

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5845365号  
(P5845365)

(45) 発行日 平成28年1月20日(2016.1.20)

(24) 登録日 平成27年11月27日(2015.11.27)

(51) Int.Cl.		F I			
<b>G06T</b>	<b>7/20</b>	<b>(2006.01)</b>	G06T	7/20	300A
<b>G06F</b>	<b>3/01</b>	<b>(2006.01)</b>	G06T	7/20	B
			G06F	3/01	310C

請求項の数 17 (全 20 頁)

(21) 出願番号	特願2014-559265 (P2014-559265)	(73) 特許権者	512168630
(86) (22) 出願日	平成25年6月6日(2013.6.6)		ソフトキネティック ソフトウェア
(65) 公表番号	特表2015-510203 (P2015-510203A)		ベルギー国 ベ - 1050 ブリュッセル、ブルヴァール ドゥ ラ プレヌ 11
(43) 公表日	平成27年4月2日(2015.4.2)	(74) 代理人	110000855
(86) 国際出願番号	PCT/EP2013/061711		特許業務法人浅村特許事務所
(87) 国際公開番号	W02014/001058	(72) 発明者	ジーク、ローラン
(87) 国際公開日	平成26年1月3日(2014.1.3)		ベルギー国、ブリュッセル、リュ エミール クラウス 47
審査請求日	平成26年8月27日(2014.8.27)	(72) 発明者	フェイ、オリヴィエ
(31) 優先権主張番号	12173489.1		ベルギー国、ブリュッセル、アブニュ デ ムーニエール 22
(32) 優先日	平成24年6月25日(2012.6.25)		
(33) 優先権主張国	欧州特許庁 (EP)		

最終頁に続く

(54) 【発明の名称】 3次元の近接インタラクションにおける、又は3次元の近接インタラクションに関連する改善

(57) 【特許請求の範囲】

【請求項1】

3次元シーンの深度画像を用いて、該シーン内のユーザ(210)の肢端に関連付けられる少なくとも1つの特徴を検出し、追跡する方法であって、前記方法は、

a) 各深度画像(200)内の少なくとも1つの対象領域(210)をセグメント化するステップと、

b) 前記セグメント化された深度画像内の前記ユーザ(210)の肢部に対応する少なくとも1つの対象物(220、230)を特定するステップと、

c) 各肢部(220、230)を経時的に追跡するステップと、

d) 各肢部(220、230)に関連付けられる肢端(260、270)を識別するステップと、

e) 識別された各肢端(260、270; 400; 500)を経時的に追跡するステップと、

f) 識別された各肢端(260、270; 400; 500)をその関連付けられる肢部(220、230)から分離するステップと、

g) 分離された各肢端(260、270; 400; 500)に関連するパラメータ(410、420、430、440、450)を求めるステップと、

h) 分離された各肢端(260、270; 400; 500)内の少なくとも1つの特徴(510、520、530、540、550、560)を識別するステップと、

i) 識別された各特徴(510、520、530、540、550、560)を経時的

10

20

に追跡するステップとを含む、3次元シーンの深度画像を用いて、該シーン内のユーザ(210)の肢端に関連付けられる少なくとも1つの特徴を検出し、追跡する方法。

【請求項2】

前記ステップa)は、前記深度画像(200)内の各ピクセルの深度の関数であるように構成される少なくとも1つの構造化要素を利用することを、請求項1に記載の方法。

【請求項3】

前記ステップc)は、肢部(220、230)ごとに時間的関連性を特定することを含む、請求項1又は2に記載の方法。

【請求項4】

前記ステップe)は、肢部(220、230)ごとに評価された時間的関連性に基づいて、各肢端(260、270;400;500)を追跡することを含む、請求項3に記載の方法。

【請求項5】

前記ステップd)は、前記深度画像(200)内の各肢部(220、230)の複数の点の加重平均を用いて、各肢端(260、270;400;500)を識別することを更に含む、請求項1から4までのいずれか一項に記載の方法。

【請求項6】

前記ステップf)は

f1)前記ユーザの各肢部に関連付けられる各肢端(400)の中心(410)を特定するステップと、

f2)前記肢端の前記中心(410)からその境界(450)までの距離を求めるステップとを含む、請求項1から5までのいずれか一項に記載の方法。

【請求項7】

前記ステップf)は、切断プロセスを用いて、前記肢部(220、230)から前記肢端(260、270;400;500)を分離することを更に含み、前記切断プロセスは前記肢端(260、270;400;500)の前記特定された中心(410)を中心にして行われる、請求項6に記載の方法。

【請求項8】

前記ステップf)は、成長プロセスを用いて、前記肢部(220、230)から前記肢端(260、270;400;500)を分離することを更に含む、請求項6又は7に記載の方法。

【請求項9】

前記ステップg)は、前記肢端(400;500)に関連付けられる少なくとも1つのパラメータ(410、420、430、440、450)を求める前記ステップを含む、請求項6から8までのいずれか一項に記載の方法。

【請求項10】

前記ステップg)は、前記肢端(400;500)に関連付けられる求められた各パラメータ(410、420、430、440、450)を用いて、前記肢端(400;500)に関連付けられる少なくとも1つの特徴(510、520、530、540、550、560)を識別するステップを更に含む、請求項9に記載の方法。

【請求項11】

前記肢端(400;500)に関連付けられる前記少なくとも1つのパラメータは前記深度画像(200)内の前記肢端の向きを含む、請求項9又は10に記載の方法。

【請求項12】

前記ステップg)は、求められた各パラメータ(410、420、430、440)を用いて、前記肢端(400)の先端位置(450)を特定することを更に含む、請求項9から11までのいずれか一項に記載の方法。

【請求項13】

前記肢部は腕を含み、前記肢端は手(500)を含み、前記肢端に関連付けられる各特

10

20

30

40

50

徴は前記手(500)の少なくとも1つの指(510、520、530、540、550、560)を含む、請求項9から12までのいずれか一項に記載の方法。

【請求項14】

前記ステップh)は、手輪郭内のシルエット指(510、520、530、540、550)及びポインティング指(560)を識別することを含む、請求項13に記載の方法。

【請求項15】

前記ステップh)は、前記識別されたシルエット指(510、520、530、540、550)及びポインティング指(560)を合体させて、固有の1組の指を形成することを含む、請求項14に記載の方法。

10

【請求項16】

少なくとも1つの手(400;500)によって行われるジェスチャを認識するためのジェスチャ認識方法であって、請求項13から15までのいずれか一項による、3次元シーンの深度画像を用いて該シーン内のユーザの肢端に関連付けられる少なくとも1つの特徴を検出し、追跡する前記方法を含む、ジェスチャ認識方法。

【請求項17】

請求項1から16までのいずれか一項による、3次元シーンの深度画像を用いて該シーン内のユーザ(210)の肢端に関連付けられる少なくとも1つの特徴を検出し、追跡する前記方法の前記すべてのステップを実行するためのコンピュータプログラム。

【発明の詳細な説明】

20

【技術分野】

【0001】

本発明は3次元(3D:three dimensional)近接インタラクションにおける、又はそれに関連する改善に関し、より詳細には、限定はしないが、手及び/又は指に基づくジェスチャ・インタラクションを可能にする情報が特定されるように、深度検知撮像デバイスに近い距離においてユーザの手(両手)及び/又は手の主要部分を検出し、追跡することに関する。

【背景技術】

【0002】

ジェスチャ及び動作認識を用いる人間とコンピュータとのインタラクティブシステム及び方法は、コンピュータビジョンの分野から既知の技術である。特定の撮像デバイスの視野内で実行されるときにユーザの3Dジェスチャ及び動作認識を可能にする、最新の測距カメラデバイスに基づくインタラクティブシステム、例えば、3D構造化光カメラも既知である。そのようなシステムでは、人間ユーザは通常、カメラから少なくとも1mの場所に立っており、ユーザの動作及び/又はジェスチャは派手である、すなわち、その解像度があらかじめ決まっており、限られている3Dカメラによって検出されるほど十分に大きい必要がある。例えば、通常の3Dカメラ解像度はクォーター・ビデオ・グラフィックス・アレイ(QVGA:quarter video graphics array)に準拠し、すなわち、320×240ピクセルであり、それゆえ、指のような小さな物体が3D撮像デバイスからあまりにも離れて位置するとき、検出可能な動作の精度、又はそれらの物体を検出する能力を制限する。そのようなシステムでは、一般的に、そのような距離においてユーザの手を、より詳細には、手の指を確実に検出し、追跡することはできない。より厳密には、現在の3D撮像デバイスの大部分が、カメラから1m~5mの範囲内のユーザの完全な骨格を追跡できるように構築される。更に、これらの3D撮像デバイスの大部分が、カメラセンサから0m~1.5mの範囲内での測定の実行を許さない技術仕様を有する。

