

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6571108号
(P6571108)

(45) 発行日 令和1年9月4日(2019.9.4)

(24) 登録日 令和1年8月16日(2019.8.16)

(51) Int. Cl. F I
G06T 7/20 (2017.01) G O 6 T 7/20 3 0 0 A
G06F 3/01 (2006.01) G O 6 F 3/01 5 7 0

請求項の数 16 (全 21 頁)

(21) 出願番号	特願2016-562725 (P2016-562725)	(73) 特許権者	516200840
(86) (22) 出願日	平成26年12月22日 (2014.12.22)		マノモーション アーバー
(65) 公表番号	特表2017-505965 (P2017-505965A)		スウェーデン国 114 28 ストック
(43) 公表日	平成29年2月23日 (2017.2.23)		ホルム リンドステッドスヴェーゲン 2
(86) 国際出願番号	PCT/SE2014/051562		4
(87) 国際公開番号	W02015/102527	(74) 代理人	100147485
(87) 国際公開日	平成27年7月9日 (2015.7.9)		弁理士 杉村 憲司
審査請求日	平成29年11月15日 (2017.11.15)	(74) 代理人	100161148
(31) 優先権主張番号	61/923,687		弁理士 福尾 誠
(32) 優先日	平成26年1月5日 (2014.1.5)	(74) 代理人	100134577
(33) 優先権主張国・地域又は機関	米国 (US)		弁理士 石川 雅章
		(72) 発明者	シャホーズ ユセフィ
			スウェーデン国 1401 35255
			ベクショー アポテカレ レインホルツ
			ヴェーゲン 18

最終頁に続く

(54) 【発明の名称】 モバイル機器用三次元ジェスチャのリアルタイム認識及び追跡システム

(57) 【特許請求の範囲】

【請求項1】

三次元ジェスチャの認識方法であって、ジェスチャ画像のデータベース(52)へのアクセスを有する機器(20)において実行する方法において、前記機器は、前記三次元ジェスチャの画像(33)をキャプチャするよう構成されたセンサ(34)と通信し、前記ジェスチャ画像のデータベース(52)は、正規化されたジェスチャ画像のインデクス可能な特徴(54)を備え、該インデクス可能な特徴は、前記正規化されたジェスチャ画像のエッジ画像の各画素に対して位置及び方向を備え、該方法は、

前記三次元ジェスチャの前記画像(33)を、前記センサを介してキャプチャするステップ(110)と、

比較を可能とするために、前記データベース(52)の前記正規化されたジェスチャ画像に従って、キャプチャされた画像を正規化するステップ(120)と、

キャプチャされ、正規化された前記画像(33)から、インデクス可能な特徴(36)を得るステップ(130)であって、該インデクス可能な特徴(36)は、前記キャプチャされ、正規化された画像のエッジ画像の各画素に対して位置及び方向を備えるステップ(130)と、

得られた前記インデクス可能な特徴(36)を、前記データベースの前記インデクス可能な特徴(54)と、相似機能を使用して比較するステップ(140)と、

前記データベース(52)において、前記三次元ジェスチャとマッチするジェスチャ画像を、前記比較に基づいて決定するステップ(150)であって、前記ジェスチャ画像の

データベース(52)における各エントリーに、3つの方向及び3つの位置のパラメータを備える、関連する三次元の動きのパラメータをタグ付けするステップと、

前記三次元ジェスチャとマッチする、決定されたジェスチャ画像に関連する三次元の動きのパラメータを、前記データベース(52)からリトリブするステップと、を含む方法。

【請求項2】

請求項1に記載の方法であって、

得られた前記インデクス可能な特徴を比較する前記ステップ(140)は、更に、

前記データベースにおいて、前記キャプチャされた画像とマッチする複数のジェスチャ画像を決定するために、直接相似分析を使用するステップ(141)と、

前記複数のジェスチャ画像の1つとするために、前記データベース(52)において、前記三次元ジェスチャとマッチするジェスチャ画像を決定するステップ(150)と、を含む方法。

【請求項3】

請求項2に記載の方法であって、

得られた前記インデクス可能な特徴を比較する前記ステップ(140)は、

前記キャプチャされた画像とマッチする、前記複数のジェスチャ画像のサブセットを決定するために、前記複数のジェスチャ画像に対して逆相似分析を使用するステップ(142)と、

前記複数のジェスチャ画像の前記サブセットの1つとするために、前記データベース(52)において、前記三次元ジェスチャとマッチするジェスチャ画像を決定するステップ(150)と、

を含む方法。

【請求項4】

請求項1～3の何れか一項に記載の方法であって、更に、

前記キャプチャされ、正規化された画像から、前記エッジ画像の各画素の近接画素に対して位置及び方向を備える追加的なインデクス可能な特徴を得るステップと、

得られた前記追加的なインデクス可能な特徴を、前記データベースの追加的なインデクス可能な特徴と、前記相似機能を使用して比較するステップと、を含む、

前記データベース(52)において、前記三次元ジェスチャとマッチするジェスチャ画像を、前記追加的なインデクス可能な特徴との比較にも基づいて決定する方法。

【請求項5】

請求項1～4の何れか一項に記載の方法であって、

前記三次元ジェスチャとマッチするジェスチャ画像を、ジェスチャ画像のシーケンスにおいて、互いに近いジェスチャ画像を示すジェスチャマップに基づいて決定する方法。

【請求項6】

請求項1～5の何れか一項に記載の方法であって、

前記画像をキャプチャする前記ステップ(110)は、前記三次元ジェスチャの画像のシーケンスをキャプチャするステップを含む方法。

【請求項7】

請求項6に記載の方法であって、更に、

ユーザのジェスチャを、前記画像のシーケンスに基づいて追跡するステップを含み、

前記データベース(52)において、前記三次元ジェスチャとマッチするジェスチャ画像を、追跡された前記ユーザのジェスチャにも基づいて、決定する方法。

【請求項8】

請求項1～7の何れか一項に記載の方法であって、更に、

ディスプレイ上に提示されたビジュアルコンテンツを修正するために、前記三次元ジェスチャとマッチする、決定されたジェスチャ画像を使用するステップ(160)を含む方法。

10

20

30

40

50

【請求項 9】

三次元ジェスチャを認識する機器(20)であって、正規化されたジェスチャ画像の、インデクス可能な特徴(54)を備えるジェスチャ画像のデータベース(52)へのアクセスを有するよう構成される機器(20)において、

前記インデクス可能な特徴は、前記正規化されたジェスチャ画像のエッジ画像の各画素に対して位置及び方向を備え、前記機器(20)は、前記三次元ジェスチャの画像(33)をキャプチャするよう構成されたセンサ(34)に接続可能であり、及び処理ユニット(24)を備え、

前記処理ユニット(24)は、

前記センサ(34)を介して前記三次元ジェスチャの前記画像(33)をキャプチャし

10

、
比較を可能とするために、前記データベース(52)の前記正規化されたジェスチャ画像に従って、キャプチャされた画像を正規化し、

キャプチャされ、正規化された前記画像(33)から、インデクス可能な特徴(36)を得て、前記インデクス可能な特徴(36)は、前記キャプチャされ、正規化された画像のエッジ画像の各画素に対して位置及び方向を備え、

得られた前記インデクス可能な特徴(36)を、前記データベースの前記インデクス可能な特徴(54)と、相似機能を使用して比較し、及び

前記データベース(52)において、前記三次元ジェスチャとマッチするジェスチャ画像を、前記比較に基づいて決定し、前記ジェスチャ画像のデータベース(52)における各エントリは、3つの方向及び3つの位置のパラメータを備える、関連する三次元の動きのパラメータをタグ付けされ、

20

前記三次元ジェスチャとマッチする、決定されたジェスチャ画像に関連する三次元の動きのパラメータを、前記データベース(52)からリトリブする、よう構成される機器(20)。

【請求項 10】

請求項 9 に記載の機器(20)であって、

前記処理ユニット(24)は更に、前記データベースにおいて、前記キャプチャされた画像とマッチする、複数のジェスチャ画像を決定するために、得られた前記インデクス可能な特徴を、直接相似分析を使用して比較するよう構成され、

30

前記処理ユニット(24)は更に、前記複数のジェスチャ画像の1つとするために、前記データベース(52)において、前記三次元ジェスチャとマッチするジェスチャ画像を決定する、よう構成される機器(20)。

【請求項 11】

請求項 10 に記載の機器(20)であって、

前記処理ユニット(24)は更に、前記キャプチャされた画像とマッチする、前記複数のジェスチャ画像のサブセットを決定するために、得られた前記インデクス可能な特徴を、前記複数のジェスチャ画像に対する逆相似分析を使用して比較するよう構成され、

