し、前記ノードの第2のセット
を前記第2の手にグループ化するような、ノードの組み合わせを生成し、

両方の手の特徴に基づいて、前記ノードの組み合わせを評価するための指令であること
を特徴とする請求項8に記載の媒体。

【請求項12】

前記ノードの組み合わせを評価する指令は、

同じ手に分類された第1のノードと第2のノードとのリンクを表す第1のスコアを求め

、セグメントが分類された手のひらと前記セグメントとの距離を表す第2のスコアを求め

、前記同じ手に分類された、リンクするセグメントどうしの角度を表す第3のスコアを求
め、

分岐ノードと前記分岐ノードが分類される手のひらとの距離を表す第4のスコアを求
め、この場合の前記分岐ノードは、3つ以上のノードにリンクするノード、もしくは手のひ

10

20

30

40

50

らと2つ以上のノードにリンクするノードを表し、

前記第1のスコアと、前記第2のスコアと、前記第3のスコアと、前記第4のスコアとの合計値が最も小さい組み合わせを決定する指令を含むことを特徴とする請求項8に記載の媒体。

【請求項13】

前記媒体は、

前記第1の手または前記第2の手を表す画像画素に基づいてスケルトン線を生成し、

前記スケルトン線と直交し、かつ前記第1の手と前記第2の手の輪郭となる画像画素のエッジと交差する線の極大長さを求めて、前記第1の手のひらまたは前記第2の手のひらの中央部を含んでいる線の候補を決定し、

10

前記第1の手または前記第2の手の指を表す画像画素に最も近接する前記線の候補の中央部を、前記手のひらの中央部として決定することによって、

前記第1の手のひらと前記第2の手のひらを検出するための指令を更に含むことを特徴とする請求項8に記載の媒体。

【請求項14】

前記第1の手および前記第2の手を表す前記画像画素を細線化することによって、前記リンク構造を生成する

ことを特徴とする請求項8に記載の媒体。

【請求項15】

手話認識用のコンピュータベースシステムであって、

20

前記システムは、対象の奥行画像を生成する1つ以上のカメラを備え、前記奥行画像は、対象の部位と前記1つ以上のカメラとの距離を表す画像画素を含み、

前記1つ以上のカメラに接続されて、前記奥行画像における前記対象の第1の手と前記第2の手とが重なり合うか隣接していると決定する場合に、前記奥行画像から前記第1の手および前記第2の手の手のひらを検出し、前記一組の互いにリンクするセグメントを含むリンク構造を生成し、前記一組の互いにリンクするセグメント各々を前記第1の部位又は前記第2の部位に割り当てることで、前記奥行画像を、前記対象の第1の手を表す第1の部位と前記対象の第2の手を表す第2の部位とに分類する画像前処理モジュールと、

前記画像前処理モジュールに接続され、前記第1の部位と前記第2の部位との形状を、記憶している手形状にマッチングさせて、前記奥行画像において前記第1の手と前記第2の手とで表されるサインを出力する形状マッチングモジュールとを備えることを特徴とするコンピュータベースシステム。

30

【請求項16】

前記形状マッチングモジュールは、

前記2つ以上のセグメントを前記第1の手と前記第2の手とに分類するためのコストを示す、複数の隣接するセグメントが複数の異なる手に分類される場合、増加する項を備えるコスト関数を生成し、

前記コスト関数を最小化し、さらに、手の特性を表す制約条件を満たし、前記制約条件が、近接するセグメントどうしは、同じ手に属するセグメントになる可能性が高い第1条件を備えるような、前記第1の手と前記第2の手とについての手形状を決定する最適化モジュールを備える

40

ことを特徴とする請求項15に記載のコンピュータベースシステム。

【請求項17】

前記制約条件は、

画像画素の第1クラスタから生成した第1セグメントは、前記第1クラスタの第1の厚さが閾値よりも大きい場合には、2本以上の指を含む第2条件と、

画像画素の第2クラスタから生成した第2セグメントは、前記第2クラスタの第2の厚さが前記閾値以下である場合には、1本の指を含む第3条件と、

前記第1の手と前記第2の手とに分類されるセグメントの長さの合計は、最大値を超えない第4条件とを更に含む

50

ことを特徴とする請求項15に記載のコンピュータベースシステム。

【請求項18】

前記形状マッチングモジュールは、前記リンク構造に基づいて、セグメントを表すノードを含み、かつ前記ノード間の連結間関係を示すような非循環有向グラフの構造を生成し、

前記ノードの第1のセットを前記第1の手にグループ化し、前記ノードの第2のセットを前記第2の手にグループ化するような、ノードの組み合わせを生成し、

両方の手の特徴に基づいて、前記ノードの組み合わせを評価するようなツリー検索モジュールを備える

ことを特徴とする請求項15に記載のコンピュータベースシステム。 10

【請求項19】

前記ツリー検索モジュールは、

同じ手に分類される第1のノードと第2のノードとのリンクを表す第1のスコアと、セグメントが分類される手のひらと前記セグメントとの距離を表す第2のスコアと、前記同じ手に分類されたリンクするセグメントどうしの角度を表す第3のスコアと、分岐ノードと前記分岐ノードが分類される手のひらとの距離を表す第4のスコアとを足して、前記ノードの組み合わせを評価し、

この場合の前記分岐ノードは、3つ以上のノードにリンクするノードもしくは手のひらと2つ以上のノードにリンクするノードを表す

ことを特徴とする請求項15に記載のコンピュータベースシステム。 20

【請求項20】

前記画像前処理モジュールは、

前記第1の手または前記第2の手を表す画像画素に基づいてスケルトン線を生成し、

前記スケルトン線と直交し、かつ前記第1の手と前記第2の手の輪郭となる画像画素のエッジと交差する線の極大長さを求めてことで、前記第1の手のひらまたは前記第2の手のひらの中央部を含んでいる線の候補を決定し、

前記第1の手または前記第2の手の指を表す画像画素に最も近接する前記線の候補の中央部を、前記手のひらの中央部として決定することによって、

前記第1の手のひらと前記第2の手のひらとを検出することを特徴とする請求項15に記載のコンピュータベースシステム。 30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般的には人間 - 機械インタラクションに関し、具体的には手話の手サインの認識に関する。

【背景技術】

【0002】

超高速演算システムと高効率デジタル画像システムの到来によって、人間 - 機械インタラクションに基づくコンピュータビジョンの分野は、著しい技術進歩の時代を経験している。 40

モーションによって機械（監視システムなど）からの応答を誘発するような簡便なモーション検出システムから、高度に複雑化した三次元（3D）画像サイン認識システム（imaging sign recognition systems）に至るまでの技術は、近年の著しい発展の主題である。例えば、サインに基づく人間 - 機械コミュニケーションの領域では、人間の手話を認識することは、人間 - 機械コミュニケーションにおける有望な技術として、近年、多くの研究においてテーマとなっている。

この他に、手と手の動きを、他の身体部位（腕部、胴体部、頭部など）の関係から探し出して追跡するような各種方法に基づいたサイン認識システムや、より複雑なジェスチャ認識システムが開発されている。

【0003】

サインおよびジェスチャの認識には、一般に、マシンビジョンシステムでサインやジェスチャのソースを見つけ出すための支援として、マーカー、特定色、背景あるいはグローブが必要になる。例えば、手検出に関する従来の手法によっては、システムに対して身振りをする片手や両手に対応する画像領域を決定するために、色やモーションの情報を用いる。しかし、このような手法では、照明条件の変化にともなって手のモーションの追跡の信頼性が著しく低下する。さらに、システムによってはグローブなどの特殊な装置を用いたり、タスクを実行可能にするために特殊色の背景を用いたりするものもある。