30

40

【0003】

しかしながら、その中に3D飛行時間(ToF:time-of-flight)カメラを含む、幾つかの新たな技術が開発されるにつれて、カメラセンサから1m未満、典型的には0.5m未満の距離にあるユーザの手及び/又は指のような特徴を含む特徴をカメラセンサから種々の距離において検出し、追跡できるようになってきた。これは「近接インタラクション

50

(close interaction)」と呼ばれ、デスクトップ環境において、又は自動車環境内のような他の環境において、手及び指に基づく非接触のインタラクションを可能にするように特に適応する。

#### 【 0 0 0 4 】

皮膚色の検出を用いてピクセルごとにカラーチャネルを埋め込むことにより、画像内の人間の複数の部位の存在を検出することが知られている。そのような方法が、A. A. Argyros及びM. I. A. Lourakisによる「Real Time Tracking of Multiple Skin Colored Objects with a Possibly Moving Camera」(Proceedings of the European Conference on Computer Vision (ECCV '04), Springer-Verlag, vol. 3, pages 368 to 379, May 11 to 14, 2004, Prague, Czech Republic)と題する論文において記述されている。ここでは、  
10  
ベイズ分類器に基づいて斑点が検出され、そのベイズ分類器は、小さなトレーニングデータセットでブートストラップされ、色測定が環境照明によって妨害される場合があるときに分類器が照明変化に対処できるようにする皮膚 - 色確率を含む。この検出方法は基本的に、近接インタラクション環境においてユーザの両手及び顔を検出できるようにし、画像内の斑点の位置について情報を与えることができるように、これらの検出された皮膚 - 色斑点の何らかの追跡能力と更に組み合わせることができる。その際、そのような情報を用いて、ジェスチャ認識システムへの入力を与えることができ、例えば、物理的に接触することなく、コンピュータのマウスを簡単なやり方で制御することができる。

#### 【 0 0 0 5 】

別の開示では、ジェスチャ認識のための2D / 3D画像の融合に基づく高速で、ロバストなセグメント化技法が、S.E. Ghobadi、O.E. Loeprich、K.Hartmann及びO. Loffeldによる「Hand Segmentation using 2D/3D Images」(Proceedings of Image and Vision Computing New Zealand 2007, pages 64 to 69, Hamilton, New Zealand, December 2007)と題する論文において記述されている。フォトニックミキサデバイス(PMD: photonic mixer device)センサのピクセルごとの距離情報とともに、強度画像を生成するために3D ToFカメラが用いられる。強度及び距離データを融合して、K平均と期待値最大化技法との組み合わせに基づくセグメント化アルゴリズム(KEM: K-Means mixed with expectation maximisation technique)への入力情報を与える。KEMは、K平均クラスタリング技法と、期待値最大化技法とを組み合わせ、K平均を用いる初期ステップにおいて、自然クラスタの中心が特定され、第2のステップにおいて、期待値最大化の柔軟な確率的割当てを用いて極大値を反復的に見つけ、それにより、見つけられたクラスタの位置及び境界を精緻化する。背景が均一である単純な環境において、且つセグメント化されることになる物体から十分な距離において、融合されたデータは、いかなる照明条件のシーンであっても、ユーザの体、顔、腕又は他の物体からユーザの手をリアルタイムに適切にセグメント化できるようにする。検討されている環境におけるその方法のロバスト性は、2D / 3Dジェスチャ追跡及びジェスチャ認識における処理技法の最初のステップとして使用する要件を満たすと言われている。I. Oikonomidis、N. Kyriazis及びA.A. Argyrosによる「Tracking the Articulated Motion of Two Strongly Interacting Hands」([http://www.ics.forth.gr/~argyros/mypapers/2012\\_06\\_cvpr\\_twohands.pdf](http://www.ics.forth.gr/~argyros/mypapers/2012_06_cvpr_twohands.pdf))と題する論文では、粒子群最適化(PSO: particle swarm optimisation)を用いて、RGB-Dセンサ、すなわち、深度情報も与えるカラーセンサからの信号を処理する。その方法は、マーカレス目視観測を実行して、複雑で自由に互いに相互作用する両手の全関節を追跡する。記述される方法は、感覚情報を、既知のサイズ及び運動学の2つの合成及び対称3D手モデルの接合関節に直接起因すると考えることによって、2つの相互作用する手の完全な検出及び追跡を達成する。2つの手の所与の関節の場合に、RGB-Dセンサであれば感知することになる信号の予測が、画像収集プロセスをシミュレートすることによって、例えば、特定のカメラシーン較正のための合成深度マップを生成することによって得られる。実際の入力に相当するデータを生成するパラメトリックプロセスを確立した後に、実際の入力に最も類似している深度マップを生成するパラメータを探索することによって追跡が実行される。追跡がオンラインで実行され、その場合、ステップごとに、且つ新た  
20  
30  
40  
50

な入力ごとに、最適化ステップが実行され、PSOの変形形態を用いて、実際のRGB-D入力と、仮定された関節から生成されたシミュレーションによる深度マップとの間の相違が最小化され、最もスコアが高い仮定が現在の入力のための解を構成する。各最適化ステップにおいて利用される計算効率及び時間的連続性によってロバスト性(robustness)が達成されるように、相違指標は注意深く定式化される。

【0006】

M.K. Bhuyanら「Fingertip Detection for Hand Pose Recognition」(International Journal on Computer Science and Engineering (ISCSE), Vol. 4, No. 3, March 2012)において、前腕及び手のセグメント化も記述されている。ここでは、指先検出及び指タイプ認識のための方法が提案され、その方法では、ベイズの規則に基づく皮膚色セグメント化を用いて、手領域において指先の場所が特定される。その後、ある重要な幾何学的特徴を観測することによって、セグメント化された手領域においてモルフォロジー演算が実行される。その後、指の動きの幾何学的特徴の最終的な確率的モデル化は、手話認識及びジェスチャアニメーションにおける指綴りのための有益な情報を与える状況において特に、指タイプ認識プロセスが著しくロバストにされるように実行される。

10

【0007】

S. Malassiotis及びM.G. Strintzisによる「Real-time hand posture recognition using range data」(Image and Vision Computing, 26 (2008), pages 1027 to 1037)と題する論文では、アルファベットの手話において直面するような複雑な手姿勢を認識するために、腕セグメント化、手-前腕セグメント化、手姿勢推定及びジェスチャ分類が実行される方法が記述される。その方法によれば、3次元の手の幾何学的形状を利用して、手話姿勢及び/又はジェスチャを認識することによって、3次元画像データが取り込まれ、処理される。

20

【0008】

WO-A-03/071410号は、ジェスチャ及び姿勢を用いて人が電子デバイスと相互作用できるようにするための別の方法を記述する。その方法は、その人の体の一部の複数の離散領域に対する位置情報を入手することを含む。その位置情報は、基準位置に対する、その体部分の各離散領域の深度値を含む。3Dセンサシステムを用いて、3Dシーンの深度画像が取り込まれ、そのシーンの背景から1つ又は複数の対象領域をセグメント化するように処理され、各対象領域は、そのジェスチャが検出されることになる手を含む。各ジェスチャが分類され、1組の異なる手のジェスチャと比較される。