前記処理ユニット(24)は更に、前記複数のジェスチャ画像のサブセットの1つとするために、前記データベース(52)において、前記三次元ジェスチャとマッチするジェスチャ画像を決定する、よう構成される機器(20)。

40

【請求項 12】

請求項 9 ~ 11 の何れか一項に記載の機器(20)であって、

前記処理ユニット(24)は、更に、

前記キャプチャされ、正規化された画像から、前記エッジ画像の各画素の近接画素に対して位置及び方向を備える追加的なインデクス可能な特徴を得て、

得られた前記追加的なインデクス可能な特徴を、前記データベースの追加的なインデクス可能な特徴と、前記相似機能を使用して比較し、

前記データベース(52)において、前記三次元ジェスチャとマッチするジェスチャ画像を、前記追加的なインデクス可能な特徴との比較にも基づいて決定する、よう構成され

50

る機器（20）。

【請求項13】

請求項9～12の何れか一項に記載の機器（20）であって、

前記処理ユニット（24）は更に、前記三次元ジェスチャとマッチするジェスチャ画像を、ジェスチャ画像のシーケンスにおいて、互いに近いジェスチャ画像を示すジェスチャマップに基づいて決定する、よう構成される機器（20）。

【請求項14】

請求項9～13の何れか一項に記載の機器（20）であって、

前記処理ユニット（24）は更に、前記三次元ジェスチャの画像のシーケンスを、前記センサ（34）を介してキャプチャする、よう構成される機器（20）。

10

【請求項15】

請求項14に記載の機器（20）であって、

前記処理ユニット（24）は、更に、

ユーザのジェスチャを、前記画像のシーケンスに基づいて追跡し、

前記データベース（52）において、前記三次元ジェスチャとマッチするジェスチャ画像を、追跡された前記ユーザのジェスチャにも基づいて決定する、よう構成される機器（20）。

【請求項16】

請求項9～15の何れか一項に記載の機器（20）であって、

前記処理ユニット（24）は更に、ディスプレイ上に提示されたビジュアルコンテンツを修正するために、前記三次元ジェスチャとマッチする、決定されたジェスチャ画像を使用する、よう構成される機器（20）。

20

【発明の詳細な説明】

【技術分野】

【0001】

本開示はジェスチャ認識に関し、より具体的には、三次元ジェスチャを認識する機器及び方法に関する。

【背景技術】

【0002】

人間の手は、27の自由度（Dof：degree of freedom）を有する。つまり、伸縮と屈曲のために3つ、及び外転と内転のために1つで、各指に4つの自由度がある。親指は更に複雑であって、5つの自由度を有する。手首の回転及び並進のためには、6つの自由度が残されている。手指の動きをビデオシーケンスでキャプチャすることは、手のキネマティクスにおいて極めて多数の自由度が存在するため、高度に挑戦的な課題である。こうした過程は、手で持つスマート機器においては、電力が限られ、かつ計算コストが高いため、更により複雑になる。

30

【0003】

基本的に、一般的に存在する解決策においては、図1に示すステップに従う。センサによりキャプチャされた画像シーケンスは、ユーザの手/指をセグメント化すべく、分析される。背景除去、分類、特徴検出等の画像分析アルゴリズムは、手/指を検出するために使用される。実際には、手の追跡及びジェスチャ認識のために存在するアルゴリズムは、2つのカテゴリーにグループ分け可能である。つまり、外観に基づくアプローチと三次元ハンドモデルに基づくアプローチ（米国特許出願公開第2010053151A1号明細書、米国特許出願公開第2010159981A1号明細書、国際公開第2012135545A1号パンフレット、及び米国特許出願公開第2012062558A1号明細書）である。前者は、手のジェスチャを、二次元画像特徴と直接に比較することに基づいている。人間のジェスチャを検出するのに使用される、定評のある画像特徴には、手の色及び形状、局所的な手の特徴等が含まれる。特徴に基づくアプローチの欠点は、手の特徴を抽出するために、一般的には、鮮明な画像セグメンテーションが要求されることである。これは、例えばクラッタ背景の場合には簡単な課題ではない。さらに、人間の手は関節

40

50

により高度に結ばれており、自己遮蔽のために、局所的な手の特徴を見出すことが困難であることが多く、多種多様な手のジェスチャを処理するには、ある種のヒューリスティックが必要である。手を直接に表示するために、二次元画像特徴を使用するのに替えて、三次元ハンドモデルに基づくアプローチにおいては、ハンドポーズをレンダリングするために、三次元キネマティックハンドモデルを使用する。合成による分析 (A B S : Analysis-by-synthesis) 戦略を使用して、手の動きのパラメータを回復するために、三次元ハンドモデルによって投影された外観を、カメラからの観察画像に位置合わせする。一般に、二次元画像特徴がより単純であるために、外観に基づくアプローチによってリアルタイムに実行する方が、より容易である。しかしながらこの種のアプローチが処理可能なのは、指先の検出や追跡というような単純な手のジェスチャのみである。それとは対照的に、三次元ハンドモデルに基づくアプローチによっては豊かな描写が提供され、潜在的に、広範なクラスの手のジェスチャを扱うことが可能である。主たる挑戦的課題は、三次元の手が、27の自由度を有する複雑なモデルだということである。従って、手の画像の全ての特徴を異なる視野で網羅するためには、極めて大きいデータベースが必要となる。ビデオ入力からのクエリ画像を、データベース内の全ての手の画像とマッチングするのは、時間がかかり、かつ計算コストも高い。こうしたことから、三次元ハンドモデルに基づく既存のアプローチの多くは、照明及び背景状態を制限しつつ、汎用的な手の動きをリアルタイムで追跡することに注力している。

10

【先行技術文献】

【特許文献】

20

【0004】

【特許文献1】米国特許出願公開第2010053151A1号明細書

【特許文献2】米国特許出願公開第2010159981A1号明細書

【特許文献3】国際公開第2012135545A1号パンフレット

【特許文献4】米国特許出願公開第2012062558A1号明細書

【発明の概要】

【発明が解決しようとする課題】

【0005】

課題は、上記に概略的に述べた問題のいくつかに対処し、及びリアルタイムのジェスチャ認識を計算的に効率化する解決策を提供することである。この課題等は、独立請求項に記載の方法及び装置、ならびに従属請求項に記載の実施形態により解決される。

30

【課題を解決するための手段】

【0006】

第1の態様により、三次元ジェスチャのリアルタイム認識方法が提供される。方法は、ジェスチャ画像のデータベースへのアクセスを有する機器において実行する。機器は、三次元ジェスチャの画像をキャプチャするよう構成されたセンサと通信する。ジェスチャ画像のデータベースは、正規化されたジェスチャ画像の、インデクス可能な特徴を備える。インデクス可能な特徴は、正規化されたジェスチャ画像のエッジ画像の各画素に対して位置及び方向を備える。方法は、センサを介して三次元ジェスチャの画像をキャプチャするステップ、及びキャプチャされた画像を、データベースの正規化されたジェスチャ画像に従って正規化するステップを含む。方法はまた、キャプチャされ、正規化された画像からインデクス可能な特徴を得るステップを含む。インデクス可能な特徴は、キャプチャされ、正規化されたジェスチャ画像のエッジ画像の各画素に対して位置及び方向を備える。方法は更に、得られたインデクス可能な特徴を、データベースのインデクス可能な特徴と、相似機能を使用して比較するステップ、及びデータベースにおいて、三次元ジェスチャとマッチするジェスチャ画像を、比較に基づいて決定するステップを含む。

40

【0007】

第2の態様により、三次元ジェスチャを認識する機器が提供される。機器は、正規化されたジェスチャ画像の、インデクス可能な特徴を備えるジェスチャ画像のデータベースへのアクセスを有するよう構成される。インデクス可能な特徴は、正規化されたジェスチャ

50

画像のエッジ画像の各画素に対して位置及び方向を備える。機器は、三次元ジェスチャの画像をキャプチャするよう構成されたセンサに接続可能である。機器は、処理ユニットを備える。処理ユニットは、センサを介して三次元ジェスチャの画像をキャプチャし、キャプチャされた画像を、データベースの正規化されたジェスチャ画像に従って正規化し、及びキャプチャされ、正規化された画像から、インデクス可能な特徴を得るよう構成される。インデクス可能な特徴は、キャプチャされ、正規化されたジェスチャ画像のエッジ画像の各画素に対して位置及び方向を備える。処理ユニットはまた、得られたインデクス可能な特徴を、データベースのインデクス可能な特徴と、相似機能を使用して比較するよう構成される。さらに処理ユニットは、データベースにおいて、三次元ジェスチャとマッチするジェスチャ画像を、比較に基づいて決定するよう構成される。