【0004】

また、従来技術の別群では、サインやジェスチャを認識するために、ステレオビジョンカメラや飛行時間センサで生成した奥行画像を用いる。奥行画像を解析して、手や腕を表す画像画素を抽出する。抽出した画像画素をさらに処理し、記憶している手形状とマッチングさせて、手と腕を表す画像画素で表される手サインを認識する。また、手と腕の軌跡を追跡して、画像シーケンスにおいて手形状と手の動きで表されるジェスチャや手話を決定することもある。奥行画像を用いた従来技術は、管理されない実生活の環境下での実装に便利であるため、マーカー、特殊な色や背景あるいはグローブを必要とする技術と比較して、有利である。さらに、奥行画像を用いる従来技術は、手話を介してコミュニケーションする人物に機器や装置を取付けたり、あるいは、その人物の周辺に機器や装置を設置する必要がない点でも有利である。

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、奥行画像を用いる従来の手サイン認識システムは、1本または2本の手が隣接したり重なり合わないような手話をにおいてのみ認識可能である。奥行画像とは、一般に、カメラから対象までの距離を示す画像画素を有するグレースケール画像である。このような奥行画像では、奥行画像中で2つ以上のオブジェクト（手など）が重なり合ったり隣接している場合は、どの画像画素がどのオブジェクト（手など）を表しているのかが曖昧になる。

【0006】

例えば、図1(a)ないし図1(c)は、手が重なり合ったり隣接するような日本手話（JSL：Japanese Sign Language）を示したものである。図1(a)は「手紙」を意味するサインを、図1(b)は「井戸」を意味するサインを、図1(c)は「合う(fit)」を意味するサインを日本手話で示している。従来の奥行画像を用いたサイン認識システムは、手どうしが重り合う場合には、片方の手を表す画像画素をもう片方の手から分離させる能力を備えていないため、このように重なり合ったり隣接するような手形状や手の動きを認識し、特定することはできない。したがって、従来の奥行画像を用いたサイン認識システムが認識できるのは、手話で一般的に用いられるサインの一部のみである。

【課題を解決するための手段】

【0007】

本発明の実施形態では、重なり合ったり隣接する手を含んでいる手話サインを認識するための方法とシステムを提供する。サイン認識システムは、対象の奥行画像を受信し、対象の手を表す奥行画像の部分を抽出し、さらに、サイン認識システムは、奥行画像の手が重なり合ったり隣接していると決定した場合には、奥行画像を対象の第1の手を表す第1の部位と対象の第2の手を表す第2の部位とに分類する。

【0008】

一実施形態では、サイン認識システムは、第1の手と第2の手を表すリンク構造を生成する。次に、リンク構造のセグメントを、第1の手に割当てるための第1のセグメントのセットと、第2の手に割当てるための第2のセグメントのセットとに分類する。次に、第1のセグメントのセットと第2のセグメントのセットをそれぞれ、記憶している手形状にマッチングさせて、第1の手と第2の手によって表されるサインを決定する。

【0009】

10

20

30

40

50

本発明の一実施形態では、リンク構造のセグメントを、制約条件付最適化技術を用いて第1の手と第2の手とに分類する。1つのセグメントのセットを第1の手と、別のセグメントのセットを第2の手とに分類するためのコストを表すコスト関数を生成する。さらに、手の物理的な構造および特徴に関する制約条件を生成する。例えば、制約条件には、互いに隣接するセグメントどうしは同じ手に分類される可能性が高い条件が含まれる。本実施形態に係るサイン認識システムは、上記のコスト関数と制約条件とを用いて、サイン認識をラベル付与問題に変換する。次に、サイン認識システムは、リンク構造のセグメントを第1の手と第2の手とに分類する。

【0010】

本発明の一実施形態では、ツリー検索技術を用いて、リンク構造のセグメントを第1の手と第2の手とに分類する。ツリー検索技術は、リンク構造に基づいて（セグメントを表す）ノードを有する非循環有向グラフを生成し、これらのノードどうしの連結関係を示すことを含む。上記のノードは、第1の手に割当てた第1のノードのセットと、第2の手に割当てた第2のノードのセットとに分類される。1つ以上の基準を用いて、ノードの各種組み合わせを評価する。

10

【0011】

本発明の一実施形態では、第1の手のひらと第2の手のひらとを検出する。第1の手のひらと第2の手のひらは、まず、第1の手または第2の手を表す画像画素のスケルトン線を生成する。第1の手のひらまたは第2の手のひらの中央部を含んでいる可能性のある線の候補を求めるには、スケルトン線と直交する線の、第1の手や第2の手の輪郭となる画像画素のエッジ間の長さを測定することで求める。第1の手のひらまたは第2の手のひらの中央部を含む線として、指に最も近接する線の候補を選ぶようにしてもよい。

20

【0012】

本明細書に記載される特徴および利点はすべて包含される必要はなく、特に、当業者であれば、図や明細書あるいは特許請求の範囲に照らして、さらなる特徴および利点を追加することが可能であることは明らかである。

さらに、本明細書中の説明で用いられた文言は、読みやすさや例示的な目的から主に用いられたものであって、本発明の主題の範囲を詳説したり、限定するために選択されたものではない。

【図面の簡単な説明】

30

【0013】

【図1】(a)ないし(c)は、重なり合ったり隣接する手を含む日本手話のサインと例示している。

【図2】本発明の一実施形態に係るサイン認識システムの概略ブロック図を例示している。

【図3】(a)は本発明の一実施形態に係る形状マッチングモジュールのブロック図を、(b)は本発明の一実施形態に係る形状マッチングモジュールのブロック図を、それぞれ例示している。

【図4】本発明の一実施形態に係る手話を認識する方法のフローチャートを例示している。

40

【図5】本発明の一実施形態に係る手のひらを検出する方法のフローチャートを例示している。

【図6】本発明の一実施形態に係る手のひらを検出するための手の画像を例示している。

【図7】(a)は本発明の一実施形態に係るサイン認識システムで処理する奥行画像の事例を、(b)は本発明の一実施形態に係る、(a)の奥行画像から生成したスケルトン構造を、(c)は本発明の一実施形態に係る、(b)のスケルトン構造から生成したリンク構造を、それぞれ例示している。

【図8】本発明の一実施形態に係る、制約条件最適化法を用いてリンク構造のセグメントを分類する方法を例示している。

【図9】本発明の一実施形態に係る、ツリー検索法を用いてリンク構造のセグメントを分

50

類する方法を例示している。

【図10】本発明の一実施形態に係る、非循環有向グラフを用いて手形状を生成する方法のフローチャートを例示している。

【図11】本発明の一実施形態に係る、図7(c)のリンク構造から生成したノードを有するツリー構造を例示している。

【図12】本発明の一実施形態に係る、右手と左手に分類したときの奥行画像の画像画素の事例を示している。

【発明を実施するための形態】

【0014】

図面を参照しつつ、以下に、本発明の好ましい実施形態について説明するのが、類似する参考番号は同一もしくは機能的に類似する構成要素を表す。さらに、図面の各参考番号の最左桁は、その参考番号が最初に用いられる図に対応する。10