30

【0009】

論文において手検出及びモデル化が広く検討されているにもかかわらず、上記の文献はいずれも、いなかかる色情報も用いることなく、そしていかなる手モデルも、若しくは関節を有する骨格表現も用いることなく、3D撮像センサからの深度マップを入力情報として用いて、撮像デバイスに近い距離において指、掌及び向きを含む、手の複数の部分及びそれらのパラメータを検出し、追跡することを提供しない。

【先行技術文献】

【特許文献】

【0010】

【特許文献1】WO-A-03/071410号

【非特許文献】

【0011】

【非特許文献1】A. A. Argyros及びM.I. A. Lourakis「Real Time Tracking of Multiple Skin Colored Objects with a Possibly Moving Camera」(Proceedings of the European Conference on Computer Vision (ECCV '04), Springer-Verlag, vol. 3, pages 368 to 379, May 11 to 14, 2004, Prague, Czech Republic)

【非特許文献2】S.E. Ghobadi、O.E. Loepprich、K.Hartmann及びO. Loffeld「Hand Segmentation using 2D/3D Images」(Proceedings of Image and Vision Computing New Zealand 2007, pages 64 to 69, Hamilton, New Zealand, December 2007)

40

50

【非特許文献3】I. Oikonomidis、N. Kyriazis及びA.A. Argyrosによる「Tracking the Articulated Motion of Two Strongly Interacting Hands」([http://www.ics.forth.gr/~argyros/mypapers/2012\\_06\\_cvpr\\_twohands.pdf](http://www.ics.forth.gr/~argyros/mypapers/2012_06_cvpr_twohands.pdf))

【非特許文献4】M.K. Bhuyanら「Fingertip Detection for Hand Pose Recognition」(International Journal on Computer Science and Engineering (ISCSE), Vol. 4, No. 3, March 2012)

【非特許文献5】S. Malassiotis及びM.G. Strintzis「Real-time hand posture recognition using range data」(Image and Vision Computing, 26 (2008), pages 1027 to 1037)

【非特許文献6】L. Guigues「Comparison of image segmentations using a hierarchical model for n-m regions matchings」(Proc. Of 2nd IAPR-TC15 Workshop on Graph-Based Representations, Haindorf, Austria 1999) 10

【非特許文献7】Y. Bulalov、S. Jambawalikar、P. Kumar及びS. Sethia「Hand recognition using geometric classifiers」([http://www.ams.sunysb.edu/~saurabh/research/hand\\_workshop.pdf](http://www.ams.sunysb.edu/~saurabh/research/hand_workshop.pdf))

【非特許文献8】A. Kirillov「Hands gesture recognition」(<http://www.codeproject.com/Articles/26280/Hands-Gesture-Recognition>, May 2008)

【非特許文献9】Thiago R. Trigo及びSergio Roberto M. Pellegrino「An Analysis of Features for Hand-Gesture Classification」(IWSSIP 2010, 17th international Conference on Systems, Signals and Image Processing) 20

【発明の概要】

【発明が解決しようとする課題】

【0012】

それゆえ、本発明の目的は、十分な解像度を有する飛行時間(ToF)カメラのような、3D検知デバイスの錐台によって画定されるような所定の体積において、ユーザの手、詳細には、手の指、指先、掌及びパラメータを検出し、位置を特定し、セグメント化するための方法を提供することである。

【0013】

本発明の別の目的は、入力深度マップに当てはまるか、又は一致することになる任意のモデル又は骨格表現を用いることなく、ユーザの手、詳細には、手の指、指先、掌及びパラメータを検出し、位置を特定し、セグメント化するための方法を提供することである。 30

【0014】

本発明の別の目的は、空間内の手姿勢にかかわらず、ユーザの手、詳細には、手の指、指先、掌及びパラメータを検出し、位置を特定し、セグメント化するための方法を提供することである。

【0015】

本発明の更なる目的は、至近距離モード検出能力を有する3D検知デバイスによって与えられる入力情報を用いて、シーン内に立っている少なくとも1人のユーザの複数の手を検出し、位置を特定し、セグメント化するための方法を提供することである。

【0016】 40

本発明の更に別の目的は、3D検知デバイスを用いる近接インタラクションのために手及び指のジェスチャ及び/又は姿勢を利用するヒューマン・マシン・インターフェースへの入力データを与えることである。

【0017】

本発明の更なる目的は、コンピュータシステムに関連付けられる3D検知システムの錐台内で動作する手のみを用いて、コンピュータシステムを制御することである。

【課題を解決するための手段】

【0018】

本発明の第1の態様によれば、3次元シーンの深度画像を用いてそのシーン内のユーザの肢端に関連付けられる少なくとも1つの特徴を検出し、追跡する方法が提供され、その 50

方法は、

- a) 各深度画像内の少なくとも1つの対象領域をセグメント化するステップと、
- b) セグメント化された深度画像内のユーザの肢部に対応する少なくとも1つの対象物を特定するステップと、
- c) 各肢部を経時的に追跡するステップと、
- d) 各肢部に関連付けられる肢端を識別するステップと、
- e) 識別された各肢端を経時的に追跡するステップと、
- f) 識別された各肢端をその関連付けられる肢部から分離するステップと、
- g) 分離された各肢端に関連するパラメータを特定するステップと、
- h) 分離された各肢端内の少なくとも1つの特徴を識別するステップと、
- e) 識別された各特徴を経時的に追跡するステップとを含む。

10

【0019】

ステップa)は、深度画像内の各ピクセルの深度の関数であるように構成される少なくとも1つの構造化要素を利用することを含むことができる。

【0020】

ステップc)は肢部ごとの時間的関連性を特定することを含むことができる。この場合、ステップe)は、肢部ごとに評価された時間的関連性に基づいて各肢端を追跡することを含む。

【0021】

一実施例では、ステップd)は、深度画像内の各肢部の点の加重平均を用いて、各肢端を識別することを更に含むことができる。

20

【0022】

別の実施例では、ステップf)は、

- f1) ユーザの各肢部に関連付けられる各肢端の中心を特定するステップと、
- f2) 肢端の中心から境界までの距離を求めるステップとを含むことができる。

【0023】

更に、ステップf)は、切断プロセスを用いて、肢部から肢端を分離することを更に含むことができ、切断プロセスは肢端の特定された中心を中心にして行われる。

【0024】

代替的には、又は更には、ステップd)は成長プロセスを用いて、肢部から肢端を分離することを更に含むことができる。

30

【0025】

ステップg)は、肢端に関連付けられる少なくとも1つのパラメータを求めるステップを含むことができる。肢端に関連付けられるそのような特定された各パラメータを用いて、肢端に関連付けられる少なくとも1つ特徴を識別することができる。一実施例では、肢端に関連付けられる少なくとも1つパラメータは、深度画像内の肢端の向きを含むことができる。

【0026】

一実施例では、ステップg)は、求められた各パラメータを用いて、肢端の先端位置を特定することを更に含むことができる。

40

【0027】

一実施例では、肢部は腕を含み、肢端は手を含み、肢端に関連付けられる各特徴は手の少なくとも1つの指を含む。この場合、ステップh)は手輪郭内のシルエット指及びポインティング指を識別することを含む。別の実施例では、ステップh)は、各指先を識別することを含む。これは、識別されたシルエット指及びポインティング指を合体させて、固有の1組の指を形成することによって達成することができる。

【0028】

本発明の別の態様によれば、少なくとも1つの手によって行われるジェスチャを認識するためのジェスチャ認識方法が提供され、その方法は、上記のようにシーンの深度画像を用いて3次元シーン内のユーザの肢端に関連付けられる少なくとも1つの特徴を検出し、

50

追跡する方法を含む。

【0029】

本発明の更なる態様によれば、上記のように3次元シーンの深度画像を用いて3次元シーン内のユーザの肢端に関連付けられる少なくとも1つの特徴を検出し、追跡する方法のステップを実行するためのコンピュータプログラムが提供される。

