

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4332649号
(P4332649)

(45) 発行日 平成21年9月16日 (2009.9.16)

(24) 登録日 平成21年7月3日 (2009.7.3)

(51) Int.Cl.

F I

G O 6 T 1/00 (2006.01)

G O 6 T 1/00 3 4 O Z

G O 9 B 21/00 (2006.01)

G O 9 B 21/00 E

G O 6 T 7/20 (2006.01)

G O 6 T 7/20 3 O O A

請求項の数 30 (全 61 頁)

(21) 出願番号 特願平11-321580

(22) 出願日 平成11年11月11日 (1999.11.11)

(65) 公開番号 特開2001-56861 (P2001-56861A)

(43) 公開日 平成13年2月27日 (2001.2.27)

審査請求日 平成18年8月24日 (2006.8.24)

(31) 優先権主張番号 特願平11-161557

(32) 優先日 平成11年6月8日 (1999.6.8)

(33) 優先権主張国 日本国 (JP)

(73) 特許権者 301022471

独立行政法人情報通信研究機構

東京都小金井市貫井北町4-2-1

(73) 特許権者 000005821

パナソニック株式会社

大阪府門真市大字門真1006番地

(74) 代理人 100098291

弁理士 小笠原 史朗

(72) 発明者 今川 和幸

大阪府門真市大字門真1006番地 松下

電器産業株式会社内

(72) 発明者 松尾 英明

大阪府門真市大字門真1006番地 松下

電器産業株式会社内

最終頁に続く

(54) 【発明の名称】 手の形状と姿勢の認識装置および手の形状と姿勢の認識方法並びに当該方法を実施するプログラムを記録した記録媒体

(57) 【特許請求の範囲】

【請求項 1】

光学的読取手段によって取得された手の画像（以下、入力手画像という）の形状および姿勢を認識する装置であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向、画像の大きさ、画像の明るさ）に正規化した手形状画像を、それぞれ生成する第1の手画像正規化手段と、

前記手形状画像を、当該手形状画像に関する形状情報および姿勢情報と共に、それぞれ格納する手形状画像情報記憶手段と、

固有空間法による解析を行って、前記手形状画像から固有値と固有ベクトルとを、それぞれ計算する固有空間計算手段と、

前記固有ベクトルの集合を格納する固有ベクトル記憶手段と、

前記手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求め、前記手形状画像情報記憶手段に格納する第1の固有空間投影手段と、

前記入力手画像を入力し、当該入力手画像の手首領域を削除した後前記手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第2の手画像正規化手段と、

前記入力手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第2の固有空間投影手段と、

10

20

前記第 2 の固有空間投影手段が求めた前記投影座標と、前記手形状画像情報記憶手段に格納した前記投影座標とをそれぞれ比較し、前記入力手形状画像に最も近い前記手形状画像を求める手形状画像選択手段と、

前記最も近い手形状画像の前記形状情報と前記姿勢情報とを、前記手形状画像情報記憶手段から取得して出力する形状・姿勢出力手段とを備える、手の形状と姿勢の認識装置。

【請求項 2】

光学的読取手段によって取得された手の画像（以下、入力手画像という）の形状および姿勢を認識する装置であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向、画像の大きさ、画像の明るさ）に正規化した手形状画像を、それぞれ生成する第 1 の手画像正規化手段と、

前記手形状画像を、当該手形状画像に関する形状情報および姿勢情報と共に、それぞれ格納する手形状画像情報記憶手段と、

固有空間法による解析を行って、前記手形状画像から固有値と固有ベクトルとを、それぞれ計算する固有空間計算手段と、

前記固有ベクトルの集合を格納する固有ベクトル記憶手段と、

前記手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求め、前記手形状画像情報記憶手段に格納する第 1 の固有空間投影手段と、

前記投影座標を、クラスタ分析によりそれぞれグループ分けし、各前記手形状画像がどのクラスタに属するかを求めて前記手形状画像情報記憶手段に格納すると共に、各クラスタに関する統計情報を求めるクラスタ分析手段と、

前記統計情報を、対応するクラスタと共に、それぞれ格納するクラスタ情報記憶手段と、

前記入力手画像を入力し、当該入力手画像の手首領域を削除した後前記手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第 2 の手画像正規化手段と、

前記入力手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第 2 の固有空間投影手段と、

前記第 2 の固有空間投影手段が求めた前記投影座標と、前記クラスタ情報記憶手段に記憶した前記統計情報とをそれぞれ比較し、最も近いクラスタを求める最尤クラスタ判別手段と、

前記最も近いクラスタに属する前記手形状画像と前記入力手形状画像とをそれぞれ比較し、当該入力手形状画像に最も近い前記手形状画像を求める画像比較手段と、

前記最も近い手形状画像の前記形状情報と前記姿勢情報とを、前記手形状画像情報記憶手段から取得して出力する形状・姿勢出力手段とを備える、手の形状と姿勢の認識装置。

【請求項 3】

画像比較手段は、

前記最尤クラスタ判別手段が求めたクラスタに属する前記手形状画像と、前記第 2 の手画像正規化手段が生成した前記入力手形状画像とを比較する際に、前記手形状画像を同一クラスタ内の同一形状ごとにグループ化する同一形状分類手段と、

前記グループを表現する統計量を求める形状グループ統計量計算手段と、

前記入力手形状画像と前記統計量との距離を計算し、最も近いグループに属する形状を出力する最尤形状判別手段とで構成されることを特徴とする、請求項 2 に記載の手の形状と姿勢の認識装置。

【請求項 4】

前記クラスタ分析手段は、各クラスタに対する前記手形状画像と前記形状情報とを前記手形状画像情報記憶手段から取得し、各前記手形状画像を判別するための部分領域をそれぞれ計算して前記クラスタ情報記憶手段に格納し、

前記画像比較手段は、前記最尤クラスタ判別手段で求めたクラスタに属する前記手形状画像と、前記第 2 の手画像正規化手段が生成した入力手形状画像とを比較する際に、前記ク

ラストに対応する前記部分領域内のみで比較することを特徴とする、請求項 2 に記載の手の形状と姿勢の認識装置。

【請求項 5】

前記入力手画像が、認識対象となる手を複数の視点から撮影した複数の画像である場合、

前記第 2 の手画像正規化手段は、複数の前記入力手画像のそれぞれについて前記入力手形状画像を生成し、

前記第 2 の固有空間投影手段は、前記第 2 の手画像正規化手段が生成した複数の前記入力手形状画像について固有空間内での投影座標をそれぞれ求め、

前記最尤クラスタ判別手段は、前記第 2 の固有空間投影手段が求めた各前記投影座標と前記統計情報とを比較して、最も近いクラスタをそれぞれ求め、

前記画像比較手段は、前記最尤クラスタ判別手段が求めた複数の前記最も近いクラスタを統合し、各クラスタに属する前記手形状画像の前記形状情報および前記姿勢情報から矛盾しない形状・姿勢を推定することを特徴とする、請求項 2 に記載の手の形状と姿勢の認識装置。

【請求項 6】

光学的読取手段によって取得された一連の意味ある動作を行う連続した手の画像（以下、手振り動作画像という）の意味を認識する装置であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向、画像の大きさ、画像の明るさ）に正規化した手形状画像を、それぞれ生成する第 1 の手画像正規化手段と、

前記手形状画像を、当該手形状画像に関する形状情報および姿勢情報と共に、それぞれ格納する手形状画像情報記憶手段と、

固有空間法による解析を行って、前記手形状画像から固有値と固有ベクトルとを、それぞれ計算する固有空間計算手段と、

前記固有ベクトルの集合を格納する固有ベクトル記憶手段と、

前記手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求め、前記手形状画像情報記憶手段に格納する第 1 の固有空間投影手段と、

前記投影座標を、クラスタ分析によりそれぞれグループ分けし、各前記手形状画像がどのクラスタに属するかを求めて前記手形状画像情報記憶手段に格納すると共に、各クラスタに関する統計情報を求めるクラスタ分析手段と、

前記統計情報を、対応するクラスタと共に、それぞれ格納するクラスタ情報記憶手段と、前記手振り動作画像を入力し、当該手振り動作画像の各画像の中から手領域をそれぞれ検出する手領域検出手段と、

前記検出した手領域において前記手振り動作画像の手の動きをそれぞれ求め、手の動きに従い手動作の分節点を求める手動作分節手段と、

前記手振り動作画像中の手動作分節点である画像から、前記検出した手領域の部分を切り出す手画像切出手段と、

前記手画像切出手段において前記手振り動作画像から切り出された 1 つ以上の手画像（以下、手画像系列という）について、当該手画像の手首領域を削除した後前記手形状画像と同等の画像形態となるように正規化した入力手形状画像を、それぞれ生成する第 2 の手画像正規化手段と、

前記入力手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第 2 の固有空間投影手段と、

前記第 2 の固有空間投影手段が求めた前記投影座標と、前記クラスタ情報記憶手段に記憶した前記統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルをそれぞれ出力する最尤クラスタ判別手段と、

前記最尤クラスタ判別手段が出力する前記手画像系列に対応したシンボル（以下、シンボル系列という）を、当該手画像系列の元となる前記手振り動作画像の意味と共に系列識別

10

20

30

40

50

辞書手段に登録する系列登録手段と、

前記手振り動作画像の意味と対応する前記シンボル系列とを記憶する前記系列識別辞書手段と、

前記最尤クラスタ判別手段が出力する前記シンボル系列に該当する意味を、前記系列識別辞書手段から取得して出力する識別演算手段とを備える、手の形状と姿勢の認識装置。

【請求項 7】

前記手振り動作画像を入力し、当該画像の動作主体の動きと位置とから意味の候補を出力する大局動作認識手段と、

一連の意味ある動作に基づいて、入力された前記手振り動作画像の意味を制約する制約条件を予め記憶する制約条件記憶手段とをさらに備え、

前記識別演算手段は、前記最尤クラスタ判別手段が出力する前記シンボル系列に該当する意味を、前記制約条件に従いつつ前記系列識別辞書手段から取得して出力することを特徴とする、請求項 6 に記載の手の形状と姿勢の認識装置。

【請求項 8】

前記手領域検出手段は、

入力する前記手振り動作画像の各画像の中から、手領域の候補となる領域をそれぞれ切り出す領域候補切出手段と、

手画像の候補となる領域だけを矩形領域から抜き出すための領域マスクを記憶する領域マスク記憶手段と、

前記手振り動作画像から切り出された手領域候補の領域に、前記領域マスクを付加し、さらに前記固有ベクトルを計算する時に用いた手画像と同等の画像形態となるように正規化した画像をそれぞれ生成する手領域画像正規化手段と、

前記手領域候補の領域を正規化した画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める手領域固有空間投影手段と、

前記手領域固有空間投影手段が求めた前記投影座標と、前記クラスタ情報記憶手段に記憶した前記統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルと比較対象のクラスタとの近さを示す評価値とをそれぞれ出力する手領域最尤クラスタ判別手段と、

前記評価値に基づいて、最も高い前記評価値を持つ前記手領域候補の位置情報とそのクラスタとを出力する領域決定手段とで構成されることを特徴とする、請求項 6 または 7 に記載の手の形状と姿勢の認識装置。

【請求項 9】

前記第 1 の手画像正規化手段および前記第 2 の手画像正規化手段は、それぞれ、

入力する手の画像から抽出すべき前記手領域を、色分布として予め記憶する色分布記憶手段と、

前記色分布に従い、入力する手の画像から手領域を抽出する手領域抽出手段と、

手首方向を求め、当該手首方向に従い前記手領域から手首領域を削除する手首領域削除手段と、

前記手首領域を削除した前記手領域を、画像上の予め定義した位置に移動させる領域移動手段と、

前記手領域内の手が、予め定めた一定方向に向くように回転角を求める回転角計算手段と、

前記回転角に従い、手が一定方向に向くように前記手領域を回転させる領域回転手段と、回転させた前記手領域の大きさを、予め定めた一定の大きさに正規化する大きさ正規化手段とで構成されることを特徴とする、請求項 1, 2 または 6 に記載の手の形状と姿勢の認識装置。

【請求項 10】

前記形状情報および前記姿勢情報に対応する命令をそれぞれ格納する命令記憶手段と、前記形状・姿勢出力手段が出力する前記形状情報および前記姿勢情報を入力し、当該形状情報および当該姿勢情報に対応する命令を前記命令記憶手段から取得して出力する命令出

10

20

30

40

50

力手段とをさらに備える、請求項 1 に記載の手の形状と姿勢の認識装置。

【請求項 1 1】

光学的読取手段によって取得された手の画像（以下、入力手画像という）の形状および姿勢を認識する方法であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向，画像の大きさ，画像の明るさ）に正規化した手形状画像を、それぞれ生成する第 1 の正規化ステップと、

固有空間法による解析を行って、前記手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

前記手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第 1 の投影ステップと、

前記入力手画像を入力し、当該入力手画像の手首領域を削除した後前記手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第 2 の正規化ステップと、

前記入力手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第 2 の投影ステップと、

前記手形状画像について求めた前記投影座標と、前記入力手形状画像について求めた前記投影座標とをそれぞれ比較し、前記入力手形状画像に最も近い前記手形状画像を求める比較ステップと、

前記最も近い手形状画像の前記形状情報と前記姿勢情報とを出力するステップとを備える、手の形状と姿勢の認識方法。

【請求項 1 2】

光学的読取手段によって取得された手の画像（以下、入力手画像という）の形状および姿勢を認識する方法であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向，画像の大きさ，画像の明るさ）に正規化した手形状画像を、それぞれ生成する第 1 の正規化ステップと、

固有空間法による解析を行って、前記手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

前記手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第 1 の投影ステップと、

前記投影座標を、クラスタ分析によりそれぞれグループ分けし、各前記手形状画像がどのクラスタに属するかと、各クラスタに関する統計情報とを求める分析ステップと、

前記入力手画像を入力し、当該入力手画像の手首領域を削除した後前記手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第 2 の正規化ステップと、

前記入力手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第 2 の投影ステップと、

前記入力手形状画像について求めた前記投影座標と、前記統計情報とをそれぞれ比較し、最も近いクラスタを求める判別ステップと、

前記最も近いクラスタに属する前記手形状画像と前記入力手形状画像とをそれぞれ比較し、当該入力手形状画像に最も近い前記手形状画像を求める比較ステップと、

前記最も近い手形状画像の前記形状情報と前記姿勢情報とを出力するステップとを備える、手の形状と姿勢の認識方法。

【請求項 1 3】

前記比較ステップは、

前記判別ステップが求めたクラスタに属する前記手形状画像と、前記第 2 の正規化ステップが生成した前記入力手形状画像とを比較する際に、前記手形状画像を同一クラスタ内の同一形状ごとにグループ化するステップと、

前記グループを表現する統計量を求めるステップと、

前記入力手形状画像と前記統計量との距離を計算し、最も近いグループに属する形状を出力するステップとを含むことを特徴とする、請求項 1 2 に記載の手の形状と姿勢の認識方法。

【請求項 1 4】

前記分析ステップは、各クラスタに対する前記手形状画像と前記形状情報とに基づいて、各前記手形状画像を判別するための部分領域をそれぞれ計算し、

前記比較ステップは、前記判別ステップで求めたクラスタに属する前記手形状画像と、前記第 2 の正規化ステップが生成した前記入力手形状画像とを比較する際に、前記クラスタに対応する前記部分領域内のみで比較することを特徴とする、請求項 1 2 に記載の手の形状と姿勢の認識方法。

10

【請求項 1 5】

前記入力手画像が、認識対象となる手を複数の視点から撮影した複数の画像である場合、

前記前記第 2 の正規化ステップは、複数の前記入力手画像のそれぞれについて前記入力手形状画像を生成し、

前記第 2 の投影ステップは、前記第 2 の正規化ステップが生成した複数の前記入力手形状画像について固有空間内での投影座標をそれぞれ求め、

前記判別ステップは、前記第 2 の投影ステップが求めた各前記投影座標と前記統計情報とを比較して、最も近いクラスタをそれぞれ求め、

前記比較ステップは、前記判別ステップが求めた複数の前記最も近いクラスタを統合し、各クラスタに属する前記手形状画像の前記形状情報および前記姿勢情報から矛盾しない形状・姿勢を推定することを特徴とする、請求項 1 2 に記載の手の形状と姿勢の認識方法。

20

【請求項 1 6】

光学的読取手段によって取得された一連の意味ある動作を行う連続した手の画像（以下、手振り動作画像という）の意味を認識する方法であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向、画像の大きさ、画像の明るさ）に正規化した手形状画像を、それぞれ生成する第 1 の正規化ステップと、

固有空間法による解析を行って、前記手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

30

前記手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第 1 の投影ステップと、

前記投影座標を、クラスタ分析によりそれぞれグループ分けし、各前記手形状画像がどのクラスタに属するかと、各クラスタに関する統計情報とを求める分析ステップと、

前記手振り動作画像を入力し、当該手振り動作画像の各画像の中から手領域をそれぞれ検出する検出ステップと、

前記検出した手領域において前記手振り動作画像の手の動きをそれぞれ求め、手の動きに従い手動作の分節点を求める分節ステップと、

前記手振り動作画像中の手動作分節点である画像から、前記検出した手領域の部分を切り出す切出ステップと、

40

前記手振り動作画像から切り出された 1 つ以上の手画像（以下、手画像系列という）について、当該手画像の手首領域を削除した後前記手形状画像と同等の画像形態となるように正規化した入力手形状画像を、それぞれ生成する第 2 の正規化ステップと、

前記入力手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第 2 の投影ステップと、

前記入力手形状画像について求めた前記投影座標と、前記統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルをそれぞれ出力する判別ステップと、

判別した前記手画像系列に対応したシンボル（以下、シンボル系列という）を、当該手画像系列の元となる前記手振り動作画像の意味と共に記憶するステップと、

50

入力する前記手振り動作画像を識別する場合、判別した前記シンボル系列に該当する意味を、記憶している前記シンボル系列とその意味とに基づいて出力する識別ステップとを備える、手の形状と姿勢の認識方法。

【請求項 17】

前記手振り動作画像を入力し、当該画像の動作主体の動きと位置とから意味の候補を出力する認識ステップと、
一連の意味ある動作に基づいて、入力された前記手振り動作画像の意味を制約する制約条件を予め記憶する記憶ステップとをさらに備え、
前記識別ステップは、判別した前記シンボル系列に該当する意味を、前記制約条件に従いつつ記憶している前記シンボル系列とその意味とに基づいて出力することを特徴とする、
請求項 16 に記載の手の形状と姿勢の認識方法。

10

【請求項 18】

前記検出ステップは、
入力する前記手振り動作画像の各画像の中から、手領域の候補となる領域をそれぞれ切り出す切出ステップと、
手画像の候補となる領域だけを矩形領域から抜き出すための領域マスクを記憶する記憶ステップと、
前記手振り動作画像から切り出された手領域候補の領域に、前記領域マスクを付加し、さらに前記固有ベクトルを計算する時に用いた手画像と同等の画像形態となるように正規化した画像をそれぞれ生成する正規化ステップと、
前記手領域候補の領域を正規化した画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める投影ステップと、
前記投影座標と前記統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルと比較対象のクラスタとの近さを示す評価値とをそれぞれ出力する判別ステップと、
前記評価値に基づいて、最も高い前記評価値を持つ前記手領域候補の位置情報とそのクラスタとを出力する決定ステップとを含むことを特徴とする、請求項 16 または 17 に記載の手の形状と姿勢の認識方法。

20

【請求項 19】

前記第 1 の正規化ステップおよび前記第 2 の正規化ステップは、それぞれ、
入力する手の画像から抽出すべき前記手領域を、色分布として予め記憶する色記憶ステップと、
前記色分布に従い、入力する手の画像から手領域を抽出するステップと、
手首方向を求め、当該手首方向に従い前記手領域から手首領域を削除するステップと、
前記手首領域を削除した前記手領域を、画像上の予め定義した位置に移動させるステップと、
前記手領域内の手が、予め定めた一定方向に向くように回転角を求めるステップと、
前記回転角に従い、手が一定方向に向くように前記手領域を回転させるステップと、
回転させた前記手領域の大きさを、予め定めた一定の大きさに正規化するステップとを含むことを特徴とする、請求項 11、12 または 16 に記載の手の形状と姿勢の認識方法。

30

40

【請求項 20】

前記形状情報および前記姿勢情報に対応する命令をそれぞれ格納する命令記憶ステップと、
前記出力するステップが出力する前記形状情報および前記姿勢情報を入力し、当該形状情報および当該姿勢情報に対応する命令を前記命令記憶ステップから取得して出力するステップとをさらに備える、請求項 11 に記載の手の形状と姿勢の認識方法。

【請求項 21】

光学的読取手段によって取得された手の画像（以下、入力手画像という）の形状および姿勢を認識する方法を、コンピュータ装置において実行するためのプログラムを記録した記録媒体であって、

50

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向，画像の大きさ，画像の明るさ）に正規化した手形状画像を、それぞれ生成する第１の正規化ステップと、固有空間法による解析を行って、前記手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

前記手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第１の投影ステップと、

前記入力手画像を入力し、当該入力手画像の手首領域を削除した後前記手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第２の正規化ステップと、

10

前記入力手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第２の投影ステップと、

前記手形状画像について求めた前記投影座標と、前記入力手形状画像について求めた前記投影座標とをそれぞれ比較し、前記入力手形状画像に最も近い前記手形状画像を求める比較ステップと、

前記最も近い手形状画像の前記形状情報と前記姿勢情報とを出力するステップとを含む動作環境を、前記コンピュータ装置上で実現するためのプログラムを記録した、記録媒体。

【請求項２２】

光学的読取手段によって取得された手の画像（以下、入力手画像という）の形状および姿勢を認識する方法を、コンピュータ装置において実行するためのプログラムを記録した記録媒体であって、

20

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向，画像の大きさ，画像の明るさ）に正規化した手形状画像を、それぞれ生成する第１の正規化ステップと、

固有空間法による解析を行って、前記手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

前記手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第１の投影ステップと、

前記投影座標を、クラスタ分析によりそれぞれグループ分けし、各前記手形状画像がどのクラスタに属するかと、各クラスタに関する統計情報とを求める分析ステップと、

30

前記入力手画像を入力し、当該入力手画像の手首領域を削除した後前記手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第２の正規化ステップと、

前記入力手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第２の投影ステップと、

前記入力手形状画像について求めた前記投影座標と、前記統計情報とをそれぞれ比較し、最も近いクラスタを求める判別ステップと、

前記最も近いクラスタに属する前記手形状画像と前記入力手形状画像とをそれぞれ比較し、当該入力手形状画像に最も近い前記手形状画像を求める比較ステップと、

40

前記最も近い手形状画像の前記形状情報と前記姿勢情報とを出力するステップとを含む動作環境を、前記コンピュータ装置上で実現するためのプログラムを記録した、記録媒体。

【請求項２３】

前記比較ステップは、

前記判別ステップが求めたクラスタに属する前記手形状画像と、前記第２の正規化ステップが生成した前記入力手形状画像とを比較する際に、前記手形状画像を同一クラスタ内の同一形状ごとにグループ化するステップと、

前記グループを表現する統計量を求めるステップと、

前記入力手形状画像と前記統計量との距離を計算し、最も近いグループに属する形状を出力するステップとを含むことを特徴とする、請求項２２に記載の記録媒体。

【請求項２４】

50

前記分析ステップは、各クラスタに対する前記手形状画像と前記形状情報とに基づいて、各前記手形状画像を判別するための部分領域をそれぞれ計算し、

前記比較ステップは、前記判別ステップで求めたクラスタに属する前記手形状画像と、前記第2の正規化ステップが生成した前記入力手形状画像とを比較する際に、前記クラスタに対応する前記部分領域内のみで比較することを特徴とする、請求項22に記載の記録媒体。

【請求項25】

前記入力手画像が、認識対象となる手を複数の視点から撮影した複数の画像である場合、

前記前記第2の正規化ステップは、複数の前記入力手画像のそれぞれについて前記入力手形状画像を生成し、

前記第2の投影ステップは、前記第2の正規化ステップが生成した複数の前記入力手形状画像について固有空間内での投影座標をそれぞれ求め、

前記判別ステップは、前記第2の投影ステップが求めた各前記投影座標と前記統計情報とを比較して、最も近いクラスタをそれぞれ求め、

前記比較ステップは、前記判別ステップが求めた複数の前記最も近いクラスタを統合し、各クラスタに属する前記手形状画像の前記形状情報および前記姿勢情報から矛盾しない形状・姿勢を推定することを特徴とする、請求項22に記載の記録媒体。

【請求項26】

光学的読取手段によって取得された一連の意味ある動作を行う連続した手の画像（以下、手振り動作画像という）の意味を認識する方法を、コンピュータ装置において実行するためのプログラムを記録した記録媒体であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向、画像の大きさ、画像の明るさ）に正規化した手形状画像を、それぞれ生成する第1の正規化ステップと、

固有空間法による解析を行って、前記手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

前記手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第1の投影ステップと、

前記投影座標を、クラスタ分析によりそれぞれグループ分けし、各前記手形状画像がどのクラスタに属するかと、各クラスタに関する統計情報とを求める分析ステップと、

前記手振り動作画像を入力し、当該手振り動作画像の各画像の中から手領域をそれぞれ検出する検出ステップと、

前記検出した手領域において前記手振り動作画像の手の動きをそれぞれ求め、手の動きに従い手動作の分節点を求める分節ステップと、

前記手振り動作画像中の手動作分節点である画像から、前記検出した手領域の部分を切り出す切出ステップと、

前記手振り動作画像から切り出された1つ以上の手画像（以下、手画像系列という）について、当該手画像の手首領域を削除した後前記手形状画像と同等の画像形態となるように正規化した入力手形状画像を、それぞれ生成する第2の正規化ステップと、