10

【0008】

実施形態の効果により、高解像度のジェスチャ認識を、計算的リソースを低減しつつ、リアルタイムで可能である。

【0009】

実施形態に関する、他の目的、効果及び特徴は、別添の図面及び請求項に関連する考察において、以下に詳述される。

【図面の簡単な説明】**【0010】**

【図1】従来技術によるジェスチャ追跡及び認識方法を示す、概略的なフローチャートである。

20

【図2A】本発明の実施形態による、モバイルプラットフォームにおける三次元ユーザインタフェースシステムの概略図である。

【図2B】本発明の実施形態による、ウェアラブル機器を使用した三次元ユーザインタフェースの概略図である。

【図2C】本発明の実施形態による、固定プラットフォームにおける三次元ユーザインタフェースの概略図である。

【図3】本発明の実施形態による方法及びシステムの概略図である。

【図4】本発明の実施形態による、ジェスチャエントリのデータベースへの記憶方法を示す、概略的なフローチャートである。

【図5】本発明の実施形態による、ジェスチャエントリを検索し、クエリ入力に対するマッチを見出す方法を示す、概略的なフローチャートである。

30

【図6】本発明の実施形態による、画像クエリの処理方法を示す概略的なフローチャートである。

【図7】本発明の実施形態による、インタフェースレベルの方法を示す、概略図である。

【図8】本発明の実施形態による、図2Aに示すモバイル機器20の概略図である。

【図9】本発明の実施形態による、図2Bに示すウェアラブル機器20の概略図である。

【図10】本発明の実施形態による、図2Cに示す固定機器20の概略図である。

【図11a】本発明の実施形態による機器が実行する方法を概略的に示す、フローチャートである。

【図11b】本発明の実施形態による機器が実行する方法を概略的に示す、フローチャートである。

40

【図12】本発明の実施形態による機器の概略図である。

【発明を実施するための形態】**【0011】**

人間と未来のモバイル機器の間のインタラクションを設計する上で、三次元ジェスチャ認識は、要求度合いが極めて高い特徴である。具体的には、仮想現実環境又は拡張現実環境において、物質世界との直感的なインタラクションが不可避であり、及び三次元ジェスチャによるインタラクションは、軌道パッド及びタッチスクリーンのような現在の入力装置に対して、最も効果的な代替装置となり得るだろう。本発明の実施形態において、三次元ジェスチャ認識及び追跡のための解決策が提供される。提供される方法論及びシステム

50

は、極めて大きいジェスチャデータベースにおいて、マッチを見出すことに基づくものである。このデータベースは、回転及び位置において可能な全てのバリエーションと、対応する位置/方向パラメータを含む、多様な種類の手のジェスチャをキャプチャしたエントリーを含む。クエリ入力とデータベースエントリー間の属性に関して、相似分析を行なう。システムは、取得されたクエリ入力に対して、データベースエントリーと注釈付情報を含む、マッチをリトリートする。

【0012】

大量の電力、計算及びメモリを必要とする古典的なコンピュータビジョンアプローチと異なり、新たなフレームワークは、同一の問題ではあるが、それを全く異なるアプローチを使用して解決するよう定義される。提案される技術は、例えば自由度の高い複雑な手の動きを、大規模検索フレームワークによって処理可能である。一方で現在の技術は、低解像度のジェスチャ認識及び追跡に限定される。

10

【0013】

一般的なモバイル機器へ適用するには、手のジェスチャ/体のジェスチャの全ての範囲をカバーする必要がある、人間のジェスチャが存在する高次元空間を包括的に検索するという挑戦的な問題を処理するために、ジェスチャ画像を大規模に検索するための、効果的なインデキシングアルゴリズムを提案する。開示するシステムの利点は、自由度の高い手の動きを、多様な照明条件においてノイズやクラッタの存在を許容しつつ処理可能な、膨大な個数のデータベース画像を、極めて迅速にリトリート可能なことである。解決策は、リアルタイム、低複雑性及び頑強性を有すると同時に、高解像度での追跡と精度を有する等の、モバイル機器への適用に関する特殊要件に対して適合している。

20

【0014】

本発明の実施形態により、ビジョンセンサ、又はモバイルカメラ、ウェブカメラ、深度センサ又は超音波センサのような他のタイプのセンサを装備した、任意のモバイル機器、ウェアラブル機器、又は固定機器によっても、例えば手、頭又は体のジェスチャといった人間のジェスチャを、三次元空間において決定可能、又は認識可能である。ジェスチャ追跡は、クエリ入力のシークエンスにおいて決定又は認識されたジェスチャを使用して実行する。認識及び追跡は、注釈付ジェスチャエントリーの極めて大きいデータベース(DB)を検索する、高度な検索システムに基づいている。データベースは、三次元空間における全ての变形とバリエーションを取り込んだ、可能な全ての手のジェスチャを含むものであり、それは数百万エントリーに対応するかもしれない。提案するシステムは常に、任意のクエリジェスチャに対して自動的にデータベースを検索し、ベストマッチをリトリートする。この結果、リアルタイムの三次元ジェスチャ追跡が実行される。この技術により、直感的な三次元インタラクションを使用可能であろうリアルタイムの適用において、ユーザデバイスでのインタラクションが促進される。本発明の実施形態は、スマートフォン及び拡張現実メガネのような、モバイル機器/ウェアラブル機器でのインタラクションを補助するよう設計され、固定機器、モバイル機器及び他のデジタル機器に対しても使用可能である。

30

【0015】

図3は、スマートフォン、モバイル機器、ウェアラブルスマート機器、固定システム及びデジタルガジェットに適用可能な方法論及びシステムを含む、本発明の実施形態を示す。図3の実施形態は、4つの主たる部品を含む。つまり、前処理済の注釈付及びインデクス付ジェスチャデータベース50、クエリジェスチャを受信する、画像クエリ処理ユニット30、クエリジェスチャを受信し、ジェスチャのデータベースから自動的にベストマッチをリトリートする、リアルタイムのジェスチャ検索エンジン70、そして最後に、検索エンジンからの出力を受信し、それを継続中のアプリケーションに適用する、インタフェースレベル90である。必要とされるハードウェアのプラットフォームは、任意のデジタル機器とする。

40

【0016】

(システムに関する説明)

50

図2Aは、本発明の実施形態による、三次元ユーザインタフェースシステム200Aの概略図である。ユーザインタフェースは、任意の種類(例えば二次元/三次元カメラ、超音波、三次元深度カメラ、IRカメラ)を装備した、任意の種類(モバイル、固定型、ウェアラブル等)のスマート機器20に基づく。センサ34は、機器20の後ろ、正面、及び/又は機器20の周囲の三次元シーン情報をキャプチャする。この三次元シーン情報は、人間であるユーザ10の、例えば手、頭又は体のジェスチャであるジェスチャ32を含む。このジェスチャ32(手/頭/体のジェスチャ)を検出/認識すべく、スマート機器20は、十分な解像度でジェスチャ画像をキャプチャし、ジェスチャ32(手/頭/体のジェスチャ)、ならびにその特定の位置及び方向の抽出を可能にする。位置は、画像におけるジェスチャ中心(x, y)の空間的座標に、ジェスチャスケール(z方向のセンサからの距離)を加えて表す。そして方向は、手のジェスチャの、センサの三次元座標(x, y, z)に対する相対的な方向である。ジェスチャ32(手/頭/体のジェスチャ)に加え、キャプチャされた画像又はクエリ画像33は通常、体の他の部分、及び/又はクラッタ背景を含む。

【0017】

図2Aにおいて、システム200Aは、ユーザのジェスチャ32(手/頭/体のジェスチャ)を含むクエリ画像33のシーケンスをキャプチャして処理する。ユーザ10がジェスチャ32(手/頭/体のジェスチャ)を実行する間、システム200Aはユーザのジェスチャ32(手/頭/体のジェスチャ)を、クエリ画像33のシーケンスに亘って追跡する。機器20における処理ユニット24上で動作するソフトウェア、及び/又はキャプチャリングセンサ34は、画像シーケンスを処理し、各クエリ画像33におけるユーザのジェスチャ32の、インデクス可能な特徴36をリトリブする。詳細は以下に説明する。ソフトウェアは、クエリ画像33に対するベストマッチを見出すために、抽出されたインデクス可能な特徴36を、インデクス付特徴の大規模語彙テーブル72とマッチさせる。詳細は以下に説明する。大規模語彙テーブルは、データベース画像からの、インデクス可能な特徴の大規模マトリクスである。

【0018】

データベース52は、手のジェスチャの数百万の画像から構成される。手のジェスチャの画像は、特定の三次元の動きのパラメータ(3つの位置及び3つの方向のパラメータ)58を注釈付けされる。詳細は以下に説明する。クエリ入力33に対して、データベース52におけるベストの手のジェスチャを見出すことで、クエリ入力33の三次元の動きのパラメータが供給される。

