

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第6144759号
(P6144759)

(45) 発行日 平成29年6月7日 (2017.6.7)

(24) 登録日 平成29年5月19日 (2017.5.19)

(51) Int.Cl.

F I

G O 6 T 7/00 (2017.01)

G O 6 T 7/20 (2017.01)

G O 6 T 7/00 3 0 0 C

G O 6 T 7/20 3 0 0 A

請求項の数 31 (全 30 頁)

(21) 出願番号	特願2015-516033 (P2015-516033)	(73) 特許権者	507364838
(86) (22) 出願日	平成25年5月20日 (2013.5.20)		クアルコム、インコーポレイテッド
(65) 公表番号	特表2015-524115 (P2015-524115A)		アメリカ合衆国 カリフォルニア 921
(43) 公表日	平成27年8月20日 (2015.8.20)		21 サン ディエゴ モアハウス ドラ
(86) 国際出願番号	PCT/US2013/041766		イブ 5775
(87) 国際公開番号	W02013/184333	(74) 代理人	100108453
(87) 国際公開日	平成25年12月12日 (2013.12.12)		弁理士 村山 靖彦
審査請求日	平成28年4月27日 (2016.4.27)	(74) 代理人	100163522
(31) 優先権主張番号	61/657,565		弁理士 黒田 晋平
(32) 優先日	平成24年6月8日 (2012.6.8)	(72) 発明者	フィッツジェラルド・ジョン・アーチバル
(33) 優先権主張国	米国 (US)		ド
(31) 優先権主張番号	13/843,907		アメリカ合衆国・カリフォルニア・921
(32) 優先日	平成25年3月15日 (2013.3.15)		21-1714・サン・ディエゴ・モアハ
(33) 優先権主張国	米国 (US)		ウス・ドライブ・5775

最終頁に続く

(54) 【発明の名称】 高速ポーズ検出器

(57) 【特許請求の範囲】

【請求項 1】

ジェスチャがソース画像のシーケンス内で実行されているかどうかを決定するための方法であって、

参照画像の中の前記ジェスチャを検出するステップと、

複数のジェスチャモデルから、検出された前記ジェスチャに最も密接に一致する第1のジェスチャモデルを選択するステップであって、前記第1のジェスチャモデルは、検出された前記ジェスチャと同じタイプを有するジェスチャを含む画像の合成を含む、ステップと、

前記ジェスチャが実行されているかどうかを決定するために、前記第1のジェスチャモデル以外を使用せずに前記第1のジェスチャモデルを使用して、ソース画像の前記シーケンスをスキャンするステップであって、前記ソース画像のシーケンスは、前記参照画像の後に取得された複数の画像を含み、前記ソース画像のシーケンスをスキャンすることが、前記参照画像の後に取得されたソース画像と前記第1のジェスチャモデルを比較し、前記ソース画像内で前記ジェスチャが実行されているかどうかを決定することを含む、ステップと、

前記第1のジェスチャモデルを使用して前記画像のシーケンスをスキャンし前記ジェスチャが実行されているかどうかを決定することが、前記ジェスチャが実行されているとの決定をもたらす場合、前記ジェスチャが実行されているかどうかを決定するための、前記複数のジェスチャモデルのうちの別のジェスチャモデルを使用せずに前記第1のジェスチャ

10

20

ャを使用した前記ソース画像のシーケンスのスキャンを終了するステップとを備える、方法。

【請求項 2】

前記複数のジェスチャモデルが手のポーズを備える、請求項 1 に記載の方法。

【請求項 3】

前記複数のジェスチャモデルが顔の表情を備える、請求項 1 に記載の方法。

【請求項 4】

前記複数のジェスチャモデルが、左の開いた手のモデル、右の開いた手のモデル、または拳のモデルを備える、請求項 1 に記載の方法。

【請求項 5】

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の検出された前記ジェスチャの位置を、他の位置よりも優先するステップと、

前記検出されたジェスチャの優先された位置を使用して、前記ジェスチャが実行されているかどうかを決定するために、ソース画像の前記シーケンスをスキャンするステップとをさらに備える、請求項 1 に記載の方法。

【請求項 6】

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の検出された前記ジェスチャのスケールを、他のスケールよりも優先するステップと、

前記検出されたジェスチャの優先された前記スケールを使用して、前記ジェスチャが実行されているかどうかを決定するために、ソース画像の前記シーケンスをスキャンするステップとをさらに備える、請求項 1 に記載の方法。

【請求項 7】

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の前記検出されたジェスチャの位置を、他の位置よりも優先するステップと、

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の前記検出されたジェスチャのスケールを、他のスケールよりも優先するステップと、

前記検出されたジェスチャの優先された前記位置、および前記検出されたジェスチャの前記優先されたスケールの組合せを使用して、前記ジェスチャが実行されているかどうかを決定するために、ソース画像の前記シーケンスをスキャンするステップとをさらに備える、請求項 1 に記載の方法。

【請求項 8】

ソース画像の前記シーケンス内であらかじめ定められた数の前記ジェスチャが検出されると、ソース画像の前記シーケンスのフルスキャンを完了する前に、ソース画像の前記シーケンス内の前記ジェスチャのスキャンを終了するステップとをさらに備える、請求項 1 に記載の方法。

【請求項 9】

前記参照画像内に第 2 のジェスチャが検出される可能性が、前記ジェスチャよりも低いと決定するステップと、

前記第 2 のジェスチャが検出される可能性がより低いと決定するステップに基づいて、前記ジェスチャを検出した後、前記第 2 のジェスチャの前記参照画像をスキャンするステップとをさらに備える、請求項 1 に記載の方法。

【請求項 10】

前記ジェスチャの前記参照画像をスキャンするステップと、

連続する数の段階内で順序付けられた複数のカスケード接続された段階において、前記ジェスチャの前記参照画像のスキャンを絞り込むステップであって、前記カスケード接続された段階の各段階が、前記ジェスチャを検出するために、ソース画像の前記シーケンス内の少なくとも 1 つの画像内のピクセルの周期数をスキャンするステップを備え、ピクセルの前記周期数が、ピクセルの前記周期数内で互いのピクセルから離れた一定の垂直方向距離および水平方向距離を有し、任意の段階におけるピクセルの前記周期数が、以前の段階のピクセルの前記周期数以下であるステップと、

10

20

30

40

50

前記連続する数の段階の最後の段階が、前記ジェスチャの少なくとも一部を検出するかどうかを決定することによって、前記ジェスチャを識別するステップとをさらに備える、請求項 1 に記載の方法。

【請求項 1 1】

ジェスチャがソース画像のシーケンス内で実行されているかどうかを決定するための装置であって、

参照画像、およびソース画像の前記シーケンスを格納するように構成されたメモリと、前記メモリに通信可能に結合されたプロセッサであって、

参照画像の中の前記ジェスチャを検出すること、

複数のジェスチャモデルから、検出された前記ジェスチャに最も密接に一致する第 1 のジェスチャモデルを選択することであって、記第 1 のジェスチャモデルは、検出された前記ジェスチャと同じタイプを有するジェスチャを含む画像の合成を含む、こと、

前記ジェスチャが実行されているかどうかを決定するために、前記第 1 のジェスチャモデル以外を使用せずに前記第 1 のジェスチャモデルを使用して、ソース画像の前記シーケンスをスキャンすることであって、前記ソース画像のシーケンスは、前記参照画像の後に取得された複数の画像を含み、前記ソース画像のシーケンスをスキャンすることが、前記参照画像の後に取得されたソース画像と前記第 1 のジェスチャモデルを比較し、前記ソース画像内で前記ジェスチャが実行されているかどうかを決定することを含む、こと、および