前記入力手形状画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第2の投影ステップと、

前記入力手形状画像について求めた前記投影座標と、前記統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルをそれぞれ出力する判別ステップと、

判別した前記手画像系列に対応したシンボル（以下、シンボル系列という）を、当該手画像系列の元となる前記手振り動作画像の意味と共に記憶するステップと、

入力する前記手振り動作画像を識別する場合、判別した前記シンボル系列に該当する意味を、記憶している前記シンボル系列とその意味とに基づいて出力する識別ステップとを含む動作環境を、前記コンピュータ装置上で実現するためのプログラムを記録した、記録媒

10

20

30

40

50

体。

【請求項 27】

前記手振り動作画像を入力し、当該画像の動作主体の動きと位置とから意味の候補を出力する認識ステップと、
一連の意味ある動作に基づいて、入力された前記手振り動作画像の意味を制約する制約条件を予め記憶する記憶ステップとをさらに含み、
前記識別ステップは、判別した前記シンボル系列に該当する意味を、前記制約条件に従いつつ記憶している前記シンボル系列とその意味とに基づいて出力することを特徴とする、請求項 26 に記載の記録媒体。

【請求項 28】

前記検出ステップは、
入力する前記手振り動作画像の各画像の中から、手領域の候補となる領域をそれぞれ切り出す切出ステップと、
手画像の候補となる領域だけを矩形領域から抜き出すための領域マスクを記憶する記憶ステップと、
前記手振り動作画像から切り出された手領域候補の領域に、前記領域マスクを付加し、さらに前記固有ベクトルを計算する時に用いた手画像と同等の画像形態となるように正規化した画像をそれぞれ生成する正規化ステップと、
前記手領域候補の領域を正規化した画像を前記固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める投影ステップと、
前記投影座標と前記統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルと比較対象のクラスタとの近さを示す評価値とをそれぞれ出力する判別ステップと、
前記評価値に基づいて、最も高い前記評価値を持つ前記手領域候補の位置情報とそのクラスタとを出力する決定ステップとを含むことを特徴とする、請求項 26 または 27 に記載の記録媒体。

【請求項 29】

前記第 1 の正規化ステップおよび前記第 2 の正規化ステップは、それぞれ、
入力する手の画像から抽出すべき前記手領域を、色分布として予め記憶する色記憶ステップと、
前記色分布に従い、入力する手の画像から手領域を抽出するステップと、
手首方向を求め、当該手首方向に従い前記手領域から手首領域を削除するステップと、
前記手首領域を削除した前記手領域を、画像上の予め定義した位置に移動させるステップと、
前記手領域内の手が、予め定めた一定方向に向くように回転角を求めるステップと、
前記回転角に従い、手が一定方向に向くように前記手領域を回転させるステップと、
回転させた前記手領域の大きさを、予め定めた一定の大きさに正規化するステップとを含むことを特徴とする、請求項 21、22 または 26 に記載の記録媒体。

【請求項 30】

前記形状情報および前記姿勢情報に対応する命令をそれぞれ格納する命令記憶ステップと、
前記出力するステップが出力する前記形状情報および前記姿勢情報を入力し、当該形状情報および当該姿勢情報に対応する命令を前記命令記憶ステップから取得して出力するステップとをさらに含む、請求項 21 に記載の記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、手の形状と姿勢の認識装置および手の形状と姿勢の認識方法並びに当該方法を実行するプログラムを記録した記録媒体に関し、より特定的には、データグローブなどのケーブル等をつけた手袋を用いることなく人間の手の形状と姿勢推定の認識を行い、当該

10

20

30

40

50

認識結果に基づいて人間と機器との間のインタフェース装置や手話認識装置等に利用可能な手の形状と姿勢の認識装置およびその方法並びに当該方法を実行するプログラムを記録した記録媒体に関する。

【0002】

【従来の技術】

現在、新たなヒューマンインタフェースの手法として、人間の手の形状を認識し、手を用いて人間が提示する情報を取得する装置の研究開発が盛んに行われている。また、聴覚障害者と健常者との間のコミュニケーション支援を目指して、手話動作中で提示される手の形状と姿勢を認識する研究も盛んに行われている。

【0003】

一般に、人間の手の形を取得する方法として、手にデータグローブ等のセンサを付けて各状態を測定する手法、例えば、電気学会計測研究会資料(1994 第49項～第56項)(以下、第1の文献という)が有名である。この第1の文献には、指に沿ってグローブ光ファイバを設置し、ファイバが伸びているときと曲がっているときの綱領の変化を利用して、指関節の角度を推定する手法が述べられている。

【0004】

また、上記第1の文献に述べられているようにセンサ付き手袋を用いない手法としては、カメラを用いて手の形状を認識する方法、例えば、渡辺，岩井，八木，谷内田による論文「カラーグローブを用いた指文字の認識」(電子情報通信学会誌 Vol. J80-D-2, No. 10, 第2713頁～第2722頁)(以下、第2の文献という)が存在する。この第2の文献には、複数の色を付けた手袋(マーカ)を手に着用して、この手袋の画像から手の形状を認識する手法が述べられている。

【0005】

一方、手に手袋等のマーカを着用しないでカメラによって手の形状と姿勢を認識する手法としては、例えば、特開平8-263629号公報「物体の形状・姿勢検出装置」(以下、第3の文献という)に開示されているものが存在する。この第3の文献には、少なくとも3台のカメラを用いて手を撮影し、手を平面とみなして、その手がどのカメラに対して向いているかを判別し、正面に向いているカメラの画像から形状を認識し、姿勢を推定する手法が記載されている。

【0006】

さらに、正面に向いているカメラの画像から形状を認識する手法として、石淵，岩崎，竹村，岸野による論文「画像処理を用いた実時間手振り推定とヒューマンインタフェースへの応用」(電子情報通信学会論文誌 Vol. J79-D-2, No. 7, 第1218項～第1229項)(以下、第4の文献という)が存在する。この第4の文献には、複数のカメラから得られる手の画像から手首-中指方向(以下、掌主軸と呼ぶ)を求め、同時に伸びている指先の位置を求めて伸展指が何本あるかを認識する手法が述べられている。

【0007】

さて、一般に顔や車などの物体の姿勢と種類を認識するために、近年、見かけの画像を用いた手法と固有空間法とを組み合わせた画像認識法が注目されている。ここで、見かけの画像に基づく手法とは、予め取得した三次元物体の二次元の見かけ画像のみを用いて、物体の姿勢や種類を認識するものである。一方、固有空間法とは、古くから行われている手法で、画像集合の共分散行列(または、自己相関行列)の固有ベクトルから構成される固有空間を用いるものであり、主成分分析若しくはKL展開による手法を用いるものが有名である。

以下、上記主成分分析を画像について適用する手法を、簡単に述べる。

【0008】

主成分分析は、多次元空間上の特徴点を、より見やすく、あるいは扱いやすくするために、固有空間を利用して少ない次元で表現しようとする統計手法で、多変量解析の一手法としてよく使われている。原理は、多次元空間上の特徴点を分散の大きい少数の低次元の直交部分空間に線形射影するというものである。

10

20

30

40

50

この主成分分析手法を画像に適用する場合、まず、 $n \times m$ 画素の画像をラスタスキャンして得た列ベクトルを U とし、 p 個の画像が属する画像群を、

$$\{U_1, U_2, U_3, \dots, U_p\}$$

で表現する。

次に、この画像群の各要素から画像集合の平均画像 c の要素を引いた列ベクトルからなる $n \times m \times p$ 行列を A とすると、行列 A

$$A = [U_1 - c, U_2 - c, U_3 - c, \dots, U_p - c]$$

によって、画像集合の共分散行列 Q は、下記式(1)で計算される。なお、行列 A^T は、行列 A の転置行列を表す。

$$Q = A A^T \quad \dots (1)$$

そして、この共分散行列 Q を用いて、下記式(2)の固有方程式を解く。

$$Q e_i = \lambda_i e_i \quad \dots (2)$$

【0009】

ここで、求めるべき部分空間は、その次元を k とすれば、 k 個の大きな固有値に対応する固有ベクトル

$$e_1, e_2, \dots, e_k \quad (1 \quad 2 \quad \dots \quad k \quad \dots \quad p)$$

を基底ベクトルとすることにより得られる部分空間となる。

従って、ある画像 x を、下記式(3)に従って固有ベクトルによる部分空間に線形射影することにより、 $n \times m$ 次元の画像を k 次元特徴ベクトル y に次元圧縮することができる。

$$y = [e_1, e_2, \dots, e_k]^T x \quad \dots (3)$$

【0010】

一方、主成分分析若しくはKL展開を用いて、人間の顔の様な複雑で多種多様な実体を検出認識する手法としては、例えば、特開平8-339445号公報「確率的固有空間解析を用いた複雑な対象物の検出、認識、および符号化方法および装置」(以下、第5の文献という)に開示されているものが存在する。この第5の文献では、従来から行われている上記手法を複雑な対象物、特に顔に対して適用していることに特徴がある。上記第5の文献には、複雑な対象物の例として手の形状の認識に適用した実施例が示されており、以下にその手法を説明する。

【0011】

まず、手振り若しくはジェスチャーで用いられる手の画像集合を、黒の背景に対して撮影する。次に、手の二次元輪郭を、Cannyのエッジ・オペレータを用いて抽出する。そして、得られたエッジ画像集合に対してKL展開を行って部分空間を求めるわけだが、2値のエッジマップをそのまま使用すると、画像間で互いにほとんど相関がなく、それにより非常に部分空間の次元 k をかなり大きくしなければならなかった。そこで、上記第5の文献に記載されている実施例では、2値のエッジマップ上で、拡散処理を介してエッジをぼかして部分空間を求めることにより、部分空間の次元を小さくすることを提案している。また、入力画像から手の位置を求めるために、全画像を、ある一定の大きさで探索することにより手の位置を求め、認識を行っている。

【0012】

【発明が解決しようとする課題】

しかしながら、人の手の形状を認識しようとするとき、上記第1の文献に述べられているような、手にデータグローブを装着して手形状認識を行う場合、手にコードをつけるため動作が制約されたり、装着感に課題が残る。

【0013】

また、カメラを用いて手の形状を認識する場合でも、上記第2の文献に述べられているような、手袋などのマーカを装着することを前提とする場合、手袋がないときには手形状認識ができない上、装着時の違和感が問題となっている。

さらに、上記第3の文献に記載されているような、手袋やマーカを装着しないで複数カメラにより手の形状と姿勢を認識する場合、手を平面とみなして、その手がどのカメラに対して向いているかにより姿勢を判別するが、実際に手は様々な形状を表現することが可能

10

20

30

40

50

であり、その中には平面に近似できない形状も多数存在する。従って、認識対象として伸展指の本数などの単純な形状には対処できるが、より複雑な形状（例えば、親指と人差指を接触させて穴をつくった形状）などには適用不可能である。

【0014】

一方、上記第4の文献に述べられているような、より一般的な固有空間解析に基づいた手法では、手のみの正規化された画像をどのようにして取得するのかが明らかでない。固有空間解析に基づいた手法では、認識する対象物体の画像領域をいかに切り出して正規化をするかということが重要である。一般に、剛体で認識する対象物体が明らかに違う場合、この正規化は大きさや明るさの正規化だけで十分であるが、手や顔といった複雑な物体の場合、その部分を切り出す処理が重要である。

10

例えば、顔の認識にこの方法を用いる場合は、目、鼻の位置をある一定の位置に移動させ、顎（あご）や髪の毛を削除する手法がよく使われている。一方、手の場合、手首の領域を何らかの手法で削除し、手を一定の位置に移動させ正規化を行う必要があり、その処理をしないで複数の手の形状と姿勢の認識に固有空間解析に基づく手法を用いても認識率が悪いという課題がある。

【0015】

また、上記第5の文献に述べられているような、実際に手の画像に固有空間解析を適用する場合も、手の画像のエッジによる輪郭を求め、さらにエッジをぼかす必要があった。そのため、指1本が伸展している画像と指2本を接触させた形で伸展させた画像では、その違いが画像上区別できなくなり、結果として、より複雑な形状などには適用不可能であった。

20

【0016】

それ故、本発明の目的は、予め認識すべき手形状に対して、様々な姿勢を示す画像から手首領域を削除して正規化し、正規化した画像に対して固有空間解析に基づく方法を適用することにより、より複雑な手形状の画像に対してもその形状と姿勢を認識する手の形状と姿勢の認識装置および手の形状と姿勢の認識方法並びに当該方法を実行するプログラムを記録した記録媒体を提供することである。

【0017】

【課題を解決するための手段および発明の効果】

第1の発明は、光学的読取手段によって取得された手の画像（以下、入力手画像という）の形状および姿勢を認識する装置であって、

30

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態（手の方向、画像の大きさ、画像の明るさ）に正規化した手形状画像を、それぞれ生成する第1の手画像正規化手段と、

手形状画像を、当該手形状画像に関する形状情報および姿勢情報と共に、それぞれ格納する手形状画像情報記憶手段と、

固有空間法による解析を行って、手形状画像から固有値と固有ベクトルとを、それぞれ計算する固有空間計算手段と、

固有ベクトルの集合を格納する固有ベクトル記憶手段と、

手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求め、手形状画像情報記憶手段に格納する第1の固有空間投影手段と、

40

入力手画像を入力し、当該入力手画像の手首領域を削除した後手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第2の手画像正規化手段と、

入力手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第2の固有空間投影手段と、

第2の固有空間投影手段が求めた投影座標と、手形状画像情報記憶手段に格納した投影座標とをそれぞれ比較し、入力手形状画像に最も近い手形状画像を求める手形状画像選択手段と、

最も近い手形状画像の形状情報と姿勢情報とを、手形状画像情報記憶手段から取得して出力する形状・姿勢出力手段とを備える。

50

【 0 0 1 8 】

上記のように、第 1 の発明によれば、様々な手形状と手姿勢とを持つ複数の手の画像と認識対象である入力手画像の両方から手首領域を削除するため、単純に大きさと明るさを正規化するよりも精密に手画像の正規化を行うことができる。これにより、固有空間に基づく手法を手の形状および姿勢の認識に用いても、十分に精度の高い結果を得ることが可能になる。

また、固有空間に基づく手法を手の形状および姿勢の認識に用いることにより、伸展指の個数を数えるなどの幾何的な特徴による手法により、幾何的な特徴がとれにくい、より複雑な手の形状に対しても認識が可能である。

【 0 0 1 9 】

第 2 の発明は、光学的読取手段によって取得された入力手画像の形状および姿勢を認識する装置であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態に正規化した手形状画像を、それぞれ生成する第 1 の手画像正規化手段と、

手形状画像を、当該手形状画像に関する形状情報および姿勢情報と共に、それぞれ格納する手形状画像情報記憶手段と、

固有空間法による解析を行って、手形状画像から固有値と固有ベクトルとを、それぞれ計算する固有空間計算手段と、

固有ベクトルの集合を格納する固有ベクトル記憶手段と、

手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求め、手形状画像情報記憶手段に格納する第 1 の固有空間投影手段と、

投影座標を、クラスタ分析によりそれぞれグループ分けし、各手形状画像がどのクラスタに属するかを求めて手形状画像情報記憶手段に格納すると共に、各クラスタに関する統計情報を求めるクラスタ分析手段と、

統計情報を、対応するクラスタと共に、それぞれ格納するクラスタ情報記憶手段と、

入力手画像を入力し、当該入力手画像の手首領域を削除した後手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第 2 の手画像正規化手段と、

入力手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第 2 の固有空間投影手段と、

第 2 の固有空間投影手段が求めた投影座標と、クラスタ情報記憶手段に記憶した統計情報とをそれぞれ比較し、最も近いクラスタを求める最尤クラスタ判別手段と、

最も近いクラスタに属する手形状画像と入力手形状画像とをそれぞれ比較し、当該入力手形状画像に最も近い手形状画像を求める画像比較手段と、

最も近い手形状画像の形状情報と姿勢情報とを、手形状画像情報記憶手段から取得して出力する形状・姿勢出力手段とを備える。

【 0 0 2 0 】

上記のように、第 2 の発明によれば、手形状画像情報記憶手段に記憶した複数の手形状画像を固有空間内におけるクラスタ分析によりグループ化し、入力手画像を認識するときに、最初にどのグループに属するかを求め、そのグループ内でどの手形状画像に最も近いかを求める。これにより、画像比較回数を減らすことができ、さらに高速に処理を行うことが可能となる。また、違う形状でも似た画像が存在する場合でも正確に手の形状および姿勢を求めることが可能になる。

【 0 0 2 1 】

第 3 の発明は、第 2 の発明に従属する発明であって、

画像比較手段は、

最尤クラスタ判別手段が求めたクラスタに属する手形状画像と、第 2 の手画像正規化手段が生成した入力手形状画像とを比較する際に、手形状画像を同一クラスタ内の同一形状ごとにグループ化する同一形状分類手段と、

グループを表現する統計量を求める形状グループ統計量計算手段と、

入力手形状画像と統計量との距離を計算し、最も近いグループに属する形状を出力する最尤形状判別手段とで構成されることを特徴とする。

【0022】

上記のように、第3の発明によれば、第2の発明において、姿勢まで出力する必要がない場合に、姿勢と形状の両方を認識する場合よりも正確に手の形状を求めることが可能になる。

【0023】

第4の発明は、第2の発明に従属する発明であって、
クラスタ分析手段は、各クラスタに対する手形状画像と形状情報とを手形状画像情報記憶手段から取得し、各手形状画像を判別するための部分領域をそれぞれ計算してクラスタ情報記憶手段に格納し、
画像比較手段は、最尤クラスタ判別手段で求めたクラスタに属する手形状画像と、第2の手画像正規化手段が生成した入力手形状画像とを比較する際に、クラスタに対応する部分領域内のみで比較することを特徴とする。

【0024】

上記のように、第4の発明によれば、第2の発明において、画像を判別するための部分領域を予め定めておき、この部分領域内のみで手形状画像と入力手形状画像とを比較する。これにより、第4の発明に比べ、画像比較回数を減らすことができると共に、認識対象となる手の画像を認識するときに、違う形状でも似た画像が存在する場合でも正確かつ高速に手の形状および姿勢を求めることが可能となる。

【0025】

第5の発明は、第2の発明に従属する発明であって、
入力手画像が、認識対象となる手を複数の視点から撮影した複数の画像である場合、
第2の手画像正規化手段は、複数の入力手画像のそれぞれについて入力手形状画像を生成し、
第2の固有空間投影手段は、第2の手画像正規化手段が生成した複数の入力手形状画像について固有空間内での投影座標をそれぞれ求め、
最尤クラスタ判別手段は、第2の固有空間投影手段が求めた各投影座標と統計情報とを比較して、最も近いクラスタをそれぞれ求め、
画像比較手段は、最尤クラスタ判別手段が求めた複数の最も近いクラスタを統合し、各クラスタに属する手形状画像の形状情報および姿勢情報から矛盾しない形状・姿勢を推定することを特徴とする。

【0026】

上記のように、第5の発明によれば、第2の発明において、複数の視点から撮影した入力手画像から求めたクラスタに基づき、最も近いクラスタを統合して、入力手画像の形状および姿勢を求める。これにより、1方向からの画像だけでは形状および姿勢が決定できないような場合（例えば、横方向の手の画像など）でも、正確に手の形状および姿勢を求めることが可能となる。

【0027】

第6の発明は、光学的読取手段によって取得された一連の意味ある動作を行う連続した手の画像（以下、手振り動作画像という）の意味を認識する装置であって、
予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態に正規化した手形状画像を、それぞれ生成する第1の手画像正規化手段と、
手形状画像を、当該手形状画像に関する形状情報および姿勢情報と共に、それぞれ格納する手形状画像情報記憶手段と、
固有空間法による解析を行って、手形状画像から固有値と固有ベクトルとを、それぞれ計算する固有空間計算手段と、
固有ベクトルの集合を格納する固有ベクトル記憶手段と、
手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座

標をそれぞれ求め、手形状画像情報記憶手段に格納する第1の固有空間投影手段と、
投影座標を、クラスタ分析によりそれぞれグループ分けし、各手形状画像がどのクラスタ
に属するかを求めて手形状画像情報記憶手段に格納すると共に、各クラスタに関する統計
情報を求めるクラスタ分析手段と、

統計情報を、対応するクラスタと共に、それぞれ格納するクラスタ情報記憶手段と、
手振り動作画像を入力し、当該手振り動作画像の各画像の中から手領域をそれぞれ検出する
手領域検出手段と、

検出した手領域において手振り動作画像の手の動きをそれぞれ求め、手の動きに従い手動作
の分節点を求める手動作分節手段と、

手振り動作画像中の手動作分節点である画像から、検出した手領域の部分を切り出す手画
像切出手段と、

手画像切出手段において手振り動作画像から切り出された1つ以上の手画像（以下、手画
像系列という）について、当該手画像の手首領域を削除した後手形状画像と同等の画像形
態となるように正規化した入力手形状画像を、それぞれ生成する第2の手画像正規化手段
と、

入力手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投
影座標をそれぞれ求める第2の固有空間投影手段と、

第2の固有空間投影手段が求めた投影座標と、クラスタ情報記憶手段に記憶した統計情報
とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルをそれ
ぞれ出力する最尤クラスタ判別手段と、

最尤クラスタ判別手段が出力する手画像系列に対応したシンボル（以下、シンボル系列と
いう）を、当該手画像系列の元となる手振り動作画像の意味と共に系列識別辞書手段に登
録する系列登録手段と、

手振り動作画像の意味と対応するシンボル系列とを記憶する系列識別辞書手段と、
最尤クラスタ判別手段が出力するシンボル系列に該当する意味を、系列識別辞書手段から
取得して出力する識別演算手段とを備える。

【0028】

上記のように、第6の発明によれば、ジェスチャー単語や手話単語等のような一連の意味
ある手振り動作画像に対して、手動作の分節点となる画像のクラスタ系列をその意味と共
に予め記憶し、手振り動作画像を認識するときに、求めたクラスタ系列に基づいて記憶し
ている意味を出力する。これにより、ジェスチャー単語や手話単語等のような一連の意味
ある動作に対して、より誤認識を減らし、正確に意味を求めることが可能となる。

【0029】

第7の発明は、第6の発明に従属する発明であって、
手振り動作画像を入力し、当該画像の動作主体の動きと位置とから意味の候補を出力する
大局動作認識手段と、

一連の意味ある動作に基づいて、入力された手振り動作画像の意味を制約する制約条件を
予め記憶する制約条件記憶手段とをさらに備え、

識別演算手段は、最尤クラスタ判別手段が出力するシンボル系列に該当する意味を、制約
条件に従いつつ系列識別辞書手段から取得して出力することを特徴とする。

【0030】

上記のように、第7の発明によれば、第6の発明において、手の大局的動作の特徴に基づ
く制約条件をさらに加えて、手振り動作画像の意味を導き出す。これにより、手振り動作
画像の誤認識を減らすことができる。

【0031】

第8の発明は、第6、第7の発明に従属する発明であって、

手領域検出手段は、

入力する手振り動作画像の各画像の中から、手領域の候補となる領域をそれぞれ切り出す
領域候補切出手段と、

手画像の候補となる領域だけを矩形領域から抜き出すための領域マスクを記憶する領域マ

10

20

30

40

50

スク記憶手段と、

手振り動作画像から切り出された手領域候補の領域に、領域マスクを付加し、さらに固有ベクトルを計算する時に用いた手画像と同等の画像形態となるように正規化した画像をそれぞれ生成する手領域画像正規化手段と、

手領域候補の領域を正規化した画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める手領域固有空間投影手段と、

手領域固有空間投影手段が求めた投影座標と、クラスタ情報記憶手段に記憶した統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルと比較対象のクラスタとの近さを示す評価値とをそれぞれ出力する手領域最尤クラスタ判別手段と、

10

評価値に基づいて、最も高い評価値を持つ手領域候補の位置情報とそのクラスタとを出力する領域決定手段とで構成されることを特徴とする。

【0032】

上記のように、第8の発明によれば、第6、第7の発明において、手領域を検出するときに手領域候補の領域を固有空間へ投影して、該当クラスタを求めることによって手領域を検出する。このため、手領域を検出すると同時にその手領域の該当クラスタを求めることができるので、手領域検出と手の形状と姿勢の認識、若しくは手領域検出と手振り動作認識処理を一つの処理として統合することが可能となる。

【0033】

第9の発明は、第1、第2、第6の発明に従属する発明であって、

20

第1の手画像正規化手段および第2の手画像正規化手段は、それぞれ、

入力する手の画像から抽出すべき手領域を、色分布として予め記憶する色分布記憶手段と、

色分布に従い、入力する手の画像から手領域を抽出する手領域抽出手段と、

手首方向を求め、当該手首方向に従い手領域から手首領域を削除する手首領域削除手段と、

手首領域を削除した手領域を、画像上の予め定義した位置に移動させる領域移動手段と、

手領域内の手が、予め定めた一定方向に向くように回転角を求める回転角計算手段と、

回転角に従い、手が一定方向に向くように手領域を回転させる領域回転手段と、

回転させた手領域の大きさを、予め定めた一定の大きさに正規化する大きさ正規化手段とで構成されることを特徴とする。

30

【0034】

上記のように、第9の発明によれば、第1、第2、第6の発明において、手画像を正規化する際に、手首領域を削除するだけではなく、肌色により手領域を抽出する。これにより、一般的で自然な背景で撮影した手の画像から手領域を抽出することができ、さらに、正確に手の形状および姿勢を認識することが可能となる。

【0035】

第10の発明は、第1の発明に従属する発明であって、

形状情報および姿勢情報に対応する命令をそれぞれ格納する命令記憶手段と、

形状・姿勢出力手段が出力する形状情報および姿勢情報を入力し、当該形状情報および当該姿勢情報に対応する命令を命令記憶手段から取得して出力する命令出力手段とをさらに備える。