【0015】

明細書中の「一実施形態」あるいは「実施形態」という記述は、当該の実施形態について記載された特定の特徴、構造、特性が本発明の少なくとも1つの実施形態に含まれることを意味する。明細書の随所に現れる「一実施形態では」という表現は、必ずしも同一の実施形態について言及するものではない。

【0016】

詳細な記載の部分によっては、コンピュータメモリ内のデータビットの操作に関するアルゴリズムや符号表現によって提示される。このようなアルゴリズム的な記載や表現は、データ処理技術を扱う当業者によって用いられる手段であり、他の当業者に仕事の実体を効果的に伝達するためのものである。ここで言う一般的なアルゴリズムとは、所望の結果を導くためのステップ(指令)を自己無頓着に連続したものと認識されるものである。このようなステップでは、物理量を物理的に操作することを必要とする。必ずしもではないが通常、このような物理量は、記憶、伝送、結合、比較、もしくは処理が可能な電気信号あるいは磁気信号の形態をとる。主に一般用途の理由から、このような信号をビット、値、要素、シンボル、文字、用語、数などで呼ぶほうがしばしば利便である。さらに、物理的な数量の物理的な操作を必要とするステップに関する特定の構成を、モジュールや符号装置などと呼ぶほうが、一般性を失うことなく、しばしば利便である。20

【0017】

しかしながら、このような表現あるいは類似の表現は所定の物理量に関連する便宜上のラベルにすぎない。説明から明らかな場合を除いて特に言及しない限りは、記載全体を通じて「処理する」、「演算する」、「算出する」、「決定する」あるいは「表示する」などの表現を用いる説明では、コンピュータシステムの記録部や記憶部に記憶される物理的(電子的)な数量として表されるデータを操作して、コンピュータシステムの記録部、記憶部、送信部、表示部などに記憶される、同様に物理量として表される別のデータに変換するようなコンピュータシステムあるいは類似の電子演算装置の実行や処理について述べている。

【0018】

本発明の特定の側面は、ここではアルゴリズムの形態で記載される処理ステップや指令を含む。なお、本発明のこれらの処理ステップや指令は、ソフトウェア、ファームウェアあるいはハードウェアにおいて具現可能であり、ソフトウェアにおいて具現する場合、各種のオペレーションシステムで用いられる異なるプラットフォーム上に搭載され、そこから操作したりするためにダウンロードが可能である。40

【0019】

また、本発明は、ここに記載される動作を実行する装置に関する。この装置は、必要な用途に合わせて特定の構成を備えていてもよいし、コンピュータに記憶されるコンピュータプログラムによって選択的に起動されたり、再構成されるような汎用コンピュータから構成されていてもよい。かかるコンピュータプログラムは、以下のようなコンピュータ読み取り可能な媒体に記憶されてもよいが、これらに限定されるものではない：フレキシブル50

ディスク、光ディスク、CD-ROM、磁気光ディスクなどを含むあらゆる種類のディスク、ROM、RAM、EPROM、EEPROM、磁気カードや光カード、ASIC (Application Specific Integrated Circuits)、あるいは電子命令 (electronic instructions) の記憶に適したあらゆる種類の媒体、および各々がコンピュータシステムバスに連結されたもの。さらに、本明細書に記載されるコンピュータには単一のプロセッサが含まれ、あるいは、演算能力を向上させるために複数のプロセッサデザインを用いた構造であってもよい。

【0020】

ここでいうアルゴリズムや表示は、特定のコンピュータやこの他の装置のみを指すものではない。各種の汎用システムは、ここに教示される内容に従ったプログラムとともに用いてもよく、あるいは、必要な方法ステップを実行するために専用の装置を構築するほうが利便な場合もある。これらの各種システムに必要な構造は、以下の記載から明らかである。さらに、本発明は、特定のプログラミング言語を指して記載されるものではない。各種のプログラミング言語を用いて、以下に記載される本発明の教示を実装するようにしてよく、さらに、以下の特定言語は、本発明の実施可能性と最良の形態を開示するために記載されたものであることは理解されるであろう。

10

【0021】

さらに、本明細書中の説明で用いられた文言は、読みやすさや例示的な目的から主に用いられたものであって、本発明の主題の範囲を詳説したり、限定するために選択されたものではない。したがって、本発明の開示は、例示を意図したものであって、以下の特許請求の範囲に記載される本発明の範囲を限定するものではない。

20

【0022】

手サイン認識の一実施形態によれば、サイン認識システムにより認識、特定できるサインの数を増大させることができる。具体的には、サイン認識システムは、奥行画像において両手が重なり合ったり隣接する場合であっても、両手の形状を決定するものである。重なり合ったり隣接する両手の形状は、(i)コスト関数と制約条件を用いて、リンク構造のセグメントを分類する問題を線形計画最適化問題 (linear programming optimization problem) の形態のラベル付与問題に変換する制約条件付最適化法 (constrained optimization method)、あるいは(ii)複数のノードを含む非循環有向グラフ (directed acyclic graph) を用いて、奥行画像に表される両手について最も可能性が高そうな形状を求めるツリー検索法 (tree search method) によって決定するようにしてよい。両手の形状を決定した後、リンク構造のセグメントを、記憶している手形状とマッチングさせて、奥行画像に表されるサインを決定する。

30

【0023】

ここで言うリンク構造とは、重なり合ったり隣接する両手の形状を決定するために用いる画像画素から得たリンク線のセットを言う。このようなリンク構造には、奥行画像の手の特徴を表す2つ以上のセグメントが含まれる。リンク構造は、例えば、重なり合ったり隣接する手を表す画像画素のクラスタを細線化し、さらに、手形状を特定、認識するための処理を行うことで求める。このように、リンク構造を用いれば、手話サインを認識、特定する際に画像画素を処理する必要がなくなるため、手話サインを認識するための演算やアルゴリズムを簡略化することができる。

40

【0024】

図面と以下の記載は、実例のみを用いた本発明の好ましい実施形態に関する。以下の記載から、ここに開示される構成や方法の代替の実施形態は、請求項に係る発明の原則から逸脱しない範囲で実現可能な代替案として容易に認識されるであろう。

【0025】

サイン認識システムの構造

図2は、本発明の一実施形態に係るサイン認識システム200の概略ブロック図である。サイン認識システム200は、構成部品の中でも特に、画像キャプチャ装置210と画像処理モジュール220とを備える。画像キャプチャ装置210は、対象252の奥行画像212を取得し、この奥行画像212を画像処理モジュール220に送信する。画像処

50

理モジュール 220 は、奥行画像 212 を処理し、特定した対象 252 の手サインを示すデータ 240 を出力する。

【0026】

一実施形態では、画像キャプチャ装置 210 は、手形状を分析するために十分な解決策として機能する飛行時間センサを備え、例えば、スイス連邦チューリッヒのCSEM SA社製、イスラエル国ヨークナムの3DV Systems社製、あるいはカリフォルニア州サンノゼのCasta, Inc社製のカメラなどである。飛行時間センサは、対象が動き回ると予測される範囲の領域を包含する奥行のある窓に設置する。なお、背景などの非対象物は奥行画像 212 の一部をなさない。別の実施形態では、画像キャプチャ装置 210 は、距離を隔てて設置した 2 つ以上の画像センサに基づいて、奥行画像 212 を生成するステレオビジョンカメラを備える。符号化光や時間 - 空間ステレオ画像を用いて奥行画像 212 を生成するようにもよい。10