前記第 1 のジェスチャモデルを使用して前記画像のシーケンスをスキャンし前記ジェスチャが実行されているかどうかを決定することが、前記ジェスチャが実行されているとの決定をもたらす場合、前記ジェスチャが実行されているかどうかを決定するための、前記複数のジェスチャモデルのうちの別のジェスチャモデルを使用せずに前記第 1 のジェスチャを使用した前記ソース画像のシーケンスのスキャンを終了することを行うように構成されたプロセッサとを備える、装置。

【請求項 1 2】

前記複数のジェスチャモデルが手のポーズを備える、請求項 1 1 に記載の装置。

【請求項 1 3】

前記複数のジェスチャモデルが顔の表情を備える、請求項 1 1 に記載の装置。

【請求項 1 4】

前記複数のジェスチャモデルが、左の開いた手のモデル、右の開いた手のモデル、または拳のモデルを備える、請求項 1 1 に記載の装置。

【請求項 1 5】

前記プロセッサが、

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の検出された前記ジェスチャの位置を、他の位置よりも優先し、

前記検出されたジェスチャの優先された前記位置を使用して、前記ジェスチャが実行されているかどうかを決定するために、ソース画像の前記シーケンスをスキャンするようにさらに構成される、請求項 1 1 に記載の装置。

【請求項 1 6】

前記プロセッサが、

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の前記検出されたジェスチャのスケールを、他のスケールよりも優先し、

前記検出されたジェスチャの前記優先されたスケールを使用して、前記ジェスチャが実行されているかどうかを決定するために、ソース画像の前記シーケンスをスキャンするようにさらに構成される、請求項 1 1 に記載の装置。

【請求項 1 7】

前記プロセッサが、

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の前記検出されたジェスチャの位置を、他の位置よりも優先し、

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の前記検出されたジェスチャのスケールを、他のスケールよりも優先し、

前記検出されたジェスチャの優先された位置、および前記検出されたジェスチャの優先されたスケールの組合せを使用して、前記ジェスチャが実行されているかどうかを決定するために、ソース画像の前記シーケンスをスキャンするようにさらに構成される、請求項 11 に記載の装置。

【請求項 18】

前記プロセッサがソース画像の前記シーケンス内であらかじめ定められた数のジェスチャが検出されると、ソース画像の前記シーケンスのフルスキャンを完了する前に、ソース画像の前記シーケンス内の前記ジェスチャのスキャンを終了するようにさらに構成される、請求項 11 に記載の装置。

【請求項 19】

前記プロセッサが、

前記参照画像内に第 2 のジェスチャが検出される可能性が、前記ジェスチャよりも低いと決定し、

前記第 2 のジェスチャが検出される可能性がより低いと決定するステップに基づいて、前記ジェスチャを検出した後、前記第 2 のジェスチャの前記参照画像をスキャンするようにさらに構成される、請求項 11 に記載の装置。

【請求項 20】

前記プロセッサが、

前記ジェスチャの前記参照画像をスキャンすること、

連続する数の段階内で順序付けられた複数のカスケード接続された段階において、前記ジェスチャの前記参照画像のスキャンを絞り込むことであって、前記カスケード接続された段階の各段階が、前記ジェスチャを検出するために、ソース画像の前記シーケンス内の少なくとも 1 つの画像内のピクセルの周期数をスキャンするステップを備え、ピクセルの前記周期数が、ピクセルの前記周期数内で互いのピクセルから離れた一定の垂直方向距離および水平方向距離を有し、任意の段階におけるピクセルの前記周期数が、以前の段階のピクセルの前記周期数以下である、こと、および

前記連続する数の段階の最後の段階が、前記ジェスチャの少なくとも一部を検出するかどうかを決定することによって、前記ジェスチャを識別することを行うようにさらに構成される、請求項 11 に記載の装置。

【請求項 21】

ジェスチャが、ソース画像のシーケンス内で実行されているかどうかを決定するための装置であって、

参照画像の中の前記ジェスチャを検出するための手段と、

複数のジェスチャモデルから、検出された前記ジェスチャに最も密接に一致する第 1 のジェスチャモデルを選択するための手段であって、前記第 1 のジェスチャモデルは、検出された前記ジェスチャと同じタイプを有するジェスチャを含む画像の合成を含む、手段と

前記ジェスチャが実行されているかどうかを決定するために、前記第 1 のジェスチャモデル以外を使用せずに前記第 1 のジェスチャモデルを使用して、ソース画像の前記シーケンスをスキャンするための手段であって、前記ソース画像のシーケンスは、前記参照画像の後に取得された複数の画像を含み、前記ソース画像のシーケンスをスキャンすることが、前記参照画像の後に取得されたソース画像と前記第 1 のジェスチャモデルを比較し、前記ソース画像内で前記ジェスチャが実行されているかどうかを決定することを含む、手段と、

前記第 1 のジェスチャモデルを使用して前記画像のシーケンスをスキャンし前記ジェスチャが実行されているかどうかを決定することが、前記ジェスチャが実行されているとの決定をもたらす場合、前記ジェスチャが実行されているかどうかを決定するための、前記複数のジェスチャモデルのうちの別のジェスチャモデルを使用せずに前記第 1 のジェスチャ

ャを使用した前記ソース画像のシーケンスのスキンを終了するための手段とを備える、装置。

【請求項 2 2】

前記複数のジェスチャモデルが手のポーズを備える、請求項 2 1 に記載の装置。

【請求項 2 3】

前記複数のジェスチャモデルが顔の表情を備える、請求項 2 1 に記載の装置。

【請求項 2 4】

前記複数のジェスチャモデルが、左の開いた手のモデル、右の開いた手のモデル、または拳のモデルを備える、請求項 2 1 に記載の装置。

【請求項 2 5】

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の前記検出されたジェスチャの位置を、他の位置よりも優先するための手段と、

前記検出されたジェスチャの前記優先された位置を使用して、前記ジェスチャが実行されているかどうかを決定するために、ソース画像の前記シーケンスをスキャンするための手段とをさらに備える、請求項 2 1 に記載の装置。

【請求項 2 6】

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の前記検出されたジェスチャのスケールを、他のスケールよりも優先するための手段と、

前記検出されたジェスチャの優先された前記スケールを使用して、前記ジェスチャが実行されているかどうかを決定するために、ソース画像の前記シーケンスをスキャンするための手段とをさらに備える、請求項 2 1 に記載の装置。

【請求項 2 7】

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の前記検出されたジェスチャの位置を、他の位置よりも優先するための手段と、

ソース画像の前記シーケンス内の前記ジェスチャを検索するために、前記参照画像内の前記検出されたジェスチャのスケールを、他のスケールよりも優先するための手段と、

前記検出されたジェスチャの優先された位置、および前記検出されたジェスチャの優先された前記スケールの組合せを使用して、前記ジェスチャが実行されているかどうかを決定するために、ソース画像の前記シーケンスをスキャンする手段とをさらに備える、請求項 2 1 に記載の装置。

【請求項 2 8】

ソース画像の前記シーケンス内であらかじめ定められた数のジェスチャが検出されると、ソース画像の前記シーケンスのフルスキャンを完了する前に、ソース画像の前記シーケンス内の前記ジェスチャのスキンを終了するための手段とをさらに備える、請求項 2 1 に記載の装置。

【請求項 2 9】

前記参照画像内に第 2 のジェスチャが検出される可能性が、前記ジェスチャよりも低いと決定するための手段と、

前記第 2 のジェスチャが検出される可能性がより低いと決定するステップに基づいて、前記ジェスチャを検出した後、前記第 2 のジェスチャの前記参照画像をスキャンするための手段とをさらに備える、請求項 2 1 に記載の装置。