40

【0036】

上記のように、第10の発明によれば、第1の発明の認識装置を、求めた手の形状および姿勢に従い、他の機器の制御を行うインタフェースとして機能させることができる。

【0037】

第11の発明は、光学的読取手段によって取得された入力手画像の形状および姿勢を認識する方法であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態に正規化した手形状画像を、それぞれ生成する第1の正規化ステ

50

ップと、

固有空間法による解析を行って、手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第1の投影ステップと、

入力手画像を入力し、当該入力手画像の手首領域を削除した後手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第2の正規化ステップと、

入力手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第2の投影ステップと、

手形状画像について求めた投影座標と、入力手形状画像について求めた投影座標とをそれぞれ比較し、入力手形状画像に最も近い手形状画像を求める比較ステップと、

最も近い手形状画像の形状情報と姿勢情報とを出力するステップとを備える。

10

【0038】

上記のように、第11の発明によれば、様々な手形状と手姿勢とを持つ複数の手の画像と認識対象である入力手画像の両方から手首領域を削除するため、単純に大きさと明るさを正規化するよりも精密に手画像の正規化を行うことができる。これにより、固有空間に基づく手法を手の形状および姿勢の認識に用いても、十分に精度の高い結果を得ることが可能になる。

また、固有空間に基づく手法を手の形状および姿勢の認識に用いることにより、伸展指の個数を数えるなどの幾何的な特徴による手法により、幾何的な特徴がとれにくい、より複雑な手の形状に対しても認識が可能である。

20

【0039】

第12の発明は、光学的読取手段によって取得された入力手画像の形状および姿勢を認識する方法であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態に正規化した手形状画像を、それぞれ生成する第1の正規化ステップと、

固有空間法による解析を行って、手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第1の投影ステップと、

30

投影座標を、クラスタ分析によりそれぞれグループ分けし、各手形状画像がどのクラスタに属するかと、各クラスタに関する統計情報とを求める分析ステップと、

入力手画像を入力し、当該入力手画像の手首領域を削除した後手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第2の正規化ステップと、

入力手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第2の投影ステップと、

入力手形状画像について求めた投影座標と、統計情報とをそれぞれ比較し、最も近いクラスタを求める判別ステップと、

最も近いクラスタに属する手形状画像と入力手形状画像とをそれぞれ比較し、当該入力手形状画像に最も近い手形状画像を求める比較ステップと、

40

最も近い手形状画像の形状情報と姿勢情報とを出力するステップとを備える。

【0040】

上記のように、第12の発明によれば、複数の手形状画像を固有空間内におけるクラスタ分析によりグループ化し、入力手画像を認識するときに、最初にどのグループに属するかを求め、そのグループ内でどの手形状画像に最も近いかを求める。これにより、画像比較回数を減らすことができ、さらに高速に処理を行うことが可能となる。また、違う形状でも似た画像が存在する場合でも正確に手の形状および姿勢を求めることが可能になる。

【0041】

第13の発明は、第12の発明に従属する発明であって、

50

比較ステップは、
判別ステップが求めたクラスタに属する手形状画像と、第2の正規化ステップが生成した入力手形状画像とを比較する際に、手形状画像を同一クラスタ内の同一形状ごとにグループ化するステップと、
グループを表現する統計量を求めるステップと、
入力手形状画像と統計量との距離を計算し、最も近いグループに属する形状を出力するステップとを含むことを特徴とする。

【0042】

上記のように、第13の発明によれば、第12の発明において、姿勢まで出力する必要がない場合に、姿勢と形状の両方を認識する場合よりも正確に手の形状を求めることが可能になる。

10

【0043】

第14の発明は、第12の発明に従属する発明であって、
分析ステップは、各クラスタに対する手形状画像と形状情報とに基づいて、各手形状画像を判別するための部分領域をそれぞれ計算し、
比較ステップは、判別ステップで求めたクラスタに属する手形状画像と、第2の正規化ステップが生成した入力手形状画像とを比較する際に、クラスタに対応する部分領域内のみで比較することを特徴とする。

【0044】

上記のように、第14の発明によれば、第12の発明において、画像を判別するための部分領域を予め定めておき、この部分領域内のみで手形状画像と入力手形状画像とを比較する。これにより、第23の発明に比べ、画像比較回数を減らすことができると共に、認識対象となる手の画像を認識するときに、違う形状でも似た画像が存在する場合でも正確かつ高速に手の形状および姿勢を求めることが可能となる。

20

【0045】

第15の発明は、第12の発明に従属する発明であって、
入力手画像が、認識対象となる手を複数の視点から撮影した複数の画像である場合、
第2の正規化ステップは、複数の入力手画像のそれぞれについて入力手形状画像を生成し、
第2の投影ステップは、第2の正規化ステップが生成した複数の入力手形状画像について固有空間内での投影座標をそれぞれ求め、
判別ステップは、第2の投影ステップが求めた各投影座標と統計情報とを比較して、最も近いクラスタをそれぞれ求め、
比較ステップは、判別ステップが求めた複数の最も近いクラスタを統合し、各クラスタに属する手形状画像の形状情報および姿勢情報から矛盾しない形状・姿勢を推定することを特徴とする。

30

【0046】

上記のように、第15の発明によれば、第12の発明において、複数の視点から撮影した入力手画像から求めたクラスタに基づき、最も近いクラスタを統合して、入力手画像の形状および姿勢を求める。これにより、1方向からの画像だけでは形状および姿勢が決定できないような場合でも、正確に手の形状および姿勢を求めることが可能となる。

40

【0047】

第16の発明は、光学的読取手段によって取得された手振り動作画像の意味を認識する方法であって、
予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態に正規化した手形状画像を、それぞれ生成する第1の正規化ステップと、
固有空間法による解析を行って、手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、
手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座

50

標をそれぞれ求める第1の投影ステップと、
投影座標を、クラスタ分析によりそれぞれグループ分けし、各手形状画像がどのクラスタに属するかと、各クラスタに関する統計情報とを求める分析ステップと、
手振り動作画像を入力し、当該手振り動作画像の各画像の中から手領域をそれぞれ検出する検出ステップと、
検出した手領域において手振り動作画像の手の動きをそれぞれ求め、手の動きに従い手動作の分節点を求める分節ステップと、
手振り動作画像中の手動作分節点である画像から、検出した手領域の部分を切り出す切出ステップと、
手振り動作画像から切り出された手画像系列について、当該手画像の手首領域を削除した後手形状画像と同等の画像形態となるように正規化した入力手形状画像を、それぞれ生成する第2の正規化ステップと、
入力手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第2の投影ステップと、
入力手形状画像について求めた投影座標と、統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルをそれぞれ出力する判別ステップと、
判別した手画像系列に対応したシンボル系列を、当該手画像系列の元となる手振り動作画像の意味と共に記憶するステップと、
入力する手振り動作画像を識別する場合、判別したシンボル系列に該当する意味を、記憶しているシンボル系列とその意味とに基づいて出力する識別ステップとを備える。

10

20

【0048】

上記のように、第16の発明によれば、ジェスチャー単語や手話単語等のような一連の意味ある手振り動作画像に対して、手動作の分節点となる画像のクラスタ系列をその意味と共に予め記憶し、手振り動作画像を認識するときに、求めたクラスタ系列に基づいて記憶している意味を出力する。これにより、ジェスチャー単語や手話単語等のような一連の意味ある動作に対して、より誤認識を減らし、正確に意味を求めることが可能となる。

【0049】

第17の発明は、第16の発明に従属する発明であって、
手振り動作画像を入力し、当該画像の動作主体の動きと位置とから意味の候補を出力する認識ステップと、
一連の意味ある動作に基づいて、入力された手振り動作画像の意味を制約する制約条件を予め記憶する記憶ステップとをさらに備え、
識別ステップは、判別したシンボル系列に該当する意味を、制約条件に従いつつ記憶しているシンボル系列とその意味とに基づいて出力することを特徴とする。

30

【0050】

上記のように、第17の発明によれば、第16の発明において、手の大局的動作の特徴に基づく制約条件をさらに加えて、手振り動作画像の意味を導き出す。これにより、手振り動作画像の誤認識を減らすことができる。

【0051】

第18の発明は、第16、第17の発明に従属する発明であって、
検出ステップは、
入力する手振り動作画像の各画像の中から、手領域の候補となる領域をそれぞれ切り出す切出ステップと、
手画像の候補となる領域だけを矩形領域から抜き出すための領域マスクを記憶する記憶ステップと、
手振り動作画像から切り出された手領域候補の領域に、領域マスクを付加し、さらに固有ベクトルを計算する時に用いた手画像と同等の画像形態となるように正規化した画像をそれぞれ生成する正規化ステップと、
手領域候補の領域を正規化した画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める投影ステップと、

40

50

投影座標と統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルと比較対象のクラスタとの近さを示す評価値とをそれぞれ出力する判別ステップと、

評価値に基づいて、最も高い評価値を持つ手領域候補の位置情報とそのクラスタとを出力する決定ステップとを含むことを特徴とする。

【0052】

上記のように、第18の発明によれば、第16、第17の発明において、手領域を検出するときに手領域候補の領域を固有空間へ投影して、該当クラスタを求めることによって手領域を検出する。このため、手領域を検出すると同時にその手領域の該当クラスタを求めることができるので、手領域検出と手の形状と姿勢の認識、若しくは手領域検出と手振り動作認識処理を一つの処理として統合することが可能となる。

10

【0053】

第19の発明は、第11、第12、第16の発明に従属する発明であって、第1の正規化ステップおよび第2の正規化ステップは、それぞれ、入力する手の画像から抽出すべき手領域を、色分布として予め記憶する色記憶ステップと、

色分布に従い、入力する手の画像から手領域を抽出するステップと、手首方向を求め、当該手首方向に従い手領域から手首領域を削除するステップと、手首領域を削除した手領域を、画像上の予め定義した位置に移動させるステップと、手領域内の手が、予め定めた一定方向に向くように回転角を求めるステップと、回転角に従い、手が一定方向に向くように手領域を回転させるステップと、回転させた手領域の大きさを、予め定めた一定の大きさに正規化するステップとを含むことを特徴とする。

20

【0054】

上記のように、第19の発明によれば、第11、第12、第16の発明において、手画像を正規化する際に、手首領域を削除するだけではなく、肌色により手領域を抽出する。これにより、一般的で自然な背景で撮影した手の画像から手領域を抽出することができ、さらに、正確に手の形状および姿勢を認識することが可能となる。

【0055】

第20の発明は、第11の発明に従属する発明であって、形状情報および姿勢情報に対応する命令をそれぞれ格納する命令記憶ステップと、出力するステップが出力する形状情報および姿勢情報を入力し、当該形状情報および当該姿勢情報に対応する命令を命令記憶ステップから取得して出力するステップとをさらに備える。

30

【0056】

上記のように、第20の発明によれば、第11の発明の認識装置を、求めた手の形状および姿勢に従い、他の機器の制御を行うインタフェースとして機能させることができる。

【0057】

第21の発明は、光学的読取手段によって取得された入力手画像の形状および姿勢を認識する方法を、コンピュータ装置において実行するためのプログラムを記録した記録媒体であって、

40

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態に正規化した手形状画像を、それぞれ生成する第1の正規化ステップと、

固有空間法による解析を行って、手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第1の投影ステップと、

入力手画像を入力し、当該入力手画像の手首領域を削除した後手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第2の正規化ステップと、

50

入力手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第2の投影ステップと、
手形状画像について求めた投影座標と、入力手形状画像について求めた投影座標とをそれぞれ比較し、入力手形状画像に最も近い手形状画像を求める比較ステップと、
最も近い手形状画像の形状情報と姿勢情報とを出力するステップとを含む動作環境を、コンピュータ装置上で実現するためのプログラムを記録している。

【0058】

第22の発明は、光学的読取手段によって取得された入力手画像の形状および姿勢を認識する方法を、コンピュータ装置において実行するためのプログラムを記録した記録媒体であって、

予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態に正規化した手形状画像を、それぞれ生成する第1の正規化ステップと、

固有空間法による解析を行って、手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第1の投影ステップと、

投影座標を、クラスタ分析によりそれぞれグループ分けし、各手形状画像がどのクラスタに属するかと、各クラスタに関する統計情報とを求める分析ステップと、

入力手画像を入力し、当該入力手画像の手首領域を削除した後手形状画像と同等の画像形態となるように正規化した入力手形状画像を、生成する第2の正規化ステップと、

入力手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める第2の投影ステップと、

入力手形状画像について求めた投影座標と、統計情報とをそれぞれ比較し、最も近いクラスタを求める判別ステップと、

最も近いクラスタに属する手形状画像と入力手形状画像とをそれぞれ比較し、当該入力手形状画像に最も近い手形状画像を求める比較ステップと、

最も近い手形状画像の形状情報と姿勢情報とを出力するステップとを含む動作環境を、コンピュータ装置上で実現するためのプログラムを記録している。

【0059】

第23の発明は、第22の発明に従属する発明であって、

比較ステップは、

判別ステップが求めたクラスタに属する手形状画像と、第2の正規化ステップが生成した入力手形状画像とを比較する際に、手形状画像を同一クラスタ内の同一形状ごとにグループ化するステップと、

グループを表現する統計量を求めるステップと、

入力手形状画像と統計量との距離を計算し、最も近いグループに属する形状を出力するステップとを含むことを特徴とする。

【0060】

第24の発明は、第22の発明に従属する発明であって、

分析ステップは、各クラスタに対する手形状画像と形状情報とに基づいて、各手形状画像を判別するための部分領域をそれぞれ計算し、

比較ステップは、判別ステップで求めたクラスタに属する手形状画像と、第2の正規化ステップが生成した入力手形状画像とを比較する際に、クラスタに対応する部分領域内のみで比較することを特徴とする。

【0061】

第25の発明は、第22の発明に従属する発明であって、

入力手画像が、認識対象となる手を複数の視点から撮影した複数の画像である場合、

第2の正規化ステップは、複数の入力手画像のそれぞれについて入力手形状画像を生成し、

10

20

30

40

50

第2の投影ステップは、第2の正規化ステップが生成した複数の入力手形状画像について固有空間内での投影座標をそれぞれ求め、

判別ステップは、第2の投影ステップが求めた各投影座標と統計情報とを比較して、最も近いクラスをそれぞれ求め、

比較ステップは、判別ステップが求めた複数の最も近いクラスを統合し、各クラスに属する手形状画像の形状情報および姿勢情報から矛盾しない形状・姿勢を推定することを特徴とする。

【0062】

第26の発明は、光学的読取手段によって取得された手振り動作画像の意味を認識する方法を、コンピュータ装置において実行するためのプログラムを記録した記録媒体であって

10

、
予め様々な形状と姿勢の手を撮影した複数の画像を入力し、当該画像の手首領域を削除した後予め定めた画像形態に正規化した手形状画像を、それぞれ生成する第1の正規化ステップと、

固有空間法による解析を行って、手形状画像から固有値と固有ベクトルとを、それぞれ計算する解析ステップと、

手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第1の投影ステップと、

投影座標を、クラス分析によりそれぞれグループ分けし、各手形状画像がどのクラスに属するかと、各クラスに関する統計情報とを求める分析ステップと、

20

手振り動作画像を入力し、当該手振り動作画像の各画像の中から手領域をそれぞれ検出する検出ステップと、

検出した手領域において手振り動作画像の手の動きをそれぞれ求め、手の動きに従い手動作の分節点を求める分節ステップと、

手振り動作画像中の手動作分節点である画像から、検出した手領域の部分を切り出す切出ステップと、

手振り動作画像から切り出された手画像系列について、当該手画像の手首領域を削除した後手形状画像と同等の画像形態となるように正規化した入力手形状画像を、それぞれ生成する第2の正規化ステップと、

入力手形状画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める第2の投影ステップと、

30

入力手形状画像について求めた投影座標と、統計情報とをそれぞれ比較し、最も近いクラスを求め、当該クラスを特定するシンボルをそれぞれ出力する判別ステップと、

判別した手画像系列に対応したシンボル系列を、当該手画像系列の元となる手振り動作画像の意味と共に記憶するステップと、

入力する手振り動作画像を識別する場合、判別したシンボル系列に該当する意味を、記憶しているシンボル系列とその意味とに基づいて出力する識別ステップとを含む動作環境を、コンピュータ装置上で実現するためのプログラムを記録している。

【0063】

第27の発明は、第26の発明に従属する発明であって、

40

手振り動作画像を入力し、当該画像の動作主体の動きと位置とから意味の候補を出力する認識ステップと、

一連の意味ある動作に基づいて、入力された手振り動作画像の意味を制約する制約条件を予め記憶する記憶ステップとをさらに含み、

識別ステップは、判別したシンボル系列に該当する意味を、制約条件に従いつつ記憶しているシンボル系列とその意味とに基づいて出力することを特徴とする。

【0064】

第28の発明は、第26、第27の発明に従属する発明であって、

検出ステップは、

入力する手振り動作画像の各画像の中から、手領域の候補となる領域をそれぞれ切り出す

50

切出ステップと、

手画像の候補となる領域だけを矩形領域から抜き出すための領域マスクを記憶する記憶ステップと、

手振り動作画像から切り出された手領域候補の領域に、領域マスクを付加し、さらに固有ベクトルを計算する時に用いた手画像と同等の画像形態となるように正規化した画像をそれぞれ生成する正規化ステップと、

手領域候補の領域を正規化した画像を固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標をそれぞれ求める投影ステップと、

投影座標と統計情報とをそれぞれ比較し、最も近いクラスタを求め、当該クラスタを特定するシンボルと比較対象のクラスタとの近さを示す評価値とをそれぞれ出力する判別ステップと、

10

評価値に基づいて、最も高い評価値を持つ手領域候補の位置情報とそのクラスタとを出力する決定ステップとを含むことを特徴とする。

【 0 0 6 5 】

第 2 9 の発明は、第 2 1 , 第 2 2 , 第 2 6 の発明に従属する発明であって、

第 1 の正規化ステップおよび第 2 の正規化ステップは、それぞれ、

入力する手の画像から抽出すべき手領域を、色分布として予め記憶する色記憶ステップと、

色分布に従い、入力する手の画像から手領域を抽出するステップと、

手首方向を求め、当該手首方向に従い手領域から手首領域を削除するステップと、

20

手首領域を削除した手領域を、画像上の予め定義した位置に移動させるステップと、

手領域内の手が、予め定めた一定方向に向くように回転角を求めるステップと、

回転角に従い、手が一定方向に向くように手領域を回転させるステップと、

回転させた手領域の大きさを、予め定めた一定の大きさに正規化するステップとを含むことを特徴とする。

【 0 0 6 6 】

第 3 0 の発明は、第 2 1 の発明に従属する発明であって、

形状情報および姿勢情報に対応する命令をそれぞれ格納する命令記憶ステップと、

出力するステップが出力する形状情報および姿勢情報を入力し、当該形状情報および当該姿勢情報に対応する命令を命令記憶ステップから取得して出力するステップとをさらに含む。

30

【 0 0 6 7 】

上記のように、第 2 1 ~ 第 3 0 の発明は、上記第 1 1 ~ 第 2 0 の発明の手の形状と姿勢の認識方法を実行するためのプログラムを記録した記録媒体である。これは、既存の装置に対し、上記第 1 1 ~ 第 2 0 の発明の手の形状と姿勢の認識方法を、ソフトウェアの形態で供給することに対応させたものである。

【 0 0 6 8 】

【発明の実施の形態】

以下、本発明の各実施形態を説明するに先立ち、説明内で用いる「手形状」および「手姿勢」という用語についての定義を行う。

40

【 0 0 6 9 】

人間がジェスチャーや手話等のように、手に何らかの意味を持たせて提示する場合、指と掌による関節の曲げ角によって決められる手のポーズ（例えば、「グー」、「チョキ」、「パー」等）と、手首および腕関節によって決められる手の方向（例えば、指差しの方向等）との 2 つの意味が含まれている。そこで、本発明の各実施形態の説明においては、指と掌による関節の曲げ角によって決められる手のポーズを「手形状」と呼び、手首および腕関節によって決められる手の方向を「手姿勢」と呼ぶ。

【 0 0 7 0 】

ここで、手姿勢を厳密に定義する場合、例えば、図 3 7 のように定義することができる。

まず、ある形状を示す手が存在する三次元空間において、手の手首断面中心から中指先端

50

中心へ伸ばした方向を X_i 軸（掌主軸）と、 X_i 軸と直交して手のひら平面に対して垂直方向を Y_i 軸と、 X_i 軸および Y_i 軸の双方に直交する方向を Z_i 軸としたローカル座標系 i を定義する（図37（a））。一方、カメラにおいて手の画像を取り込んで投影するカメラ座標系 c （ X_c 軸、 Y_c 軸、 Z_c 軸。なお、各軸は相互に直交する）を、予め設定する（図37（b））。なお、以下、カメラ座標系 c の Z_c 軸を光軸と表現する。

そして、カメラ座標系 c 上に投影された手の画像に対して、手のローカル座標系 i の各軸とカメラ座標系 c の各軸との差を、次のように定義する（図37（c））。

： X_c 軸を中心とする回転角度

： X_c 軸 - Z_c 軸平面における回転角度

： X_c 軸 - Y_c 軸平面における回転角度

これらの回転角度、 θ_1 、 θ_2 、 θ_3 により、手姿勢を定義する。

【0071】

なお、手姿勢の表現は、このように厳密に定義することも可能だが、例えば、手の平がカメラに対してどの程度傾いているかを「カメラに対して正面、カメラに対して左に向いている」等の定性的な表現で定義することも可能である。本発明では、これらのいずれにおいても対応可能であるが、以下に示す各実施形態では、わかりやすく説明するために、定性的な表現で姿勢を定義する場合を一例に挙げて説明することにする。

【0072】

以下、図面を参照して、本発明の各実施形態を詳細に説明する。

（第1の実施形態）

本発明の第1の実施形態は、予め記憶する様々な手形状と手姿勢とを持つ複数の画像と、入力される認識対象である手の画像とを、固有空間法に基づき認識を行う場合に、用意する手の画像から手首領域を削除することによって正規化をすることにより、より複雑な形状の手画像に対しても手形状と手姿勢とを認識する装置および方法を提供するものである。

【0073】

図1は、本発明の第1の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。図1において、第1の実施形態に係る手の形状と姿勢の認識装置は、記憶部構築系1と、姿勢・形状認識系2とで構成される。

記憶部構築系1は、様々な手形状と手姿勢を示す複数の手形状画像とその形状情報と姿勢情報とから、認識するために必要な情報を予め構築する。姿勢・形状認識系2は、入力される認識対象となる手の画像（以下、入力手画像という）に対し、記憶部構築系1により構築された記憶部に格納されている情報を用いて、当該入力手画像の手形状および手姿勢を求める。

【0074】

まず、記憶部構築系1および姿勢・形状認識系2の各構成を、それぞれ説明する。図1において、記憶部構築系1は、手画像正規化部11と、手形状画像情報記憶部12Aと、固有空間計算部13と、固有ベクトル記憶部14と、固有空間投影部15とを備える。姿勢・形状認識系2は、手画像正規化部21と、固有空間投影部22と、手形状画像選択部23と、形状・姿勢出力部24とを備える。

【0075】

手画像正規化部11は、様々な手形状と手姿勢とを持つ複数の手画像を入力し、当該手画像から手首領域を削除して予め定めた正規化を施した手形状画像をそれぞれ生成する。手形状画像情報記憶部12Aは、手画像正規化部11が生成した複数の手形状画像を、別途与えられる当該手形状画像の形状情報および姿勢情報と、手形状画像を固有空間へ投影した固有空間投影座標と共にそれぞれ格納する。固有空間計算部13は、手形状画像情報記憶部12Aが格納した手形状画像から、固有空間法による解析を行って固有値と固有ベクトルとを求める。ここで、固有空間計算部13が行う固有空間解析の手法としては、例えば、手形状画像情報記憶部12Aに記憶した手形状画像から主成分分析を行って固有空間を求める手法、手形状画像情報記憶部12Aに記憶した手形状画像と形状情報とから判別

10

20

30

40

50

分析を行ってその分析結果から手形状判別空間を求める手法等、種々の手法が考えられるが、第1の実施形態では、前者の手法で以降の動作を説明することにする。固有ベクトル記憶部14は、固有空間計算部13が求めた固有ベクトルを格納する。固有空間投影部15は、手形状画像情報記憶部12Aに格納した手形状画像を、固有ベクトル記憶部14に格納した固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求め、手形状画像情報記憶部12Aに格納する。

【0076】

手画像正規化部21は、入力手画像を入力し、当該入力画像が手形状画像情報記憶部12Aに予め記憶している手形状画像と同等のものとなるように、当該入力手画像から手首領域を削除して予め定めた正規化を施した入力手形状画像を生成する。固有空間投影部22は、手画像正規化部21が生成した入力手形状画像を、固有ベクトル記憶部14に格納した固有ベクトルを基底とする固有空間に投影して、当該固有空間内での投影座標を求める。手形状画像選択部23は、固有空間投影部22が求めた投影座標と、手形状画像情報記憶部12Aに予め記憶している固有空間投影座標とを比較し、入力手形状画像に最も近い手形状画像を求める。形状・姿勢出力部24は、手形状画像選択部23が求めた最も近い手形状画像の形状情報と姿勢情報とを出力する。