【0019】

また図5の方法では、検索プロセスを最適化し、及びそのスピードを増すために、シーケンスにおける複数のフレームに亘ってジェスチャマップ73を分析する。詳細は以下に説明する。

【0020】

システムは、ユーザのジェスチャ32をクエリ入力33のシーケンスに亘って追跡する、モーション追跡機能をも含むため、図5の方法は任意に、各2つ(以上)のフレームにおいて、一度のみの実行としてもよい。

【0021】

検出/認識された出力(アクション/ジェスチャ/三次元の動き、注釈付画像等)92は、アプリケーションプログラミングインターフェイス(API)を介して、機器20上で動作するアプリケーションプログラムに供給される。このプログラムにより、ディスプレイ100に提示された画像、三次元オブジェクト、又は他の二次元/三次元ビジュアルコンテンツ94を、実行されたジェスチャ32に反応して、例えば移動及び変更できる。

【0022】

代替案として、これらの処理機能の全て、又はいくつかを、ゲーム機器、メディアプレーヤ、スマートテレビ等の、任意の他のコンピュータ化デバイスと一体化した適切なプロセッサにより実行してもよい。キャプチャセンサ34(二次元/三次元カメラ、IRセン

10

20

30

40

50

サ、超音波等)を装備した任意のコンピュータ化装置、記憶ユニット22及び処理ユニット24は、上述の機能のうちの少なくともいくつかにおいて使用可能であり、より良好なユーザインタフェースシステムを供給可能である。

【0023】

(ジェスチャ画像のデータベースの提供)

図4は、注釈付ジェスチャ画像のデータベース52の、インデクス可能な特徴54の形成方法50を示すフローチャートである。データベースは、方向、位置及びスケールングにおける全ての潜在的バリエーションを有する手のジェスチャエントリーの、異なる実像56の大きなセットを含んでいる。データベースはまた、関節で結ばれた三次元ハンドモデル/三次元図形モデル等により合成し、既知の位置及び方向パラメータを有する、全ての手のジェスチャ図形57を含む。

10

【0024】

クエリ入力33とデータベースの間のマッチングに加えて、達成すべき重要な特徴の一つは、クエリ入力33から三次元の動きのパラメータ(三次元に対応する3つの位置及び3つの方向のパラメータ)をリトリブすることである。クエリ入力33は三次元の動きのパラメータ(3つの位置及び3つの方向のパラメータ)を含まないため、最善の解決策は、クエリ入力33の三次元の動きのパラメータ(3つの位置及び3つの方向のパラメータ)を、データベースからリトリブされたベストマッチと関連付けることである。このために、データベースエントリーに、そのグランドトゥルースな三次元の動きのパラメータ(3つの位置及び3つの方向のパラメータ)58をタグ付けする。これは、ビジョンベースのシステム、磁気センサ、IMU等、任意のモーションキャプチャシステムにより実行可能である。ジェスチャエントリーの他のソース59もまた、データベースを拡張するために使用する。手のジェスチャ画像、注釈付ジェスチャ画像のデータベース52に対して、三次元の動きのパラメータ(3つの位置及び3つの方向のパラメータ)のタグを形成する。注釈付ジェスチャ画像のデータベース52における各エントリーは、純粋なジェスチャエントリー(背景及びノイズフリー)を表す。方法50において、注釈付ジェスチャ画像のデータベース52における各エントリーに対して、インデクス可能な特徴54を抽出する。インデクス可能な特徴54は、注釈付ジェスチャ画像のデータベース52におけるエントリーから得られたエッジ画素の、正確な位置及び方向を含んだ低レベルのエッジ方向属性を含む。単一のエッジ画素の各々を、二次元画像座標における短線とみなした場合、エッジ画素の方向は、この短線の、画像座標の原点に対する角度となる。技術的には、エッジ画素の方向を、x及びy方向に対する画像の傾斜から算出可能である。

20

30

【0025】

インデクス可能な特徴54を抽出するために、注釈付ジェスチャ画像のデータベース52における全てのエントリーを正規化し、及びそれらのエントリーの対応するエッジ画像を計算する。エッジ画像は、ジェスチャ画像をフィルタリングすることで計算してよい。コンピュータビジョンの分野では、異なるエッジ検出器が知られており、それも同様に使用可能である。単一のエッジ画素の各々は、その位置及び方向によって表される。低レベルのエッジ方向特徴用の汎用的構成を作成するために、各々のエッジ特徴が起こりうる可能な全てのケースを表すよう、大規模語彙テーブル72を形成する。データベース全体を、エッジの位置及び方向に関して考慮すると、大規模語彙テーブル72は、ジェスチャの語彙全体をエッジ画素フォーマットで表現可能である。エッジ画素フォーマットは、その位置及び方向の観点から、エッジ画像の各画素を表す。

40

【0026】

(画像クエリの処理)

図6は、画像クエリの処理方法30を概略的に示すフローチャートである。クエリ入力33は、ユーザ10のジェスチャ32(手/頭/体のジェスチャ)を、センサ34(二次元/三次元カメラ、IRセンサ、超音波等)によりキャプチャされた、ジェスチャ32の特定の3つの位置及び3つの方向のパラメータで特徴付けする。センサ34は、機器20の後ろ、又は正面の三次元シーン情報をキャプチャする。スマート機器20は、クエリ入

50

力 3 3 のシーケンスをキャプチャし、及びそれら进行处理して、インデクス可能な特徴 3 6 をリトリブする。方法 3 0 は、クエリ入力 3 3 から、インデクス可能な特徴 3 6 を抽出する。インデクス可能な特徴 3 6 は、クエリ入力 3 3 から得られたエッジ画素の、正確な位置及び方向を含む低レベルのエッジ方向属性を含む。

【 0 0 2 7 】

インデクス可能な特徴 3 6 を抽出するために、クエリ入力 3 3 を正規化し、及びそれらのエントリーの対応するエッジ画素を計算する。単一のエッジ画素の各々は、その位置及び方向によって表される。

【 0 0 2 8 】

基本的に、ユーザのジェスチャ 3 2 (手/頭/体のジェスチャ) をキャプチャするクエリ入力 3 3 は、無関係のオブジェクト、環境騒音等に起因するクラッタ背景を含んでいる。従って、クエリ入力 3 3 からリトリブしたインデクス可能な特徴 3 6 は、ジェスチャ 3 2 及びノイズな背景からの、双方の特徴を含む。一方、注釈付ジェスチャ画像のデータベース 5 2 における各エントリーは、純粋なジェスチャエントリー(背景及びノイズフリー)を表すため、注釈付ジェスチャ画像のデータベース 5 2 における各エントリーからリトリブしたインデクス可能な特徴 5 4 は、純粋なジェスチャからの特徴のみを表す。従って、クエリ画像のエッジ画像は、データベース画像のエッジ画像のように正確には定義できない。

【 0 0 2 9 】

(ジェスチャ検索エンジン)

図 5 は、ジェスチャ検索エンジン 7 0 の方法を示す。注釈付ジェスチャ画像のデータベース 5 2 における各エントリーの、抽出されたインデクス可能な特徴 5 4 は、インデクス可能な特徴 7 2 の大規模語彙テーブル 7 2 を、ジェスチャ検索エンジン 7 0 において構築する。

【 0 0 3 0 】

インデクス付特徴の大規模語彙テーブル 7 2 は、各エッジ特徴が起こりうる可能な全てのケースを表すよう形成する。データベース全体を、エッジの位置及び方向に関して考慮すると、大規模語彙テーブル 7 2 は、ジェスチャの語彙全体をエッジ画素フォーマットで表現可能である。例えば、 $p \times q$ 画素の画像サイズ、及び L エッジ方向表示、ジェスチャ画像 N のデータベースに対して、語彙テーブル 7 2 は、 $p \times q \times L$ コラム及び N ロウを有する。従って、語彙テーブル 7 2 は、特定のロウ及びコラムにおける特徴を有する、全てのデータベース画像 5 2 のインデクスで満たされている。語彙テーブル 7 2 は、要求される情報を全データベース 5 2 から収集するものであり、これはジェスチャ検索エンジン 7 0 ための方法において、必須の機能である。

【 0 0 3 1 】

直接相似分析機能 7 5 は、クエリ画像 3 3 においてユーザのジェスチャ 3 2 を検出/認識するために、大規模検索テーブル 7 2、及び各クエリ画像 3 3 に対してリトリブされたインデクス可能な特徴 3 6 を用いて、注釈付ジェスチャ画像のデータベース 5 2 におけるトップ m 第 1 レベルのマッチを選択する。