【請求項 3 0】

前記ジェスチャの前記参照画像をスキャンするための手段と、

連続する数の段階内で順序付けられた複数のカスケード接続された段階において、前記ジェスチャの前記参照画像のスキンを絞り込むための手段であって、前記カスケード接続された段階の各段階が、前記ジェスチャを検出するために、ソース画像の前記シーケンス内の少なくとも 1 つの画像内のピクセルの周期数をスキャンするステップを備え、ピクセルの前記周期数が、ピクセルの前記周期数内で互いのピクセルから離れた一定の垂直方向距離および水平方向距離を有し、任意の段階におけるピクセルの前記周期数が、以前の段階のピクセルの前記周期数以下である手段と、

10

20

30

40

50

前記連続する数の段階の最後の段階が、前記ジェスチャの少なくとも一部を検出するかどうかを決定することによって、前記ジェスチャを識別するための手段とをさらに備える、請求項 2 1 に記載の装置。

【請求項 3 1】

プロセッサに、

参照画像の中のジェスチャを検出すること、

複数のジェスチャモデルから、検出された前記ジェスチャに最も密接に一致する第 1 のジェスチャモデルを選択することであって、前記第1のジェスチャモデルは、検出された前記ジェスチャと同じタイプを有するジェスチャを含む画像の合成を含む、こと、

前記ジェスチャが実行されているかどうかを決定するために、前記第1のジェスチャモデル以外を使用せずに前記第1のジェスチャモデルを使用して、ソース画像のシーケンスをスキャンすることであって、前記ソース画像のシーケンスは、前記参照画像の後に取得された複数の画像を含み、前記ソース画像のシーケンスをスキャンすることが、前記参照画像の後に取得されたソース画像と前記第1のジェスチャモデルを比較し、前記ソース画像内で前記ジェスチャが実行されているかどうかを決定することを含む、こと、および

前記第1のジェスチャモデルを使用して前記画像のシーケンスをスキャンし前記ジェスチャが実行されているかどうかを決定することが、前記ジェスチャが実行されているとの決定をもたらす場合、前記ジェスチャが実行されているかどうかを決定するための、前記複数のジェスチャモデルのうちの別のジェスチャモデルを使用せずに前記第1のジェスチャを使用した前記ソース画像のシーケンスのスキャンを終了することを
行なわせるように構成されるプロセッサ可読命令を備える、非一時的プロセッサ可読記録媒体。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、高速ポーズ検出器に関するものである。

【背景技術】

【0 0 0 2】

本開示は、記録された画像内のジェスチャ検出に関する。より詳細には、本開示は、第 1 の画像および画像のシーケンス内のジェスチャを検出する際の速度、精度、および消費されるリソースを改善するための効率的な方法および装置を議論する。当分野では、手のポーズ、腕の動き、および同類のものなどのジェスチャを検出するためのいくつかの技法が知られている。しかしながら、その多くは、大型で、電力を消費する機器内にのみ見受けられる高性能プロセッサに依存し得る。ジェスチャ検出方法はそれほど効率的ではなく、より小型のモバイルデバイスで使用する際に、および/または一定のリアルタイム画像記録に使用する際に、そのような方法を困難にする場合がある。したがって、それほど時間がかからず、より電力効率のよいジェスチャ検出方法を実装することが望ましい。

【発明の概要】

【発明が解決しようとする課題】

【0 0 0 3】

これらおよび他の問題は、本明細書に記載の様々な実施形態によって解決することができる。

【課題を解決するための手段】

【0 0 0 4】

いくつかの実施形態では、ジェスチャがソース画像のシーケンス内で実行されているかどうかを決定するための方法が提示される。本方法は、複数のジェスチャモデルのうちの 1 つまたは複数のジェスチャモデルを使用して、1 つまたは複数の参照画像のそれぞれの中のジェスチャを検出するステップを含み得る。本方法は、1 つまたは複数のジェスチャモデルから、検出されたジェスチャに最も密接に一致する第 1 のジェスチャモデルを選択するステップと、ソース画像のシーケンス内のジェスチャを検索するために、複数のジェ

スチャモデルのうちの第1のジェスチャモデルを、他のジェスチャモデルよりも優先するステップと、優先された第1のジェスチャモデルを使用して、ジェスチャが実行されているかどうかを決定するために、ソース画像のシーケンスをスキャンするステップとをさらに含み得る。ジェスチャが実行されている場合、本方法は、ジェスチャが実行されているかどうかを決定するために、複数のジェスチャモデルのうちの別のジェスチャモデルを使用する前に、スキャンを終了するステップも含み得る。

【0005】

いくつかの実施形態では、参照画像は、ソース画像のシーケンス内の以前の画像を含む。いくつかの実施形態では、複数のジェスチャモデルは手のポーズを含む。いくつかの実施形態では、複数のジェスチャモデルは顔の表情を含む。いくつかの実施形態では、複数のジェスチャモデルは、左の開いた手のモデル、右の開いた手のモデル、または拳のモデルを含む。

10

【0006】

いくつかの実施形態では、本方法は、ソース画像のシーケンス内のジェスチャを検索するために、1つまたは複数の参照画像内の検出されたジェスチャの位置を、他の位置よりも優先するステップと、検出されたジェスチャの優先された位置を使用して、ジェスチャが実行されているかどうかを決定するために、ソース画像のシーケンスをスキャンするステップとをさらに含む。

【0007】

いくつかの実施形態では、本方法は、ソース画像のシーケンス内のジェスチャを検索するために、1つまたは複数の参照画像内の検出されたジェスチャのスケールを、他のスケールよりも優先するステップと、検出されたジェスチャの優先されたスケールを使用して、ジェスチャが実行されているかどうかを決定するために、ソース画像のシーケンスをスキャンするステップとをさらに含む。

20

【0008】

いくつかの実施形態では、本方法は、ソース画像のシーケンス内のジェスチャを検索するために、1つまたは複数の参照画像内の検出されたジェスチャの位置を、他の位置よりも優先するステップと、ソース画像のシーケンス内のジェスチャを検索するために、1つまたは複数の参照画像内の検出されたジェスチャのスケールを、他のスケールよりも優先するステップと、検出されたジェスチャの優先された位置、検出されたジェスチャの優先されたスケール、および優先された第1のジェスチャモデルの組合せを使用して、ジェスチャが実行されているかどうかを決定するために、ソース画像のシーケンスをスキャンするステップとをさらに含む。

30

【0009】

いくつかの実施形態では、本方法は、ソース画像のシーケンス内であらかじめ定められた数のジェスチャが検出されると、ソース画像のシーケンスのフルスキャンを完了する前に、ソース画像のシーケンス内のジェスチャのスキャンを終了するステップとをさらに含む。

【0010】

いくつかの実施形態では、本方法は、1つまたは複数の参照画像内に第2のジェスチャが検出される可能性が、ジェスチャよりも低いと決定するステップとをさらに含む。また、本方法は、第2のジェスチャが検出される可能性がより低いと決定するステップに基づいて、ジェスチャを検出した後、第2のジェスチャの1つまたは複数の参照画像をスキャンするステップを含む。

40

【0011】

いくつかの実施形態では、本方法は、ジェスチャの1つまたは複数の参照画像をスキャンするステップと、連続する数の段階内で順序付けられた複数のカスケード接続された段階において、ジェスチャの1つまたは複数の参照画像のスキャンを絞り込むステップとをさらに含む。カスケード接続された段階の各段階は、ジェスチャを検出するために、画像内のピクセルの周期数をスキャンするステップを含み得る。ピクセルの周期数は、ピクセ