【0077】

次に、図2～図5を用いて、第1の実施形態に係る手の形状と姿勢の認識装置が行う手形状/手姿勢の認識方法を、処理の順に説明する。図2は、図1の手画像正規化部11が行う処理の概略を説明する図である。図3は、図1の手形状画像情報記憶部12Aが有する格納テーブルの一例を示す図である。図4は、図1の固有空間計算部13において、固有空間を求める手法の一例の概略を説明する図である。なお、図4においては、上述した主成分分析による手法を用いた場合を一例として記載している。図5は、図1の固有空間投影部15において、固有空間投影座標を求める手法の概略を説明する図である。

【0078】

最初に、記憶部構築系1が行う処理を説明する。

記憶部構築系1は、上述したように、姿勢・形状認識系2に入力される入力手画像と比較するための手形状画像を、様々な手形状と手姿勢とを持つ複数の手画像を用いて予め構築する。ここで、記憶部構築系1は、手形状画像に対する固有空間を求めるため、手画像の正規化を行う。

【0079】

図2を参照して、手画像正規化部11は、まず、与えられた手画像から手首方向を求める(図2(b))。次に、手画像正規化部11は、手首側の端から掌側に向けて手首と背景との境界線に従い直線を引き、その直線との距離が予め定めたしきい値以上になった地点を手首領域の終点(手首切出点)として求める(図2(c))。次に、手画像正規化部11は、手画像から求めた手首切出点までの手首領域を削除する(図2(d))。次に、手画像正規化部11は、手首領域を削除した画像から手の部分だけを抜き出し、手首-中指方向が、ある一定方向に向くように回転する(図2(e))。なお、本実施例では、その一定方向が真下であると仮定している。そして、手画像正規化部11は、回転した手画像の大きさと全体の明るさを、予め定めた値に正規化することで手形状画像を生成し(図2(f))、この手形状画像の指の状態を示す形状情報(図2の例では、伸展3本指)および手画像の掌の向きを示す姿勢情報(図2の例では、掌方向:後)と共に、手形状画像情報記憶部12Aに記憶する。なお、図2の例では、姿勢情報として言葉で表現したものをを用いたが、その他にも光軸に対する角度で表現してもかまわない。手画像正規化部11は、上記正規化処理を様々な手形状と手姿勢とを持つ複数の手画像に対してそれぞれ行い、図3に示すように、複数の手形状画像を手形状画像情報記憶部12Aに記憶する。なお、手形状画像情報記憶部12Aにおける固有空間投影座標は、固有空間投影部15によって求められた結果を格納するため、この時点では何も格納されていない。

【0080】

次に、固有空間計算部13は、手形状画像情報記憶部12Aに格納されたそれぞれの手形

10

20

30

40

50

状画像の固有空間を求める。

図4を参照して、まず、固有空間計算部13は、手形状画像情報記憶部12Aに記憶されている全ての手形状画像の平均画像 c を求める(ステップS1)。次に、固有空間計算部13は、各手形状画像ごとに、手形状画像から平均画像 c を引いた画像をラスタスキャンして一次元ベクトルで表現し(ステップS2)、全ての画像の一次元ベクトルを列ベクトルとして並べた行列 A を求める(ステップS3)。次に、固有空間計算部13は、行列 A から画像集合の共分散行列 Q を求め(ステップS4)、この共分散行列 Q の固有値と固有ベクトルを求める(ステップS5)。最後に、固有空間計算部13は、予め別途定義した k 個の大きい固有値に対応する固有ベクトル(e_1, e_2, \dots, e_k)を基底ベクトルとする固有空間を求める(ステップS6)。

10

以上の処理により、固有空間計算部13は、固有空間基底ベクトルを計算し、固有ベクトル記憶部14に固有ベクトルの集合を格納する。

【0081】

次に、固有空間投影部15は、手形状画像情報記憶部12Aに格納されたそれぞれの手形状画像に対し、各手形状画像を固有空間へ投影した固有空間投影座標を求める。

図5を参照して、固有空間投影部15は、手形状画像情報記憶部12Aに記憶されている各手形状画像ごとに、画像をラスタスキャンして一次元ベクトルを求め、当該一次元ベクトルと固有ベクトル記憶部14に格納された固有ベクトルとの掛け算を行って固有空間投影座標を求める。そして、固有空間投影部15は、この求めた各固有空間投影座標を、手形状画像情報記憶部12Aにそれぞれ格納する。

20

以上の処理によって、記憶部構築系1で予め行う処理が終了し、手形状画像情報記憶部12Aおよび固有ベクトル記憶部14に全ての情報が格納される。

【0082】

次に、姿勢・形状認識系2が行う処理を説明する。

認識対象となる入力手画像は、手画像正規化部21に入力される。手画像正規化部21は、入力手画像に対し、手画像正規化部11と同様の手法で正規化した入力手形状画像を生成する。固有空間投影部22は、手画像正規化部21において生成された入力手形状画像について、固有空間投影部15と同様に、固有ベクトル記憶部14に記憶された固有ベクトルを用いて固有空間投影座標を求める。次に、手形状画像選択部23は、固有空間投影部22が求めた入力手形状画像に関する固有空間投影座標と、手形状画像情報記憶部12Aに予め記憶している各手形状画像の固有空間投影座標との距離(例えば、ユークリッド距離)をそれぞれ求め、入力手形状画像に最も近い手形状画像を求める。そして、形状・姿勢出力部24は、求めた最も近い手形状画像の形状情報および姿勢情報を出力する。

30

以上の結果、入力手画像の手形状と手姿勢とを同時に求めることができる。

【0083】

なお、典型的なハードウェア環境では、上記第1の実施形態に係る手の形状と姿勢の認識装置は、所定のプログラムデータが格納された記憶装置(ROM, RAM, ハードディスク等)とCPU(セントラル・プロセッシング・ユニット)と入出力装置とによって構成される。図6に、本第1の実施形態に係る手の形状と姿勢の認識装置を実現するハードウェア構成の一例を示す。

40

図6において、記憶装置50は、例えば、ハードディスク等であって、手形状画像情報記憶部12Aおよび固有ベクトル記憶部14の機能を持つ。CPU51は、各部の動作を制御する中央演算装置である。メモリ52は、各部が動作するとき一時データを保存する。画像入力装置53は、例えば、ビデオキャプチャカード等であって、認識対象である入力手画像を入力する。入力装置54は、様々な手形状と手姿勢とを持つ複数の手形状画像とその形状情報および姿勢情報を入力する。出力装置55は、認識した手形状と手姿勢とを示すデータを出力する。これらのハードウェア構成をとることにより、第1の実施形態に係る手の形状と姿勢の認識装置を実現することができる。なお、このような場合、第1の実施形態に係る手の形状と姿勢の認識装置が行う各処理は、独立したプログラムデータの形態で提供される。このプログラムデータは、CD-ROMやフロッピーディスク等

50

の記録媒体を介して導入されてもよい。

【0084】

また、上記第1の実施形態に係る手の形状と姿勢の認識装置を、他の機器へのインタフェースとして使用する場合、次の機能を有する構成を付加すればよい。その構成とは、形状情報および姿勢情報に対する命令を記憶する命令記憶部と、当該命令を出力する命令出力部である。命令記憶部は、例えば、図7に示すような、形状情報および姿勢情報に対応する他の機器への命令を記憶している。図7は、オーディオ機器に対する命令が記憶されている一例を示す。そして、命令出力部は、形状・姿勢出力部24が求めた形状情報・姿勢情報に従い、命令記憶部から形状情報・姿勢情報に対応する命令を他の機器に出力する。例えば、図7において、形状・姿勢出力部24が「伸展5本指」の形状情報、かつ、「全姿勢」の姿勢情報を求めた場合には、命令出力部は、オーディオ機器を「スタート」する命令を出力するのである。このようにして、他の機器のインタフェースとして、上記第1の実施形態に係る手の形状と姿勢の認識装置を用いることが可能になる。

10

【0085】

以上のように、本発明の第1の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、様々な手形状と手姿勢とを持つ複数の手画像と認識対象である入力手画像の両方から手首領域を削除するため、単純に大きさと明るさを正規化するよりも精密に手画像の正規化を行うことができる。これにより、固有空間に基づく手法を手形状と手姿勢との認識に用いても、十分に精度の高い結果を得ることが可能になる。

また、固有空間に基づく手法を手形状と手姿勢との認識に用いることにより、伸展指の個数を数えるなどの幾何的な特徴による手法により、幾何的な特徴がとれにくい、より複雑な手の形状に対しても対処が可能である。

20

さらに、複数の手形状画像と入力手形状画像とを比較する場合、全ての手形状画像とのマッチングを行おうとすると、画像数が膨大になることが考えられる。しかし、本第1の実施形態のように、手画像を正規化して固有空間内への投影してその投影座標を予め求めておき、その固有空間内で入力手形状画像と比較することにより、画像自体を比較するよりも計算量が減り、高速に処理を行うことができる。このため、様々な手形状と手姿勢とを示す複数の手形状画像のように膨大な量の手形状画像が考えられる場合には、非常に実用的な手法となる。

【0086】

30

なお、上記第1の実施形態では、様々な手形状と手姿勢とを持つ複数の手形状画像として、実際の手画像を格納していることを仮定していたが、複数視点の画像を何らかの理由で撮影できない場合が考えられる。例えば、手の形状を認識する場合、手はいろいろな姿勢をとるため、予め様々な手姿勢で提示する手画像を用意する必要がある。実際の手をターンテーブルの上に載せて撮影することは不可能であり、また、人間にある姿勢を保つように命令し、その姿勢を撮影する場合、人間が示す姿勢の精度には限界があるため、実際には手の周囲を覆うような形で撮影する特別な機材を用意する必要があると考えられる。そこで、予め、CADやCGなどで使われる三次元モデルとして手のモデルを用意し、そのモデルの複数視点の投影像を格納することにより、より、精度良く手形状画像とその時の手形状および手姿勢との関係を定義することができる。また、マネキン等で使われているような実際の手モデル等を用いても、同等に定義することが可能である。なお、本第1の実施形態においては、三次元モデルの投影像を用意した場合でも実際の手画像を用意した場合でも、全く同じ構成および手法を用いることにより実現が可能である。

40

【0087】

また、上記第1の実施形態は、手形状と手姿勢を1つだけ出力することを基本にしていたが、画像の解像度などの関係で、区別がつかないほど似ている場合等には、1つに絞れない場合があり得る。その場合は、複数の手形状と手姿勢の候補を出力することもあると考えられる。その場合も、上記第1の実施形態と全く同じ構成および手法を用いることにより実現が可能である。さらに、上記第1の実施形態は、手形状画像および入力手形状画像の両方に対し、濃淡画像を仮定したが、これらの画像がシルエット画像であっても、カラー画像

50

であっても、上記第 1 の実施形態と全く同じ構成および手法を用いることにより実現が可能である。

【 0 0 8 8 】

(第 2 の実施形態)

一般に、手形状と手姿勢とが与えられている複数の手の画像を分類する場合、手形状ごとに、若しくは、手姿勢ごとに分類することが考えられる。しかしながら、手の画像の場合、「手形状は違うが似た画像（例えば、指を 1 本、若しくは 2 本出した形状を横から見た場合）」、「手姿勢は違うが似た画像（例えば、握った形状）」等が考えられる。そのため、手形状や手姿勢に従って分類すると、手形状と手姿勢を認識する場合に不都合な場合が多い。

10

【 0 0 8 9 】

そこで、本発明の第 2 の実施形態は、上記第 1 の実施形態で述べた固有空間法に基づく手の形状と姿勢の認識装置および方法において、手形状画像情報記憶部 1 2 A に記憶した全ての手形状画像の固有空間投影座標をクラスタ分析により自動的にグループ分けし、認識対象である入力手画像が与えられたときに、最初にどのグループに属するかを求め、そして、そのグループ内のどの手形状画像に近いかを求めることにより、比較回数を減らし高速に処理を行う装置および方法を提供するものである。

【 0 0 9 0 】

図 8 は、本発明の第 2 の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。図 8 において、第 2 の実施形態に係る手の形状と姿勢の認識装置は、上記第 1 の実施形態に係る手の形状と姿勢の認識装置と同様、記憶部構築系 1 と、姿勢・形状認識系 2 とで構成される。

20

図 8 において、記憶部構築系 1 は、手画像正規化部 1 1 と、手形状画像情報記憶部 1 2 B と、固有空間計算部 1 3 と、固有ベクトル記憶部 1 4 と、固有空間投影部 1 5 と、クラスタ分析部 1 6 と、クラスタ情報記憶部 1 7 A とを備える。姿勢・形状認識系 2 は、手画像正規化部 2 1 と、固有空間投影部 2 2 と、最尤クラスタ判別部 2 5 と、画像比較部 2 6 と、形状・姿勢出力部 2 4 とを備える。

【 0 0 9 1 】

図 8 に示すように、第 2 の実施形態に係る手の形状と姿勢の認識装置は、上記第 1 の実施形態に係る手の形状と姿勢の認識装置に対し、記憶部構築系 1 において、手形状画像情報記憶部 1 2 A を手形状画像情報記憶部 1 2 B に代えて、クラスタ分析部 1 6 とクラスタ情報記憶部 1 7 A とをさらに加えた構成であり、姿勢・形状認識系 2 において、手形状画像選択部 2 3 を、最尤クラスタ判別部 2 5 および画像比較部 2 6 に代えた構成である。

30

なお、第 2 の実施形態に係る手の形状と姿勢の認識装置におけるその他の構成は、上記第 1 の実施形態に係る手の形状と姿勢の認識装置の構成と同様であり、当該構成については、同一の参照番号を付してその説明を省略する。

【 0 0 9 2 】

まず、第 2 の実施形態における記憶部構築系 1 および姿勢・形状認識系 2 の各構成を、上記第 1 の実施形態と異なる部分を中心に説明する。

手形状画像情報記憶部 1 2 B は、手画像正規化部 1 1 が生成した複数の手形状画像を、その形状情報および姿勢情報と手形状画像を固有空間へ投影した固有空間投影座標と共にそれぞれ格納する。ここで、手形状画像情報記憶部 1 2 B は、上記第 1 の実施形態で述べた手形状画像情報記憶部 1 2 A と異なり、複数の手形状画像を自動的にクラスタリングしたときのクラスタインデックス（以下、クラスタ ID という）を格納する。クラスタ分析部 1 6 は、手形状画像情報記憶部 1 2 B に格納された固有空間投影座標をクラスタ分析によりクラスタ化し、各手形状画像がどのクラスタに属するかを求め、クラスタを特定するクラスタ ID を手形状画像情報記憶部 1 2 B に格納すると共に、各クラスタに関する統計情報を求める。クラスタ情報記憶部 1 7 A は、クラスタ分析部 1 6 が求めたクラスタ ID および統計情報を格納する。

40

【 0 0 9 3 】

50

最尤クラスタ判別部 25 は、固有空間投影部 22 が求めた固有空間投影座標に最も近い投影座標を有するクラスタを取得する。画像比較部 26 は、最尤クラスタ判別部 25 が取得したクラスタに属する手形状画像情報記憶部 12B に記憶された手形状画像から、手画像正規化部 21 が生成した入力手形状画像に最も近い手形状画像を求める。

【0094】

次に、図 9 ~ 図 11 を用いて、第 2 の実施形態に係る手の形状と姿勢の認識装置が行う手形状 / 手姿勢の認識方法を、処理の順に説明する。図 9 は、図 8 の手形状画像情報記憶部 12B が有する格納テーブルの一例を示す図である。図 10 は、図 8 のクラスタ分析部 16 が行う処理手順の一例を示すフローチャートである。なお、図 10 においては、クラスタ分析の一手法である ISODATA 法を用いた場合を挙げて説明している。図 11 は、図 8 の画像比較部 26 において行う比較手法の概念の一例を示す図である。なお、図 11 においては、単純なパターンマッチングによる比較手法を用いた場合を挙げて説明している。

10

【0095】

最初に、記憶部構築系 1 が行う処理を説明する。

まず、手画像正規化部 11 は、上記第 1 の実施形態と同様に、様々な姿勢を示す複数の手の画像から手首領域を削除して正規化を行うことで手形状画像をそれぞれ生成し、図 9 に示すように、複数の手形状画像、形状情報および姿勢情報を手形状画像情報記憶部 12B に記憶する。なお、手形状画像情報記憶部 12B における固有空間投影座標とクラスタ ID とについては、固有空間投影部 15 およびクラスタ分析部 16 によって求められた投影座標とクラスタ ID とを格納するため、この時点では何も格納されていない。

20

次に、固有空間計算部 13、固有ベクトル記憶部 14 および固有空間投影部 15 が、上記第 1 の実施形態と同様に固有空間法に基づいて固有空間を求め、固有空間へ手形状画像情報記憶部 12B に記憶した手形状画像を投影し、投影によって求めた固有空間投影座標を手形状画像情報記憶部 12B にそれぞれ格納する。

【0096】

次に、クラスタ分析部 16 は、手形状画像情報記憶部 12B に格納された固有空間投影座標についてクラスタ分析を行い、手形状画像が近いもの同士を同一グループに分類されるようにグループ分けを行う。このクラスタ分析部 16 で行うクラスタ分析の手法には、単純な再配置法 (k - 平均法) や ISODATA 法等の様々な手法が存在するが、ここでは、ISODATA 法によるクラスタリングの手法を一例に挙げて説明する。

30

【0097】

ISODATA 法は、非階層的クラスタリングの中では代表的な手法で、再配置法によるクラスタリングに加えて、クラスタの分割と統合の手続からなっている。

図 10 を参照して、クラスタ分析部 16 は、まず、初期パラメータを設定する (ステップ S101)。初期パラメータとしては、例えば、最終クラスタの数、再配置の収束条件、微少クラスタ・孤立データの判定条件、分裂・融合の分岐条件、反復計算の終了条件といったものがある。次に、クラスタ分析部 16 は、初期クラスタの中心を決定する (ステップ S102)。この初期クラスタは、手形状画像の投影座標集合に対し、適当に初期クラスタの中心となる画像を選択して決定することで代用できる。

40

【0098】

次に、クラスタ分析部 16 は、再配置法によりクラスタリングを行う。まず、クラスタ分析部 16 は、固有空間内での各手形状画像とクラスタとの距離を計算し、それぞれの画像を距離が最小となるクラスタに配置する (ステップ S103)。次に、クラスタ分析部 16 は、再配置された画像の固有空間投影座標に従い、各クラスタの中心を再計算する (ステップ S104)。そして、クラスタ分析部 16 は、所属するクラスタを変えた画像数が、予め定めたしきい値以下である (収束した) か否かを判断する (ステップ S105)。このステップ S105 の判断において、所属するクラスタを変えた画像数があるしきい値以下であれば、クラスタ分析部 16 は、再配置法によるクラスタリング処理を終了し、それ以外の場合は、上記ステップ S103 に戻って処理を繰り返す。

50

【 0 0 9 9 】

上記ステップ S 1 0 5 の判断において、収束したと判定した場合、クラスタ分析部 1 6 は、個体の数が著しく少ないクラスタと、他の個体から著しく離れた個体とを、以後のクラスタリングから除外する（ステップ S 1 0 6）。次に、クラスタ分析部 1 6 は、クラスタの数が最終クラスタ数を中心とする一定の範囲内にあり、クラスタ中心間の距離の最小値が予め定めたしきい値以下であるか否かを判断する（ステップ S 1 0 7）。このステップ S 1 0 7 の判断において、クラスタ中心間の距離の最小値が予め定めたしきい値以下である場合、クラスタ分析部 1 6 は、クラスタリングは収束したとして、各クラスタの情報（クラスタ ID、固有空間におけるクラスタの平均値、分散などの統計情報）をクラスタ情報記憶部 1 7 A に記憶し、各手形状画像がどのクラスタに属するかを示すクラスタ ID を手形状画像情報記憶部 1 2 B に記憶する（ステップ S 1 0 8）。一方、上記ステップ S 1 0 7 の判断において、クラスタ中心間の距離の最小値が予め定めたしきい値以下でない場合、クラスタ分析部 1 6 は、クラスタの分裂または融合を行う（ステップ S 1 0 9）。このステップ S 1 0 9 では、クラスタ分析部 1 6 は、クラスタ数が最終クラスタに対し一定の範囲を越えて大きいときは、クラスタの分裂を行い、小さいときは融合を行う。クラスタ数が一定の範囲にあるときは、反復回数が偶数なら融合を行い、奇数なら分裂を行う。

10

【 0 1 0 0 】

クラスタ分析部 1 6 は、クラスタの融合では、クラスタ中心間の距離の最小値が予め定めたしきい値以下なら、そのクラスタ対を融合し新しいクラスタ中心を求める。次に、クラスタ分析部 1 6 は、再び中心間距離を計算し、最小値がしきい値以上となるまで融合を続ける。

20

一方、クラスタ分析部 1 6 は、クラスタの分裂では、クラスタの分散の最大値が予め定めたしきい値以上なら、そのクラスタを第 1 主成分に沿って 2 分し、新しいクラスタ中心と分散とを計算する。分散の最大値がしきい値以下となるまで分裂を繰り返す。

そして、クラスタ分析部 1 6 は、上記ステップ S 1 0 9 による分裂または融合が終了すると、再びステップ S 1 0 3 に戻って処理を繰り返す。

【 0 1 0 1 】

上記処理を行うことにより、クラスタ分析が終了し、各クラスタの情報であるクラスタ ID、固有空間におけるクラスタの平均値、分散等の統計情報がクラスタ情報記憶部 1 7 A に、各手形状画像がどのクラスタに属するかを示すクラスタ ID が手形状画像情報記憶部 1 2 B に格納される。なお、上記パラメータについて、実験などにより、最適なパラメータを随時選択することも考えられるが、それ以外に、ある情報量基準（例えば、AIC、MDL 等）に従い、最終クラスタの数、クラスタの分割・統合基準を指定することも可能である。なお、本実施例では、ISODATA 法によるクラスタ分析について説明したが、単純な再配置法によるクラスタ分析であっても、しきい値等のパラメータを適切に設定することで、ISODATA 法と同等の効果を奏する。

30

以上の処理によって、記憶部構築系 1 で予め行う処理が終了し、手形状画像情報記憶部 1 2 B、固有ベクトル記憶部 1 4 およびクラスタ情報記憶部 1 7 A に全ての情報が格納される。

【 0 1 0 2 】

40

次に、姿勢・形状認識系 2 が行う処理を説明する。

認識対象となる入力手画像は、手画像正規化部 2 1 に入力される。手画像正規化部 2 1 および固有空間投影部 2 2 は、上記第 1 の実施形態と同様に、正規化した入力手形状画像および固有空間投影座標を求める。最尤クラスタ判別部 2 5 は、固有空間投影部 2 2 が求めた固有空間投影座標とクラスタ情報記憶部 1 7 A に記憶されているクラスタ情報との距離を求め、入力手形状画像に最も近い手形状画像が属するクラスタを求める。なお、最も近いクラスタを求める手法としては、各クラスタの平均とのユークリッド距離による手法、各クラスタとのマハラノビス距離による手法、最尤法により各クラスタと尤度を求め最も尤度が高いクラスタを近いクラスタとする手法等が考えられるが、ここでは、最尤法により最近クラスタを求める手法を一例に挙げて説明する。

50

【 0 1 0 3 】

まず、最尤クラスタ判別部 2 5 は、クラスタの統計情報として、クラスタ情報記憶部 1 7 A にあるクラスタに属する画像の固有空間投影座標 u から平均 μ を求め、クラスタ中心座標とする。さらに、最尤クラスタ判別部 2 5 は、各画像の固有空間投影座標 u とクラスタ中心座標とから共分散行列 Σ を求め、これらの値から下記式 (4) に従って、クラスタ i に関する尤度関数 $G_i(u)$ を定義する。なお、下記式 (4) 中の Σ_i は、画像の固有空間投影座標 u とクラスタ i とのマハラノビス距離を示す。

【 数 1 】

$$G_i(u) = -\frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} x^2(u; \mu_i, \Sigma_i) \quad \cdots (4)$$

10

この尤度関数 $G_i(u)$ により、最も尤度が高いクラスタを求める。

なお、これ以外の前述した手法（ユークリッド距離による手法、マハラノビス距離による手法）においても、登録している形状の数が少ない場合には、同等の効果を奏することができる。

【 0 1 0 4 】

次に、画像比較部 2 6 は、手形状画像情報記憶部 1 2 B に格納されているクラスタ ID を参照して、最尤クラスタ判別部 2 5 が求めたクラスタに属する手形状画像のみと、手画像正規化部 2 1 が生成した入力手形状画像とを比較し、入力手形状画像に最も近い手形状画像を求める。なお、画像比較部 2 6 において行う入力手形状画像と手形状画像との比較手法として各種考えられるが、例えば、単純なパターンマッチングによる手法で比較すればよい。そして、形状・姿勢出力部 2 4 は、画像比較部 2 6 が求めた手形状画像の形状情報および姿勢情報を出力する。

20

【 0 1 0 5 】

以上のように、本発明の第 2 の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、記憶部構築系 1 において、手形状画像情報記憶部 1 2 B に記憶した複数の手形状画像を固有空間内におけるクラスタ分析によりグループ化し、姿勢・形状認識系 2 では、入力手画像を認識するときに、まず、最初にどのグループに属するかを求め、そして、そのグループ内でどの手形状画像に最も近いかを求めることにより、画像比較回数を減らすことができ、さらに高速に処理を行うことが可能となる。

30

また、固有空間内でのグループ分けのときに、手形状ごと若しくは手姿勢ごとといったグループ分けでなく、固有空間内で近い画像、すなわち似ている画像が同一グループとなるようにクラスタリングするので、異なる形状だが似た画像が存在する場合でも、正確にその手形状と手姿勢を求めることが可能になる。