【 0 0 3 2 】

エッジ画素フォーマットの各クエリ入力 3 3 は、エッジポイントのセットを含む。エッジポイントのセットは、ロウ コラム位置及び特定の方向により表現可能である。直接相似機能分析 7 5 は、クエリ入力 3 3 に関して、リトリブしたインデクス可能な特徴 3 6 の相似性を、インデクス付特徴の大規模語彙テーブル 7 2 を使用し、全てのエッジ特徴の位置及び特定の方向に基づいて計算する。直接相似分析機能は、対となったデータ値にスコアを割り当てる機能である。スコアは、クエリのインデクス付特徴の、データベースにおける各エントリーのインデクス付特徴に対する相似性を示す。もし、ある状態がクエリ入力 3 3 においてリトリブしたインデクス可能な特徴 3 6、及び注釈付ジェスチャ画像のデータベース 5 2 においてリトリブしたインデクス可能な特徴 5 4 の双方に対して満たされた場合、直接相似分析機能 7 5 は、 $+K$ 1 ポイントを、そうした特定のロウ コラ

10

20

30

40

50

ムポイントにおいて類似の方向を有するエッジを持つ、全てのデータベース画像52に対して割り当てる。直接相似分析機能75は、クエリ入力33の単一のエッジ画素フォーマットの各々に対して、上記のプロセスを実行する。

【0033】

直接相似分析機能75の第1ステップは、クエリ入力33及びデータベース画像52からの2つのエッジパターンが、互いに正確に重なり合ったケースを満たすものである。一方、実際の多くのケースでは、2つの類似したパターンは、位置においては互いに極めて近いが、これらパターンの間の重複は大きくない。直接相似分析機能75は、通常起こるこうしたケースに対して、第1及び第2レベルの近接画素に基づいてエクストラポイントを割り当てる。

10

【0034】

極めて蓋然性が高いのは、2つの極めて類似したパターンが、重なり合いはしないが、互いに近接する画素に該当するケースである。これらのケースについて考慮するために、直接相似分析機能75の第1ステップの他に、任意の単一画素に対して、データベース画像における第1レベルの8近接画素及び第2レベルの16近接画素をも考慮し、エクストラポイントを割り当てるべきである。任意の単一画素の第1レベルの8近接画素とは、単一画素を包囲する画素である。第2レベルの近接画素は、第1レベルの8近接画素を包囲する16画素を含む。第1レベル及び第2レベルの近接画素のうち、類似の方向を有するエッジを持つ全てのデータベース画像52は、各々+K2及び+K3ポイントを受け取る($K1 > K2 > K3$)。要するに、データベース画像の類似性に関して、重みの異なる3つのレベルで、クエリにおける全てのエッジ画素に対して直接相似分析75を実行するのである。最後に、各データベース画像の蓄積されたスコアを計算及び正規化し、そして最高スコアを、トップm第1レベルのマッチとして選択する。

20

【0035】

トップm第1レベルのマッチのうちで最も近いマッチを見出すために、逆相似分析76を実行する。逆相似分析76とは、注釈付ジェスチャ画像のデータベース52のエントリーに対して、クエリジェスチャ32の相似性を見出すことの他に、注釈付ジェスチャ画像のデータベース52において選択したトップmエントリーの、クエリジェスチャ32に対する逆相似性を計算すべきであることを意味する。逆相似機能は、精度向上を理由として使用する。逆相似分析を使用しない場合、リトリブの精度は低下するが、複雑性が低減される。

30

【0036】

逆相似分析76は、与えられたユーザのジェスチャ32に対して、注釈付ジェスチャ画像のデータベース52からベストnマッチ($n < m$)を返す。直接相似分析75及び逆相似分析76の組み合わせは、クエリ入力33に対して、注釈付ジェスチャ画像のデータベース52からベストマッチを返すものである。

【0037】

ジェスチャ検索エンジン70における別の任意のステップは、ジェスチャ近接性分析機能77を使用する、ジェスチャ検索のスムーズ度である。スムーズ度とは、三次元ジェスチャインタラクションのシークエンスにおいてリトリブしたベストマッチが、スムーズな動きを表すべきであることを意味する。スムーズにリトリブするために、注釈付ジェスチャ画像のデータベース52におけるエントリーを分析し、及びジェスチャマップ73を検出するべく、高次元空間に位置を割り当てる。ジェスチャマップ73は、どのジェスチャが互いに近いが、及びどのジェスチャが高次元において同一の近接位置に該当するかを示すものである。従って、シークエンスにおけるクエリ入力33に対して、直接相似分析機能75を実行した後に、逆相似分析機能76によって逆相似性を計算し、トップマッチを選択する。その後、方法70がジェスチャマップ73を検索し、これらのトップマッチのうちの何れが以前のフレームマッチにより近いかを確認する。そして、最も近いエントリーを、注釈付画像のデータベース52から最後のベストマッチとして選択する。その後、ベストマッチに対してタグ付けされた三次元の動きのパラメータ58(3つの位置及び

40

50

3つの方向のパラメータ)を、ディスプレイ100上で動作する様々なアプリケーションのシナリオを促進させるために、即座に使用可能である。

【0038】

(インタフェース)

図7は、検索エンジン70の検出/認識の出力(アクション/ジェスチャ/三次元の動き、注釈付画像等)92を受信する、インタフェースレベル90の方法を示す概略図である。検出/認識されたパラメータ(アクション/ジェスチャ/三次元の動き、注釈付画像等)92は、アプリケーションプログラミングインターフェイス(API)を介して、機器20上で動作するアプリケーションプログラムに供給される。アプリケーションは、ディスプレイ100上に提示された二次元/三次元ビデオゲーム、二次元/三次元オブジェクトモデリング/レンダリング、フォトブラウジング、マップ、ナビゲーション等を含んでよい。ユーザ10は、ディスプレイ100上に提示されたビジュアルコンテンツの出力(二次元/三次元)94を知覚する。この出力は、ユーザの実行するジェスチャ32に反応して、連続的に修正される。

10

【0039】

(機器の詳細な説明)

図8は、図2Aのモバイル機器20を示す。モバイル機器20は、記憶ユニット22、処理ユニット24、センサ34(例えば二次元/三次元カメラ、IRセンサ、超音波等)及びディスプレイ100からなる。センサ34は、機器20の正面の三次元のシーン情報をキャプチャする。機器20は、機器20の後ろの三次元のシーン情報をキャプチャする、リアセンサ34(例えば二次元/三次元カメラ、IRセンサ、超音波等)を含んでもよい。モバイル機器20は、クエリ入力33のシークエンスをキャプチャし、及びそれら进行处理して、インデクス可能な特徴36をリトリブする。記憶ユニット22は、注釈付ジェスチャ画像のデータベース52、インデクス付特徴の大規模語彙テーブル72及びジェスチャマップ73を記憶する。処理ユニット24は、画像クエリの処理方法30及び検索エンジン70の方法を実行する。処理ユニット24はまた、ディスプレイ100上に提示されたビジュアルコンテンツの出力(二次元/三次元)94を、ユーザの実行するジェスチャ32に反応して修正する。ディスプレイ100は、機器20上で動作するアプリケーションを表示する。アプリケーションは、ディスプレイ100上に提示された二次元/三次元ビデオゲーム、二次元/三次元オブジェクトモデリング/レンダリング、フォトブラウジング、マップ、ナビゲーション等を含んでよい。ユーザ10は、ディスプレイ100上に提示されたビジュアルコンテンツの出力(二次元/三次元)94を知覚する。この出力は、ユーザの実行するジェスチャ32に反応して、連続的に修正されている。

20

30

【0040】

図9は、図2Bのウエアラブル機器20を示す。ウエアラブル機器20は、記憶ユニット22、処理ユニット24、センサ34(例えば二次元/三次元カメラ、IRセンサ、超音波等)及びディスプレイ100からなる。センサ34は、ウエアラブル機器20の正面の三次元のシーン情報をキャプチャする。ウエアラブル機器20は、クエリ入力33のシークエンスをキャプチャし、及びそれら进行处理して、インデクス可能な特徴36をリトリブする。記憶ユニット22は、注釈付ジェスチャ画像のデータベース52、インデクス付特徴の大規模語彙テーブル72及びジェスチャマップ73を記憶する。処理ユニット24は、画像クエリの処理方法30及び検索エンジン70の方法を実行する。処理ユニット24はまた、ディスプレイ100上に提示されたビジュアルコンテンツの出力(二次元/三次元)94を、ユーザの実行するジェスチャ32に反応して修正する。ディスプレイ100は、ウエアラブル機器20上で動作するアプリケーションを表示する。アプリケーションは、ディスプレイ100上に提示された二次元/三次元ビデオゲーム、二次元/三次元オブジェクトモデリング/レンダリング、フォトブラウジング、マップ、ナビゲーション等を含んでよい。ユーザ10は、ディスプレイ100上に提示されたビジュアルコンテンツの出力(二次元/三次元)94を知覚する。この出力は、ユーザの実行するジェスチャ32に反応して、連続的に修正されている。