50

ルの周期数内で、互いのピクセルから一定の垂直方向距離および水平方向距離だけ離れていてよく、任意の段階におけるピクセルの周期数は、以前の段階のピクセルの周期数以下でよい。また、本方法は、連続する数の段階の最後の段階が、ジェスチャの少なくとも一部を検出するかどうかを決定することによって、ジェスチャを識別するステップを含み得る。

【 0 0 1 2 】

いくつかの実施形態では、ジェスチャがソース画像のシーケンス内で実行されているかどうかを決定するための装置が提示される。本装置は、1つまたは複数の参照画像、およびソース画像のシーケンスを格納するように構成されたメモリを含み得る。また、本装置は、メモリに通信可能に結合されたプロセッサを含み得る。本プロセッサは、複数のジェスチャモデルのうちの1つまたは複数のジェスチャモデルを使用して、1つまたは複数の参照画像のそれぞれの中のジェスチャを検出し、1つまたは複数のジェスチャモデルから、検出されたジェスチャに最も密接に一致する第1のジェスチャモデルを選択し、ソース画像のシーケンス内のジェスチャを検索するために、複数のジェスチャモデルのうちの第1のジェスチャモデルを他のジェスチャモデルよりも優先し、優先された第1のジェスチャモデルを使用して、ジェスチャが実行されているかどうかを決定するために、ソース画像のシーケンスをスキャンし、ジェスチャが実行されている場合、別のジェスチャモデルを使用する前に、スキャンを終了するように構成され得る。

10

【 0 0 1 3 】

いくつかの実施形態では、装置が提示される。本装置は、複数のジェスチャモデルのうちの1つまたは複数のジェスチャモデルを使用して、1つまたは複数の参照画像のそれぞれの中のジェスチャを検出するための手段と、1つまたは複数のジェスチャモデルから、検出されたジェスチャに最も密接に一致する第1のジェスチャモデルを選択するための手段と、ソース画像のシーケンス内のジェスチャを検索するために、複数のジェスチャモデルのうちの第1のジェスチャモデルを、他のジェスチャモデルよりも優先するための手段と、優先された第1のジェスチャモデルを使用して、ジェスチャが実行されているかどうかを決定するために、ソース画像のシーケンスをスキャンするための手段と、ジェスチャが実行されている場合、ジェスチャが実行されているかどうかを決定するために、複数のジェスチャモデルのうちの別のジェスチャモデルを使用する前に、スキャンを終了するための手段とを含み得る。

20

30

【 0 0 1 4 】

いくつかの実施形態では、非一時的プロセッサ可読媒体が提示される。非一時的プロセッサ可読媒体は、プロセッサに、複数のジェスチャモデルのうちの1つまたは複数のジェスチャモデルを使用して、1つまたは複数の参照画像のそれぞれの中のジェスチャを検出させ、1つまたは複数のジェスチャモデルから、検出されたジェスチャに最も密接に一致する第1のジェスチャモデルを選択させ、ソース画像のシーケンス内のジェスチャを検索するために、複数のジェスチャモデルのうちの第1のジェスチャモデルを他のジェスチャモデルよりも優先させ、優先された第1のジェスチャモデルを使用して、ジェスチャが実行されているかどうかを決定するために、ソース画像のシーケンスをスキャンさせ、ジェスチャが実行されている場合、本方法は、ジェスチャが実行されているかどうかを決定するために、複数のジェスチャモデルのうちの別のジェスチャモデルを使用する前に、スキャンを終了させるように構成されたプロセッサ可読命令を含み得る。

40

【 0 0 1 5 】

いくつかの実施形態では、手のポーズなどの、1つまたは複数のジェスチャを検出するための方法および装置が提示される。いくつかの実施形態では、方法は、たとえばカメラや視覚装置の正面に伸ばされた開いた手のひらなどの、開いた手のひらを検出するステップを含む。しかしながら、これらの実施形態は限定的なものではなく、当業者は、以下で説明する実施形態は、他のポーズおよび/またはジェスチャ、あるいは動きを検出するために使用され得ることが認識できるだろう。いくつかの実施形態では、ポーズまたはジェスチャを検出するために、カスケード接続された弱分類器が使用される。いくつかの実施

50

形態では、本検出方法は回転集約的でよく、たとえば、検出はマルチスケールビデオフレーム、および各フレームの各ピクセルに焦点を当ててよい。

【0016】

いくつかの実施形態では、提示される方法は、手追跡アルゴリズムの第1のステップでよい。いくつかの実施形態では、たとえば、最初の係合位置 (engagement position) または再係合位置 (re-engagement position)、および/あるいは最初の係合位置に対する現在の位置の追跡が提供され得る。いくつかの実施形態では、ポーズが一定の時間、たとえば500ミリ秒の間静止している場合、係合が認識される。いくつかの実施形態では、本方法は、デバイスの電源をオンにするためにジェスチャを認識することができるようにするために、閲覧するためのデバイスがたとえスタンバイモードでも実行すること可能である。加えて、いくつかの方法は、偽の、および/または失った追跡から回復 (再係合) するために、低周期で実行することができる。

10

【0017】

いくつかの実施形態では、方法は、スケーリング、前処理、応答計算、および矩形結果処理の少なくとも4つの部分を含み得る。

【0018】

様々な実施形態の性質および利点の理解は、以下の図面を参照することによって実現され得る。添付の図面において、同様の構成要素または特徴は、同じ参照ラベルを有し得る。さらに、同じタイプの様々な構成要素は、同様の構成要素を区別するダッシュおよび第2のラベルによる参照ラベルに従うことによって区別され得る。本明細書において第1の参照ラベルのみが使用される場合、説明は、第2の参照ラベルに関係なく同じ第1の参照ラベルを有する同様の構成要素のいずれにも適用可能である。

20

【図面の簡単な説明】

【0019】

【図1】いくつかの実施形態による、例示的なデバイスを示す図である。

【図2A】いくつかの実施形態による、画像記録デバイスによって検出されるジェスチャの例示的なシナリオを示す図である。

【図2B】いくつかの実施形態による、画像記録デバイスによって検出されるジェスチャの例示的なシナリオを示す図である。

30

【図3】いくつかの実施形態による、例示的なジェスチャおよびジェスチャモデルを示す図である。

【図4】いくつかの実施形態による、例示的な処理フローを示す図である。

【図5】いくつかの実施形態による、ジェスチャを検出するための例示的な流れ図である。

【図6】いくつかの実施形態による、ジェスチャを検出するための別の例示的な流れ図である。

【図7】いくつかの実施形態による、例示的なカスケード接続された段階のサンプル分布を示す図である。

【図8】いくつかの実施形態による、ビデオフレーム上でスライドする例示的なモデルを示す図である。

40

【図9A】いくつかの実施形態による、例示的な流れ図である。

【図9B】いくつかの実施形態による、例示的な流れ図である。

【図10】本開示の1つまたは複数の態様が実装され得る、例示的なコンピューティングシステムを示す図である。

【発明を実施するための形態】

【0020】

本明細書では、「例示的 (exemplary)」という単語は、「例、事例、または例示として役立つ」を意味するために使用される。本明細書で「例示的 (exemplary)」として説明されるあらゆる実施形態または設計は、必ずしも他の実施形態または