【0027】

画像処理モジュール 220 は、奥行画像 212 から手サインを認識するための専用装置であってもよい。あるいは、画像処理モジュール 220 は、パーソナルコンピュータなどの汎用演算装置の一部であってもよい。画像処理モジュール 220 は、構成部品の中でも特に、コンピュータプロセッサや、RAM、ROM、フラッシュメモリ、ディスクドライブ等のコンピュータ読み取り可能媒体などの従来のデータ処理装置やデータ記憶装置を備える。好ましくは、コンピュータプロセッサは、実行可能なソフトウェアプログラムからのプログラム指令が実行されると画像処理機能を具現化するような能力を有する従来の汎用プロセッサである。一実施形態では、画像処理モジュール 220 は、ジェスチャの特徴を認識し、関連する変化を表示装置（図示せず）に表示するために、1 つ以上のコンピュータプログラムにアクセスできるようなコンピュータ読み取り可能媒体（図示せず）を備える。20

【0028】

一実施形態では、サイン認識システム 200 は、例えば、日本国東京の本田技研工業株式会社製の最新人間型ロボット「ASIMO」（登録商標）などに用いるロボット用人間 - 機械インタラクションシステムの一部である。本実施形態では、画像キャプチャ装置 210 は、ロボットビジョンユニットの一部であってもよく、画像処理モジュール 220 はロボットの制御理論の一部であってもよい。ロボット用人間 - 機械インタラクションシステムにおけるサイン認識システム 200 は、例えば、会話や音声による指令が有効でないような騒々しい環境でロボットに指示を与えるときに用いるようにしてよい。30

【0029】

画像処理モジュール 220 は画像キャプチャ装置 210 に接続しており、奥行画像 212 を受信する。画像処理モジュール 220 は、構成部品の中でも特に、画像バッファ 222 と、画像前処理モジュール 224 と、形状マッチングモジュール 226 とを備える。画像バッファ 222 は、画像キャプチャリング装置 210 から奥行画像 212 を受信するように構成されている。画像バッファ 222 は、周辺インターフェースハードウェア、ネットワークインターフェースハードウェア（有線または無線）などの一部であってもよい。一実施形態では、画像バッファ 222 から受信した奥行画像 212 を、ハードディスク、フラッシュメモリ、RAMなどの内部データ記憶装置（図示せず）に記憶する。40

【0030】

画像前処理モジュール 224 では、各種の画像処理作業を行い、その中でも特に、例えば、米国特許出願番号第11/129164号、発明の名称「Sign Based Human-Machine Interaction」、出願日2005年5月12日（米国特許公開番号第2005/0271279号として公開）、および米国特許出願番号第11/869435号、発明の名称「Human Pose Estimation and Tracking Using Label Assignment」、出願日2007年10月9日に記載されるような、対象 252 の特定の部位を表す画像画素のモーション追跡やクラスタ抽出などを行う。なお、前記文献に記載されるすべての内容をここに引用する。サイン認識については、画像前処理モジュール 224 は、図 4 を参照しつつ後述するように、処理作業の中でも特に、両50

手を表す画像画素のクラスタの抽出、両手のひらの検出、およびリンク構造の生成の処理を行う。

【0031】

形状マッチングモジュール226は画像前処理モジュール224に接続しており、リンク構造と両手のひらの位置に関する情報を受信する。形状マッチングモジュール226は、リンク構造のどのセグメントをどちらの手に割り当てるべきかを決定し、この決定した手の形状と記憶している手形状とをマッチングさせて、対象252が示す手形状を決定する。リンク構造のどのセグメントをどちらの手に割り当てるべきかを決定すれば、どの画像画素をどちらの手に割り当てるべきかを決定することができる。一実施形態では、画像画素を隣接のセグメントと同じ手に割り当てる。

10

【0032】

図3(a)および(b)は、本発明の実施形態に係る形状マッチングモジュール226のブロック図を示している。図3(a)に示す実施形態では、形状マッチングモジュール226Aは、オーバーラップ手形状検出モジュール316と、非オーバーラップ手形状マッチングモジュール320と、制約条件付最適化モジュール340と、形状データベース350とを備える。オーバーラップ手形状検出モジュール316は、手が重なり合ったり隣接していることを示す画像画素が奥行画像212に含まれているか否かを決定する。

【0033】

オーバーラップ手形状検出モジュール316は、奥行画像の手が重なり合ったり隣接していないと決定した場合には、非オーバーラップ手形状マッチングモジュール320を用いて、例えば、米国特許出願番号第11/129164号、発明の名称「Sign Based Human-Machine Interaction」(米国特許公開番号第2005/0271279号)に開示されるような方法で手サインを認識する。一方、奥行画像212の手が重なり合ったり隣接していると決定した場合には、図8ないし図10を参照しつつ詳細に後述するように、制約条件付最適化モジュール340を用いて手形状を決定するとともに、手形状が示すサインを決定する。

20

【0034】

例えば、米国特許出願第11/129164号、発明の名称「Sign Based Human-Machine Interaction」(米国特許公開番号第2005/0271279号)に開示されるように、形状データベース350には、非オーバーラップ手形状マッチングモジュール320あるいは制約条件付最適化モジュール340で決定した手形状とマッチングさせるための手形状のプロファイルが記憶されている。なお、前記文献に記載されるすべての内容をここに引用する。

30

【0035】

図3(b)の実施形態は、制約条件付最適化モジュール340の代わりにツリー検索モジュール344を用いる点を除いて、実質的に図3(a)の実施形態と同じである。図3(b)のオーバーラップ手形状検出モジュール316の機能と、非オーバーラップ手形状マッチングモジュール320の機能と、形状データベース350の機能は、本質的に図3(a)の実施形態と同じである。しかしながら、図3(b)の実施形態のツリー検索モジュール344は、非循環有向グラフを用いて、図9ないし図11を参照しつつ詳細に後述するような方法で手形状を決定する。図3(a)および(b)の形状マッチングモジュール226Aおよび226Bは、以下においては、まとめて形状マッチングモジュール226と呼ぶ。

40

【0036】

手サイン認識方法

図4は、本発明の一実施形態に係る手サイン認識方法のフローチャートを示している。まず、画像前処理モジュール224は奥行画像212から両手を表す画像画素のクラスタを抽出する(S410)。両手を表す画像画素のクラスタを決定するには、例えば、米国特許出願番号第11/869435号、発明の名称「Human Pose Estimation and Tracking Using Label Assignment」、出願日2007年10月9日に開示されるように、奥行画像および色画像の一部分を、対象の異なる部位に分類し、ラベリングすることで決定できる。なお、前記文献に記載されるすべての内容をここに引用する。次に、画像前処理モジュール2

50

24は、抽出した画像画素のクラスタに重なり合ったり隣接する両手が含まれているか否かを決定する(S420)。

【0037】

画像画素のクラスタに含まれる手が重なり合っていたり隣接していると決定した場合には、図5および図6を参照しつつ詳細に後述するように、前処理モジュール224は両手のひらを検出する。続いて、前処理モジュール224は、例えば、手を表す画像画素のクラスタについて中間軸変換(medial axis transform)を行い、リンク構造(図7Cを参照)を生成する(S440)。一実施形態においては、中間軸変換は、クラスタに細線化処理を施すことによって実行する。前処理モジュール224で生成したリンク構造は、画像画素のクラスタのスケルトンを近似化したものである。10

【0038】

次に、形状マッチングモジュール226は、制約条件付最適化もしくはツリー検索を行って、図8および図10を参照しつつ詳細に後述するように、手形状の曖昧さを解像する(S450)。手の形状の曖昧さを、リンク構造もしくはノードのセグメントを片方の手かもう片方の手に割り当てることで解像する。