【 0 1 0 6 】

なお、上記第 2 の実施形態では、様々な手形状と手姿勢とを示す複数の手形状画像として、実際の手画像を格納していることを仮定していたが、上記第 1 の実施形態と同様に、予め CAD や CG などで作られる三次元モデルとして手のモデルを用意し、そのモデルの複数視点の投影像を格納することも考えられる。その場合、精度良く、投影像を取得したときのモデルの手姿勢を定義することができる。また、マネキン等で使われているような実際の手のモデル等を用いても、同等に定義することが可能である。

40

また、上記第 2 の実施形態は、手形状と手姿勢を 1 つだけ出力することを基本にしていたが、画像の解像度などの関係で、区別がつかないほど似ている場合等には、1 つに絞れない場合があり得る。その場合は、複数の手形状と手姿勢の候補を出力することも考えられる。その場合も、上記第 2 の実施形態と全く同じ構成および方法を用いることにより実現が可能である。さらに、上記第 2 の実施形態は、似た画像を分類するために画像比較部 2 6 を用いたが、状況により手形状まで出力すれば十分な場合もあり得る。その場合は、クラスタに属する手形状ごとに平均画像や、分散画像などの統計量に基づく手形状画像を求め、それらの画像と入力手形状画像とを比較することにより、手形状だけを求めることも

50

可能である。また、第 2 の実施形態に係る手の形状と姿勢の認識装置をハードウェアで実現する場合には、上記図 6 で示したものと同等の構成を用いればよい。

【0107】

また、上記第 2 の実施形態に係る手の形状と姿勢の認識装置の画像比較部 26 を、最尤クラスタ判別部 25 が求めたクラスタに属する手形状画像情報記憶部 12B に記憶した手形状画像と、手画像正規化部 21 が生成した入力手形状画像とを比較する際に、手形状画像を同一クラスタ内の同一手形状ごとにグループ化する同一形状分類部と、分類した各グループを表現する統計量を求める形状グループ統計量計算部と、入力手形状画像と形状グループ統計量計算部が求めた統計量との距離を計算し、最も近いグループに属する手形状を出力する最尤形状判別部とからなる構成に置き換えてもよい。このようにすれば、さらに画像比較回数を減らすことができ、より高速に処理を行うことが可能となる。

10

【0108】

(第 3 の実施形態)

さて、上記第 2 の実施形態に述べたように、分析した各クラスタには、手形状や手姿勢によって分類された画像ではなく、似た画像が同一クラスタに分類されている。従って、例えば、図 13 に示すような人差指と中指の 2 本の指を並べて立てる手画像と、人差指と中指の 2 本の指を重ねて立てる手画像とは、同一クラスタに分類される。このような手形状の違いは、例えば、現実的に手話の指文字を区別する場合に存在する。これらの手形状を判別する場合、上記第 2 の実施形態で述べたように画像の全体で違いを判別するのではなく、異なっている部分だけを抽出して判別する必要がある。

20

【0109】

そこで、本発明の第 3 の実施形態は、上記第 2 の実施形態における画像比較部 26 が、入力手形状画像と手形状画像情報記憶部 12B に格納されている手形状画像とをパターンマッチングにより直接全体比較するのではなく、予め各クラスタにおける判別枠を求めておき、この判別枠内で手形状を判別する手法を提供するものである。

【0110】

図 12 は、本発明の第 3 の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。図 12 において、第 3 の実施形態に係る手の形状と姿勢の認識装置は、上記第 2 の実施形態に係る手の形状と姿勢の認識装置と同様、記憶部構築系 1 と、姿勢・形状認識系 2 とで構成される。

30

図 12 において、記憶部構築系 1 は、手画像正規化部 11 と、手形状画像情報記憶部 12B と、固有空間計算部 13 と、固有ベクトル記憶部 14 と、固有空間投影部 15 と、クラスタ分析 / 判別部 18 と、クラスタ情報記憶部 17B とを備える。姿勢・形状認識系 2 は、手画像正規化部 21 と、固有空間投影部 22 と、最尤クラスタ判別部 25 と、画像比較部 27 と、形状・姿勢出力部 24 とを備える。

【0111】

図 12 に示すように、第 3 の実施形態に係る手の形状と姿勢の認識装置は、上記第 2 の実施形態に係る手の形状と姿勢の認識装置に対し、記憶部構築系 1 において、クラスタ分析部 16 をクラスタ分析 / 判別部 18 に代え、クラスタ情報記憶部 17A をクラスタ情報記憶部 17B に代え、姿勢・形状認識系 2 において、画像比較部 26 を画像比較部 27 に代えた構成である。

40

なお、第 3 の実施形態に係る手の形状と姿勢の認識装置におけるその他の構成は、上記第 2 の実施形態に係る手の形状と姿勢の認識装置の構成と同様であり、当該構成については、同一の参照番号を付してその説明を省略する。

【0112】

以下、第 3 の実施形態における記憶部構築系 1 および姿勢・形状認識系 2 の各構成および処理動作を、図 12 および図 14 を参照して、上記第 2 の実施形態と異なる部分を中心に説明する。図 14 は、図 12 のクラスタ分析 / 判別部 18 が行う形状判別枠位置の算出方法の一例を示す図である。

【0113】

50

クラスタ分析／枠判別部 18 は、まず、手形状画像情報記憶部 12B に格納された固有空間投影座標についてクラスタ分析を行い、手形状画像が近いもの同士を同一グループに分類されるようにグループ分けを行う。この処理は、上記第 1 の実施形態で述べたクラスタ分析部 16 と同じである。

次に、クラスタ分析／枠判別部 18 は、各クラスタに対して、形状判別枠の位置を算出する。図 14 を参照して、クラスタ分析／枠判別部 18 は、まず、1つのクラスタ内に存在する同一手形状の手形状画像を複数抽出して平均化し、それぞれの手形状の平均画像を求める。次に、クラスタ分析／枠判別部 18 は、予め定めた一定の枠（枠の形状は、任意に定めることができる。なお、図 14 では、方形枠を用いている）を用い、それぞれの平均画像上で枠を移動させながら、枠内における双方の平均画像間の差を順次求め、最も差が大きい位置を形状判別枠の位置とする。そして、クラスタ分析／枠判別部 18 は、この求めた形状判別枠の位置をクラスタ情報記憶部 17B に記憶する。

10

【0114】

画像比較部 27 は、まず、手形状画像情報記憶部 12B に格納されているクラスタ ID を参照して、最尤クラスタ判別部 25 が求めたクラスタに属する手形状画像のみと、手画像正規化部 21 が生成した入力手形状画像とを取得する。一方、画像比較部 27 は、クラスタ情報記憶部 17B から、最尤クラスタ判別部 25 が求めたクラスタに対応する形状判別枠の位置を取得する。そして、画像比較部 27 は、形状判別枠の位置内のみにおいて、取得した手形状画像と入力手形状画像とを比較し、入力手形状画像に最も近い手形状画像を求める。

20

【0115】

以上のように、本発明の第 3 の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、形状判別枠位置を予め定めておき、この形状判別枠内のみで手形状画像と入力手形状画像とを比較する。これにより、上記第 2 の実施形態に比べ、画像比較回数を減らすことができ、さらに高速かつ正確に処理を行うことが可能となる。

【0116】

（第 4 の実施形態）

本発明の第 4 の実施形態は、上記第 2 の実施形態において、最尤クラスタ判別部 25 が求めたクラスタから、手形状と手姿勢を求めるときに、画像比較部 26 によって、手形状画像情報記憶部 12B に記憶している手形状画像と、入力手形状画像とを直接比較する代わりに、ある手形状と手姿勢とを示す手を複数のカメラを用いて複数の視点から撮影し、それぞれのカメラで撮影した手画像から最尤クラスタ判別部 25 が求めたクラスタの中の形状情報を統合することにより、手形状と手姿勢とを求める手法を提供するものである。

30

【0117】

なお、本発明の第 4 の実施形態に係る手の形状と姿勢の認識装置の構成は、上記第 2 の実施形態に係る手の形状と姿勢の認識装置の構成と同様であるため図面を省略する。また、第 4 の実施形態における姿勢・形状認識系 2 の各構成および処理動作を、図 8 および図 15 を参照して、上記第 2 の実施形態と異なる部分を中心に説明する。図 15 は、本発明の第 4 の実施形態に係る手の形状と姿勢の認識装置において、複数台のカメラの入力手画像から手形状画像を求める概念の一例を説明する図である。なお、図 15 においては、3台

40

【0118】

まず、前提として、図 15 に示すように、ある手形状と手姿勢とを示す手を、3台のカメラを用いて異なる 3 方向から撮影し、3つの入力手画像を得る。この 3 方向から撮影した 3 つの入力手画像は、手画像正規化部 21，固有空間投影部 22 および最尤クラスタ判別部 25 によって各々処理され、対応する最も近いクラスタがそれぞれ求められる。これに対し、画像比較部 26 は、3台のカメラの入力手画像から得られた 3 つのクラスタと、当該 3 つのクラスタに属する 3 つの手形状画像の形状情報および姿勢情報とから、以下の条件（1），（2）に従い、3つの入力手画像について最も近い手形状画像を求める。

（1）同一手形状であること

50

(2) カメラの位置関係と姿勢とが矛盾しないこと

【 0 1 1 9 】

すなわち、画像比較部 2 6 は、まず上記条件 (1) に従って、3つのクラスタに属する手形状の中で、全てのクラスタに属する手形状 (図 1 5 に示す例では、伸展 1 本指となる) を抽出する。次に、画像比較部 2 6 は、上記条件 (2) に従って、それぞれ抽出した手形状に対応する手姿勢から、各カメラの位置関係に従って矛盾しない手形状画像を導き出す (統合する) 。図 1 5 に示す例では、第 1 カメラで手の甲側の画像を選択した場合、第 2 カメラでは掌が下向きの画像、第 3 カメラでは手が手前を向いている画像を選択すれば矛盾しなくなる。

上記の処理を行うことで、それぞれのカメラからの入力手画像に従い、最も条件に合う手形状画像が選択され、認識対象である入力手画像に対する手形状と手姿勢とを導くことができる。

10

【 0 1 2 0 】

以上のように、本発明の第 4 の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、複数カメラからの入力手画像から求めたクラスタに基づき、各クラスタに属する手形状画像の形状情報と姿勢情報とをカメラの位置関係を用いて統合して、入力手画像の手形状および手姿勢を求める。これにより、1方向からの画像だけでは、手形状および手姿勢が決定できないような場合 (例えば、横方向の手の画像など) でも、正確に手形状および手姿勢を求めることが可能となる。

【 0 1 2 1 】

20

なお、上記第 4 の実施形態では、各カメラの画像から求めたクラスタにおいて、全てが矛盾しないように統合するように説明したが、各カメラからの結果に従い、多数決等によって一部のカメラのクラスタ選択することで、最も可能性が高い手形状および手姿勢を出力することも可能である。また、上記第 4 の実施形態では、3台のカメラを用いた場合を一例に説明したが、他の複数台のカメラを用いた場合であっても、上記と同様に実施することができる。

【 0 1 2 2 】

(第 5 の実施形態)

上記第 2 の実施形態では、認識対象の手画像が静止画像 (例えば、人差指を 1 本だけ伸ばして数字「1」を表現するだけの場合) であることを前提に、入力手画像に対応する手形状および手姿勢を出力する手の形状と姿勢の認識装置について述べた。しかし、ジェスチャーや手話等で行われる手振り動作には、動画像としての一連の動作を完了することで 1 つの意味を表す場合がある (例えば、人に行き先を教えるとき等によく用いられる人差指を 1 本だけ伸ばして指差す方向を変化させるような場合である) 。このような動画像である手振り動作に関しては、上記第 2 の実施形態に係る手の形状と姿勢の認識装置では、手振り動作の意味を求めることができない。

30

【 0 1 2 3 】

そこで、本発明の第 5 の実施形態に係る手の形状と姿勢の認識装置は、認識対象の手画像が一連の意味ある手振り動作を行う手を撮影した動画像 (以下、手振り動作画像という) である場合に対応するものであり、様々な手振り動作について各々特徴点を抽出してその意味と共に予め記憶し、入力する手振り動作画像の特徴点と記憶する特徴点とを比較することにより、手振り動作の意味を求める手法を提供するものである。

40

以下、第 5 の実施形態においては、手振りを行う人物の上半身若しくは全身が撮影されている手振り動作画像が入力される場合を想定して説明する。なお、人物を撮影する方向としては、正面、斜め上、横等の様々な方向が考えられるが、第 5 の実施形態では、そのいずれの方向から撮影した画像に対しても有用な効果を奏することが可能である。

【 0 1 2 4 】

図 1 6 は、本発明の第 5 の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。図 1 6 において、第 5 の実施形態に係る手の形状と姿勢の認識装置は、上記第 2 の実施形態に係る手の形状と姿勢の認識装置と同様、記憶部構築系 1 と、姿勢・形状

50

認識系 2 とで構成される。

図 16 において、記憶部構築系 1 は、手画像正規化部 11 と、固有ベクトル記憶部 14 と、固有空間計算部 13 と、手形状画像情報記憶部 12B と、固有空間投影部 15 と、クラスタ情報記憶部 17A と、クラスタ分析部 16 とを備える。また、姿勢・形状認識系 2 は、手領域検出部 28 と、手動作分節部 29 と、手画像切出部 30 と、手画像正規化部 21 と、固有空間投影部 22 と、最尤クラスタ判別部 25 と、識別演算部 33A と、系列登録部 31 と、系列識別辞書 32 と、データ経路制御部 34A とを備える。

【0125】

図 16 に示すように、第 5 の実施形態に係る手の形状と姿勢の認識装置は、上記第 2 の実施形態に係る手の形状と姿勢の認識装置に対し、姿勢・形状認識系 2 において、手領域検出部 28、手動作分節部 29 および手画像切出部 30 を手画像正規化部 21 の前段に加え、画像比較部 26 を系列登録部 31、系列識別辞書 32、識別演算部 33A およびデータ経路制御部 34A に代えた構成である。

なお、第 5 の実施形態に係る手の形状と姿勢の認識装置におけるその他の構成は、上記第 2 の実施形態に係る手の形状と姿勢の認識装置の構成と同様であり、当該構成については、同一の参照番号を付してその説明を省略する。

ここで、第 5 の実施形態で言う記憶部構築系 1 とは、上記第 2 の実施形態に係る手の形状と姿勢の認識装置に関する記憶部構築系 1 を意味し、系列識別辞書 32 が記憶部構築系 1 側に含まれない形で構成される。しかし、第 5 の実施形態において用いる「記憶部構築系 1」および「姿勢・形状認識系 2」という系の名称は、あくまでも上記第 2 の実施形態との関連性を示すためにのみ存在するものであり、実際の内部処理、例えば姿勢・形状認識系 2 において辞書（系列識別辞書 32）を作成することについて拘束するものではないことをここに明記する。

【0126】

まず、第 5 の実施形態における姿勢・形状認識系 2 の各構成を、上記第 2 の実施形態と異なる部分を中心に説明する。

手領域検出部 28 は、手振り動作画像を入力し、各々の画像から手領域をそれぞれ検出する。手動作分節部 29 は、手振り動作画像から手形状と手姿勢の変化点を求め、変化点を含む 1 つまたは 2 つ以上の画像のみで構成される手振り動作画像系列を作成する。手画像切出部 30 は、手動作分節部 29 が作成した手振り動作画像系列から手が含まれる周辺領域をそれぞれ切り出して手画像系列を作成し、手画像正規化部 21 に出力する。系列登録部 31 は、手振り動作画像（手振り動作画像系列）を登録する場合において、最尤クラスタ判別部 25 が出力する手画像系列に対応するクラスタ系列を、その手振り動作画像の意味と共に系列識別辞書 32 に登録する。系列識別辞書 32 は、系列登録部 31 が出力するクラスタ系列を、対応して与えられる手振り動作画像の意味と共に格納する。識別演算部 33A は、手振り動作画像を認識する場合において、最尤クラスタ判別部 25 が出力するクラスタ系列と、系列識別辞書 32 に登録されているクラスタ系列とを比較することで、手振り動作画像の意味を認識する。データ経路制御部 34A は、最尤クラスタ判別部 25 から出力されるクラスタ系列が、登録する場合には系列登録部 31 へ、認識する場合には識別演算部 33A へ出力されるように制御する。

【0127】

次に、図 17 ~ 図 20 を用いて、第 5 の実施形態に係る手の形状と姿勢の認識装置が行う認識方法を、処理の順に説明する。図 17 は、図 16 の手領域検出部 28、手動作分節部 29 および手画像切出部 30 が行う処理の概念を示す図である。図 18 は、図 16 の手画像系列および当該手画像系列から求められるクラスタ系列の一例を示す図である。図 19 および図 20 は、図 16 の系列識別辞書 32 が有する記憶形態の一例を示す図である。なお、図 19 においては、単純なテーブル形式における記憶形態の例を示し、図 20 においては、隠れマルコフモデルに基づいた記憶形態の例を示している。

【0128】

第 5 の実施形態において、記憶部構築系 1 は、上記第 2 の実施形態に係る手の姿勢と形状

の認識装置と同一の処理を行うため、ここでの説明を省略する。

姿勢・形状認識系 2 は、以下に述べる 2 つのモードの処理を行う。

1．登録モード（第 1 の登録モード）

入力される手振り動作画像から得られるクラスタ系列を、その意味と共に系列識別辞書 3 2 へ登録するモードである。

2．認識モード

入力される手振り動作画像から得られるクラスタ系列に基づいて、その意味を認識するモードである。この認識モードは、上記第 2 の実施形態で行う姿勢・形状認識に該当する処理であり、固有ベクトル記憶部 1 4 およびクラスタ情報記憶部 1 7 A および系列識別辞書 3 2 を用いて、手振り動作の意味を認識する。

これらの各モードは、データ経路制御部 3 4 A に対して、どちらのモードを選択するかを入力することにより切り替えることができる。以下、それぞれのモードに従い、順を追って説明する。

【0129】

最初に、各モードにおいて共通に行われる手領域検出部 2 8、手動作分節部 2 9、手画像切出部 3 0 および手画像正規化部 2 1 の動作について説明する。

複数の画像から構成される手振り動作画像（図 1 7（a））は、手領域検出部 2 8 に入力される。手領域検出部 2 8 は、入力される手振り動作画像に対して、画像中で手が存在する領域（手領域）をそれぞれ検出する。ここでは、撮影される手振り動作画像が、背景から手領域を分離しやすい画像であると仮定し、単純に画像を二値化し、手の領域に近い面積を持つ領域を手領域として検出する。

【0130】

手動作分節部 2 9 は、手領域検出部 2 8 が出力する手振り動作画像に対して、手形状と手姿勢に関してキーとなる画像（以下、キーフレームと呼ぶ）を求める。ここでいうキーフレームとは、手形状と手姿勢とを人間が認識できる画像を指す。一般に手振り動作の場合、手が動いている間は、残像などの影響で人間は手形状と手姿勢とを認識していない。そこで、手動作分節部 2 9 では、相対的に手の動きが小さな画像（フレーム）を求め、その画像をキーフレームとする。手動作分節部 2 9 が求めた 1 つまたは 2 つ以上のキーフレームは、手振り動作画像系列（図 1 7（b））として、手画像切出部 3 0 に出力される。

なお、上記相対的な手の動きを求める手法としては、例えば、手領域検出部 2 8 によって得られた手領域の手振り動作画像中における変位量や手領域内部の変動を求める手法や、手領域から手振り動作画像中における手の位置を追跡して手の軌跡から手が相対的に止まっている点（これには手の動作軌跡における曲率が相対的に大きなフレームも含まれる）を求める手法や、手振り動作画像から時間差分画像を求め、その時間差分画像の情報から手が相対的に止まっている点を求める手法等が考えられる。また、手振り動作画像の全ての画像をキーフレームとする場合もある。

【0131】

手画像切出部 3 0 は、手動作分節部 2 9 が求めた手振り動作画像系列の各キーフレームから、手領域検出部 2 8 が求めた手領域部分をそれぞれ切り出して、手の部分が含まれる手画像系列（図 1 7（c）、図 1 8（a））を作成する。この手画像系列を構成する個々の手画像は、上記第 2 の実施形態において入力される手画像と同等の画像である。この手画像切出部 3 0 が作成した手画像系列は、手画像正規化部 2 1 へ出力される。

【0132】

そして、手画像正規化部 2 1、固有空間投影部 2 2 および最尤クラスタ判別部 2 5 は、手画像系列を構成する各キーフレームに対し、上記第 2 の実施形態で述べた各処理を行ってキーフレームに対応する最も近いクラスタをそれぞれ求め、クラスタ系列（図 1 8（b））として出力する。

以上の処理が、各モードの前処理部として共通に行われ、手振り動作画像から対応するクラスタ系列が求められる。

【0133】

次に、各モード個別の処理について説明する。

まず、登録モードにおける処理について説明する。

この登録モードでは、最尤クラスタ判別部 25 が出力するクラスタ系列が、手振り動作を特徴づける系列であると定義して、手振り動作が示す意味とともに系列識別辞書 32 に登録（格納）する処理を行う。

【0134】

登録モードでは、データ経路制御部 34A は、最尤クラスタ判別部 25 が出力するクラスタ系列が系列登録部 31 に入力されるように、経路を切り換える。

系列登録部 31 は、最尤クラスタ判別部 25 から入力するクラスタ系列を、別途与えられる当該クラスタ系列に対応する手振り動作の意味と共に、系列識別辞書 32 に登録する。系列識別辞書 32 にデータを登録する場合の記憶形式としては幾つかの手法が存在するが、図 19 および図 20 の場合の 2 通りを一例に挙げて説明する。

図 19 は、最尤クラスタ判別部 25 で得られたクラスタ系列に対し、そのクラスタ系列を意味と共にそのまま登録した例である。なお、図 19 のように、1 つの意味に対して複数のクラスタ系列が存在するのは、同じ意味の手振り動作でも実行する人間によって速度や形状等が微妙に異なっているためであり、同じ意味の手振り動作に対して登録処理を複数回行うことで作成される。

図 20 は、状態遷移モデルの一例として隠れマルコフモデル（HMM）の形で登録した例である。この隠れマルコフモデルとは、音声認識の分野等で知られている技術であり、図 19 に示したような 1 つの意味に対して複数存在するクラスタ系列を、1 つの状態遷移モデルに統合した形で表すものである。隠れマルコフモデルに関する詳細な技術内容は、例えば、技術文献“中川著「確立モデルによる音声認識」コロナ社、電子情報通信学会編”に記載されており、図 20 はこの文献に準じて書き表したものである。なお、図 20 中、スカラ値は S1 ~ S3 への状態遷移確率を、ベクトル値はクラスタ 1 ~ 5 の状態遷移による条件付きの出力確率を示している。

【0135】

なお、系列識別辞書 32 の構築方法として、画像から得られた手形状と手姿勢とをそのまま登録する場合が一般的である。しかしながら、その場合、上記第 2 の実施形態で述べたように、手の画像には「手形状は違うが似た画像」、「手姿勢は違うが似た画像」があるため、上記第 3 および第 4 の実施形態のように画像を比較したり、複数のカメラからの画像を用いなければ、誤認識が生じやすい。そこで、第 5 の実施形態では、そのような手法とは違い、見た目が近い画像を同一クラスタとしたクラスタ系列を系列識別辞書 32 に登録することにより、誤認識がより少ない形で認識できるようにしている。

【0136】

次に、認識モードにおける処理について説明する。

この認識モードでは、入力された手振り動作画像に対して、系列識別辞書 32 を用いて実際にその意味を求める処理を行う。

認識モードでは、データ経路制御部 34A は、最尤クラスタ判別部 25 が出力するクラスタ系列が識別演算部 33A に入力されるように、経路を切り換える。

識別演算部 33A は、最尤クラスタ判別部 25 から入力するクラスタ系列と、系列識別辞書 32 に登録されている複数のクラスタ系列とを比較し、同一または最も近いクラスタ系列を判断する。そして、識別演算部 33A は、同一または最も近いと判断したクラスタ系列の意味を系列識別辞書 32 から抽出して、出力する。

【0137】

以上のように、本発明の第 5 の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、上記第 2 の実施形態と同様のクラスタ情報を用いるにあたり、ジェスチャー単語や手話単語等のような一連の意味ある手振り動作画像に対して、手動作の分節点となる画像のクラスタ系列をその意味と共に予め記憶し、手振り動作画像を認識するときに、求めたクラスタ系列に基づいて記憶している意味を出力する。

これにより、ジェスチャー単語や手話単語等のような一連の意味ある動作に対して、より

誤認識を減らし、正確に意味を求めることが可能となる。

【0138】

なお、上記第5の実施形態では、キーフレームにおける手画像を認識する方法を記載した。しかし、本発明の第5の実施形態は、この他にも、全てのフレームをキーフレームにした場合、一定間隔にサンプリングしたフレームをキーフレームにした場合、手振り動作の開始時と終了時のフレームのみをキーフレームにした場合等においても、上記処理を行うことで同様の効果を奏することができる。

【0139】

(第6の実施形態)

本発明の第6の実施形態は、上記第5の実施形態の記憶部構築系1において、手形状画像情報記憶部12Bに、予め様々な姿勢と形状の手画像を記憶する代わりに、手振り動作画像から得られる手画像系列の各画像とその意味とを記憶するようにしたものである。

10

【0140】

図21は、本発明の第6の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。図21において、第6の実施形態に係る手の形状と姿勢の認識装置は、上記第5の実施形態に係る手の形状と姿勢の認識装置のように記憶部構築系1と姿勢・形状認識系2とを区別することなく一つの統合された形で構成される。

図21において、第6の実施形態に係る手の形状と姿勢の認識装置は、手領域検出部28と、手動作分節部29と、手画像切出部30と、手画像正規化部21と、固有空間投影部22と、最尤クラスタ判別部25と、識別演算部33Aと、系列識別辞書32と、データ経路制御部34Bと、手画像登録部35と、系列再構成部36と、固有空間計算部13と、固有ベクトル記憶部14と、手形状画像情報記憶部12Cと、クラスタ分析部16と、クラスタ情報記憶部17Aとを備える。