50

設計よりも好ましいまたは有利であると解釈されるべきではない。

【 0 0 2 1 】

本明細書で使用される場合、「ジェスチャ (g e s t u r e) 」は、人体の一部を使って行われる非言語コミュニケーションの形式を指すことができ、スピーチなどの言語コミュニケーションと対比される。たとえば、ジェスチャは、第 1 の位置、ポーズ、または表情と、第 2 の位置、ポーズ、または表情との間の、動作、変化、または変形によって定義され得る。本明細書で使用される場合、「ジェスチャ」は、人体の一部を使って行われる静止ポーズも含み得る。たとえば、「ジェスチャ」は、行われた動きの一部のみを示す静止ポーズを含み得る。日常会話で使用される一般的なジェスチャは、たとえば、「エアクォーツ (a i r q u o t e) 」ジェスチャ、お辞儀ジェスチャ、ひざを曲げるお辞儀、頬へのキス、指または手の動き、ひざまずくこと、頭を上下するまたは動かすこと、ハイタッチ、うなづくこと、悲しい顔、拳を上げること、敬礼、親指を挙げる動き、ピンチジェスチャ、開いた手のひら、握った拳、拳を揺らすこと、指差し、「ピース (p e a c e) 」サイン、あるいは手または体をねじる任意の動きを含み得る。ジェスチャは、カメラを使用して、ユーザの画像を分析することによって、傾斜センサを使用してユーザがデバイスを保持または傾けている角度を検出することなどによって、あるいは他の任意の手法によって検出され得る。当業者は、上記の説明および以下のさらなる説明から、ジェスチャは、たとえば空中で行われる手の動作などの、非タッチ、タッチレス、またはタッチフリーのジェスチャを備え得ることが理解できるであろう。そのような非タッチ、タッチレス、またはタッチフリーのジェスチャは、たとえば、いくつかの実施形態では、タッチスクリーン上にパターンを描くことによって実行され得る様々な「ジェスチャ」から区別され得る。いくつかの実施形態では、ジェスチャは、デバイスを保持しながら空中で行うことができ、加速度計などのデバイス内の 1 つまたは複数のセンサを使用してジェスチャを検出することができる。

【 0 0 2 2 】

ユーザは、体の部分の位置を変えること（すなわち、振る動き）によってジェスチャを行う（または「ジェスチャする (g e s t i c u l a t e) 」）こともでき、体の部分を一定の位置に保持したままジェスチャする（すなわち、握り拳ジェスチャを行うことによって）こともできる。いくつかの構成では、カメラ入力を介して機能を制御するために手および腕のジェスチャを使用することができ、他の構成では、他のタイプのジェスチャをさらに、または代替で使用する事ができる。さらに、または代替で、1 つまたは複数のジェスチャを行う際に、手および/または他の体の部分（たとえば、腕、頭、顔、胴体、脚、足、等）を動かすことができる。たとえば、あるジェスチャは、片方または両方の手を動かすことによって行うことができ、他のジェスチャは、片方または両方の手を、片方または両方の腕、片方または両方の脚等と組み合わせて動かすことによって行うことができる。いくつかの実施形態では、ジェスチャは、しきい値の時間量にわたって保持される特定のポーズ、たとえば手または体のポーズを含み得る。

【 0 0 2 3 】

撮像デバイスの視野内のオブジェクトのジェスチャまたはポーズを効率的に検出するための方法および装置が提示される。オブジェクトのジェスチャまたはポーズを検出するステップは、従来する方法を使用するプロセッサ集約的でもよく、メモリ集約的でもよい。ジェスチャアルゴリズムおよび方法を採用する、いくつかの利用可能なジェスチャデバイスは、たとえば片手を追跡しようとするときに極めて C P U 集約的である。高い C P U 使用率のため、ジェスチャ検出方法を採用しようとするモバイルデバイスは、ユーザに悪い経験を与えることがある。高い C P U 使用率が、ジェスチャ検出のより低いフレームレートをもたらすことがあり、場合によっては、遅く、使用不可能で、信頼できない追跡結果、高い電池消費、および過熱デバイスをもたらすことさえある。しかしながら、様々な実施形態によれば、ジェスチャまたはポーズを検出するステップは、より少ない消費電力で、より短い時間内にジェスチャまたはポーズをうまく検出することができる。たとえば、実施形態は、従来する方法と比較して精度を低下させずに、30 % 短い時間を使用して、ユー

ザの右手の開いた手の動きを 176×144 ピクセルフレームで検出することができる。

【0024】

いくつかの実施形態では、ポーズまたはジェスチャをより効率的に検出するために、ポーズまたはジェスチャのモデルが使用され得る。たとえば、当該の現在の画像に関する以前の参照画像（ソース画像と呼ばれる）では、ポーズまたはジェスチャは、より遅い、またはよりプロセッサ集約的な画像認識技法を使用して識別され得る。次いで、参照画像内で検出されたポーズまたはジェスチャのモデルが識別され得る。本明細書の開示は、このモデルをジェスチャモデルと呼ぶ場合がある。ジェスチャモデルは、同じポーズまたはジェスチャの数百または数千の画像の合成であってよい。いくつかの実施形態では、ジェスチャモデルは、同じポーズまたはジェスチャの数百または数千の画像に基づく理想的なポーズまたはジェスチャである。次いで、ソース画像内の同じタイプのジェスチャまたはポーズを検出するために、識別されたジェスチャモデルが使用され得る。いくつかの実施形態では、多くの場合、参照画像内で検出されたジェスチャまたはポーズはソース画像内で検出されたものと同じである可能性が高いので、この技法は50%のリソースを節約することができる。いくつかの実施形態では、ジェスチャモデルおよび説明した技法を使用する検出は、ソース画像のシーケンスのために行われ得る。たとえば、いくつかの実施形態では、一旦ジェスチャモデルが選択されると、画像の一部を他のモデルと比較する前に、それらの画像の一部を、全体的なジェスチャモデル、またはジェスチャモデルの特徴と比較することによって、ソース画像がスキャンされ得る。これらの比較によって、たとえば、比較における類似性が何らかのしきい値を超えた場合、または比較があらかじめ定められた制約内であった場合、画像内でジェスチャが実行されていることを決定することができる。比較は、ジェスチャモデルを使用して参照画像内のジェスチャを検出するために使用される技法と同じタイプであってよい。しかしながら、優先されたジェスチャモデルがあるソース画像をスキャンする際の1つの利点は、優先されたジェスチャモデルは、ソース画像内のジェスチャと一致する可能性がより高く、したがって他のモデルで始まる画像をスキャンする必要がないのでリソースが節約されるということである。

【0025】

図1を参照すると、本開示の1つまたは複数の態様を実装することができる例示的なデバイスが示されている。たとえば、コンピューティングデバイス100は、パーソナルコンピュータ、セットトップボックス、カメラ、電子ゲーム機デバイス、ラップトップコンピュータ、スマートフォン、タブレットコンピュータ、携帯情報端末、あるいは、コンピューティングデバイス100が、動き、および/または他の感知された状態をユーザ入力の形態でキャプチャすることを可能にする1つまたは複数のセンサを備えた他のモバイルデバイスでよい。たとえば、コンピューティングデバイス100は、1つまたは複数のカメラ、マイク、近接センサ、ジャイロスコープ、加速度計、圧力センサ、グリップセンサ、タッチスクリーン、および/または他のセンサを備えてもよく、通信可能に結合されてもよく、および/または他の方法で含んでもよい。以下でより詳細に説明するように、1つまたは複数のセンサを含むことに加えて、コンピューティングデバイス100は、1つまたは複数のプロセッサ、メモリユニット、および/または他のハードウェアコンポーネントも含み得る。いくつかの実施形態では、デバイス100は、自動車、たとえば自動車の中央コンソールに組み込まれている。

【0026】

1つまたは複数の構成では、コンピューティングデバイス100は、ジェスチャ、たとえばユーザがデバイス100をタッチすることを含まない場合がある、デバイスの1人または複数のユーザによって実行されるジェスチャを認識するために、これらのセンサのうちのいずれか、および/またはすべてを単独で、あるいは組み合わせて使用することができる。たとえば、コンピューティングデバイス100は、カメラ110などの1つまたは複数のカメラを使用して、他の可能な動作の中でも特に、手を振ること、またはスワイプの動きなどの、ユーザによって実行される手および/または腕の動作をキャプチャすることができる。加えて、たとえば、ユーザによって実行される体全体の動作（たとえば、歩