【0030】

本発明を更に理解してもらうために、ここで、一例にすぎないが、添付の図面が参照される。

【図面の簡単な説明】

【0031】

【図1】本発明による方法のステップを示す流れ図である。

【図2】至近距離モード測距カメラデバイスによって取り込まれた3Dシーンを表す深度マップ内の背景からの人間ユーザのセグメント化のグレースケール表現を示す図である。

【図3】本発明による、至近距離モード測距カメラデバイスによって取り込まれた3Dシーンを表す深度マップ内の肢部セグメント化のグレースケール表現を示す図である。

【図4】本発明による、肢部が合体するときのセグメント化された肢部の追跡ステップを示す図である。

【図5】本発明による、肢部が合体するときのセグメント化された肢部の追跡ステップを示す図である。

【図6】本発明による、肢部の肢端を「切断して」手を分離するステップを示す図である

。

【図7】本発明による、肢部の肢端を「切断して」手を分離するステップを示す図である

。

【図8】分離された手の中心を特定する際のステップを示す図である。

【図9】分離された手の中心を特定する際のステップを示す図である。

【図10】本発明による、測距測定値の3D点群表現における手モデル化のステップを示す図である。

【図11】本発明による、測距測定値の3D点群表現における手モデル化のステップを示す図である。

【図12】本発明による、指検出及び追跡のステップを示す図である。

【図13】本発明による、指検出及び追跡のステップを示す図である。

【図14】本発明による、指検出及び追跡のステップを示す図である。

【発明を実施するための形態】

【0032】

本発明は、特定の実施例に関して、特定の図面を参照しながら説明されることになるが、本発明はそれには限定されない。説明される図面は概略にすぎず、限定するものではない。図面では、要素のサイズのうちの幾つかが誇張される場合があり、例示のために、縮尺どおりに描かれてない場合がある。

【0033】

本明細書において、「垂直」及び「水平」という用語は、図の特定の向きを指すために用いられ、これらの用語が、本明細書において説明される具体的な実施例への制限でないことは理解されよう。

【0034】

本発明は、近接インタラクションの状況において人とコンピュータとのインタラクション(相互作用)を可能にするために、シーン内の少なくとも1人のユーザの少なくとも1つの手の少なくとも一部を検出し、追跡することを対象とする。その検出及び追跡は、更なるステップにおいて、手によって、又はこの手の一部によって、詳細にはこの手の指によって行われる静止姿勢及び動的ジェスチャを検出し、認識できるように、経時的にインタラクションを実行することになる手又は手の一部のそれぞれの位置を特定することを目的とする。深度関連情報を用いることによって、取り込まれた種々の画像にわたって、シ

10

20

30

40

50



ーン内の複数の対象物を特定し、セグメント化することができる。深度関連情報は、近接インタラクション能力を有する、すなわち、1 m未満の距離、好ましくは0.5 m未満の距離を有するToFカメラを用いて得られる。カメラ又は撮像デバイスが、ユーザが立っている距離における指検出及び追跡に準拠している解像度を有するセンサを備える場合には、その方法は、ユーザがカメラ又は撮像デバイスから離れて、すなわち、カメラから1 mより遠く離れて立っている状況において用いるのにも適している。

#### 【0035】

本明細書において用いられるときに、「姿勢」という用語は特定の時点における手を表す1組のキーポイントの配置を指している。本明細書において用いられるときに、「ジェスチャ」という用語は、手のキーポイントの配置の経時的变化を指している。更には、又は代替的には、「ジェスチャ」という用語は、キーポイントの不変配置の空間内位置の経時的变化を指す場合もある。

10

#### 【0036】

姿勢が経時的に変化しない場合には、すなわち、キーポイントの配置及びその空間内位置が変化しないままである場合には、その姿勢は静止姿勢と呼ばれる。キーポイントの配置が変化しないままであるが、その配置の空間内位置が経時的に変化する場合には、これは、静止姿勢を有するジェスチャと呼ぶことができる。キーポイントの配置が経時的に変化する場合には、これは異なる姿勢を有するジェスチャと呼ばれる。

#### 【0037】

本発明によれば、3D検知デバイスによって収集された深度画像のリアルタイムストリームに基づいて、ユーザの手、特に指及び掌を含む手の検出、追跡及びモデル化を提供する方法が記述される。詳細には、ToF 3Dカメラを用いて、3D点群又は深度マップの形をとる深度画像を与えることができる。更に、本発明の方法は、2つ以上の手を検出し、追跡し、モデル化するために用いることができ、本発明の好ましい実施例では、シーン内の対象物、すなわち、手を分離するためにスマートセグメント化方法を利用する。しかしながら、その方法は手には限定されず、その形が既知である他の肢端又は潜在的に相互作用する物体を追跡するために用いることができる。

20

#### 【0038】

図1は、本発明の方法におけるステップを示す流れ図100を示す。最初のステップ、ステップ105は、上記のように、3D点群又は深度マップの形をとる場合がある深度画像を収集することである。次のステップ、ステップ110は、深度画像の残りからユーザの肢部をセグメント化することである。ステップ115において、セグメント化された肢部が追跡され、その後、ステップ120において、追跡された肢部の肢端が特定される。その後、ステップ125において、これらの肢部の肢端が追跡され、ステップ130において、それらの肢端のモデル化が実行される。モデル化後に、ステップ135において、手を認識するための認識プロセスが実行され、その後、ステップ140において、手の幾何学的モデル化が実行される。その後、ステップ145において、手の指が検出され、追跡され、ステップ150において、手の開きの更なる計算値を求めることができる。例えば、入力3D点群又は深度マップを取り込むために3D撮像システムを含むコンピュータシステムとのインタラクションを制御するために、ステップ150からの出力を用いることができる。ここで、そのプロセスの各ステップが以下に更に詳細に説明される。

30

40

#### 【0039】

ステップ105において、例えば、深度画像を直接与えるか、又は決められた座標系内で3D点群の形で深度画像を与える3D撮像システム又は他の測距測定カメラ又は撮像デバイスを用いて、任意の既知の技法によって深度画像を収集することができる。深度画像はステップ110における肢部セグメント化プロセスへの入力を形成する。

#### 【0040】

3D撮像システムは、既知の物理的パラメータ、例えば、焦点距離、視野、レンズ歪みなどを有するカメラ又は撮像センサを備える。これから、その深度値とともに、取り込まれた画像内の撮像システムの錐台内の物体に対応する2D点を、カメラ基準系内の3D座

50

標に容易に変換することができる。これは、空間内の所与の位置における現実世界の座標を画像空間内の座標、例えば、深度マップ内のピクセル単位に変換できることを意味する。現実世界メトリックシステムにおいて与えられる、以下に更に詳述される形態学的開きプロセスにおいて用いられる浸食/膨張球体の半径が、ピクセルの深度がわかるときに、ピクセルごとに独立して深度画像内の半径に変換される。

【0041】

ステップ110の肢部セグメント化プロセスでは、深度画像は入力を形成し、3D空間内の所定値又はしきい値より小さな寸法を有する物体に対応する、現在の深度画像の特定の領域が描かれる。最大寸法値は、人間の肢部及び頭部の幅より大きい、人間の胸部の幅より小さいように設定される。これにより、肢部セグメント化プロセスによって、画像内に見える人間の肢部及び頭部を含むが、胴体も背景も含まないように、その領域を描くことができるようになる。

10

【0042】

肢部セグメント化ステップは、基本的な画像フィルタリングアルゴリズム、例えば、形態学的開きプロセスと、その後の、深度マップ情報の使用を必要とする、本発明の実施態様に特に適応した領域成長及び合体方法との特定の適応形態を用いて実行することができる。実数値画像に関する形態学的開きプロセスは、所定のサイズの構造化要素を用いて、結果として「浸食された画像(eroded image)」を生成する、画像の浸食を計算することと、別の構造化要素を用いて、浸食された画像の膨張を計算することを含む。

【0043】

浸食は、全ての他の形態学的演算が基にする、形態学的画像処理の2つの基本的な演算のうちの一つである。2値モルフォロジー(morphology)の基本的着想は、簡単な所定の形状で画像をプローブし、この所定の形状が、処理されている画像内の形状にいかにか合致するか、合致しないかを特定することである。この簡単な「プローブ(probe)」は、構造化要素と呼ばれ、それ自体は2値画像、すなわち、空間又はグリッド又はピクセルのサブセットである。