【0039】

制約条件付最適化法あるいはツリー検索法を用いて、リンク構造またはノードのセグメントをどちらかの手に割り当てた後、例えば、米国特許出願番号第11/129164号、発明の名称「Sign Based Human-Machine Interaction」、出願日2005年5月12日(米国特許公開番号第2005/0271279号)に開示されるように、形状マッチングモジュール226において、S450で解像した手形状を、形状データベース350に記憶している手形状プロファイルにマッチングさせる(S460)。なお、前記文献に記載されるすべての内容をここに引用する。形状マッチングモジュール226は、奥行画像212の手形状を、記憶している手形状プロファイルにマッチングさせることで、特定した手サイン240を出力する(S470)。20

【0040】

S420において、奥行画像212に重なり合ったり隣接する手が含まれないと決定した場合は、処理をS480に進めて、例えば、米国特許出願番号第11/129164号、発明の名称「Sign Based Human-Machine Interaction」、出願日2005年5月12日(米国特許公開番号第2005/0271279号)に開示されるような、距離マッチングやスケルトンマッチングなどの技法を用いて、奥行画像212の手形状を決定する。なお、前記文献に記載されるすべての内容をここに引用する。次に、処理をS470に進めて、特定した手サインを出力して処理を終了する。30

【0041】

手のひら検出

図5は、本発明の一実施形態に係る手のひらを検出する方法に関するフローチャートを示している。ツリー検索法および制約条件付最適化法の精度は、手のひらの検出の正確さに左右されるため、手のひらの位置の決定は非常に重要である。ツリー検索法と制約条件付最適化法のどちらも、セグメントを片方の手またはもう片方の手に分類する(これにより画像画素を分類する)際には、手のひらまでの相対的な距離やセグメントが手のひらにリンクしているか否かの判断に一部基づいて行う。したがって、手のひらの位置の検出が正確でないならば、セグメントを手に分類するときの精度が低下してしまう。40

【0042】

まず、従来の方法を用いて、手を表す(場合によっては前腕の一部も含む)画像画素のクラスタ600のスケルトンを生成する(S520)。図6を参照しつつ、画像画素のクラスタ600のスケルトンを表す線630を描く。次に、線630(つまりスケルトン)と直交するクラスタの幅を測定する(S530)。クラスタの幅は、線630と直交する線が手の輪郭線604と交わる点どうしの距離を測定することで決定する。図6は、クラスタ600の幅は、線630に沿って移動するときに直交する線(例えば、線610、620、640、650)を用いて測定する事例を示している。極大値となる長さ(クラス50

タの幅)を有する直交する線(線610、620、640、650)を、手のひらの中央部を含んでいる可能性のある線の候補として多数選択する。

【0043】

図5に戻って、閾値よりも長さが短い線の候補を捨てる(S550)。長さの短い線の候補は、手のひらではなく指を表している可能性が高いため候補から外す。例えば、図6の線610および620はこの段階で捨てる。

【0044】

残りの線の候補(例えば線640と650)の中で、指に最も近い候補である線640を手のひらの中央部を含んでいる線として選択する(S560)。一実施形態では、指の検出は、線周辺のクラスタが、同じスケルトン中の他の線のクラスタよりも幅細であるような線から検出する。手のひらの中央部を含んでいる線として、線640を選択すると、手のひらの中央部として、線640の中央部を決定する(S570)。一実施形態では、手のひらの形状は、半径または軸の最大値がクラスタ内に収まるような円や橢円を選択して決定する。

【0045】

リンク構造の事例

図7(a)は、本発明の一実施形態に係るサイン認識システム200でこれから処理しようとするサインを含んだ奥行画像を例示したものである。

図7(a)は、日本手話(JSN)の「B」という文字を表している対象の画像から抽出した奥行画像である。図7(a)の奥行画像に含まれる、対象の画像の背景とその他の物品は画像前処理モジュール224で削除している。

【0046】

図7(b)は、本発明の一実施形態に係る図7(a)の奥行画像から生成したスケルトン構造を例示したものである。

右の手のひら710と左の手のひら720を、例えば、図5および図6を参照しつつ説明したような方法で検出する。図7(b)のスケルトン構造は、複数のセグメントに区分する前の状態である。

図7(b)のスケルトン構造には、手の特徴になり得ないような短い断片(例えば、ぶら下がり線730)を含んでいる可能性がある。このような短い断片は、デジタル画像中のノイズとなったり、あるいは画像処理中に発生するアーチファクトとなることが多い。

【0047】

図7(c)は、本発明の一実施形態に係る、図7(b)から生成したリンク構造700を例示している。リンク構造700は、複数のノードでセグメント(S0-S8)に分化されている点で、スケルトン構造とは異なる。

リンク構造700には、図7(b)のぶら下がり線730のような短い断片は含まれない。リンク構造700を生成するために、スケルトン構造を次のようなノードで区分するようにしてもよい:

(i) 3本以上の線が集まる分岐ノード

このようなノードは、指が始まる、隣接する、終端となるような点である可能性が高い。

(ii) 線の角度が急に変化するような点(例えば、図7(b)の点740)

このような点では手の指やその他の部位と交差する可能性が高い。

(iii) 所定の閾値よりも長い線の途中で最大曲率の変化が発生するような点(例えば、図7(c)の点734)

手のセグメントであれば、このような閾値を超える長さを有する可能性は低い。

【0048】

制約条件付最適化法

図8は、本発明の一実施形態に係る、制約条件付最適化法を用いてリンク構造のセグメントをグループ化する方法を例示したものである。

制約条件付最適化モジュール340は、まず、リンク構造のセグメントを、右手と左手

10

20

30

40

50

に分類するためのコスト関数を生成する。一実施形態では、次のようなコスト関数Q(f)を用いる。

【0049】

【数1】

$$Q(f) = \sum_{i=1}^N \sum_{j=1}^M c(i,j) A_{ij} + \sum_{e \in E} w_e Z_e \dots \text{式 (1)}$$

【0050】

10

ここで、Nはリンク構造中のセグメントの合計数であり、Mは手の数（つまりM = 2）、c(i, j)はセグメント s_i を手 j ($j = 1$ は左手、 $j = 2$ は右手、またはその反対) に割り当てる費用を表し、 A_{ij} はセグメント s_i が右手と左手のどちらに割り当てられているかを表す（下の表1に示す関連表の中に記載されるように）0または1の値を有するパラメータを表し、 w_e は隣接する2つのセグメント間の奥行差を表す重みを表し、 Z_e はセグメント分類の問題をラベル付与問題に変換するために導入されるパラメータである。最適化処理の全体の目的は、コスト関数 $Q(f)$ を最小化する A_{ij} のセットを見つけることである。

【0051】

20

式(1)の重み w_e は、セグメント s_i ($i = 1, 2, \dots, N$) 間の関係の強さを表す。互いに隣接するセグメントどうしは強い関係を有し、同じラベル付部位に分類される可能性が高い。このような強い関係はより高い重み w_e で表される。対照的に、互いに隣接しないセグメントどうしは弱い関係を有する。このように弱い関係はより低い重み w_e で表される。重み w_e に関するパラメータ Z_e は、セグメント分類の問題を、線形計画法による最適化問題の形式のラベル付与問題に変換するために導かれるものである。

【0052】

【表1】

手 (j) セグメント	1 (右)	2 (左)
S_1	A_{11}	A_{12}
S_2	A_{21}	A_{22}
...
S_N	A_{N1}	A_{N2}