行、ダンス等)などの、より複雑な、および/またはスケールの大きい動作が、1つまたは複数のカメラ(および/または、他のセンサ)によって同様にキャプチャされて、続いてコンピューティングデバイス100によってジェスチャとして認識され得る。さらに別の例では、コンピューティングデバイス100は、タッチスクリーン120などの1つまたは複数のタッチスクリーンを使用して、他の可能な動きの中でも特に、ピンチ、スワイプ、および回転などの、ユーザによって提供されるタッチベースのユーザ入力をキャプチャすることができる。ここでは、単独でジェスチャと考えられ得る、および/またはより複雑なジェスチャを形成するために他の動作または行為と結合され得る、これらのサンプルの動作は例として説明されているが、他の任意の種類の動き、動作、行為、または他のセンサでキャプチャされるユーザ入力は、同様に、コンピューティングデバイス(以下、
「デバイス」と省略する場合がある。)100などの本開示の1つまたは複数の態様を実装するコンピューティングデバイスによってジェスチャ入力として受信されてもよく、および/またはジェスチャとして認識されてもよい。

10

【0027】

いくつかの実施形態では、たとえば、ユーザのジェスチャまたはジェスチャの変化の認識に基づいてコンピュータまたはメディアハブを制御するために、深度カメラなどのカメラが使用され得る。指紋の不明瞭な悪影響という欠点があるいくつかのタッチスクリーンシステムとは異なり、カメラベースのジェスチャ入力は、ユーザの自然な体の動作またはポーズに基づいて、写真、ビデオ、または他の画像を明確に表示する、または他の方法で出力すること可能にする。この利点を念頭に置いて、ユーザが画像オブジェクトに対して、閲覧、パン(すなわち、移動)、サイズ変更、回転、および他の操作を実行することを可能にするジェスチャを認識することができる。

20

【0028】

構造化光カメラ、または飛行時間型カメラなどの深度カメラは、赤外線エミッタおよびセンサを含み得る。深度カメラは、赤外線光のパルスを生成して、続いて光がオブジェクトに移動してセンサに戻ってくるまでに要する時間を測定し得る。距離は、移動時間に基づいて計算され得る。以下でより詳細に説明するように、入力を検出または受信するために、および/あるいはジェスチャの検出を助けるために、他の入力デバイスおよび/またはセンサが使用され得る。

【0029】

図2Aおよび図2Bを参照すると、例示的なシナリオとしていくつかの実施形態にしたがったジェスチャを行っているユーザを記録する画像記録デバイスとしてカメラ200を示している。画像記録デバイスは、図1における記述、およびデバイス100と一致し得る。図2Aで、ユーザは、記録されている間、カメラ200に向かってジェスチャ210を行う。ジェスチャ210は、たとえば、ユーザの左手でカメラに向かって指差す動きでよい。カメラ200は、ジェスチャを行っているユーザの単一の写真だけを記録することもでき、手を上げてジェスチャ210を行うユーザの画像のシーケンスを記録することもできる。記録した間、または記録した後、カメラ200は、ジェスチャ210がユーザによって実行されているかどうかを検出するために、いくつかの実施形態による方法を実行することができる。

30

40

【0030】

図2Bで、ユーザは、今度は右手で異なるジェスチャ220を行っている。ユーザは腕を上げて、右手を開く動きをカメラ200に示してよい。他の場合、図2Bで、ユーザは右手をカメラに向かって振ってよい。いくつかの実施形態による方法は、右手によるジェスチャと左手によるジェスチャを区別することができる。いくつかの実施形態による方法は、ユーザのジェスチャは何でも検出することができる。図2Aおよび図2Bは、ジェスチャのタイプの簡単な例にすぎない。

【0031】

画像のシーケンス内のジェスチャ210または220を検出するための例示的な処理は、次の通りでよい。いくつかの実施形態では、1つまたは複数の参照画像内の第1のジェ

50

スチャが検出され得る。参照画像は、カメラ 200 によって記録された画像のシーケンスの第 1 の画像内で検出されるジェスチャ 210 などの、画像のシーケンスの第 1 の画像でよい。いくつかの実施形態では、第 1 のジェスチャを効率的に検出するための方法を以下でさらに説明する。他の場合では、参照画像内の第 1 のジェスチャを検出するための方法は、当分野で知られている、単一の画像のジェスチャ検出方法によって実行され得る。たとえば左手でカメラを指差すなどの、第 1 のジェスチャを決定した後、いくつかの実施形態では、決定された第 1 のジェスチャに最も密接に一致するジェスチャモデルが選択される。モデルジェスチャは合成画像でもよく、または数百または数千のサンプル画像、または同じタイプのジェスチャのポーズに基づくモデルでもよい。たとえば、左手での指差しのジェスチャモデルは、指差す動きを行っている、数百または数千の異なる左手に基づいてよい。次いで、画像のシーケンス内の後の画像におけるジェスチャをより効率的に検出するために、ユーザのジェスチャ 210 に最も密接に類似しているジェスチャモデルが使用され得る。たとえば、図 2A でカメラ 200 によって記録された画像のシーケンスにおいて、方法は、左手の指差しのジェスチャモデルと類似または同一に見えるジェスチャの残りの画像のスキンを開始してよい。左手の指差しのジェスチャモデルは、画像のシーケンスに対して比較されるべき第 1 のジェスチャモデルとして、残りの画像内のジェスチャを検出するために使用される他のジェスチャモデルよりも優先され得る。参照画像内で選択されたジェスチャモデルを優先することによって、いくつかの実施形態によるジェスチャを検出するための方法は、画像のシーケンス内の残りの画像内のジェスチャをより迅速に検出することによって、エネルギーと時間を節約することができる。

【0032】

いくつかの実施形態では、パフォーマンスおよび効率をさらに向上するために、他のタイプの優先順位付けが使用される。例示的な優先順位付けは、第 1 のジェスチャモデルの後に第 2 のジェスチャモデルを優先すること、ジェスチャが検出された参照画像内の位置に基づいて後続の画像上の位置を優先すること、参照画像内で検出されたジェスチャのサイズに基づいて後続の画像のサイズを優先すること、およびどのタイプのジェスチャが、他のジェスチャよりも検出される可能性がより高いかを決定するために統計分析を使用することを含む。本方法は、以下でさらに説明する。

【0033】

図 3 を参照すると、いくつかの実施形態による、ジェスチャをより効率的に検出するために、モデルジェスチャのデータベース 300 が使用され得る。ここでは、データベース 300 に格納された例示的なジェスチャ 302、304、306、308、310、312、314、316、318、および 320 が、いくつかの実施形態によって検出可能な多くのタイプのジェスチャのほんの一部を示している。上述のように、いくつかの実施形態では、データベース 300 はジェスチャモデルで構成されてよく、各ジェスチャモデルは、同じタイプのジェスチャの数百または数千の異なる画像に基づいて、異なるタイプのジェスチャ、たとえばジェスチャ 302 ~ 320 の理想化された合成でよい。いくつかの実施形態では、モデルジェスチャは、単なる手のポーズ以上のモデルを含み得る。たとえば、モデルジェスチャは、腕のポーズ、顔の表情、人全体のポーズ、およびポーズのシーケンスで構成される動きの、理想化された合成を含み得る。実施形態は、これに限定されない。