20

【0044】

形態学的画像処理の基本演算のうちのもう一つは膨張である。ここで、2値画像へのその演算子の基本的効果は、前景ピクセル、例えば、通常、2D画像における白色ピクセル、又は深度マップ内の所定の距離より近いピクセルの領域の境界を徐々に拡大することであり、それにより、前景ピクセルのエリアのサイズが成長し、一方、それらの領域内の穴が小さくなる。膨張演算子は、入力として2つのデータ、すなわち、膨張することになる画像と、構造化要素として知られる(カーネルとしても知られる)1組、標準的には小さな1組の座標点をとる。構造化要素は、入力画像への膨張の厳密な効果を決定する。

30

【0045】

形態学的開きプロセスの使用は2つの利点を有し、すなわち、浸食ステップ及び膨張ステップに対して2つの異なる構造化要素を用いることができるという利点と、標準的な2D画像処理技法の場合のように画像全体に対して決まったサイズを用いる代わりに、構造化要素のサイズを、ピクセルが含む深度値に応じて画像内のピクセルごとに適応させることができるという利点とを有する。特定の有利な実施例では、構造化要素、例えば、半径 $r$ を有する球体のサイズは、各ピクセルの深度の関数、すなわち、カメラ又は3D撮像システムからのピクセルの距離の関数になるように構成され、それにより、その要素が、カメラ座標系に対応する点の現実世界3D位置に位置決めされた場合には、その要素は一定のサイズを有することになる。これは、「視点不変性」の開きを与える。これは、物体の視点の変化が開きのサイズに影響を及ぼさないことを意味する。

40

【0046】

肢部セグメント化ステップ、すなわち、ステップ110の一部として、メートル単位の浸食半径 $r_e$ 及びメートル単位の膨張半径 $r_d$ のようなパラメータを形態学的開きプロセスの一部として決定することができる。更に、差しきい値 $t$ 及び深度の最大変化 $d$ のようなパラメータが、肢部セグメント化ステップの一部として入力深度画像に適用される。

50

これらのパラメータは肢部セグメント化ステップにおいて導出されるが、後続の処理ステップでは用いられない。

【 0 0 4 7 】

肢部セグメント化ステップ、すなわち、ステップ 1 1 0 からの出力は、浸食画像（浸食）、開き画像（膨張）、出力差画像（差）及び出力領域ラベル画像（接続領域のラベリング）から導出されるラベル画像を含む。

【 0 0 4 8 】

深度  $d$  を有する画像のピクセルごとに、ピクセルに対応する 3 D 点を中心にした半径  $r_d$  の球体を画像空間内に射影することによって、ある球体から浸食画像が求められ、この円板上の画像の最大深度値が計算され、浸食された画像の対応するピクセルに割り当てられる。

10

【 0 0 4 9 】

浸食された画像のピクセルごとに、ピクセルに対応する 3 D 点を中心にした半径  $r_d$  の球体を画像空間内に射影することによって得られた円板から出力開き画像が求められ、この円板上の浸食された画像の最大値が計算され、開き画像の対応するピクセルに割り当てられる。

【 0 0 5 0 】

深度画像のピクセルごとに、出力差画像が、その深度と開き画像内のその値との間の差として求められる。その出力差が差しきい値  $t$  より大きい場合には、ピクセルが「肢部」に属すると見なされる。

20

【 0 0 5 1 】

出力領域ラベル画像（接続領域ラベリング）は、深度値の絶対差が深度の最大変化  $d$  より小さい場合には、画像の隣接するピクセルの各対を接続されているとして記録し、接続される各構成要素を、以前に特定された「接続」関係に応じて異なるようにラベルを付けることによって求められる。

【 0 0 5 2 】

その後、「肢部」セグメント化により決定された領域から擬似小領域を除去するために、小領域を、より大きな隣接する領域に合体させる。領域ラベル画像の残りのラベルのないゾーンを接続性に基づいてセグメント化することによって、「胴体」を分離できるようになる。その後、隣接する肢部と胴体との間の接続性を解析することによって、どの肢部がどの胴体に接続されるかに関して評価できるようになる。

30

【 0 0 5 3 】

図 2 は、近接インタラクションの状況における 3 D シーン内の人間ユーザ 2 1 0 の画像 2 0 0 を示す。図示されるように、ユーザ 2 1 0 は、一對の腕 2 2 0、2 3 0 と、頭部 2 4 0 と、胴体 2 5 0 とを有する。各腕 2 2 0、2 3 0 は、関連付けられる手 2 6 0、2 7 0 を有する。上記のような肢部セグメント化ステップ、すなわち、ステップ 1 1 0 後に、ユーザ 2 0 0 の腕 2 2 0、2 3 0 は、その手 2 6 0、2 7 0 及び頭部 2 4 0 とともに、図 3 に示されるように分離することができる。胴体 2 5 0 は対象物ではなく、それゆえ、肢部セグメント化の出力の一部を形成しない。セグメント化された物体の中の対象物のソートは、そのシーン内の物体の統計学的モーメント分析に関連する深度測定値を用いて行うことができる。対象物を規定するための距離判定基準はカメラに近い物体を優先することができるので、例えば、カメラから離れた物体が大きすぎる場合には除去することができる。一方、カメラに近い同じ物体は保持されることになる。通常、近接インタラクションの状況において、よりカメラに近い 1 つ又は 2 つの物体だけを対象と見なすことができる。

40

【 0 0 5 4 】

ステップ 1 1 0 において肢部セグメント化を実行した後に、次のステップは、肢部追跡、すなわち、ステップ 1 1 5（図 1）を実行することである。このステップでは、追跡プロセスにおいて、現在の深度画像、すなわち、 $t$  における画像の「肢部」領域が、先行する深度画像、すなわち、 $t - 1$  における画像の「肢部」領域に関連付けられる。 $t - 1$  における画像内の肢部、及び  $t$  における画像内の同じ肢部に対する照合スコアを計算するこ

50

とによって、対ごとの関連付けを行うことができる。照合スコアは、表面比、又は照合表面比とすることができる。好ましい実施例では、最新の動き推定アルゴリズムによって、現在のフレームに対して、先行するフレームの各領域のピクセルの位置を推定することができる。統計値、例えば、動きの平均又は中央値を用いて、領域レベルにおいて動きを補償することもできる。

#### 【0055】

動き推定は、1つの画像から別の画像への変換、例えば、画像取込み間の変換を記述する動きベクトルを求めるプロセスを指しており、X軸、Y軸及びZ軸における回転及び並進を近似するモデルによって表すことができる。動きベクトルは、ピクセルに基づく技法（直接法とも呼ばれる）、例えば、ブロック照合、位相相関及びノ又は周波数領域法、並びにピクセル再帰アルゴリズムを用いて見つけることができる。ピクセルに基づく方法では、最小二乗誤差（MSE：minimum squared error）、絶対差の和（SAD：sum of absolute differences）、平均絶対差（MAD：mean absolute difference）及び二乗誤差の和（SSE：sum of squared errors）のうちの1つ又は複数とすることができる。間接法を使用することもでき、その場合、局所的エリア又は大域的エリアにわたって適用される統計関数、例えば、ランダム・サンプル・コンセンサス（RANSAC：random sample consensus）を用いて特徴が照合される。

#### 【0056】

その際、保持される1組の関連性が「分離」（ $n - 1$ ）関係又は「合体」（ $m - 1$ ）関係のみをモデル化できるという制約を確認する、全ての取り得る関連性のサブセットが計算される。これは、グラフ最適化問題として、L. Guigues「Comparison of image segmentations using a hierarchical model for n-m regions matchings」（Proc. Of 2<sup>nd</sup> IAPR-TC15 Workshop on Graph-Based Representations, Haindorf, Austria 1999）において記述された。先行する画像及び現在の画像の各領域がグラフ内のノードに対応し、異なるフレーム、すなわち、先行するフレーム及び現在のフレームに属する各ノード対間にエッジが設定される、加重二部グラフが定義される。エッジは照合スコアによって重み付けされる。最大でも2の直径を有するこの関連性グラフの最大加重サブグラフを見つけることによって、最適な照合が得られる。2ステップ最適化手順が提案された。好ましい実施例では、貪欲最適化アルゴリズムを用いることができる。貪欲アルゴリズムは、大域的な最適解を見つけることを期待しながら各段階において局所的に最適な選択を行う、問題解決の発見的方法に従うアルゴリズムである。場合によっては、貪欲方策が最適解を生成する必要はないが、それでも、貪欲な発見的方法が、大域的な最適解に近い最適解を局所的に生成する場合がある。