20

【0141】

図21に示すように、第6の実施形態に係る手の形状と姿勢の認識装置は、上記第5の実施形態に係る手の形状と姿勢の認識装置に対し、記憶部構築系1と姿勢・形状認識系2を統合するため、図16において、手画像正規化部11と手画像正規化部21、固有空間投影部15と固有空間投影部22をそれぞれ統合し、手形状画像情報記憶部12Bを手形状画像情報記憶部12Cに、データ経路制御部34Aをデータ経路制御部34Bに、さらに系列登録部31を手画像登録部35および系列再構成部36にそれぞれ代えた構成である。

30

なお、第6の実施形態に係る手の形状と姿勢の認識装置におけるその他の構成は、上記第5の実施形態に係る手の形状と姿勢の認識装置の構成と同様であり、当該構成については、同一の参照番号を付してその説明を省略する。

【0142】

まず、第6の実施形態に係る手の形状と姿勢の認識装置の各構成を、上記第5の実施形態と異なる部分を中心に説明する。

手画像登録部35は、手画像正規化部21から与えられる手振り動作画像に対応する手画像系列を、その系列の意味と共に手形状画像情報記憶部12Cに登録する。手形状画像情報記憶部12Cは、登録する手振り動作画像に対応する手形状画像系列(手画像系列)を、その系列の意味と共にそれぞれ格納する。また、手形状画像情報記憶部12Cは、上記第5の実施形態における手形状画像情報記憶部12Bと同様、各々の手形状画像を固有空間へ投影した投影座標およびクラスタIDをそれぞれ格納する。系列再構成部36は、手形状画像情報記憶部12Cに格納した情報に基づいて、格納したそれぞれの手形状画像系列に対応するクラスタ系列とその意味を系列識別辞書32に登録する。データ経路制御部34Bは、手画像正規化部21から出力される手画像系列が、登録する場合には手画像登録部35へ、認識する場合には固有空間投影部22へ出力されるように制御する。

40

【0143】

次に、図22を用いて、第6の実施形態に係る手の形状と姿勢の認識装置が行う認識方法を、処理の順に説明する。図22は、図21の手形状画像情報記憶部12Cが有する格納

50

テーブルの一例を示す図である。

【 0 1 4 4 】

第 6 の実施形態に係る手の姿勢と形状の認識装置は、以下に述べる 2 つのモードの処理を行う。

1 . 登録モード (第 2 の登録モード)

入力される手振り動作画像から得られるクラスタ系列を、その意味と共に系列識別辞書 3 2 へ登録するモードである。この登録モードは、手形状画像情報記憶部 1 2 C、固有ベクトル記憶部 1 4 およびクラスタ情報記憶部 1 7 A を構築するモードであり、上記第 2 の実施形態に係る記憶部構築系 1 に該当する処理である。すなわち、入力される手振り動作画像から得られる手画像系列 (手形状画像系列) をその意味と共に手形状画像情報記憶部 1 2 C に格納し、格納した手形状画像をもとに固有空間の計算およびクラスタ分析を行う。そして、求めたクラスタ系列とその意味とを系列識別辞書 3 2 へ登録する。

10

2 . 認識モード : 手振り認識

入力される手振り動作画像から得られるクラスタ系列に基づいて、その意味を認識するモードである。この認識モードは、上記第 5 の実施形態で説明した認識モードと同様、上記第 2 の実施形態で行う姿勢・形状認識に該当する処理であり、固有ベクトル記憶部 1 4 およびクラスタ情報記憶部 1 7 A および系列識別辞書 3 2 を用いて、手振り動作の意味を認識する。

これらの各モードは、データ経路制御部 3 4 B に対して、どちらのモードを選択するかを入力することにより切り替えることができる。以下、それぞれのモードに従い、順を追って説明する。

20

【 0 1 4 5 】

まず、登録モードにおける処理について説明する。

上述したように、手領域検出部 2 8、手動作分節部 2 9、手画像切出部 3 0 および手画像正規化部 2 1 が、上記第 5 の実施形態と同等の処理で入力する手振り動作画像に対応する手画像系列を求める。一方、データ経路制御部 3 4 B は、手画像正規化部 2 1 が出力する手画像系列が手画像登録部 3 5 に入力されるように、経路を切り換える。

次に、手画像登録部 3 5 は、手画像正規化部 2 1 から入力する手画像系列を、別途与えられる当該手画像系列に対応する手振り動作の意味と共に、手形状画像情報記憶部 1 2 C に格納する。図 2 2 に、手形状画像情報記憶部 1 2 C が有する格納テーブルの一例を示す。図 2 2 に示すように、手形状画像情報記憶部 1 2 C には、上記第 2 の実施形態における手形状画像情報記憶部 1 2 B の形状情報および姿勢情報の代わりに、手画像系列の番号、手画像系列に対応する手振り動作画像の意味、系列中の手形状画像が何番目に位置する画像であるかという情報 (ステップ) がそれぞれ格納される。なお、両手が接触しているような画像の場合、両手が接触している画像を 1 つの手形状画像として登録する。

30

【 0 1 4 6 】

固有空間計算部 1 3、固有ベクトル記憶部 1 4、固有空間投影部 2 2 およびクラスタ分析部 1 6 は、手形状画像情報記憶部 1 2 C に格納した各々の手形状画像について上記第 2 の実施形態で述べた処理を行い、固有ベクトル記憶部 1 4 およびクラスタ情報記憶部 1 7 A に対応する情報を格納すると共に、手形状画像情報記憶部 1 2 C に固有空間投影座標およびクラスタ ID を格納する。

40

系列再構成部 3 6 は、手形状画像情報記憶部 1 2 C への情報格納がされると、格納されている手画像系列毎にクラスタ系列とその意味を、系列識別辞書 3 2 に登録する。

【 0 1 4 7 】

次に、認識モードにおける処理について説明する。

認識モードでは、識別演算部 3 3 A は、最尤クラスタ判別部 2 5 から入力するクラスタ系列と、系列識別辞書 3 2 に登録されている複数のクラスタ系列とを比較し、同一または最も近いクラスタ系列を判断する。そして、識別演算部 3 3 A は、同一または最も近いと判断したクラスタ系列の意味を系列識別辞書 3 2 から抽出して、出力する。

【 0 1 4 8 】

50

以上のように、本発明の第 6 の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、手形状画像情報記憶部 1 2 C に記憶する画像を実際の認識時と同じ画像を使うことができるため、別の画像をわざわざ取得する必要もなく、また、同一の環境で画像を取得することが保証されるため、画像の誤認識を減らすことができる。

【 0 1 4 9 】

なお、上記第 6 の実施形態に係る手の形状と姿勢の認識装置に、上記第 5 の実施形態で述べた系列登録部 3 1 およびデータ経路制御部 3 4 A の構成をさらに加え、系列識別辞書 3 2 へのクラスタ系列とその意味の登録を、第 1 の登録モードまたは第 2 の登録モードのいずれを用いてもできるようにしてもよい。

このような構成にすることで、手形状画像情報記憶部 1 2 C が固定的なデータベースとして用いられているような場合であっても、第 1 の登録モードによって新たな手振り動作画像に関するデータ登録（系列識別辞書 3 2 の更新）を行うことが可能となる。

【 0 1 5 0 】

（第 7 の実施形態）

本発明の第 7 の実施形態は、上記第 5 または第 6 の実施形態において、認識対象の手画像がジェスチャーや手話などの一連の意味ある動作を行う手を撮影した手画像である場合、第 5 または第 6 の実施形態による認識装置をジェスチャーや手話を認識するための装置の 1 モジュールとして用いることにより、動作の意味を求める手法を提供するものである。

【 0 1 5 1 】

例えば、手話の認識のために本発明を用いる場合を考える。手話の場合、手の形状の他に手の空間位置、手の動き、手形状、手姿勢といったいくつかの構成要素から、その意味が成り立っているといわれている。さらに、手形状では、手話単語の開始時の形状と終了時の形状（右手のみ、左手のみ、左手右手両方）とを構成要素として挙げることができる。図 2 3 に、幾つかの手話単語を構成要素で記述した一例を示す。図 2 3 において、「言う」という意味の手話単語の場合、右手を用い、人差指を伸ばした手形状を、まず口元若しくは口の前に持っていき、その後その手形状のまま前方に出すという動作で表現される。一方、「好き」という意味の手話単語の場合、右手を用い、まず親指と人差指を伸ばした手形状をあごの部位に持っていき、次に指を閉じながら下方に引くという動作で表現される。

【 0 1 5 2 】

そこで、第 7 の実施形態に係る手の形状と姿勢の認識装置では、手話やジェスチャーといった動作に対し、手の空間位置や動きといった手の大局的動作に関する特徴による認識部の制約条件を加えることによって、手画像の誤認識を減らすものである。

【 0 1 5 3 】

図 2 4 は、本発明の第 7 の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。図 2 4 において、第 7 の実施形態に係る手の形状と姿勢の認識装置は、手画像登録部 3 5 と、固有ベクトル記憶部 1 4 と、固有空間計算部 1 3 と、手形状画像情報記憶部 1 2 C と、クラスタ情報記憶部 1 7 A と、クラスタ分析部 1 6 と、系列再構成部 3 6 と、手領域検出部 2 8 と、手動作分節部 2 9 と、手画像切出部 3 0 と、手画像正規化部 2 1 と、固有空間投影部 2 2 と、最尤クラスタ判別部 2 5 と、識別演算部 3 3 B と、系列識別辞書 3 2 と、データ経路制御部 3 4 B と、大局動作認識部 3 7 と、制約条件記憶部 3 8 とを備える。

図 2 4 に示す第 7 の実施形態に係る手の形状と姿勢の認識装置は、上記第 6 の実施形態に係る手の形状と姿勢の認識装置に、大局動作認識部 3 7 および制約条件記憶部 3 8 を加え、識別演算部 3 3 A を識別演算部 3 3 B に代えた構成である。なお、第 7 の実施形態に係る手の形状と姿勢の認識装置におけるその他の構成は、上記第 6 の実施形態に係る手の形状と姿勢の認識装置の構成と同様であり、当該構成については、同一の参照番号を付してその説明を省略する。

【 0 1 5 4 】

まず、制約条件記憶部 3 8 には、手話単語のような意味ある動作に基づいて手形状と手姿

10

20

30

40

50

勢とを制約する制約条件が、予め記憶されている。この制約条件とは、例えば、図 23 に示す手話単語「言う」の場合、開始の手形状と終了の手形状との双方共、人差指を伸ばした形状に、また、手姿勢、位置および動作を上述した内容のように記憶する。なお、手話単語「言う」の動作は、右手のみで行われるため、図 23 の例では左手に関する条件が記載されていない。

【0155】

手振り動作画像は、大局動作認識部 37 および手領域検出部 28 にそれぞれ入力される。大局動作認識部 37 は、入力する手振り動作画像に対して、手領域検出部 28 と同様に手領域を抽出し、当該手領域における手の軌跡と身体に対する手の位置とを求め、手の軌跡と手の位置の情報を識別演算部 33B へ出力する。この大局動作認識部 37 は、例えば、
10 本願発明者等が先に出願した「手動作認識装置」（特開平 11 - 174948 号）に記載された手法に従って、手の軌跡と手の位置とを求める。

一方、手領域検出部 28 に入力される手振り動作画像に対しては、上記第 6 の実施形態で述べた処理が手動作分節部 29、手画像切出部 30、手画像正規化部 21、固有空間投影部 22 および最尤クラスタ判別部 25 において各々行われ、手振り動作画像に対応するクラスタ系列が最尤クラスタ判別部 25 から識別演算部 33B へ出力される。

【0156】

識別演算部 33B は、まず、制約条件記憶部 38 に記憶されたデータを検索して、大局動作認識部 37 から与えられる手振り認識結果（手の軌跡と手の位置の情報）と同一の動作データの手話 / ジェスチャー単語を、1 つ以上抽出する。次に、識別演算部 33B は、最
20 尤クラスタ判別部 25 から入力するクラスタ系列と、系列識別辞書 32 に登録されている複数のクラスタ系列とを比較し、同一または最も近いクラスタ系列を判断して、当該判断したクラスタ系列の意味を系列識別辞書 32 から 1 つ以上抽出する。そして、識別演算部 33B は、1 つ以上抽出した手話 / ジェスチャー単語と、1 つ以上抽出した意味とに基づいて、入力した手振り動作画像について最も近い意味を出力する。

【0157】

以上のように、本発明の第 7 の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、手の大局的動作の特徴に基づく制約条件をさらに加えて、手振り動作画像の意味を導き出す。

これにより、手振り動作画像の誤認識を減らすことができる。

【0158】

なお、上記第 7 の実施形態では、大局動作認識部 37、制約条件記憶部 38 および識別演算部 33B を、上記第 6 の実施形態に係る手の形状と姿勢の認識装置に対して構成した場合を説明したが、上記第 5 の実施形態に係る手の形状と姿勢の認識装置に対して構成してもよいし、上記第 6 の実施形態のなお書きで述べたように第 5 の実施形態と第 6 の実施形態とを合成した手の形状と姿勢の認識装置に対して構成してもよい。

【0159】

（第 8 の実施形態）

本発明の第 8 の実施形態は、上記第 5 ~ 第 7 の実施形態の手領域検出部 28 において、手領域に関してもクラスタ情報を利用することによって、より高精度に画像中の手領域を検
40 出する手法を提供するものである。

【0160】

図 25 は、本発明の第 8 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部の詳細な構成を示すブロック図である。図 25 において、第 8 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部 48 は、領域候補切出部 39 と、領域マスク記憶部 40 と、画像正規化部 41 と、固有空間投影部 22 と、最尤クラスタ判別部 25 と、領域決定部 42 とを備える。

なお、第 8 の実施形態に係る手の形状と姿勢の認識装置におけるその他の構成は、上記第 5 ~ 第 7 の実施形態に係る手の形状と姿勢の認識装置の構成とそれぞれ同様であり、当該構成については、同一の参照番号を付してその説明を省略する。

10

20

30

40

50

【 0 1 6 1 】

まず、第 8 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部 4 8 の各構成を説明する。

領域候補切出部 3 9 は、入力する手振り動作画像から手領域の候補となる画像範囲をそれぞれ切り出す。また、領域候補切出部 3 9 は、切り出した手領域の位置情報を領域決定部 4 2 へ出力する。領域マスク記憶部 4 0 は、領域候補切出部 3 9 が切り出した候補手領域から、予め定めた領域のみを抜き出すためのマスクを記憶する。画像正規化部 4 1 は、領域候補切出部 3 9 が切り出した候補手領域に対し、大きさを正規化して領域マスク記憶部 4 0 に記憶したマスク領域を付加した後、明るさを正規化することで手領域候補画像を得る。固有空間投影部 2 2 は、上記第 5 ~ 第 7 の実施形態で述べたように、画像正規化部 4 1 によって得られる手領域候補画像を固有空間へ展開する。最尤クラスタ判別部 2 5 は、上記第 5 ~ 第 7 の実施形態で述べたように、固有空間投影部 2 2 が求めた固有空間投影座標に最も近い投影座標を有するクラスタを取得する。領域決定部 4 2 は、最尤クラスタ判別部 2 5 がクラスタを取得したときの尤度を、全ての手領域候補画像に対して比較し、最も高い尤度を持つ手領域候補画像の位置とその時のクラスタインデックスを出力する。

10

【 0 1 6 2 】

次に、図 2 6 ~ 図 2 8 を用いて、第 8 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部 4 8 が行う手領域検出方法を、処理の順に説明する。図 2 6 は、図 2 5 の領域候補切出部 3 9 が行う手領域候補を求める手法の一例を説明する図である。なお、図 2 6 においては、単純にスキャンを行う手法、色情報他の知識により手領域候補を切り出す手法、および、前時刻の手領域検出結果に従い現時刻の手領域位置を予測して手領域を切り出す手法の 3 つについて説明している。図 2 7 は、図 2 5 の画像正規化部 4 1 における処理の概要を示す図である。図 2 8 は、図 2 5 の領域マスク記憶部 4 0 に記憶されているマスク領域の一例を示す図である。

20

【 0 1 6 3 】

まず、領域候補切出部 3 9 は、手領域候補を求めて、入力する手振り動作画像から求めた手領域候補に対応する矩形領域を切り出す。この手領域候補を求める手法としては、例えば図 2 6 に示す 3 つの手法が考えられる。

第 1 の手法は、最も単純な手法であり、手領域候補として切り出す領域の大きさを予め決めておき、切り出す矩形領域を手振り動作画像上でスキャンさせて、スキャンによって順次得られる全ての領域を手領域候補とする手法である（図 2 6 (a) ）。なお、この手法の場合、手振り動作画像上の手の距離によって、スキャンさせる大きさを可変してもよい。

30

第 2 の手法は、色情報（例えば、肌色情報）などを用いることにより、手振り動作画像から、その色に対応する領域の前後のみをスキャンさせる対象として、手領域候補の矩形領域を切り出す手法である。この手法の場合、肌色情報を用いることにより、手と顔の周辺領域の画像のみを手領域候補とすることができる（図 2 6 (b) ）。

第 3 の手法は、前時刻の手領域の位置情報（領域決定部 4 2 からフィードバックされる情報）から現時刻の手領域の位置を予測し、予測した手領域の位置の周辺をスキャンすることによって手領域候補を切り出す手法である。この手法の場合、例えば、前時刻の手の位置に前時刻の手の速度を足し合せることによって現時刻の手領域を予測する手法、また、予測の時にカルマンフィルタを用いて手の位置を求める手法等がある（図 2 6 (c) ）。

40

【 0 1 6 4 】

次に、画像正規化部 4 1 は、図 2 7 に示すように、領域候補切出部 3 9 が切り出した手領域候補に対し、大きさを正規化し、領域マスク記憶部 4 0 に記憶した手領域マスクを重ね合せ、明るさを正規化する。手領域候補に対してマスク処理を行うのは、処理対象が掌や顔といった矩形領域ではない部位だからである。このため、領域マスク記憶部 4 0 に記憶されている手領域マスクの一例としては、図 2 8 (a) に示すような幾何形状マスク（単純な幾何形状（円、楕円など）を用いたマスク）や、図 2 8 (b) に示すような学習画像から作成したマスク（過去に得られた画像群を重ね合せて O R 演算をしたマスク）等が好

50

ましい。

このように、画像正規化部 4 1 は、上記手領域マスクと手領域候補の画像とを重ね合せ、明るさの正規化を行うことにより、手領域候補画像を生成する。

【0165】

その後は、上記第 5 ～ 第 7 の実施形態と同様に、固有空間投影部 2 2 は、画像正規化部 4 1 が出力する各手領域候補画像を、固有ベクトル記憶部 1 4 に従って固有空間へ投影して投影座標をそれぞれ求める。そして、最尤クラスタ判別部 2 5 は、固有空間投影部 2 2 が求めた投影座標がクラスタ情報記憶部 1 7 A に記憶しているどのクラスタに属するかを判断し、各手領域候補画像ごとに該当クラスタとそのときの尤度を領域決定部 4 2 へ出力する。

10

【0166】

そして、領域決定部 4 2 は、最尤クラスタ判別部 2 5 が出力する各手領域候補画像に対応する尤度から、最も尤度の高いときの手領域候補を求め、その時の手領域の位置（領域候補切出部 3 9 から与えられる）と大きさを手領域検出結果として、手動作分節部 2 9 へ出力する。

【0167】

以上のように、本発明の第 8 の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、手領域を検出するときに手領域候補の領域を固有空間へ投影して、該当クラスタを求めることによって手領域を検出する。

このため、手領域を検出すると同時にその手領域の該当クラスタを求めることができるので、手領域検出と手の形状と姿勢の認識、若しくは手領域検出と手振り動作認識処理を一つの処理として統合することが可能となる。

20

【0168】

なお、第 8 の実施形態では、手振り動作画像に対して上記手法を適用したが、一般の動作画像に対しても上記手法を用いることにより、動作主体の検出を行うことは可能であり、同等の効果を奏することができる。

【0169】

（第 9 の実施形態）

本発明の第 9 の実施形態は、上記第 8 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部 4 8 の画像正規化部 4 1 および領域決定部 4 2 において、前時刻のクラスタ情報を利用することにより、現時刻の手領域検出をより高精度に行う手法を提供するものである。

30

【0170】

図 2 9 は、本発明の第 9 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部の詳細な構成を示すブロック図である。図 2 9 において、第 9 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部 5 8 は、領域候補切出部 3 9 と、領域マスク記憶部 4 5 と、画像正規化部 4 1 と、固有空間投影部 2 2 と、最尤クラスタ判別部 2 5 と、領域決定部 4 2 と、クラスタ遷移情報記憶部 4 3 と、クラスタ遷移情報登録部 4 4 とを備える。

図 2 9 に示すように、第 9 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部 5 8 は、上記第 8 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部 4 8 に、クラスタ遷移情報記憶部 4 3 およびクラスタ遷移情報登録部 4 4 を付加し、領域マスク記憶部 4 0 を領域マスク記憶部 4 5 に代えた構成である。

40

なお、第 9 の実施形態に係る手の形状と姿勢の認識装置におけるその他の構成は、上記第 8 の実施形態に係る手の形状と姿勢の認識装置の構成と同様であり、当該構成については、同一の参照番号を付してその説明を省略する。

【0171】

以下、図 3 0 ～ 図 3 1 を用いて、第 9 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部 5 8 が行う手領域検出方法を、処理の順に説明する。図 3 0 は、図 2 9 のクラスタ遷移情報記憶部 4 3 に記憶されるクラスタ遷移情報の一例を示す図である。図

50

30に示すように、クラスタ遷移情報記憶部43には、ある時刻 t のクラスタが与えられたとき、次の時刻 $t+1$ にどのクラスタに遷移しやすいかを示すクラスタ遷移の頻度を記載した遷移度マップが格納されている。なお、ここではクラスタ遷移の度合いをクラスタ遷移度と呼ぶ。図31は、図29の領域マスク記憶部45に記憶されているマスク領域の一例を示す図である。図31に示すように、領域マスク記憶部45には、予め学習画像から作成したマスクが各クラスタごとに別々に登録されている。

【0172】

まず、領域候補切出部39は、上記第8の実施形態と同様、入力する手振り動作画像から手領域候補を求め、求めた候補に対応する矩形領域を切り出す。

次に、画像正規化部41は、領域候補切出部39によって得られた手領域候補の大きさを正規化し、領域マスク記憶部45に記憶した手領域マスクを重ね合せて明るさを正規化する。このとき、画像正規化部41は、前時刻の該当クラスタをもとに、クラスタ遷移情報記憶部43を参照して遷移度の高いクラスタを複数個選び、それぞれのクラスタに該当するマスクを領域マスク記憶部45から抜き出す。そして、画像正規化部41は、抜き出した複数個のマスクを重ね合せOR演算をすることにより新たなマスクを作成し、当該作成したマスクを得られた手領域候補に重ね合せ、明るさの正規化を行うことにより、手領域候補画像を生成する。

【0173】

その後は、上記第8の実施形態と同様に、固有空間投影部22は、画像正規化部41が出力する各手領域候補画像を、固有ベクトル記憶部14に従って固有空間へ投影して投影座標をそれぞれ求める。そして、最尤クラスタ判別部25は、固有空間投影部22が求めた投影座標がクラスタ情報記憶部17Aに記憶しているどのクラスタに属するかを判断し、各手領域候補画像ごとに該当クラスタとそのときの尤度を領域決定部42へ出力する。

【0174】

そして、領域決定部42は、クラスタ遷移情報記憶部43に記憶されている遷移度マップを参照して、最尤クラスタ判別部25が出力する各手領域候補画像に対応するクラスタおよび尤度に基づいて、ある値より高い遷移度のクラスタのうち、最も高い尤度を持つクラスタの手領域候補を求め、その時の手領域の位置（領域候補切出部39から与えられる）と大きさを手領域検出結果として、手動作分節部29へ出力する。また、領域決定部42は、検出した手領域のクラスタをクラスタ遷移情報登録部44へ出力する。

【0175】

クラスタ遷移情報登録部44は、領域決定部42における手領域検出結果に従い、クラスタ遷移情報記憶部43を更新することを要求する指示がある場合にのみ動作する。この指示は、このシステムを用いるユーザ、若しくはこのシステムを構築する管理者によって入力される。そして、更新することを要求する指示があった場合に、クラスタ遷移情報登録部44は、検出したクラスタと前時刻のクラスタとに従い、クラスタ遷移情報記憶部43のクラスタ遷移情報を更新する。例えば、単純に遷移度マップの該当する場所の値をある値だけ増やすことにより更新することができる。

【0176】

以上のように、本発明の第9の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、上記第10の実施形態に係る手の形状と姿勢の認識装置に対し、手領域の決定時にクラスタの遷移情報を用いる。このため、手領域の決定をより正確に行うことができる。

【0177】

なお、第9の実施形態では、手振り動作画像に対して上記手法を適用したが、一般の動作画像に対しても上記手法を用いることにより、動作主体の検出を行うことは可能であり、同等の効果を奏することができる。

【0178】

（第10の実施形態）

本発明の第10の実施形態は、上記第1～第7の実施形態の手画像正規化部11，21に

において、手画像を正規化する際に、手首領域を削除するだけでなく、肌色により手領域を抽出したり、正規化した手画像からさらに指の特徴を強調するような手法を加えることによって、一般的で自然な背景で撮影した手画像から手領域を抽出することを可能にし、さらに、より正確に手形状と手姿勢とを認識する手法を提供するものである。

【 0 1 7 9 】

図 3 2 は、本発明の第 1 0 の実施形態に係る手の形状と姿勢の認識装置を構成する手画像正規化部 1 1 , 2 1 のさらに詳細な構成を示すブロック図である。