40

50

【 0 0 4 1 】

図 1 0 は、図 2 C の固定機器 2 0 を示す。固定機器 2 0 は、記憶ユニット 2 2、処理ユニット 2 4、センサ 3 4（二次元/三次元カメラ、I R センサ、超音波等）及びディスプレイ 1 0 0 からなる。センサ 3 4 は、固定機器 2 0 の正面の三次元のシーン情報をキャプチャする。固定機器 2 0 は、クエリ入力 3 3 のシークエンスをキャプチャし、及びそれら
10
を処理して、インデクス可能な特徴 3 6 をリトリブする。記憶ユニット 2 2 は、注釈付ジェスチャ画像のデータベース 5 2、インデクス付特徴の大規模語彙テーブル 7 2 及びジェスチャマップ 7 3 を記憶する。処理ユニット 2 4 は、画像クエリの処理方法 3 0 及び検索エンジン 7 0 の方法を実行する。処理ユニット 2 4 はまた、ディスプレイ 1 0 0 上に提示されたビジュアルコンテンツの出力（二次元/三次元）9 4 を、ユーザの実行するジェスチャ 3 2 に反応して修正する。ディスプレイ 1 0 0 は、固定機器 2 0 上で動作するアプリケーションを表示する。アプリケーションは、ディスプレイ 1 0 0 上に提示された二次元/三次元ビデオゲーム、二次元/三次元オブジェクトモデリング/レンダリング、フォトブラウジング、マップ、ナビゲーション等を含んでよい。ユーザ 1 0 は、ディスプレイ 1 0 0 上に提示されたビジュアルコンテンツの出力（二次元/三次元）9 4 を知覚する。この出力は、ユーザの実行するジェスチャ 3 2 に反応して、連続的に修正されている。

【 0 0 4 2 】

（本発明の実施形態による方法及び機器）

本発明の実施形態は、リアルタイムのジェスチャ認識に使用される機器において、電力
20
が限られる状況下で、リソースを消費する計算を強いられるという問題解決に向けたものである。図 1 1 a は、本発明の実施形態による、三次元ジェスチャの認識方法を示すフローチャートである。方法は、ジェスチャ画像のデータベース 5 2 へのアクセスを有し、センサ 3 4 と通信する機器 2 0 において実行する。センサ 3 4 は、三次元のジェスチャの画像 3 3 をキャプチャするよう構成される。センサは、機器と一体型の部品とするか、又はセンサに連結可能な独立センサとしてもよい。ジェスチャ画像のデータベース 5 2 は、正規化されたジェスチャ画像のインデクス可能な特徴 5 4 を備える。インデクス可能な特徴は、正規化されたジェスチャ画像のエッジ画像の各画素に対して位置及び方向を備える。機器はデータベース 5 2 を記憶する記憶ユニット 2 2 を備えてよい。又は、機器はインタフェースユニットを備えてよく、インタフェースユニットは、例えばインターネットを介してデータベース 5 2 を記憶する、リモートデータベースノードを介して通信する。方法
30
は以下のステップを含む。

【 0 0 4 3 】

ステップ 1 1 0 では、三次元ジェスチャの画像 3 3 を、センサ 3 4 を介してキャプチャする。本発明の実施形態において、画像をキャプチャするステップは、三次元ジェスチャの画像のシークエンスをキャプチャするステップを含んでよい。データベース画像のマッチングの決定をより正確にするために、画像のシークエンスを使用してよい。詳細は以下に説明する。

【 0 0 4 4 】

ステップ 1 2 0 では、キャプチャされた画像を正規化する。正規化するステップは、比較を可能とするために、データベースの正規化されたジェスチャ画像に従って実行してよい。
40
正規化は、キャプチャされた画像を、データベース画像のサイズにサイズ変更するステップを含んでよい。データベースエントリは通常、3 2 0 × 2 4 0 画素又は 6 4 0 × 4 8 0 画素のような標準画像サイズに正規化する。従ってキャプチャされた画像を、データベースエントリの特定サイズに正規化してよい。

【 0 0 4 5 】

ステップ 1 3 0 では、キャプチャされ、正規化された画像 3 3 から、インデクス可能な特徴 3 6 を得る。インデクス可能な特徴 3 6 は、キャプチャされ、正規化された画像のエッジ画像の各画素に対して位置及び方向を備える。

【 0 0 4 6 】

ステップ 1 4 0 では、得られたインデクス可能な特徴 3 6 を、得られたデータベースの
50

インデクス可能な特徴54と、相似機能を使用して比較する。

【0047】

ステップ150では、データベース52において、三次元ジェスチャとマッチングするジェスチャ画像を、比較に基づいて決定する。

【0048】

キャプチャされ、正規化された画像のエッジ画像の各画素に対して位置及び方向を備えるインデクス可能な特徴を使用することの一つの利点は、それによって、三次元ジェスチャを認識する方法が、計算的に効率化されることにある。

【0049】

図11bは、本発明の別の実施形態による機器における方法を示す、フローチャートである。方法は、図11aに関して上述したステップを含む。しかしながら、得られたインデクス可能な特徴36を比較するステップ140は更に、以下のステップを含む。

【0050】

ステップ141では、データベースにおいて、キャプチャされた画像とマッチする、複数のジェスチャ画像を決定するために、直接相似分析を使用する。

【0051】

ステップ142では、キャプチャされた画像とマッチする、複数のジェスチャ画像のサブセットを決定するために、複数のジェスチャ画像に対して逆相似分析を使用する。

【0052】

この実施形態は、複数のジェスチャ画像のサブセットの1つとするために、データベース52において、三次元ジェスチャとマッチするジェスチャ画像を決定するステップ150を含む。しかしながら、逆相似分析を使用するステップ142は、既述したように任意のステップである。逆相似分析を実行しない場合、ステップ150で、直接相似分析により決定された、複数のジェスチャ画像の1つとするために、データベース52において、三次元ジェスチャとマッチするジェスチャ画像を決定する。直接相似分析及び逆相似分析に関しては、上記「ジェスチャ検索エンジン」の部分で詳述している。逆相似機能76は、精度向上を理由として使用してよい。しかしながら逆相似分析を使用しない場合、リトリブの精度は低下するが、利点としては複雑性が低減される。

【0053】

また図11bのフローチャートにより、方法は更に、ディスプレイ上に提示されたビジュアルコンテンツを修正するために、三次元ジェスチャとマッチする、決定されたジェスチャ画像を使用するステップ160を含んでよい。これに関しては、例えば上記「インタフェース」の部分で、例を挙げて説明している。

【0054】

2つの非常に類似したジェスチャ画像は、重なり合うエッジ画素を有することはないかもしれないが、互いに近接する画素に該当する可能性がある。これらのケースについて考慮するために、キャプチャされた画像を比較する際に、直接相似分析機能75の他に、データベース画像において、第1レベルの8近接画素及び第2レベルの16近接画素を考慮してもよい。従って本発明の実施形態においては、機器が実行する方法は更に、キャプチャされ、正規化された画像から、エッジ画像の各画素の近接画素に対して位置及び方向を備える追加的なインデクス可能な特徴を得るステップと、得られた追加的なインデクス可能な特徴を、データベースの追加的なインデクス可能な特徴と、相似機能を使用して比較するステップと、を含んでもよい。

【0055】

その後、データベース52において、三次元ジェスチャとマッチするジェスチャ画像を、追加的なインデクス可能な特徴との比較にも基づいて決定してよい。

【0056】

さらに、三次元ジェスチャとマッチするジェスチャ画像を、ジェスチャ画像のシークエンスにおいて、互いに近いジェスチャ画像を示すジェスチャマップに基づいて決定してよい。機器における方法は更に、ユーザのジェスチャを、画像のシークエンスに基づいて追

10

20

30

40

50

跡するステップを含む。そして、データベースにおいて、三次元ジェスチャとマッチするジェスチャ画像を、追跡されたユーザのジェスチャにも基づいて決定してよい。

【0057】

上述の任意の実施形態において、ジェスチャ画像のデータベース52における各エントリーに、3つの方向及び3つの位置のパラメータを備える、関連する三次元の動きのパラメータをタグ付けしてよい。従って方法は更に、三次元ジェスチャとマッチする、決定されたジェスチャ画像に関連する三次元の動きのパラメータを、データベースからリトリブするステップを含む。