#### 【0057】

貪欲最適化アルゴリズムを用いることによって、解がない状態から開始することができ、それらの関係をスコアが減少する順にソートすることができ、ソートされた関係を精査して、結果として新たな解が最大グラフ直径制約に違反しない関係を現在の解に含むことができる。同じ方法によって胴体を追跡することができ、その後、更なる処理に対して、追跡された胴体に接続される肢部間の関連性のみを考慮することができる。

#### 【0058】

ステップ120（図1）において、肢部の肢端、すなわち、この場合には手が検出され、追跡される。このステップにおいて、各肢部の特定の部分が肢部の好ましい肢端として識別される。好ましい実施例では、ステップ110（図1）においてセグメント化された肢部ごとに、肢部の3D点の加重平均を用いて、好ましい肢端が特定される。肢端は肢部の質量中心から離れた位置に位置すると判断されるので、肢部の各3D点に関連付けられる重みは、肢端から肢部の質量中心までの距離によって決まる場合がある。この距離は、3D空間内のユークリッド距離とすることができるか、又は肢部表面上の測地的距離とすることができる。

#### 【0059】

代替的には、肢部の各3D点に関連付けられる重みは、空間内の特定の方向において肢

10

20

30

40

50

端が優先されることをモデル化し、「アンカー点」と呼ぶことができる、空間内の特定の点までの距離によって決まる場合がある。アンカー点は、カメラ位置のような何らかの絶対位置又は基準位置に、又は肢部質量中心に対する何らかの位置に位置するように選択することができる。例えば、カメラ位置にあるアンカーを用いる場合には、その方法によって優先的に選択される肢端はカメラに最も近いことになる。質量中心の上方且つ前方のアンカーを用いる場合には、その方法によって優先的に選択される肢端は、肢部中心に対して上方・前方肢端になるであろう。アンカー点は、先行するフレーム内の肢端の相対的な位置に関連することができる。フレーム内で特定された肢端が先行するフレーム内の肢端の位置によって決まるので、その際、その方法は追跡方法としての役割を果たす。これは、例えば、ひじが手よりもカメラ又は撮像デバイスの近くにあるように見えるときに、肢部の肢端が手からひじに移るのを防ぐ。その追跡方法は加重平均に基づくので、肢端の位置は、追跡されている肢部に対応する3D点の位置に対して連続している。それゆえ、その追跡プロセスからの出力は、入力データ、すなわち、入力深度画像内の雑音に対して安定しており、小さな摂動に耐える。

10

#### 【0060】

肢端を検出すると、それらの肢端は、肢部追跡ステップ(ステップ125)において肢部に対して評価された時間的関連性に基づいて追跡される。動き推定を更に用いて、肢部の肢端の追跡のロバスト性を改善することができる。肢部ごとに見つけられた新たな各肢端が、肢部関連性を通して、0、1、又は複数の先行する肢端のうちの1つに関連付けられる。先行するフレームにおいて識別される同じ肢端は、0、1又は複数の新たな肢端のうちの1つに関連付けることができる。肢部の時間的追跡は、1-m個(分離)又はn-1個(合体)の関連性のみを保持し、n-m個の関連性を破棄する。

20

#### 【0061】

唯一の曖昧な事例は「分離(split)」事例及び「合体(merge)」事例であり、時間に関して特有の経路を有する肢端を時間的に追跡するために、多数の可能性から1つの特定の関連性が選択される必要がある。これが図4及び図5に示される。ここでは、ユーザの2本の腕220、230が頭部240の前方に示される。図4は、手260、270の近くで互いから離れている腕220、230を示しており、図5は、腕220の手260が腕230の手270と合体する状況を示す。図示されないが、分離事例は図5に示されるように開始し、腕220の手260が腕230の手270から分離されることは理解されよう。

30

#### 【0062】

「分離」事例の場合、(先行するフレーム内で検出された肢端を追跡する)新たなフレームにおける種々の可能性の中からの肢端の選択は、例えば、2D又は3Dにおける距離、重なりなどを用いて判断することができる。「合体」事例の場合、(現在のフレーム内で検出された肢端を追跡する)先行するフレームにおける種々の可能性の中からの肢端の選択は、例えば、距離若しくは重なりを用いることによって、又はその「エイジ(age)」パラメータ(検出してから経過した時間)若しくはこれまでに手として検出された肢端が優先されることなどの、先行するフレーム内の肢端に関連付けられる特定の判定基準を用いることによって行うことができる。

40

#### 【0063】

肢端の位置及び肢端の追跡が計算されると、肢端モデル化ステップ、すなわち、ステップ130において、特定の形態学的又は幾何学的特性を有する肢部の特定の点と一致するように、中心の位置が精緻化される。図6は、その関連付けられる手260、270と、その中心として利用される頭部240ととともに、腕220、230を示しており、図7は、頭部240に適用される中心340と、それぞれの腕220、230(図6)にある各手260、270に適用される中心360、370を示す。好ましい実施例では、その形状の中心点の位置を適用される中心として選択することができる。その形状の中心点は肢部の輪郭から最も遠い点である。この点は、その点から肢部の輪郭までの距離を計算することによって、効率的に特定することができる。一改善例では、追跡方法を用いて、肢

50

部の絶対中心点ではなく、肢端の先行する精緻化された位置に近い、距離関数の極大値である点と一致させることができる。更に、距離関数が密度と見なされる場合には、平均シフト推定のような、よりロバストなモード推定方法を用いることができる。その際、肢部の中心点を追跡するように特に設計された、位相シフトのような追跡方法を有する。

【0064】

関連する肢端領域の点を構成することができる、セグメント化された肢部の点のサブセットが選択される。この選択は、「切断(cutting)」方法、「成長(growing)」方法のいずれかによって行うことができる。

【0065】

「切断」方法では、肢端の精緻化された中心の周囲に画定される特定の領域の外部にある点が肢端領域から除外される。その特定の領域は、肢端の中心を中心にした球体、肢端中心と肢部中心との間に位置する平面によって画定される平面によって分離された半空間、すなわち3D空間の2つの部分のうちの1つ、又は肢端の中心、すなわち手の中心に対応する点と、肢部、すなわち、前腕の中心とを結ぶベクトルに対して直交する領域とすることができる。切断領域の幾何学的特性は、3Dカメラ座標において計算された、精緻化された肢端中心から肢部の輪郭までの最大距離のような、肢端に関する何らかの特性の連続学習機構から生じることができ、最大距離は、肢端が人の手である場合には掌の半径の推定値である。

【0066】

「成長」方法では、肢端領域から除外されることなる、肢部の「確実な(secure)」1組の点が識別される。この1組の点は、肢端中心から最も遠い点を含むことができる。識別された1組の点は「シード(seed)」点を形成し、その後、シード点は、成長する点が上記の「切断」方法の場合と同じようにして画定することができる「確実な」包含ゾーンに入るまで成長する。

【0067】

図8は、中心が適用されない場合のセグメント化された腕220、230を、手260、270及び頭部240とともに示しており、図9は、腕220、230(図8)から手260、270を分離する効果を示す。

【0068】

ステップ135において、手認識プロセスが実行される。手の既知の特徴を識別するために、最新技術の方法、例えば、<http://www.ams.sunysb.edu/~saurabh/research/handworkshop.pdf>においてインターネット上で公開されるY. Bulalov、S. Jambawalikar、P. Kumar及びS. Sethia「Hand recognition using geometric classifiers」及び<http://www.codeproject.com/Articles/26280/Hands-Gesture-Recognition>において入手可能なA. Kirillov「Hands gesture recognition」(2008年5月)の方法を用いることができる。このステップからの出力は、肢部追跡ステップ、すなわち、ステップ115にフィードバックすることができ、そのステップにおいて行われる任意の仮定に対する検証として用いることができる。