30

40

【0053】

一実施形態では、 $Q(f)$ を最小化する問題は次の制約条件に左右される。

【0054】

【数2】

$$\sum_j A_{ij} = 1, (i=1,2,3\cdots N), \text{if the } i^{\text{th}} \text{ segment is thin} \dots \dots \dots \text{式(2)}$$

$$\sum_j A_{ij} = 2, (i=1,2,3\cdots N), \text{if the } i^{\text{th}} \text{ segment is thick} \dots \dots \dots \text{式(3)}$$

$$Z_e = \frac{1}{2} \sum_j Z_{ej}, e \in E \dots \dots \dots \text{式(4)} \quad 10$$

$$Z_{ej} \geq A_{pj} - A_{qj}; e = (p,q); (j = 1 \text{ or } 2) \dots \dots \dots \text{式(5)}$$

$$Z_{ej} \geq A_{qj} - A_{pj}; e = (p,q); (j = 1 \text{ or } 2) \dots \dots \dots \text{式(6)}$$

$$A_{ij} \in \{0,1\}; i = 1,2,3\cdots, N; (j = 1 \text{ or } 2) \dots \dots \dots \text{式(7)}$$

$$\sum_i \text{length}(S_i) A_{ij} < \text{Max} \text{ (for } j = 1) \dots \dots \dots \text{式(8)} \quad 20$$

$$\sum_i \text{length}(S_i) A_{ij} < \text{Max} \text{ (for } j = 2) \dots \dots \dots \text{式(9)}$$

【0055】

式(2)は、周辺に幅の狭いクラスタ群を有するセグメントには指が1本のみ含まれる、という制約条件を表している。したがって、このようなセグメントは右手または左手に分類することはできるが、両手に分類することはできない。

式(3)は、幅の広い複数のクラスタを有するセグメントには2本の指が含まれる、という制約条件を表している。したがって、このようなセグメントは、左手、右手あるいは両手に分類することができる。

【0056】

式(4)ないし式(6)(Z_{ej} は絶対値 $|A_{pj} - A_{qj}|$ を表す)は、隣接するセグメントどうしは、同じラベル付け部位または隣接のラベル付け部位に分類されなければならない、という制約条件から導かれたものである。式(7)は、前記の通り、 A_{ij} は0か1である、という制約条件を表す。式(8)と式(9)は、右手または左手の指の長さの合計は最大値を越えてはならない、という制約条件を表す。

【0057】

式(1)ないし式(9)で示される最適化問題は、NP困難と呼ばれる整数計画(integer program)である。一実施形態では、 A_{ij} を非整数値にすることで、最適化問題を線形計画法に簡略化することができる。このような簡略化を用いることで、式(1)ないし式(9)で表される最適化問題は、一般に入手可能なソフトウェアライブラリ、例えば、lp_solve 5.5.0.10 (<http://lpsolve.sourceforge.net/5.5/> 等で入手可能)を使って効果的に解くことができる。

【0058】

一実施形態では、式(1)の重み w_e を無視(w_e をゼロに設定)し、その代わりに、指が鋭角度に曲げられる場合を排除する制約条件を追加する。このような制約条件は、例えば、閾値(例えば、160度)未満の角度で交差するような隣接するセグメント S_a および S_b (ここで、 $1 \leq a \leq N$ および $1 \leq b \leq N$)を求めるための下のような式(10)

10

30

40

50

) および式(11)で表すことができる。式(10)および式(11)は、セグメント S_a と S_b とはそれぞれ異なる手に分類されなければならないことを示している。

$\begin{bmatrix} 0 & 0 & 5 & 9 \end{bmatrix}$

【数3】

10

〔 0 0 6 0 〕

式(2)、(3)、(5)ないし(11)で表される制約条件下で、重み w_e がゼロに設定された式(1)の最小化問題を解くために、一般に入手可能なソフトウェアライブラリ、例えば、Ip_solve 5.5.0.10を用いてもよい。

[0 0 6 1]

一実施形態では、すべてのセグメントを右手または左手、もしくは両手に割り当てた後に、奥行画像212の画像画素を、隣接のセグメントが割りあてられている手と同じ手に割り当てる。

[0 0 6 2]

ツリー検索法

20

図9は、本発明の一実施形態に係る、ツリー検索法を用いてリンク構造のセグメントを2つの手に分類する方法に関するフローチャートである。まず、従来の方法を用いて、リンク構造700から非循環有向グラフを生成する(5920)。図11は、図7(c)のリンク構造から生成した非循環有向グラフ1100をツリーの形態で構成したもの(よって、この方法を「ツリー検索法」と呼ぶ)を示す。

[0 0 6 3]

図 11 に示す非循環有向グラフ 1100 の例では、非循環有向グラフのトップノード 8 は、右の手のひら 710 に最も近いセグメント S8 (図 7 (c) を参照) から始まって、左の手のひら 720 に向けて処理が進むにしたがって、連結するセグメントへ次々と広がっていく。非循環有向グラフ 1100 の矢印はノード間の連結関係を示す。なお、非循環有向グラフ 1100 は、代替の方法を用いて生成してもよい。

30

[0 0 6 4]

再び図9を参照しつつ、非循環有向グラフ1100を生成した後は、図10を参照しつつ詳細に後述する通りに、右手の手形状と左手の手形状の候補を生成する(S930)。

[0 0 6 5]

1つ以上の基準を用いて、これらの手形状の候補を評価する（S940）。一実施形態では、各基準についてスコアを計算し、これらスコアを足して、手形状の候補を評価する。手形状の候補を評価するための基準には、例えば、以下のようなものが含まれる：（i）手は物理的にリンクする物体であることから、左手と右手のノードがリンクするような手形状の候補には高いスコアを付与し、（ii）指は手のひらの近くにあることから、手のノードが、この手のひらから所定の距離内に位置するような手形状の候補には高いスコアを付与し、（iii）指を鋭角に曲げることはあまりないため、セグメントどうしの角度が鋭角であるような手形状の候補には低いスコアを付与し、（iv）分岐するセグメントが指から始まることはあまりない（通常、分岐するセグメントは、複数の指が連結する手のひらから始まる）ため、分岐するノードが、この分岐するノードが割り当てられた手のひらから離れた位置にあるような手形状の候補には低いスコアを付与する。基準（i）ないし（iv）までのスコアを算出し、これらのスコアを足して合計スコアを求める。なお、これらの基準は限定的、包括的なものではなく、異なる基準を用いて両手形状の候補を評価するようにしてもよい。

40

[0 0 6 6]

50

最もスコアの高い手形状の候補を、最も可能性の高い手形状として選択する。異なる基準を表す異なるスコアに対しては、異なる重みを付与するようにすることで、異なる基準を組み合わせた場合にも正確な評価を行えるようにしてよい。

【0067】

図10は、本発明の一実施形態に係る、左手と右手の手形状の候補を生成するフロー チャートを示している。可能と思われる手形状の候補をすべて生成して、手形状の候補をすべてを評価することは可能であるが、ノードとセグメントで表される手の物理的構造を考慮すれば、手形状によっては実際には不可能なものもあるため、こうしたやり方は、演算コストが高くなり非効率的である。したがって、本発明の一実施形態では、図9のステップS940でさらに評価するために、ノードどうしの特定の連結関係を満たすような手形状の候補のみを生成する。10