図 3 2 において、第 1 0 の実施形態に係る手の形状と姿勢の認識装置を構成する手画像正規化部 1 1 , 2 1 は、色分布記憶部 6 1 と、手領域抽出部 6 2 と、手首領域削除部 6 3 と、領域移動部 6 4 と、回転角計算部 6 5 と、領域回転部 6 6 と、大きさ正規化部 6 7 と、指特徴強調部 6 8 とを備える。

なお、第 1 0 の実施形態に係る手の形状と姿勢の認識装置におけるその他の構成は、上記第 1 ~ 第 7 の実施形態に係る手の形状と姿勢の認識装置の構成とそれぞれ同様であり、当該構成については、同一の参照番号を付してその説明を省略する。

【 0 1 8 0 】

まず、第 1 0 の実施形態に係る手の形状と姿勢の認識装置を構成する手画像正規化部 1 1 , 2 1 の各構成を説明する。

色分布記憶部 6 1 は、入力手画像から抽出すべき手領域を、色分布として予め記憶する。手領域抽出部 6 2 は、色分布記憶部 6 1 が記憶する色分布に従い、手領域を抽出する。手首領域削除部 6 3 は、手領域抽出部 6 2 が抽出した領域から手首方向を求め、求めた手首方向に従って当該抽出した領域から手首領域を削除する。領域移動部 6 4 は、手首領域削除部 6 3 により手首領域を削除した手領域を画像上の予め定義した位置に移動させる。回転角計算部 6 5 は、手領域から光軸に対して垂直な手の回転角を求める。領域回転部 6 6 は、回転角に従い手が一定方向に向くように回転変換を施す。大きさ正規化部 6 7 は、回転させた手領域の大きさを予め定めた一定の大きさに正規化する。指特徴強調部 6 8 は、正規化した手画像から指以外の一定領域を削除し、指の特徴を強調する。

【 0 1 8 1 】

次に、図 3 3 ~ 図 3 5 を用いて、第 1 0 の実施形態に係る手の形状と姿勢の認識装置を構成する手画像正規化部 1 1 , 2 1 が行う手画像の正規化方法を、処理の順に説明する。図 3 3 は、図 3 2 の色分布記憶部 6 1 が有する格納テーブルの構造の一例を示す図である。なお、図 3 3 においては、格納テーブルが、RGB 色空間の三次元ルックアップテーブル (LUT) である場合を一例に挙げている。図 3 4 は、図 3 2 の回転角計算部 6 5 が行う処理の概略を説明する図である。図 3 5 は、図 3 2 の指特徴強調部 6 8 が行う処理の一例を説明する図である。

【 0 1 8 2 】

最初に、色分布記憶部 6 1 に、自然な背景の画像から手領域を抜き出すために必要な肌色領域を設定する。色分布記憶部 6 1 は、図 3 3 に示すような、RGB 色空間の三次元 LUT を有している。この三次元 LUT は、各々離散値をとる 3 つの色値 R , G , B で構成された三次元色空間 CS を、各色を軸として、それぞれの軸に対し d_1 , d_2 , d_3 の幅で分割し、この分割した結果得られる個々の分割空間 DS の重心位置 (格子点) の色に対応するデータ値を保持することにより、得られるテーブルである。換言すれば、三次元 LUT は、各格子点の三次元座標 (r , g , b) をパラメータとする関数の値 $c \{ = f (r , g , b) \}$ を記憶するものである。

第 1 0 の実施形態で説明する例では、この色分布記憶部 6 1 の手の色領域、すなわち肌色領域部分に正の値を、それ以外の色領域部分に “ 0 ” の値を設定する。

【 0 1 8 3 】

まず、手領域抽出部 6 2 は、入力される画像をスキャンし、得られた画素の色と色分布記憶部 6 1 に記憶されている三次元 LUT の格子点の色との間で、最も近い距離にある格子点のデータ値を求める。これにより、画素の色が肌色であれば正の値が出力され、それ以外の色であれば “ 0 ” が出力されるので、肌色領域を抽出することができる。なお、最も

近い距離にある格子点として得られた画素の色の近傍にある 6 つの格子点からの補間演算による値を、上述した関数 f として定義しても同等の効果を奏する。

そして、手領域抽出部 6 2 は、抽出した肌色領域のうち、手の大きさに最も近い領域を手領域とし、それ以外の領域をノイズとみなし削除した手画像を、手首領域削除部 6 3 に出力する。

なお、肌色領域を色分布記憶部 6 1 に設定する手法としては、上述した手法以外にも、例えば、肌色領域を全て一定の値（例えば、255 ビット）に設定する手法（この場合、手領域抽出部 6 2 が出力する画像は、シルエット画像となる）や、肌色領域のうち、影の領域を暗い値に、反射が強い領域を明るい値となるように三次元 LUT を設定する手法や、手の画像の持つ色の度数分布を、そのまま三次元 LUT に設定する手法等を用いることが

10

【0184】

次に、手首領域削除部 6 3 は、手領域抽出部 6 2 が抽出した手画像から手首方向を求め、求めた手首方向に従って手首領域を削除する。この手首領域の削除は、図 2 で示した手法で実現することが可能である。領域移動部 6 4 は、手首領域削除部 6 3 によって手首領域が削除された手画像を入力し、残った手領域の重心点が手画像の中心になるように移動変換を行う。そして、回転角計算部 6 5 は、図 3 4 で示すように、手領域のモーメント主軸（手という図形が伸びている方向、すなわち、手首 - 中指方向）と画像上のある軸（例えば、 x 軸）との角度を計算する。

【0185】

20

今、手の画像を $f(x, y)$ と、手の重心座標を (x_g, y_g) とし、下記式 (5) に従って、 M_{11} 、 M_{20} 、 M_{02} を求める。

【数 2】

$$M_{pq} = \sum_x \sum_y (x - x_g)^p (y - y_g)^q f(x, y) \quad \dots (5)$$

この結果、モーメント主軸と x 軸とのなす角 θ は、下記式 (6) により求めることができる。

30

【数 3】

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{2M_{11}}{M_{20} - M_{02}} \right) \quad \dots (6)$$

【0186】

上記の角度計算の後、領域回転部 6 6 は、モーメント主軸が y 軸と同じ方向を向くように回転変換を施す。最後に、大きさ正規化部 6 7 は、回転変換を施した手領域が予め定めた一定の大きさになるように手画像を正規化する。

この手首領域削除部 6 3、領域移動部 6 4、回転角計算部 6 5、領域回転部 6 6 および大きさ正規化部 6 7 は、上記第 1 ~ 第 7 の実施形態における手画像正規化部 1 1、2 1 の典型的な構成例を詳細に説明したものであるが、第 1 0 の実施形態においては、より正確に画像認識を行うために、最後に、指特徴強調部 6 8 が、正規化した手画像から指以外の一定領域を削除して、指の特徴を強調する処理を行う。以下、図 3 5 を参照して、指特徴強調部 6 8 が行う処理の一例を説明する。

40

【0187】

図 3 5 において、[例 1] では、手領域の重心点（すなわち、画像の中心点）から、 $-y$ 方向（モーメント主軸の手首方向）に $\pm A$ 度の角度をなす扇形状を手画像から削除することにより、指領域を強調する。[例 2] では、手領域の重心点から、 $-y$ 方向に対して距離 D より以遠にある手首側を手画像から削除することにより、指領域を強調する。[例

50

３］では、より簡単に、手画像の手首側を一定距離分だけ削除することにより、指領域を強調する。〔例４〕では、手画像に対して極座標変換を行うことにより、指領域を強調する。

【０１８８】

以上のように、本発明の第１０の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、手画像を正規化する際に、手首領域を削除するだけではなく、肌色により手領域を抽出したり、正規化した手画像からさらに指の特徴を強調する。これにより、一般的で自然な背景で撮影した手の画像から手領域を抽出することができ、さらに、より正確に手形状と手姿勢とを認識することが可能となる。

【０１８９】

（第１１の実施形態）

本発明の第１１の実施形態は、上記第１～第１０の実施形態において、手形状画像情報記憶部１２Ａ～１２Ｃに格納された手形状画像が、掌主軸周りの回転に対する手形状画像しか格納していない場合に、認識対象となる手を複数の視点のカメラから撮影し、その撮影した入力手画像から手の方向を求め、手の方向に対しても正規化することにより、実際には、手形状画像として格納されていない手の方向の画像においても、手形状と手姿勢とを認識する手法を提供するものである。

【０１９０】

この第１１の実施形態は、上記第１～第１０の実施形態における手画像正規化部２１に、複数台のカメラから与えられる入力手画像の各々に対し、モーメント主軸を求めることで手の方向を求め、その方向に対して正規化する手法を付加することにより実現できる。なお、本発明の第１１の実施形態に係る手の形状と姿勢の認識装置を構成する手画像正規化部２１は、上記第１～第１０の実施形態で述べた手画像正規化部２１と同様の構成であるため、その図面を省略する。図３６に、複数台のカメラから手の方向を求め、正規化する手法の概念の一例を示す。なお、図３６においては、３台のカメラを用いた場合を一例に挙げて説明している。

【０１９１】

今、３台のカメラが図３６に示す位置で手を撮影するとする。

まず、手画像正規化部２１は、上記各実施形態で述べたと同様の手法で各入力手画像の手首領域を削除する。次に、手画像正規化部２１は、手首領域を削除した入力手画像の手領域を画像の中心に移動させ、手領域のモーメント主軸の方向を求める（上記第１０の実施形態の領域移動部６４および回転角計算部６５で述べた手法と同様である）。次に、手画像正規化部２１は、求めたモーメント主軸から三次元空間における主軸の方向をベクトル値として求め、求めた主軸の方向が各カメラに対して光軸に垂直な一定方向を向くような変換行列を求める。そして、手画像正規化部２１は、求めた変換行列に従い、各カメラの撮影によって入力した入力手画像を変形させる。なお、入力手画像の変形に対しては、例えば、一般に行われているアフィン変換を用いた変形手法を用いることができる。

【０１９２】

以上のように、本発明の第１１の実施形態に係る手の形状と姿勢の認識装置および認識方法によれば、手形状画像情報記憶部１２Ａ～１２Ｃに、掌主軸周りの回転に対する手形状画像しか格納していない場合でも、実際には、手形状画像として格納されていない手の方向の画像に対して、手形状と手姿勢とを認識することが可能になる。

【図面の簡単な説明】

【図１】本発明の第１の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。

【図２】図１の手画像正規化部１１が行う処理の概略を説明する図である。

【図３】図１の手形状画像情報記憶部１２Ａが有する格納テーブルの一例を示す図である。

【図４】固有空間計算部１３において固有空間を求める手法の一例の概略を説明する図である。

10

20

30

40

50

【図 5】図 1 の固有空間投影部 1 5 が行う処理の概略を説明する図である。

【図 6】本発明の第 1 の実施形態に係る手の形状と姿勢の認識装置を実現するためのハードウェア構成の一例を示す図である。

【図 7】形状情報および姿勢情報に対する命令を記憶する命令記憶部に、オーディオ機器に対する命令が記憶されている一例を示す。

【図 8】本発明の第 2 の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。

【図 9】図 8 の手形状画像情報記憶部 1 2 B が有する格納テーブルの一例を示す図である。

【図 10】図 8 のクラスタ分析部 1 6 が行う処理の一例を示すフローチャートである。

10

【図 11】図 8 の画像比較部 2 6 において行う比較手法の概念の一例を示す図である。

【図 12】本発明の第 3 の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。

【図 13】図 8 のクラスタ分析部 1 6 により同一グループに分類される似た画像の一例を示す図である。

【図 14】図 12 のクラスタ分析 / 枠判別部 1 8 が行う処理概念の一例を示す図である。

【図 15】本発明の第 4 の実施形態に係る手の形状と姿勢の認識装置において、複数台のカメラの入力手画像から手形状画像を求める概念の一例を説明する図である。

【図 16】本発明の第 5 の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。

20

【図 17】図 16 の手領域検出部 2 8、手動作分節部 2 9 および手画像切出部 3 0 が行う処理の概念を示す図である。

【図 18】図 16 の手画像系列および当該手画像系列から求められるクラスタ系列の一例を示す図である。

【図 19】図 16 の系列識別辞書 3 2 が有する記憶形態の一例を示す図である。

【図 20】図 16 の系列識別辞書 3 2 が有する記憶形態の一例を示す図である。

【図 21】本発明の第 6 の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。

【図 22】図 21 の手形状画像情報記憶部 1 2 C が有する格納テーブルの一例を示す図である。

30

【図 23】手姿勢を定義する方法の一例の概略を説明する図である。

【図 24】本発明の第 7 の実施形態に係る手の形状と姿勢の認識装置の構成を示すブロック図である。

【図 25】本発明の第 8 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部の詳細な構成を示すブロック図である。

【図 26】図 25 の領域候補切出部 3 9 における処理の一例を示す図である。

【図 27】図 25 の画像正規化部 4 1 における処理の概要を示す図である。

【図 28】図 25 の領域マスク記憶部 4 0 に記憶されているマスク領域の一例を示す図である。

【図 29】本発明の第 9 の実施形態に係る手の形状と姿勢の認識装置を構成する手領域検出部の詳細な構成を示すブロック図である。

40

【図 30】図 29 のクラスタ遷移情報記憶部 4 3 に記憶されるクラスタ遷移情報の一例を示す図である。

【図 31】図 29 の領域マスク記憶部 4 5 に記憶されているマスク領域の一例を示す図である。

【図 32】本発明の第 10 の実施形態に係る手の形状と姿勢の認識装置を構成する手画像正規化部 1 1, 2 1 のさらに詳細な構成を示すブロック図である。

【図 33】図 32 の色分布記憶部 6 1 が有する格納テーブルの構造の一例を示す図である。

【図 34】図 32 の回転角計算部 6 5 が行う処理の概略を説明する図である。

50

【図 3 5】図 3 2 の指特徴強調部 6 8 が行う処理の一例を説明する図である。

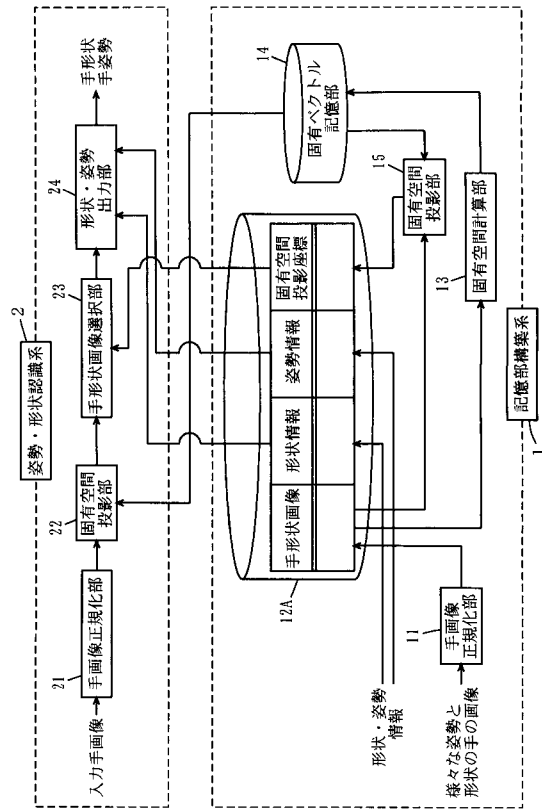
【図 3 6】本発明の第 1 1 の実施形態に係る手の形状と姿勢の認識装置において、複数台のカメラの入力手画像から手の方向を求め正規化する概念の一例を示す図である。

【図 3 7】手姿勢を定義する方法の一例の概略を説明する図である。

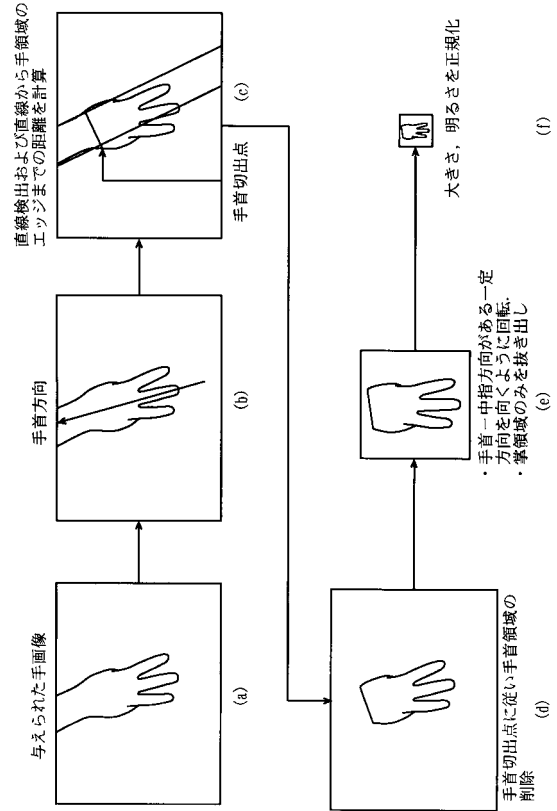
【符号の説明】

1 ... 記憶部構築系	
2 ... 姿勢・形状認識系	
1 1 , 2 1 ... 手画像正規化部	
1 2 A , 1 2 B , 1 2 C ... 手形状画像情報記憶部	
1 3 ... 固有空間計算部	10
1 4 ... 固有ベクトル記憶部	
1 5 , 2 2 ... 固有空間投影部	
1 6 ... クラスタ分析部	
1 7 A , 1 7 B ... クラスタ情報記憶部	
1 8 ... クラスタ分析 / 枠判別部	
2 3 ... 手形状画像選択部	
2 4 ... 形状・姿勢出力部	
2 5 ... 最尤クラスタ判別部	
2 6 , 2 7 ... 画像比較部	
2 8 , 4 8 , 5 8 ... 手領域検出部	20
2 9 ... 手動作分節部	
3 0 ... 手画像切出部	
3 1 ... 系列登録部	
3 2 ... 系列識別辞書	
3 3 A , 3 3 B ... 識別演算部	
3 4 A , 3 4 B ... データ経路制御部	
3 5 ... 手画像登録部	
3 6 ... 系列再構成部	
3 7 ... 大局動作認識部	
3 8 ... 制約条件記憶部	30
3 9 ... 領域候補切出部	
4 0 , 4 5 ... 領域マスク記憶部	
4 1 ... 画像正規化部	
4 2 ... 領域決定部	
4 3 ... クラスタ遷移情報記憶部	
4 4 ... クラスタ遷移情報登録部	
5 0 ... 記憶装置	
5 1 ... C P U	
5 2 ... メモリ	
5 3 ... 画像入力装置	40
5 4 ... 入力装置	
5 5 ... 出力装置	
6 1 ... 色分布記憶部	
6 2 ... 手領域抽出部	
6 3 ... 手首領域削除部	
6 4 ... 領域移動部	
6 5 ... 回転角計算部	
6 6 ... 領域回転部	
6 7 ... 大きさ正規化部	
6 8 ... 指特徴強調部	50

【図 1】



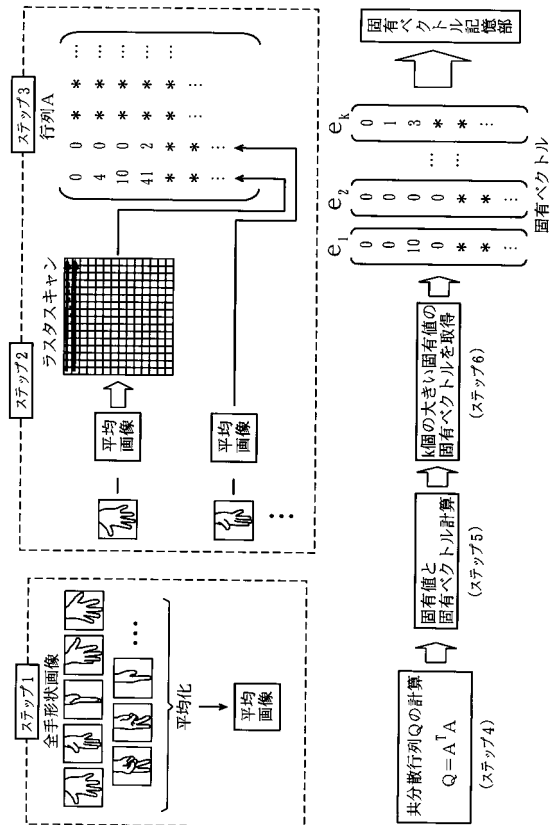
【図 2】



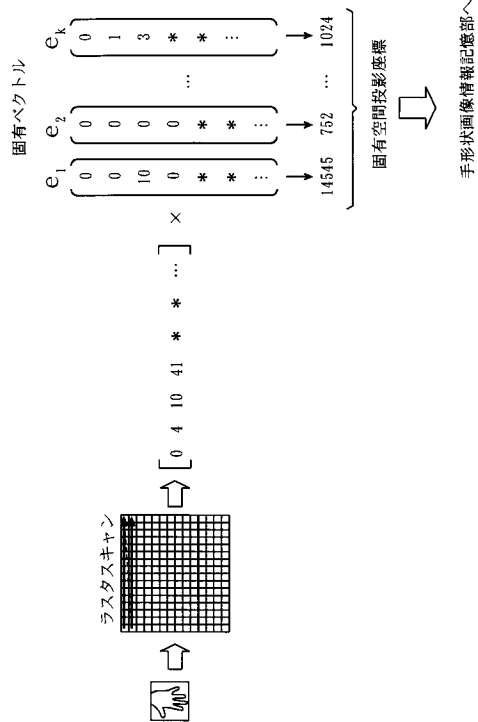
【図 3】

手形状画像	形状情報	姿勢情報	固有空間投影座標
	伸展 5 本指	光軸に垂直、掌方向：前	
	伸展 5 本指	光軸に垂直、掌方向：右前	
	伸展 5 本指	光軸に垂直、掌方向：右	
	伸展 5 本指	光軸に垂直、掌方向：右後	
	伸展 5 本指	光軸に垂直、掌方向：後	
...			
	伸展 2 本指	光軸に垂直、掌方向：前	
	伸展 2 本指	光軸に垂直、掌方向：右前	
	伸展 2 本指	光軸に垂直、掌方向：右	
...			

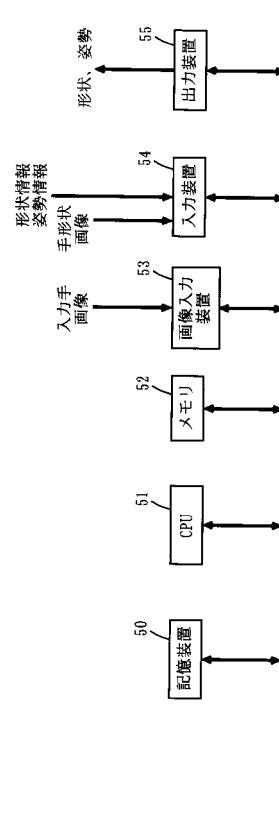
【図 4】



【図 5】



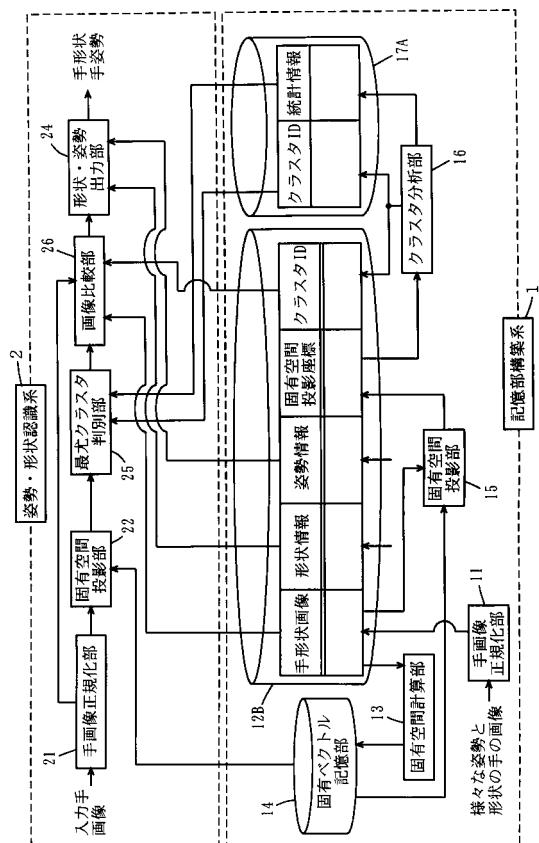
【図 6】



【図 7】

形状情報	姿勢情報	命令
全握	全姿勢	ストップ
伸展5本指	全姿勢	スタート
伸展1本指	右上方向	早送り
伸展1本指	左上方向	巻き戻し

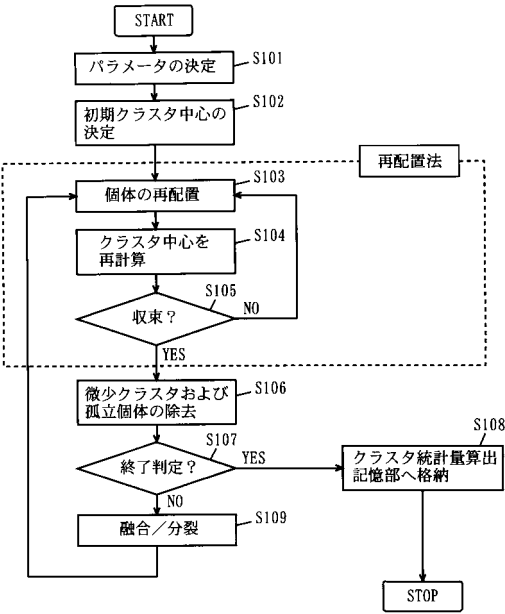
【図 8】



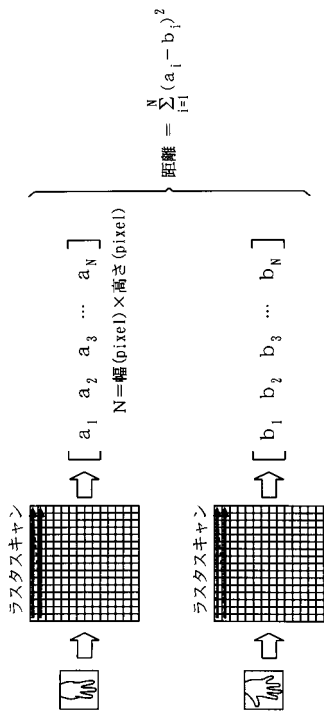
【図 9】

手形状 画像	形状情報	姿勢情報	固有空間投影座標	クラスタID
	伸展5本指	光軸に垂直, 掌方向: 前		
	伸展5本指	光軸に垂直, 掌方向: 右前		
	伸展5本指	光軸に垂直, 掌方向: 右		
	伸展5本指	光軸に垂直, 掌方向: 右後		
	伸展5本指	光軸に垂直, 掌方向: 後		
...				
	伸展2本指	光軸に垂直, 掌方向: 前		
	伸展2本指	光軸に垂直, 掌方向: 右前		
	伸展2本指	光軸に垂直, 掌方向: 右		
...				