【0069】

ステップ140において、手のモデル化が実行される。これは、少なくとも、手の向きを判断することと、手の先端位置を判断することを含む。手の向きを得るために、上記のステップ135において求められたような手に対応する3D点に関して、主成分分析(PCA: principal component analysis)が実行される。これは、共分散行列の固有システムを実効的に解く。手の特定された主軸に関する時間フィルタリングを実行するために、固有システムを解く前に、以下の式に従って、共分散行列の一次自己回帰時間平滑化を実行することができる。

$$Cov(t) = RawCov(t) * \alpha + Cov(t-1) * (1 - \alpha)$$

【0070】

ただし、 $cov(t-1)$ 及び $cov(t)$ はそれぞれ時刻 $t-1$ 及び $t$ における滑らかな共分散行列であり、 $RawCov(t)$ は時刻 $t$ において計算された生の共分散行列

10

20

30

40

50

であり、 $\alpha$  は平滑化のレベルを調整するパラメータである。

#### 【0071】

求められた第1の固有ベクトルは手の主方向又は主軸の推定値を与え、求められた第3の固有ベクトルは掌法線の推定値を与える。求められた第2の固有ベクトルは、求められた第1の固有ベクトル及び第3の固有ベクトルの両方に対して直交する。これらの固有ベクトルが図10に示される。

#### 【0072】

図10において、掌の中心に位置する中心点410とともに手400が示される。これらの軸420、430、440は中心点410を中心に位置し、軸430は掌に対して垂直であり、軸440は手400の主方向を示し、軸420は軸430及び440に対して直交する。これらの軸は上記の固有ベクトルに対応する。

10

#### 【0073】

手の向きを特定した後に、図11に示されるように、最大手先位置450、すなわち、中心点410から最も遠い位置が特定される。手の主方向における手の3D点のヒストグラムが求められ、最も遠い点が平均される。通常、これらの最も遠い点は、例えば、ヒストグラムの95%~99%の範囲内にあると判断することができる。その際、手先位置350は経時的に安定する。代替形態として、手の主方向だけの代わりに、先行する最大手先位置をそれだけで用いることができるか、又は軸440(図10)によって示されるような手の主方向と合わせて用いることができる。

#### 【0074】

20

その後、手先の正確な判断を用いて、上記のステップ135においてPCA分析によって与えられた掌法線方向の推定値を補正することができ、そして特に、第3の固有ベクトルが明確に定義されない場合に、より良好な判断を与えることができる。これは、平面モデルを用いて十分に定義された向きを得るほど手が十分に平坦でない場合に、例えば、手が握りこぶしの形をとるときに、又は指がつまむような動作をする場合に対応する。平坦度情報は、システムの固有値によって与えられる。より厳密には、平坦度情報は、第2の固有値と第3の固有値との比によって示される。更なる情報が、手先から掌の中心までの距離によって測定することができ、掌の半径によって正規化される、手の延在範囲によって与えられる。その際、手の平坦度及び延在範囲によって決まる重みを用いて、掌法線方向の推定値を手先の方向と融合することによって、その推定値が精緻化される。

30

#### 【0075】

その後、ステップ145において、手の指が個々に検出され、追跡される。本明細書において用いられるときに、「指」という用語は親指も含む。最初に、図12に示されるように、手500において「シルエット」指510、520、530、540、550が検出される。「シルエット」指は、図示されるように、手の輪郭内で見ることができる指である。「ポインティング」指、すなわち、図13における指560によって示されるように撮像デバイス又はカメラの方を指している指が検出される。検出された「シルエット」指及び「ポインティング」指を合体させて、固有の1組の指にし、シルエット構成及びポインティング構成の両方に検出された指が重複して数えられるのを回避する。この合体は、各「ポインティング」指が「シルエット」指と同じ指に対応するか否かをチェックすることによって達成される。これが当てはまる場合には、その指は破棄される。そのチェックは、2つの検出された指間の距離によることができ、最も信頼性の高い判定基準は測地的距離を用いる。その後、図14に示されるように、固有の1組の指を追跡することができる。

40

#### 【0076】

「シルエット」指510、520、530、540、550の検出は、形態学的な骨格の肢端を用いることによって、又は手輪郭の曲率の最小値を求めることによって達成することができる。「シルエット」指の検出のための1つの方法は、Thiago R. Trigo及びSergio Roberto M. Pellegrino「An Analysis of Features for Hand-Gesture Classification」(IWSSIP 2010, 17th international Conference on Systems, Signals and Image P

50

rocessing) において記述されている。ここでは、アスペクト比、真円度及び広がり度のような形状記述子が、丸み、堅さ、指の数、関係「距離 - 半径」のような他の特徴とともに用いられる。

#### 【0077】

ポインティング指560の検出は、滑らかな深度画像内の極小値を識別することによって達成することができる。指先の局所形状に対応する最小値のみを保持するために、所定の数の方向において、極小値の幾つかの1次元プロファイルが調べられる。これらの1次元プロファイルのそれぞれに対して種々の判定基準がチェックされる。ステップ145の追跡部分の場合、一致しない可能性を含む、先行するフレームの指510、520、530、540、550と現在のフレーム内の指510、520、530、540、550との間のマッピングが実行される。取り得るマッピングごとに、先行するフレーム内の指と現在のフレーム内の関連付けられる指との間の最適な剛体変換が計算される。最適な剛体変換、すなわち、3D並進及び回転が、最小二乗判定を用いて計算される。その変換の更に良好な判定は、指の基本位置と掌の位置との間のマッピングを含むことによって得ることができる。取り得る指関連性ごとに、現在のフレーム内の指から先行するフレーム内の剛性変換された指までの距離(手の変形を測定する)の加重和、剛性変換の並進及び回転のノルム(norm)(手の全体的な動きを測定する)、及び一致しない指の数(現れる指又は消える指に対する固定コスト)を含む、誤差値が計算される。最小誤差に導く関連性が保持される。

10

#### 【0078】

ステップ150における更なる手開きパラメータ計算のために、異なる測定値の結合を用いて、連続的な「開き(openness)」変数が計算される。これらの測定値は、指間の最大距離、指の掌との間の最大距離、又は手先と掌との間の最大距離とすることができる。結合は一次結合又は最大値とすることができる。その後、この連続的な「開き」変数は、最初に時間フィルタリングを実行して、雑音を除去し、その後、手が検出されてから観測された最大値で正規化することによって、ブール状態に、すなわち、手が開いている、又は手が閉じている、に2値化される。その後、ヒステリシスしきい値を適用して、望ましい2値状態を得ることができる。

20

#### 【0079】

図1を参照しながら説明されたステップを完了した後に、指追跡アプリケーションを利用するコンピュータシステムにおいて、それらの出力データ及びパラメータを用いて、ヒューマン/マシンインタラクションを実行し、詳細には、測距撮像デバイス及び処理手段を用いて、手及び指に基づく近接インタラクションを実行することができる。