【0068】

本実施形態では、図11に例示される非循環有向グラフ1100を生成した後、top-to-bottom検索 (top-to-bottom search) を行って、どちらかの手 (例えば、右手) に割り当てるためのノードの組み合わせを生成する (S1020)。具体的には、非循環有向グラフ1100のトップノード8から始まって、右手について4組のノードの組み合わせ方が可能であることが分かる：(i) ノード8のみが右手に属する (ノード8以外のノードは左手に割り当てる) 第1の手形状の候補、(ii) ノード6から分岐する枝に含まれるノードのすべてまたは一部を右手に分類する、多数の手形状の候補、(iii) ノード3から分岐する枝に含まれるノードのすべてまたは一部を右手に分類するような、多数の手形状の候補、(iv) (ノード3から分岐する枝とノード6から分岐する枝の) 両方の枝に含まれるノードのすべてまたは一部を右手に分類するような、多数の手形状の候補。ノード8、6、3を含んでいることに基づいて手形状の候補を決定した後、処理を次の段階へ進めて、ノード8のときと同じ方法で、これらの手形状の候補がノード0、5および2を含んでいるか否かを決定する。同様に、これらの手形状の候補がノード7、4および1を含んでいるか否かについても決定する。20

【0069】

右手の手形状の候補に含まれると決定したノードどうしを連結して、右手の手形状の候補をすべて生成する (S1030)。次に、どちらかの手 (例えば、右手) の手形状の候補の各々について、非循環有向グラフ1100のノードからこの (右手の) 手形状の候補のノードを削除して、反対側 (左手) の手形状の候補のノードを決定する (S1040)。続いて、反対側の手形状の候補の中にあるノードを連結して、左手の手形状の候補を生成する (S1050)。例えば、右手の手形状の候補にノード8、6、0、7が含まれる場合は、反対側の左手の手形状の候補にはノード1ないし5が含まれる。この場合の事例では、ノード8、6、0、7を連結して右手の手形状の候補を生成し、ノード1ないし5を連結して左手の手形状の候補を生成する。30

【0070】

非循環有向グラフ1100から、右手の連結関係を満たすような右手の手形状と左手の手形状との組み合わせをすべて生成した後、図9を参照して説明したような基準ごとにスコア付けを行って、各手形状の組み合わせを評価をする。40

【0071】

実験結果

図12は、本発明の各種実施形態に係る、両手に割り当てた画像画素の事例を示したものである。具体的には、図12の事例は、制約条件付最適法を用いて正確に特定した、日本手話 (JSL) のアルファベット文字のA、アルファベット文字のK、アルファベット文字のG、アルファベット文字のB、アルファベット文字のX、単語の「井戸」、単語の「会う」、単語の「手紙」を例示している。図12の各画像中の円は、図5を参照して詳述した方法を用いて検出した両手のひらを表している。

【0072】

ここで実験では、サイン認識システムは、2.13GHzペンティアム（登録商標）

50

プロセッサで動作するプロセッサと2ギガバイトのRAMを備えたラップトップ型コンピュータにより具現化した。かかるハードウェア構成のサイン認識システムは、約0.5秒のサインを正確に識別できた。このように認識時間が俊敏であることは、本発明の実施形態に係る手形状マッチングモジュールは演算コストが安価であり、サイン認識処理全体にかかる演算費用を最小化できることを意味する。

【0073】

代替の実施形態

一実施形態では、サイン認識システムは、例えば、米国特許出願番号第11/129164号、発明の名称「Sign Based Human-Machine Interaction」、出願日：2005年5月12日（米国特許公開番号第2005/0271279号）に開示されるような、手の軌道を追跡するためのジェスチャマッチングモジュールをさらに備える。なお、前記文献に記載されるすべての内容をここに引用する。このジェスチャマッチングモジュールを、本発明の各種実施形態に係る形状マッチングモジュールに対応させて、重なり合ったり隣接する手の動きを追跡するようにしてもよい。10

【0074】

一実施形態では、手形状の識別の精度を向上させるために、色画像データを用いてもよい。奥行画像だけを用いて正確に識別するのが困難な場合には、色画像データから取得した色と形状を用いて、手形状の曖昧さを解像するようにしてもよい。色画像データを用いる場合は、さらに基準や制約条件を追加するようにしてもよい。

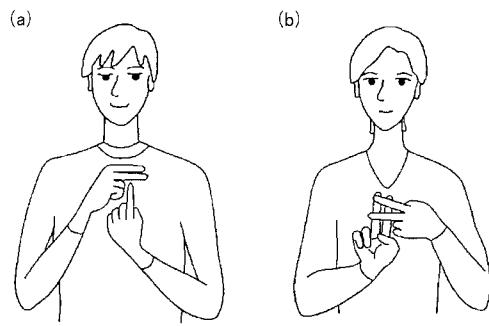
【0075】

一実施形態では、形状マッチングモジュール226は、制約条件最適化モジュールとツリー検索モジュールの両方を備えるようにしてもよい。形状マッチングモジュール226は、異なる状況であっても、より正確な結果が得られるように2つの方法（制約条件最適化法とツリー検索法）に関する情報を備えるようにして、状況に応じて制約条件最適化モジュールとツリー検索モジュールのいずれかを用いるようにしてもよい。別の実施形態では、制約条件最適化モジュールとツリー検索モジュールのそれぞれで予備的な識別サインを生成し、さらに、形状マッチングモジュール226で解析・選択の処理を行ってから、最終的な識別サインを生成するようにしてもよい。

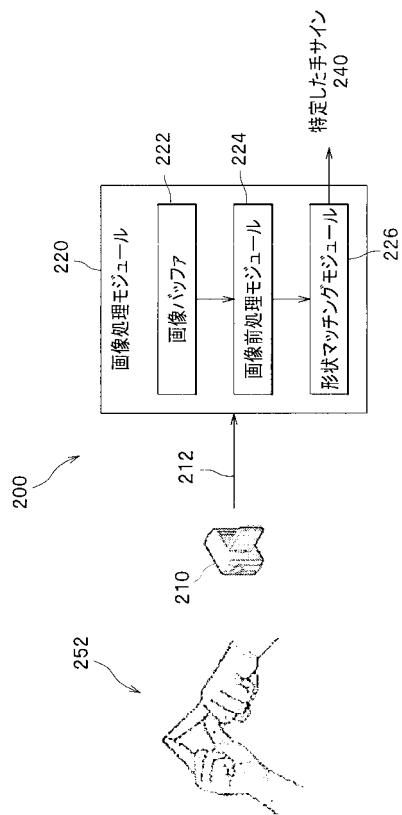
【0076】

本発明の特定の実施形態および応用例について例示し、説明したが、本発明は、ここに開示される構造や構成部品に厳密に限定されるものではなく、本発明の方法と装置の配置構成、動作および詳細について、付属の特許請求の範囲に定義されるような本発明の趣旨および範囲から逸脱しない限りにおいて各種の修正、変更、変形例が可能である。30