【0058】

図12は、本発明の実施形態による、三次元ジェスチャを認識する機器20を示す、概略的なブロック図である。機器20は、正規化されたジェスチャ画像の、インデクス可能な特徴54を備えるジェスチャ画像のデータベース52へのアクセスを有するよう構成される。インデクス可能な特徴は、正規化されたジェスチャ画像のエッジ画像の各画素に対して位置及び方向を備える。機器は、三次元ジェスチャの画像33をキャプチャするよう構成されたセンサ34に接続可能である。センサ34は、機器20内に備えられるか、又は機器から独立してもよい。機器20は処理ユニット24を備える。処理ユニット24は、センサを介して三次元ジェスチャの画像33をキャプチャし、キャプチャされた画像を正規化し、及びキャプチャされ、正規化された画像33から、インデクス可能な特徴36を取得するよう構成される。インデクス可能な特徴は、キャプチャされ、正規化された画像のエッジ画像の各画素に対して位置及び方向を備える。処理ユニット24はまた、得られたインデクス可能な特徴36を、データベースのインデクス可能な特徴54と、相似機能を使用して比較し、及びデータベース52において、三次元ジェスチャとマッチするジェスチャ画像を、比較に基づいて決定するよう構成される。

【0059】

さらに処理ユニット24は、データベースにおいて、キャプチャされた画像とマッチする複数のジェスチャ画像を決定し、及び複数のジェスチャ画像の一つとすべく、データベース52において、三次元ジェスチャとマッチするジェスチャ画像を決定するために、得られたインデクス可能な特徴を、直接相似分析を使用して比較するよう構成されてよい。

【0060】

さらに処理ユニット24は、キャプチャされた画像とマッチする複数のジェスチャ画像のサブセットを決定し、及び複数のジェスチャ画像のサブセットの一つとすべく、データベース52において、三次元ジェスチャとマッチするジェスチャ画像を決定するために、得られたインデクス可能な特徴を、ジェスチャ画像の複数に対する逆相似分析を使用して比較するよう構成されてよい。

【0061】

実施形態において、さらに処理ユニット24は、キャプチャされ、正規化された画像から、エッジ画像の各画素の近接画素に対して位置及び方向を備える追加的なインデクス可能な特徴を得るよう構成されてよい。さらに処理ユニット24は、得られた追加的なインデクス可能な特徴を、データベースの追加的なインデクス可能な特徴と、相似機能を使用して比較し、及び追加的なインデクス可能な特徴との比較にも基づいて、データベース52において、三次元ジェスチャとマッチするジェスチャ画像を決定するよう構成されてよい。

【0062】

さらに処理ユニット24は、三次元ジェスチャとマッチするジェスチャ画像を、ジェスチャ画像のシークエンスにおいて、互いに近いジェスチャ画像を示すジェスチャマップに基づいて決定するよう構成されてよい。処理ユニット24は、三次元ジェスチャの画像のシークエンスを、センサ34を介してキャプチャするよう構成されてよい。この実施形態において処理ユニットは、ユーザのジェスチャを、画像のシークエンスに基づいて追跡し、及びデータベース52において、三次元ジェスチャとマッチするジェスチャ画像もまた、追跡されたユーザのジェスチャに基づいて決定するよう構成されてよい。

【 0 0 6 3 】

さらに処理ユニット24は、ディスプレイ上に提示されたビジュアルコンテンツを修正するために、三次元ジェスチャとマッチする、決定されたジェスチャ画像を使用するよう構成されてよい。さらに、ジェスチャ画像のデータベース52における各エントリは、3つの方向及び3つの位置のパラメータを備える、関連する三次元の動きのパラメータをタグ付けされてよい。そしてさらに処理ユニット24は、三次元ジェスチャとマッチする、決定されたジェスチャ画像に関連する三次元の動きのパラメータを、データベース52からリトリブするよう構成されてよい。

【 0 0 6 4 】

機器20は、実施形態において、既述の処理ユニット24により実行可能な指示を含むメモリを備えてよい。これにより機器は、センサを介して三次元ジェスチャの画像をキャプチャし、キャプチャされた画像を、データベースの正規化されたジェスチャ画像に従って正規化し、キャプチャされ、正規化された画像からインデクス可能な特徴を得て、得られたインデクス可能な特徴を、データベースのインデクス可能な特徴と、相似機能を使用して比較し、及びデータベースにおいて、三次元ジェスチャとマッチするジェスチャ画像を、比較に基づいて決定するよう動作する。機器20はまた、処理ユニット24に接続され、及びセンサ34及び/又はデータベース53と通信するよう構成されたインタフェース回路を備えてよい。

10

【 0 0 6 5 】

図12の実施形態を代替的に説明すると、機器20は、センサを介して三次元ジェスチャの画像をキャプチャする手段、キャプチャされた画像を、データベースの正規化されたジェスチャ画像に従って正規化する手段、キャプチャされ、正規化された画像からインデクス可能な特徴を得る手段、得られたインデクス可能な特徴を、データベースのインデクス可能な特徴と、相似機能を使用して比較する手段、及びデータベースにおいて、三次元ジェスチャとマッチするジェスチャ画像を、比較に基づいて決定する手段を備えてよい。上述の手段は、ハードウェア、ソフトウェア、ファームウェア又はそれらの任意の組み合わせにおいて実装される、機能性ユニットである。一実施形態においては、手段はプロセッサ上で動作するコンピュータプログラムとして実現される。

20

【 0 0 6 6 】

図12の実施形態を、また別の方法で説明すると、機器20は、単一ユニット又は複数のユニットである中央処理装置(CPU: Central Processing Unit)を備えてよい。さらに機器20は、例えばEEPROM(Electrically Erasable Programmable Read-Only Memory)、フラッシュメモリ又はディスクドライブ等の不揮発性メモリの形状である、少なくとも1つのコンピュータプログラムプロダクト(CPP: computer program product)を備えてよい。CPPはコンピュータプログラムを備えてよく、このコンピュータプログラムは、機器20のCPU上で動作する場合、機器20に、図11a及び図11bに関連して既述した方法を実行させる、コード手段を備える。換言すれば、上述のコード手段がCPU上で動作する場合、それらのコード手段は、図12の処理ユニット24に該当する。