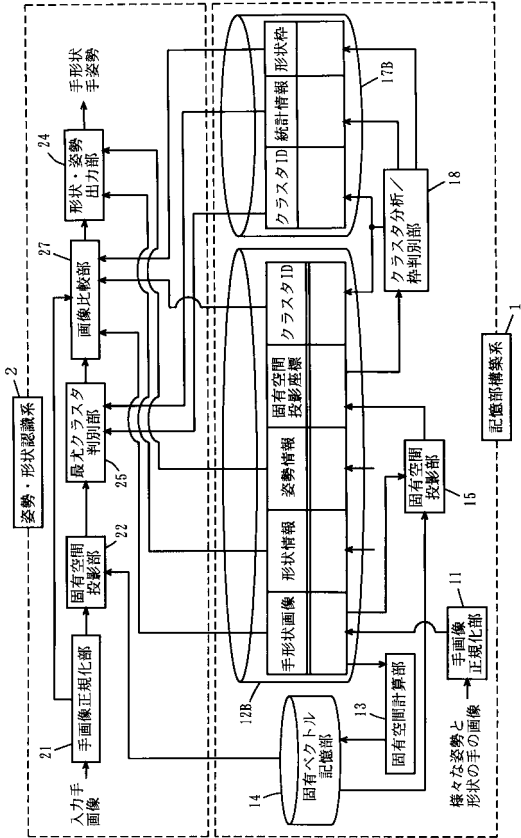
【図 10】



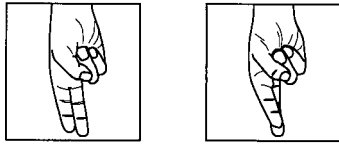
【図 11】



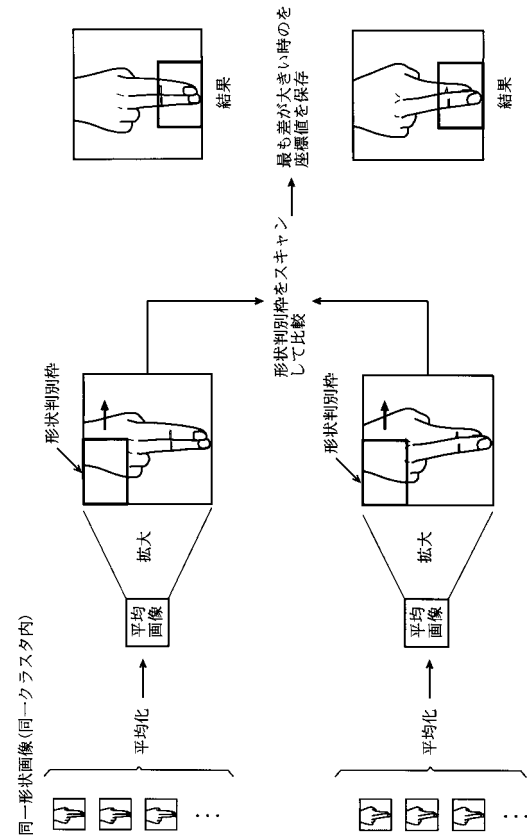
【図 12】



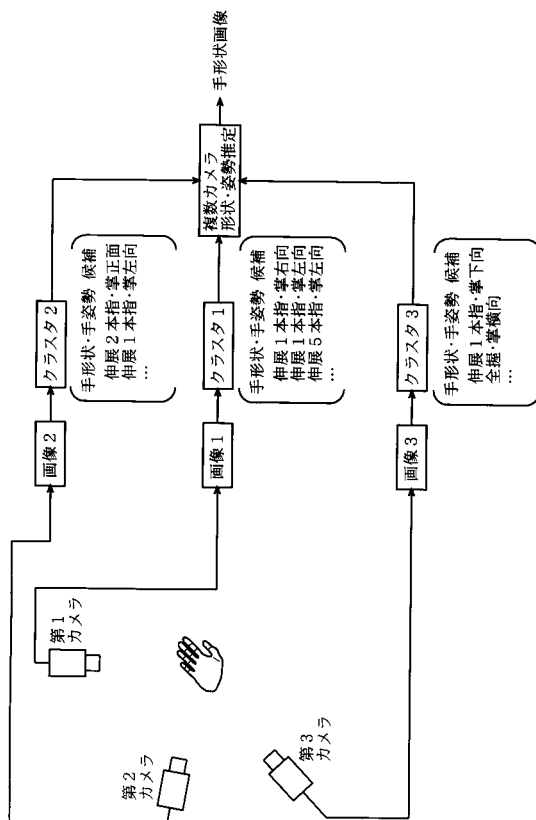
【図 13】



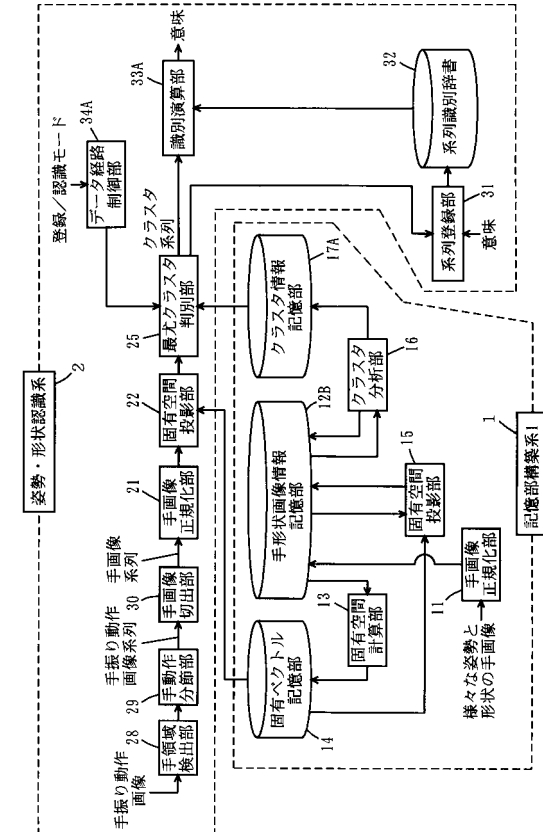
【図 14】



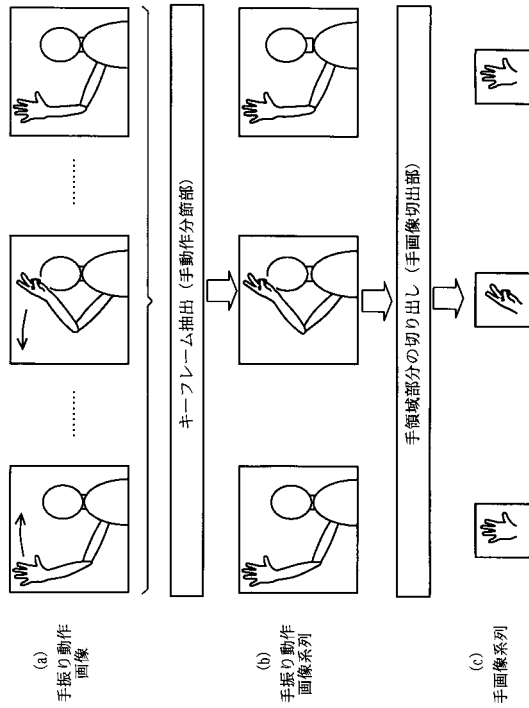
【図 15】



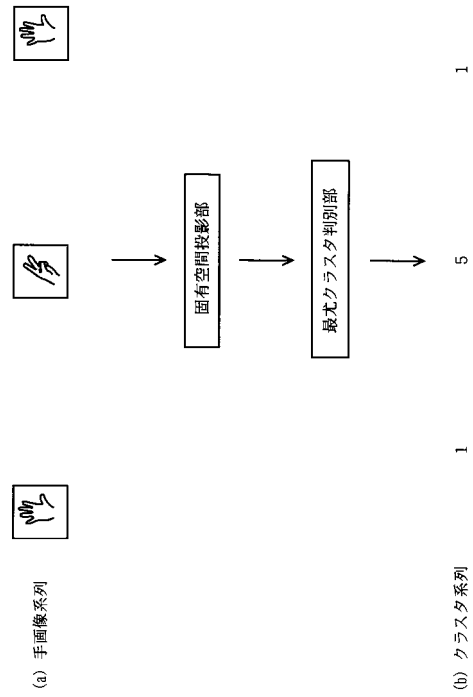
【図 16】



【図 17】



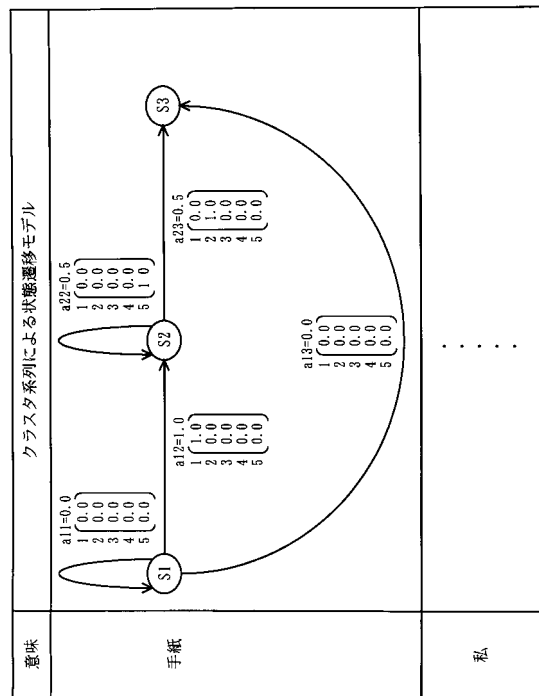
【図 18】



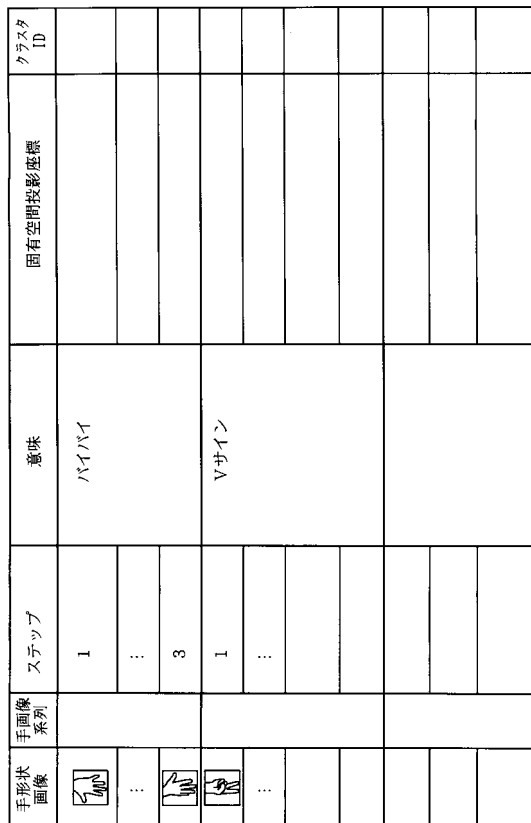
【図 19】

意味	クラスタ系列
手紙	1 5 2
手紙	1 5 5 2
私	3 3 4 2
私	3 2
住所	2 3 2
...	...

【図 20】



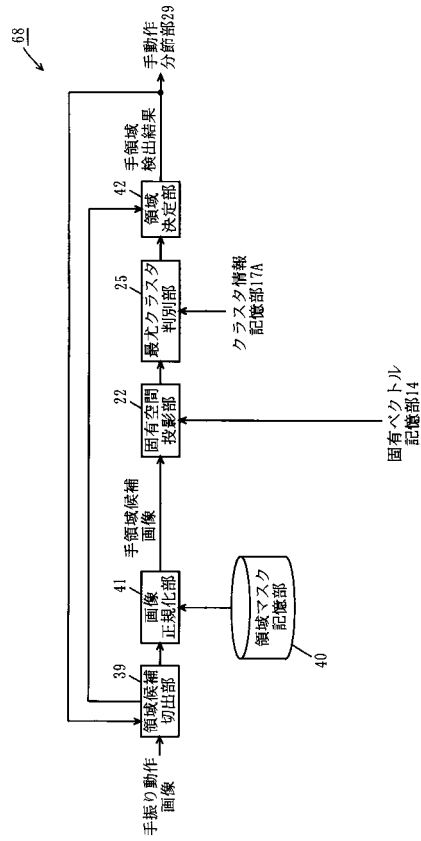
【 図 2 2 】



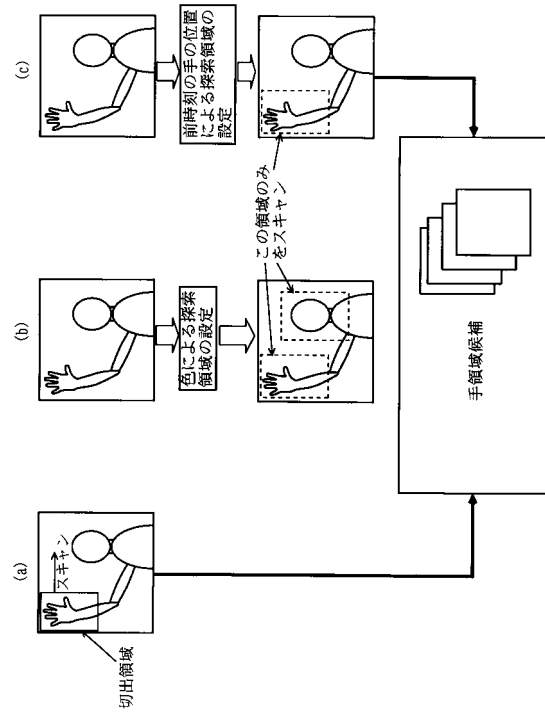
【 図 2 4 】

[illegible]

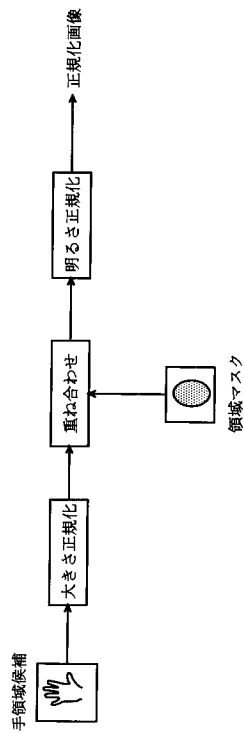
【図 25】



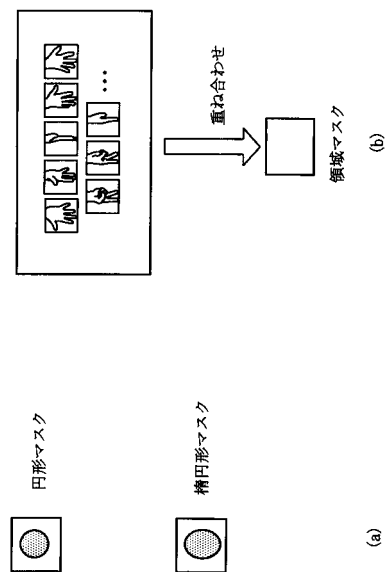
【図 26】



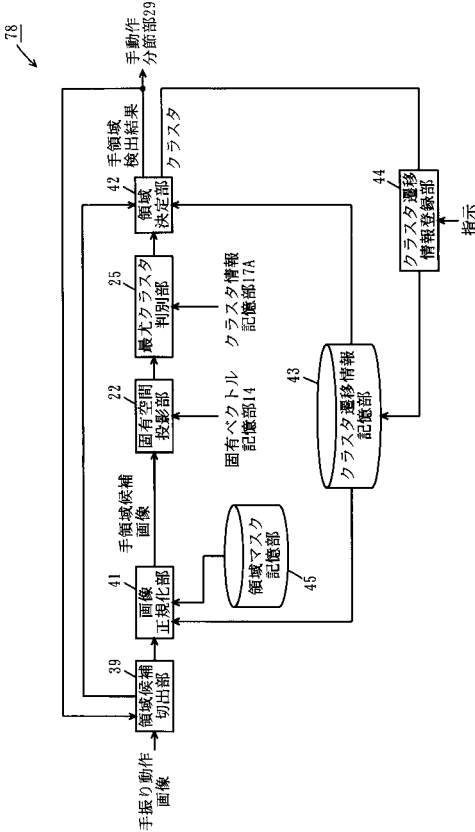
【図 27】



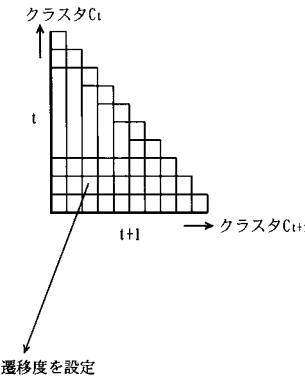
【図 28】



【図 29】



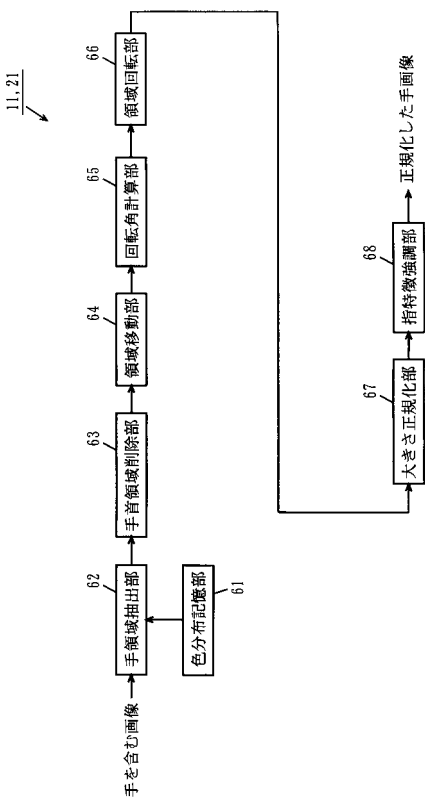
【図 30】



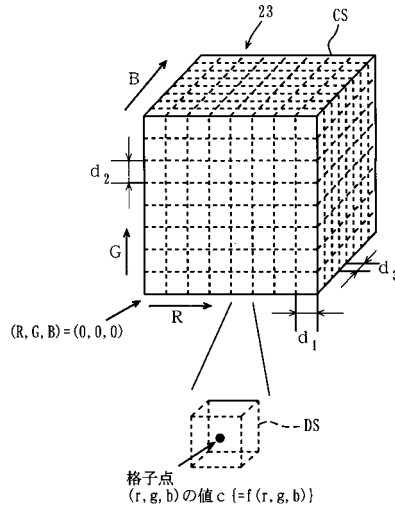
【図 31】

クラスタ	領域マスク
1	
...	
3	
4	
6	
7	
8	
...	

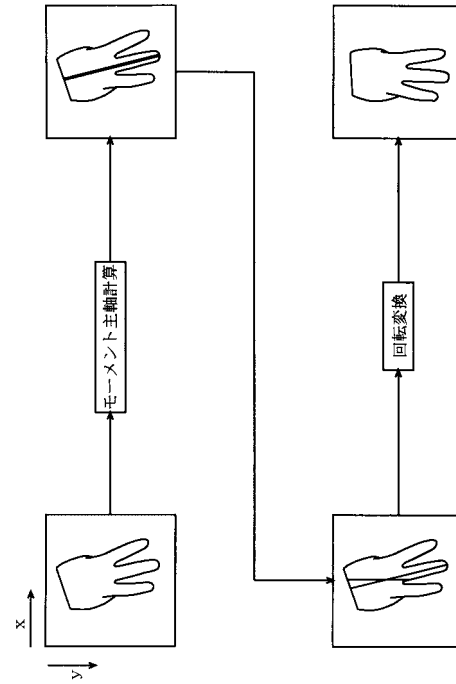
【図 32】



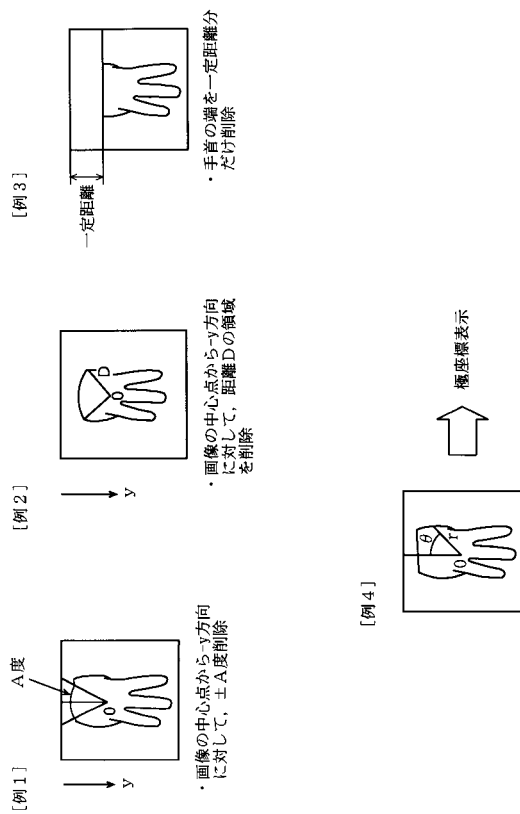
【図 3 3】



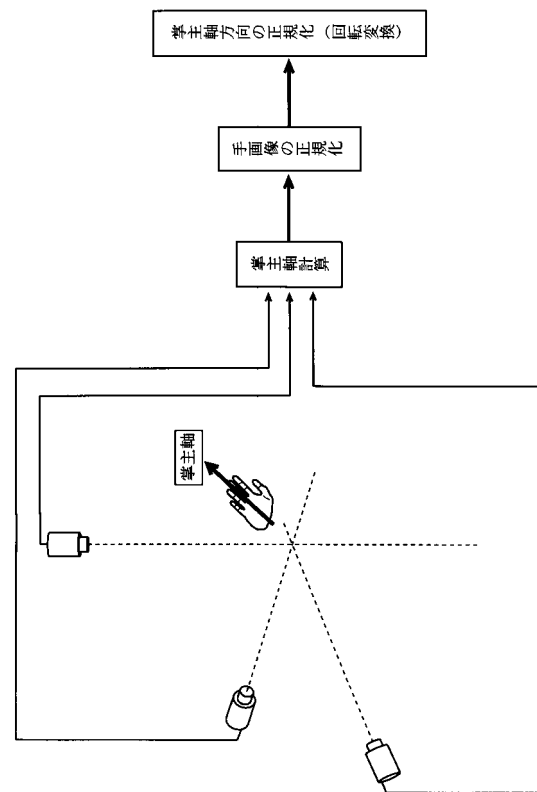
【図 3 4】



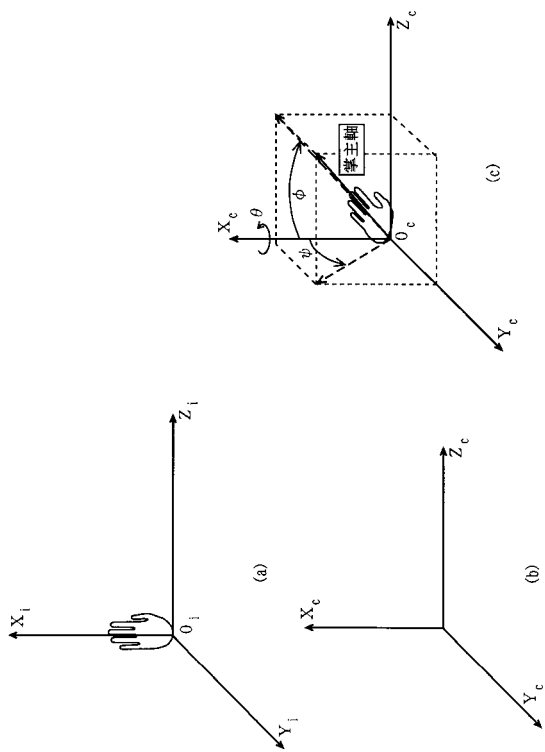
【図 3 5】



【図 3 6】



【図 37】



フロントページの続き

(72)発明者 猪木 誠二

東京都小金井市貫井北町4丁目2番1号 郵政省通信総合研究所内

(72)発明者 呂 山

東京都小金井市貫井北町4丁目2番1号 郵政省通信総合研究所内

審査官 岡本 俊威

(56)参考文献 特開平11-073512(JP,A)

特開平10-003544(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06T 1/00

G06T 7/00-7/60

G09B 21/00