30



【 図 1 】

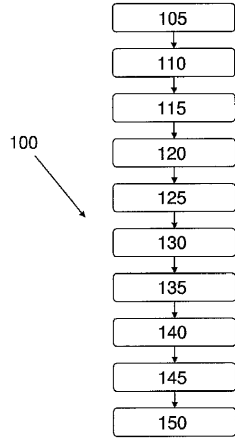


Fig. 1

【 図 2 】

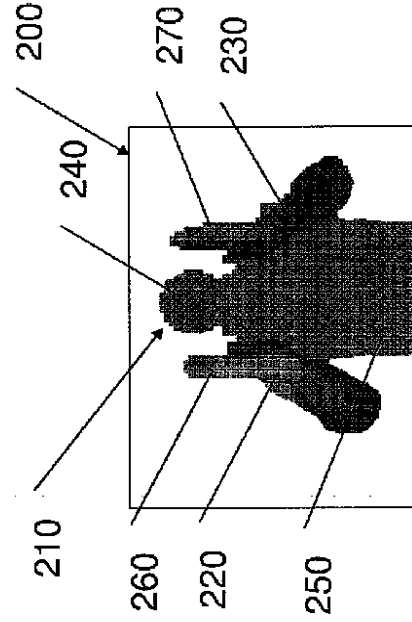


Fig. 2

【 図 3 】

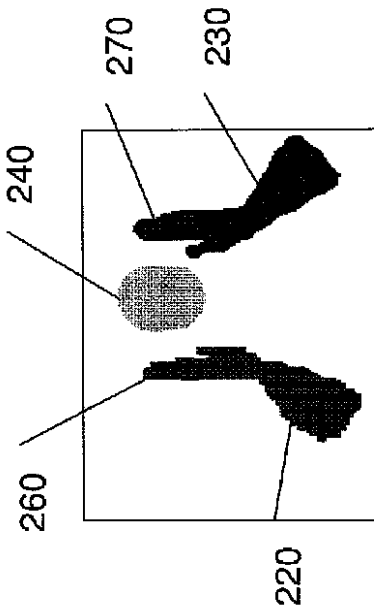


Fig. 3

【 図 4 】

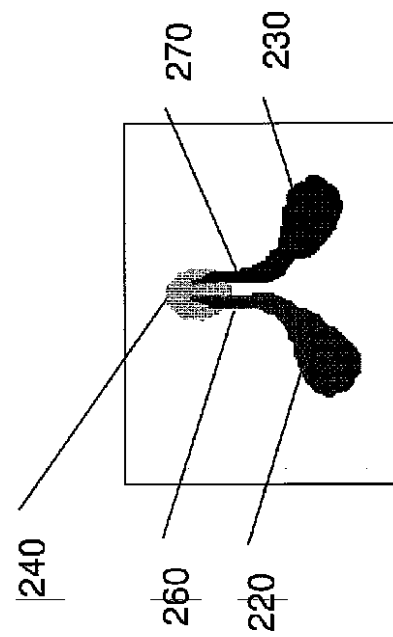
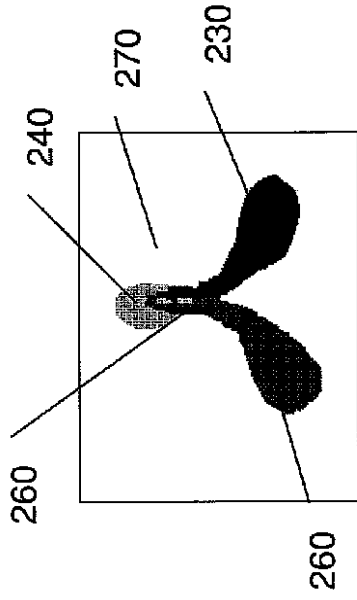


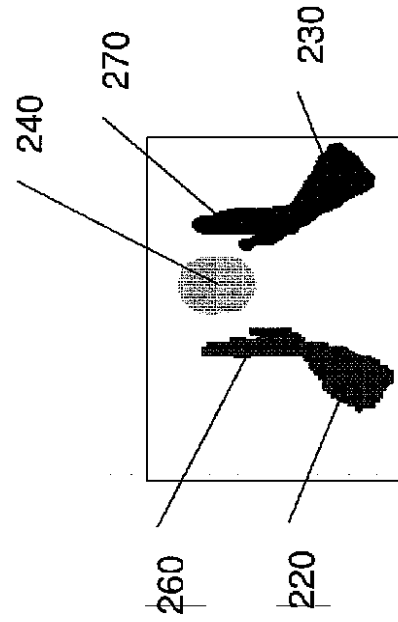
Fig. 4

【 図 5 】



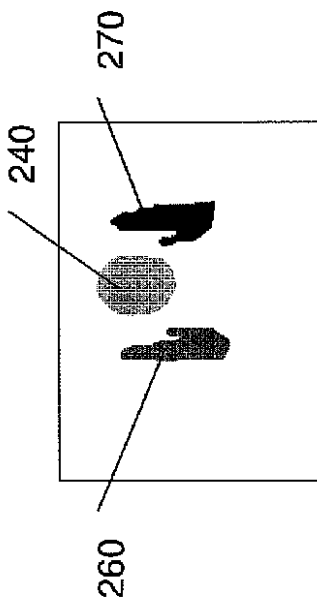
*Fig. 5*

【 図 6 】



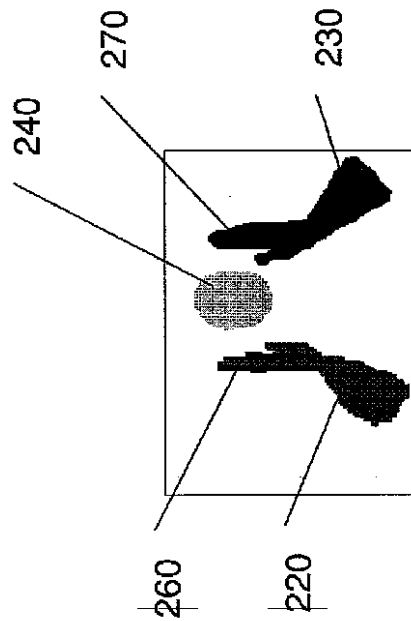
*Fig. 6*

【 図 7 】



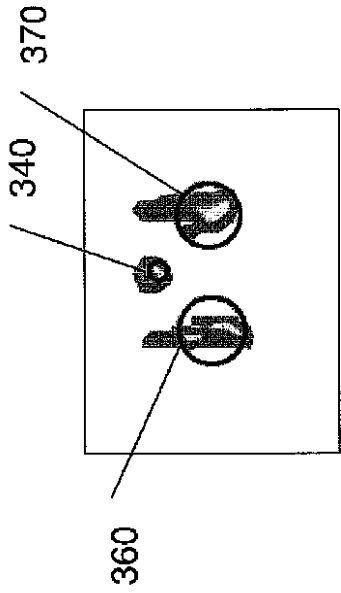
*Fig. 7*

【 図 8 】



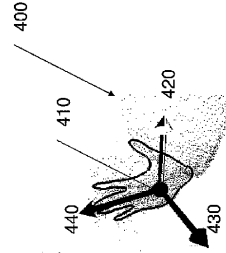
*Fig. 8*

【 9 】



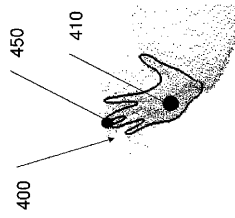
**Fig. 9**

【 10 】



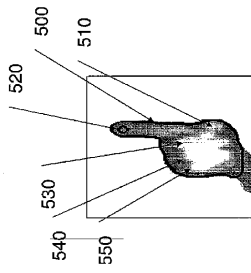
**Fig. 10**

【 11 】



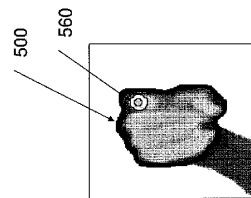
**Fig. 11**

【 12 】



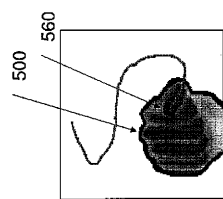
**Fig. 12**

【 13 】



**Fig. 13**

【 14 】



**Fig. 14**

## フロントページの続き

- (72)発明者 シモンズ、ケヴィン  
ベルギー国、ブリュッセル、リュ デ ラ ヴァレー 29
- (72)発明者 ホウ、ヤンシュ  
ベルギー国、ブリュッセル、リュ グザヴィエ デ ブー 22 ベー12
- (72)発明者 カモヴィッチ、アリアクサンドル  
ベルギー国、ブリュッセル、アブニュ シャトー デ ワルツィン 5

審査官 山内 裕史

- (56)参考文献 特開2012-113622(JP,A)  
特開2011-065652(JP,A)  
特開2004-094288(JP,A)  
米国特許出願公開第2003/0156756(US,A1)

- (58)調査した分野(Int.Cl., DB名)
- |      |      |
|------|------|
| G06T | 7/20 |
| G06F | 3/01 |