【0034】

いくつかの実施形態では、モデルと同じになるためのジェスチャの検索は、単にジェスチャをより効率的に検出するための第 1 の技法でよい。この技法がジェスチャを検出できない場合、以下で説明する他の技法が使用され得る。したがって、いくつかの実施形態では、このモデル利用技法が他の技法よりも優先される。いくつかの実施形態では、ジェスチャモデルを使用する検出は、他のモデルを使用する検出よりも優先される。いくつかの実施形態では、ジェスチャをより効率的に検出するために、他のタイプの優先順位付け技法が使用され得る。いくつかの例を、以下でさらに説明する。

【0035】

いくつかの実施形態では、オブジェクトを追跡する際に比較される異なるモデルを格納するために、データベースまたはルックアップテーブル（ＬＵＴ）が使用され得る。たとえば、左手の開いた手のひらは、右手の開いた手のひらとは異なる格納された画像を有する場合がある。

【 0 0 3 6 】

いくつかの実施形態では、現在のフレームについての特徴の特定の検索を優先するために、以前のフレーム内の以前の検出から得られた情報が使用され得る。トラッカー係合のために、ポーズおよび位置は静止してよい。これによって、事前情報に基づいて、操作モードの優先順位付けへの洞察を与えることができる。以前に左の開いた手のひらが検出された場合、左の開いた手のひらの検索は、右の開いた手のひらなどの他の任意のモデル応答の前に実行され得る。同様に、矩形が一致を含む位置が、後続の検索の開始点として選択され得る。加えて、以前に一致があったスケールが、開始点として選択され得る。

10

【 0 0 3 7 】

いくつかの実施形態では、あらかじめ定められたピクセルのブロックが、モデルのモデルサイズとして指定され得る（たとえば、右の開いた手のひらのモデルサイズが 40×40 ピクセルであると決定される）。このモデルサイズは、ジェスチャまたはポーズをよりよく検出するために、フレーム上で垂直および水平にスライドするブロックとして使用され得る。いくつかの実施形態では、次いで、ジェスチャモデルと適切に比較できるようにするために、モデルサイズと一致するために、ソース画像内で識別されたジェスチャまたはポーズがスケールアップされ得る。

20

【 0 0 3 8 】

いくつかの実施形態では、現在のフレーム内の最も可能性が高いジェスチャまたはポーズがどこにあるかを助けるために、以前のフレームからの検出が利用される。たとえば、その時に以前のジェスチャまたはポーズが検出されたピクセルの位置が優先され得る。したがって、処理およびメモリ利用が節約できる。いくつかの実施形態では、以前ジェスチャが検出された所与のスケール内の位置が優先され得る。たとえば、小さなオブジェクトの動きを可能にするために、この位置の周囲の一定の割合の領域、たとえば 10% が選択され得る。異なるスケールおよび検索が実行されるように、最終的なオブジェクト境界ボックスがスケールアップされ得る。これによって、いくつかのスケールが除去されることを回避することができる。

30

【 0 0 3 9 】

いくつかの実施形態では、現在のソース画像の検索において、以前検出されたジェスチャまたはポーズのスケールまたはサイズが優先され得る。たとえば、サイズ 40×40 ピクセルの開いた手のひらが以前検出された場合、サイズ 40×40 ピクセルの開いた手のひらの寸法が、現在のソース画像内でまず検索され得る。たとえば、 40×40 ピクセルに近いサイズの開いた手のひらが次に優先され得る。この技法はスケール優先順位付けと呼ばれ得る。

【 0 0 4 0 】

いくつかの実施形態では、以前ジェスチャが正確に検出されたスケールが優先され得る。所与のスケール上で一致を提供するいくつかの矩形、および／または、ジェスチャを含む矩形に対する矩形位置が、この手段として使用され得る。たとえば、検出／追跡されているオブジェクトの小さい動き（カメラ軸に沿った）を可能にするために、選択されたスケールの上下の１つのスケールが可能になってよい。

40

【 0 0 4 1 】

いくつかの実施形態では、より可能性が高いジェスチャまたはポーズをより迅速に決定するために、確率的使用統計が実装され得る。たとえば、確率的使用統計が、左手の開いた手のひらよりも右手の開いた手のひらが現れる可能性の方が高いことを示す場合、右手の開いた手のひらのジェスチャまたはポーズがまず検索され得る。

【 0 0 4 2 】

一般的に、ハンドヘルドデバイスは単一のユーザデバイスである。ユーザは、左／右利

50

き、一定の距離からのデバイスの操作、および／または開いた手のひらのサイズなどの、利用に関して関連付けられる特定の統計性質を有する場合がある。説明した処理は、より広範な検索を実行する前に、これらの特性について優先され得る。このモデルは、係合ごとに実行される最初のオブジェクト検出の処理時間を確率的に最小化することができる。

【0043】

特性ごとの確率モデルは、デバイス上のユーザのために維持され得る。確率モデルは、使用するたびに更新され得る。たとえば、ユーザが右の開いた手のひらを係合の50%以上使用すると、左の開いた手のひらの前に、右の開いた手のひらが検索される。開いた手のひらのサイズまたはスケーリングレベルの確率を使用して、検索のためのスケールレベルを優先することができる。

10

【0044】

いくつかの実施形態では、ジェスチャまたはポーズの検索は、あらかじめ定められた数のジェスチャまたはポーズがすでに検出されると、早く終了するように指定され得る。たとえば、フレーム内で3つの開いた手のひらの画像を見つけると、追跡装置がスキャンを停止することができる。この処理によって、残りのフレームをスキャンする必要がなくなるので、処理およびメモリ利用を節約することができる。

【0045】

たとえば、手の最大数を1と指定すると、検出器は1つの一致が見つかるとすぐに返すことができる。いくつかの実施形態では、係合は、モバイルデバイス内のジェスチャ認識の場合、ジェスチャを1つだけ見つけることを含み得る。たとえば、テレビやアプリケーションでは、検出および追跡される手の数は最大で10である。いくつかの実施形態では、異なる数の手が検出および／または追跡され得る。

20

【0046】

たとえば、処理を減らす、および／または電力を節約するために、検出されるべきジェスチャの優先順位付けと最大数、フレームの検索ピクセル位置、および／または、以下の第1のモデルの一致が取得されるフレームの組合せが使用され得る。いくつかの実施形態では、係合および／または再係合シナリオのためにオブジェクト／モデルが検出される第1のフレームまでのみ、オブジェクト検出器が過剰な計算を実行する場合がある。第1のフレーム処理時間は、利用統計またはユーザの習慣に依存する場合がある。実質的には、第1の検出処理負荷が継続的に低下し得る（たとえば、適切な利用統計が利用可能な場合）。

30

【0047】

すべてのスケールの検出の数がしきい値を下回る場合、優先された検索の代わりに完全検索が実行され得る。加えて、継続的に矩形の数が減少している場合（たとえば、複数のフレームにわたって）、しきい値に達すると完全検索が実行され得る。これによって、オブジェクトの移動または向きの変化による偽陰性を回避することができる。各スケールの空間的近傍に多くの一致が存在する場合があり、加えて複数の隣接するスケールからの一致がある場合があるので、複数の矩形が検出され得る。すでに検索された位置の検索は、優先検索によって回避され得る。

【0048】

同様に、肯定的な検索結果がしきい値を超える場合、完全検索の代わりに優先検索が実行され得る。

40

【0049】

いくつかの実施形態では、画像内のオブジェクトのジェスチャまたはポーズをより効率的に検出するために、これらの優先順位付けのうちの任意のまたはすべての組合せが実装され得る。たとえば、いくつかの実施形態では、ジェスチャがソース画像のシーケンス内で実行されているかどうかを決定するために、優先されたジェスチャモデル、優先された位置、および優先されたスケールの組合せが使用され得る。組合せは、たとえば3つのすべてのタイプの優先順位付けを含んでもよく、または3つのうちの2つだけを含んでもよい。優先順位付けの組合せを使用してジェスチャが実行されているかどうかを決定する場