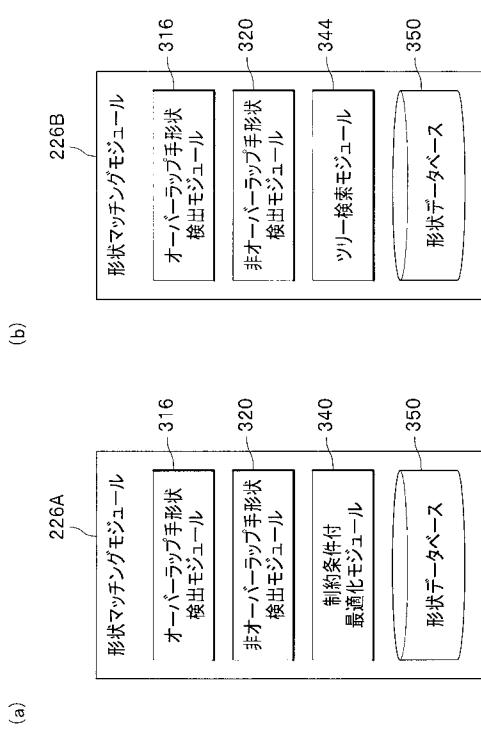
【図1】



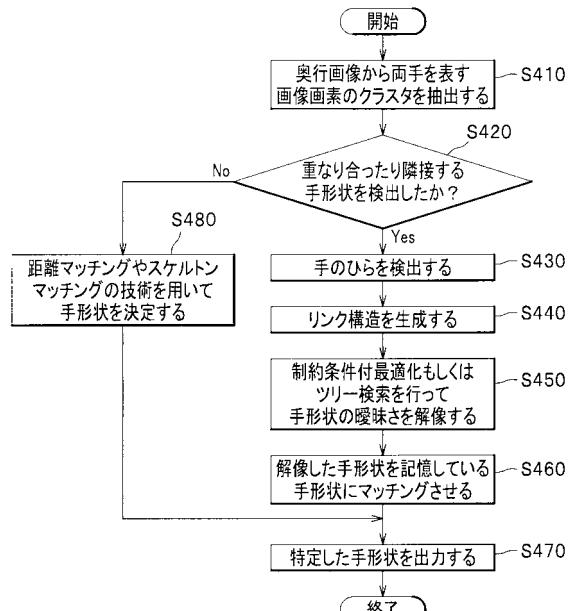
【図2】



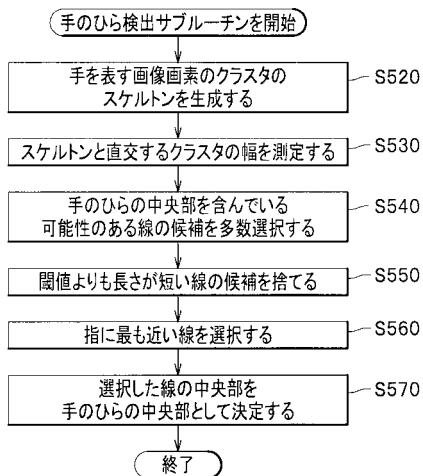
【図3】



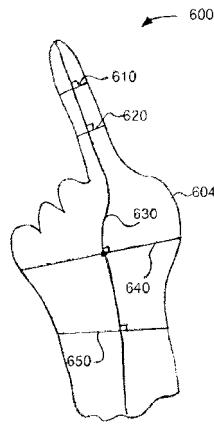
【図4】



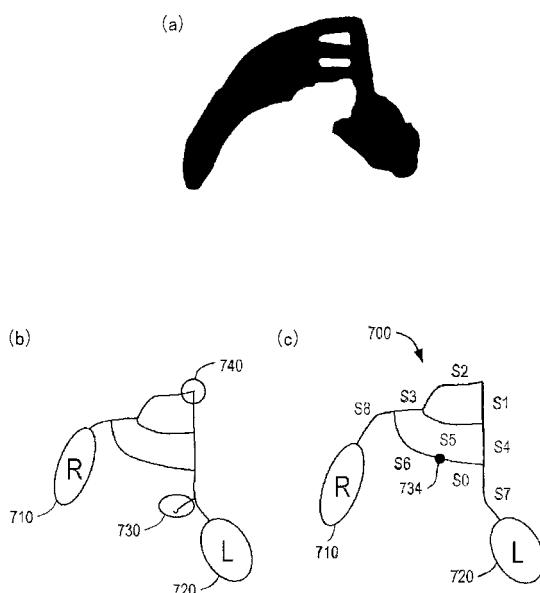
【図5】



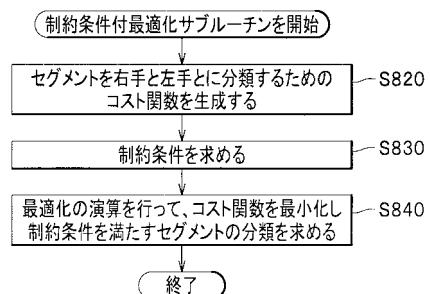
【図6】



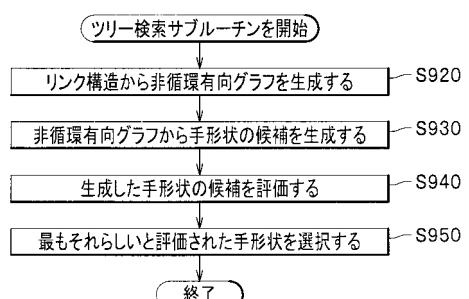
【図7】



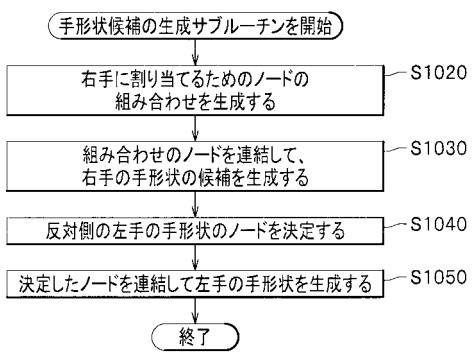
【図8】



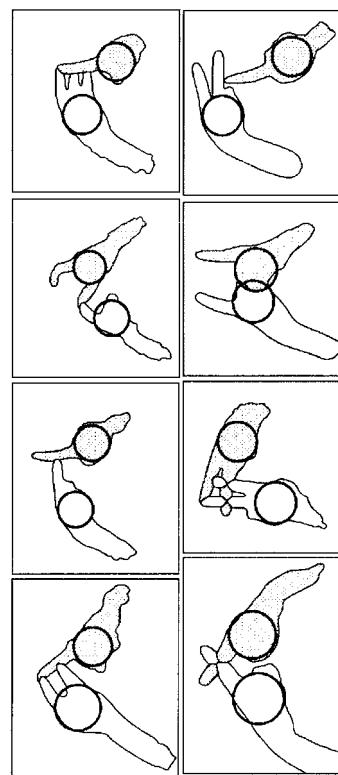
【図9】



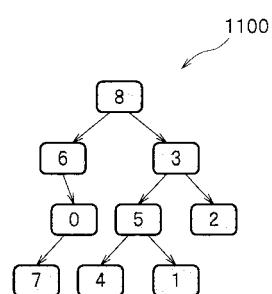
【図10】



【図12】



【図11】



フロントページの続き

(72)発明者 藤村 希久雄

アメリカ合衆国，オハイオ州 93067-9705，レイモンド，21001 ステート ルート 739，ホンダ アール アンド ディ アメリカズ インコーポレーテッド，リーガル デパートメント気付

(72)発明者 シュウ，リジェ

アメリカ合衆国，オハイオ州 93067-9705，レイモンド，21001 ステート ルート 739，ホンダ アール アンド ディ アメリカズ インコーポレーテッド，リーガル デパートメント気付

審査官 新井 則和

(56)参考文献 米国特許出願公開第2004/0151366(US, A1)

米国特許出願公開第2005/0031166(US, A1)

米国特許第05787194(US, A)

特開2002-366958(JP, A)

(58)調査した分野(Int.Cl., DB名)

G06T 7/20