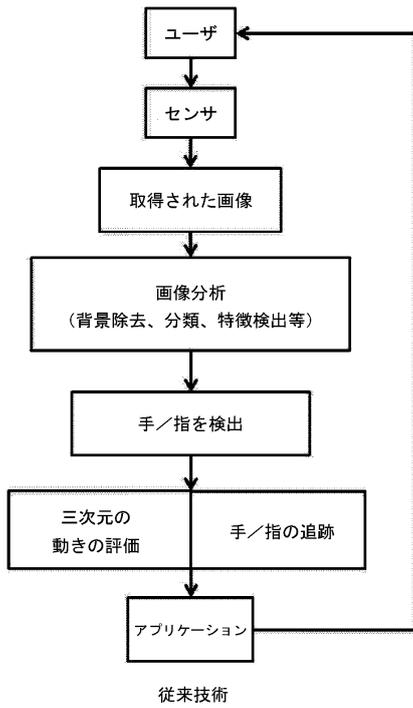
30

【 0 0 6 7 】

上記に詳述した実施形態は、例としてのみ挙げたものであり、限定的なものではない。別添の請求項の範囲内においては、他の解決策、使用、目的及び機能が可能である。

40

【図1】



【図2A】

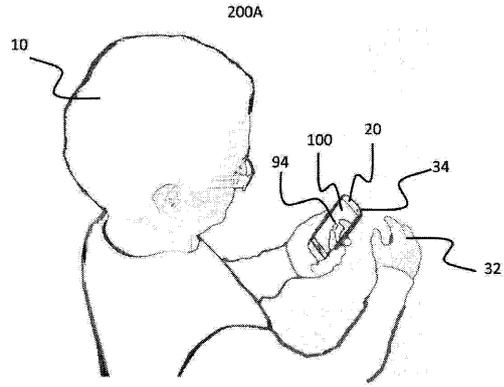


Fig.2A

【図2B】

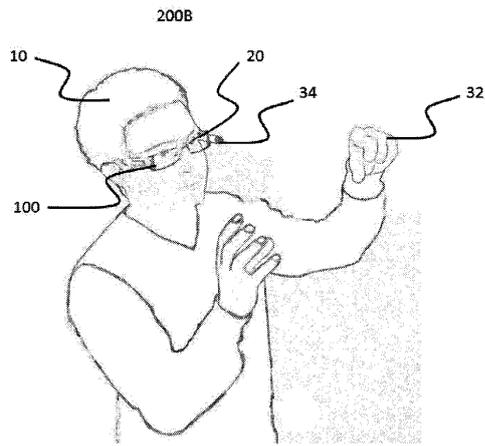


Fig.2B

【図2C】

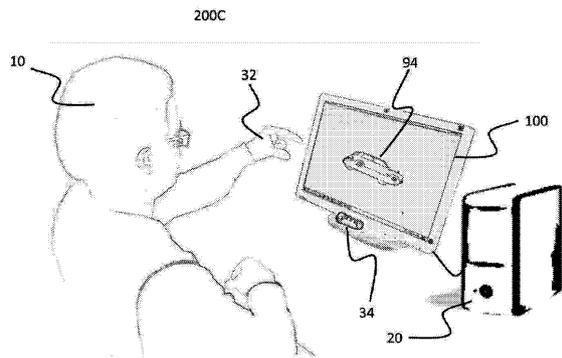
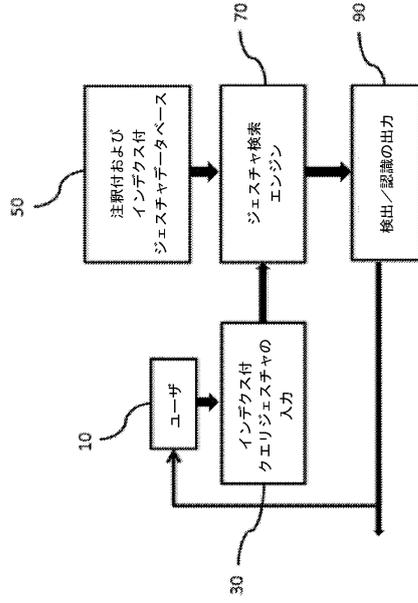
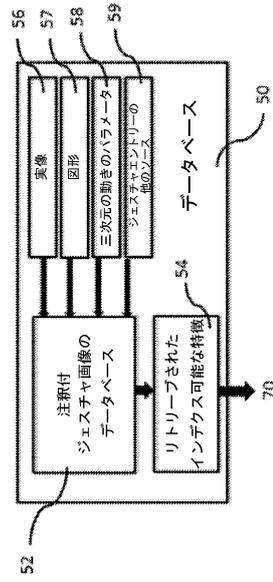


Fig.2C

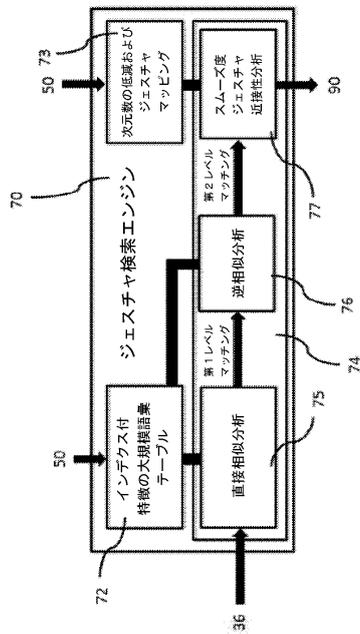
【図3】



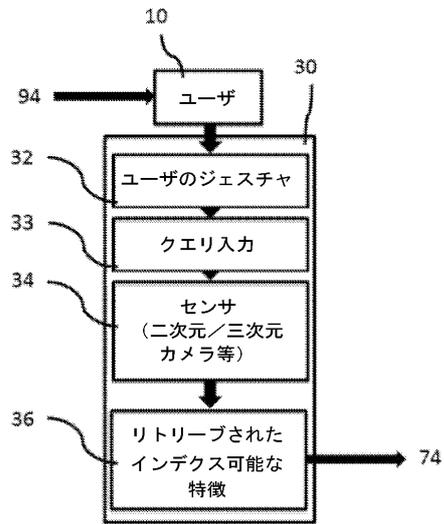
【図4】



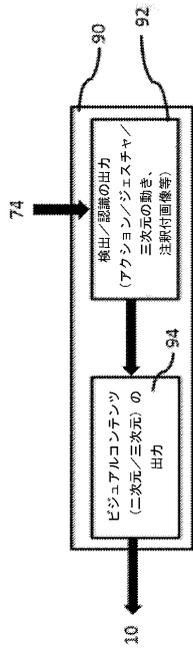
【図5】



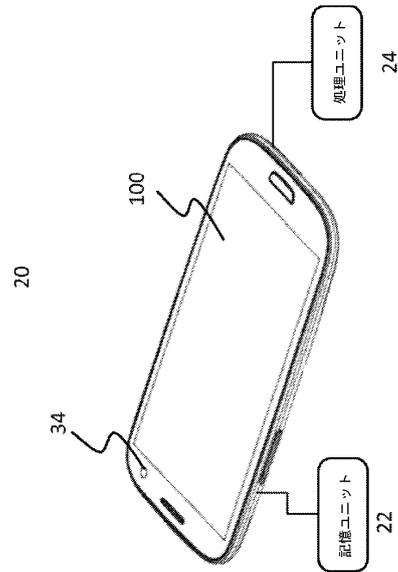
【図6】



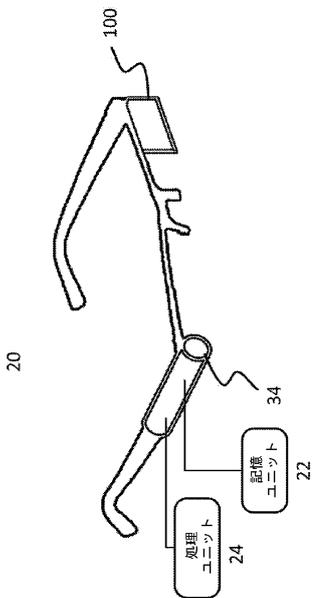
【図7】



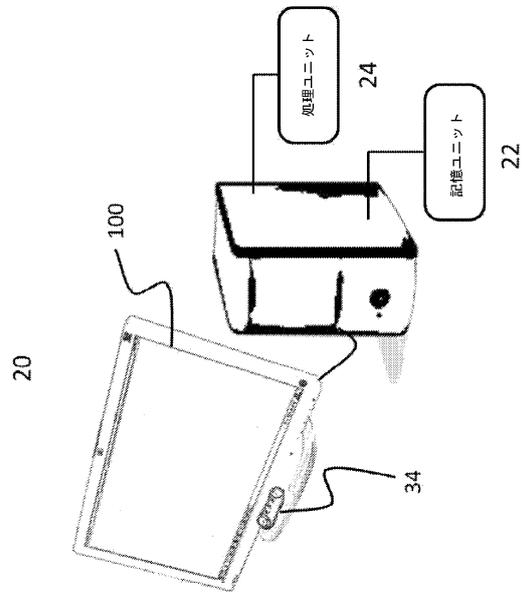
【図8】



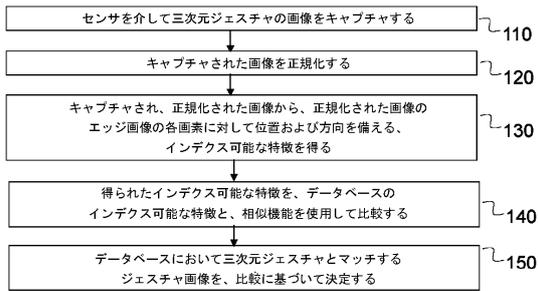
【図9】



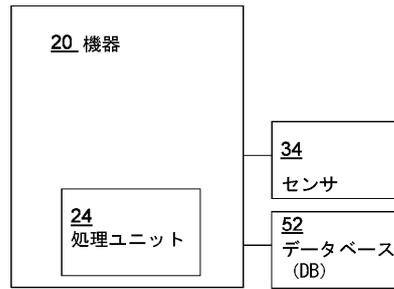
【図10】



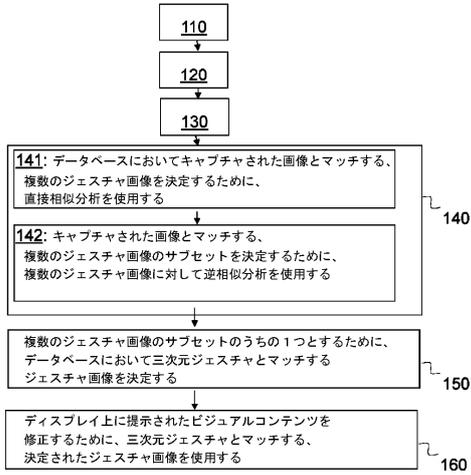
【図11a】



【図12】



【図11b】



フロントページの続き

(72)発明者 ハイボ リ

スウェーデン国 187 68 テバイ バドパークスヴェーゲン 12

(72)発明者 ファリード アベダン コンドリ

スウェーデン国 163 65 スポンガ ジュルスタ トーグ 1 リフ 1001

審査官 広 島 明芳

(56)参考文献 特表2012-506579(JP,A)

特開2011-253292(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06T 1/00 - 7/90

G06F 3/01