50

合、たとえば、まず第1のジェスチャモデルを使用してソース画像がスキャンされて、優先されたスケールでスケールアップされて、優先された位置の位置であってよい。あるいは、組合せが、第1のジェスチャモデルと優先されたスケールだけを含む場合、第1のジェスチャモデルが使用されて、優先されたスケールでサイズ変更されてよく、たとえばソース画像の左上角などの何らかのデフォルト位置でスキャンを開始してよい。当業者は、ソース画像内でジェスチャが実行されているかどうかを決定するために、他の組合せがどのように使用され得るかを容易に理解できるであろう。

【0050】

図4を参照すると、ブロック図400が、いくつかの実施形態による、参照画像内のジェスチャを検出するための例示的な方法フローを示している。ここで示されているサンプル技法は、最初にジェスチャを検出するための処理でよく、その後、後続の画像内のジェスチャを検出するために、モデル優先順位付け、または上述の他のタイプの優先順位付けがこれに基づき得る。この例示的なフローは、図示されるような4つの処理、スケールアップ、前処理、応答計算、および矩形結果処理を含み得る。スケールアップは、異なるサイズの開いた手のひら、あるいは他のジェスチャまたは制御の検出を処理することができる。ジェスチャと思われる画像の一部の検出は、あらかじめ定められたサイズのジェスチャモデル、または比較に使用される他のタイプのデータベース画像に適合するように拡大または縮小され得る。前処理段階は、しわ、影、およびジェスチャを行っている体の部分に共通する他の線などの特徴を生成することができる。第3の段階は、特定の分類器段階の選択された特徴への応答を計算することができる。最終段階は、任意の誤検出および/または重複検出を削除することができる。最初の2段階は、たとえばSIMD DSPおよびGPU内に存在する数学的ハードウェアユニットを使用する、加速のための候補でよい。いくつかの実施形態では、第3の段階（たとえば、カスケード接続された分類器）は、過剰なデータフェッチ指向を含む場合があり、処理能力およびメモリ帯域幅の大部分を消費する場合がある。第3の段階は、以下でさらに論じる。高速メモリおよびキャッシュブリフエッチを使用することによって、CPUストールに関連するデータフェッチを最小化することができる。最終段階は、検出されたジェスチャ矩形位置をグループ化することを含み、誤/重複矩形を削除して、ジェスチャを含む最終矩形位置を提供することができる。この段階は、CPUおよびメモリの使用量が少なくなるように実装され得る。

【0051】

実施形態は、図4で説明した方法を様々な装置内で実行することができる。たとえば、入力は、たとえばジェスチャ検出システムのカメラによってキャプチャされたような、ビデオフレームまたは他の画像を備え得る。図4は、たとえば、図1内の装置、または図2内のカメラによって実装され得る。また、一定の動きまたはポーズは、係合ジェスチャとして使用され得る。ユーザは、様々なジェスチャ、動き、またはポーズを実行することができ、それらはいくつかの実施形態の装置によって検出され得る。次いで、本装置は、本明細書に記載の方法のうちの少なくともいくつかを行うことができる。

【0052】

図5を参照すると、ブロック図500が、応答を計算する処理の例示的な処理フローを示している。このブロック図500は、たとえば図4の応答計算機ブロックを実装するための例示的な処理でよい。一般的に、ブロック図500は、ジェスチャが画像内にあるかどうかを決定するための方法を示している。図5は、2段階方法を使用してジェスチャを検出するための効率的な方法を説明し得る。第2の段階は、追跡詳細のレベルを第1の段階よりも上げることが伴い得る。たとえば、検出の第1の段階は、 n 個のピクセルごとに、水平および垂直に1つのピクセルを検索すること（たとえば、 $(x + / - 4, y + / - 4)$ パターン内のフレームの4番目のピクセルごとに検索すること）を伴う。これは、ブロック505および510に示されている。ブロック510は、計算応答機能によるジェスチャ検出分析が、ユーザの手または他の体の部分のサインを検出するかどうかを決定するために、しきい値テストを行う。処理フロー500は、それぞれのしきい値サイズ領域（たとえば、 4×4 ピクセル領域、 3×3 ピクセル領域等）内のピクセルについて、領域

ごとに応答が得られるまでブロック 5 0 5 および 5 1 0 を通じてループし得る。

【 0 0 5 3 】

第 2 の段階は、所望のジェスチャまたはポーズが検出された領域またはピクセルだけに焦点を当てることができ、したがって必要なスキャンの数が減少する。したがって、第 2 の段階は、詳細レベルが増加した、より焦点を絞った領域、たとえば 2 番目のピクセルごとと $(x + / - 2, y + / - 2)$ 、または 1 つ 1 つのピクセル $(x + / - 1, y + / - 1)$ を検索し得る。ブロック 5 1 5 および 5 2 0 は、別の反復ループを伴うこの分析を表している。ここでは、ジェスチャを検出した第 1 の段階におけるピクセルの周囲でのみ、より細かいレベルのジェスチャ検出が使用されている。このように、第 2 の段階はより細かいレベルの詳細を有するが、肯定的な検出をもたらした第 1 の段階からの画像の一部だけに焦点を当てている。ブロック 5 2 0 で、計算応答機能が何らかの最大しきい値を上回ると決定されると、体の一部の肯定的な検出を示し、次いで、スキャンされたピクセルでジェスチャが実行されているブロック 5 2 5 で終了し得る。

10

【 0 0 5 4 】

いくつかの実施形態では、多段階のカスケード接続された処理が使用され得る。たとえば、8 段階のカスケード処理が使用されてよく、段階 1 ~ 4 は、4 番目のピクセルごとに検索し、段階 5 ~ 6 は、段階 1 ~ 4 において肯定的な検出を有するピクセルの周囲だけの 2 番目のピクセルごとに検索し、段階 7 ~ 8 は、段階 5 ~ 6 において肯定的な検出を有するピクセルの周囲だけのすべてのピクセルを検索する。次いで、最後のカスケード段階が一定のピクセルで肯定的な検出をもたらす場合、所望のジェスチャまたはポーズがそれらの一定のピクセルでフレーム内に存在すると決定され得る。各段階は、ジェスチャ / ポーズ検出の高い信頼性があるピクセルの周囲だけの詳細レベルの増加に焦点を当てているので、ジェスチャもポーズも検出されないピクセルの高い詳細レベルを追跡しないことによって処理および時間が節約される。

20

【 0 0 5 5 】

図 6 を参照すると、図 6 0 0 は、本明細書に記載の多段階カスケード処理によるカスケード処理アルゴリズムを利用するいくつかの実施形態の例示的な処理フローを示している。たとえば、計算応答がカスケード段階 0 および 1 の 4 番目の位置ごとに計算応答が実行され得る場合、段階 0 および 1 の計算がすべてのピクセルに実行される場合と比較して、サイクルの約 7 2 % が節約され得る。図 6 で、ブロック 6 0 5、6 1 0、および 6 1 5 は、いくつかの実施形態では、たとえば 4 のステップサイズなどの第 1 のステップサイズにおいてピクセルごとに実行され得る。次いで、ブロック 6 0 5、6 1 0、および 6 1 5 は、いくつかの実施形態では、たとえば 2 のステップサイズなどの第 2 のステップサイズにおいてピクセルごとに実行され、次いで、ある実施形態では、たとえば 1 のステップサイズなどの第 3 のステップサイズについて実行され得る。

30

【 0 0 5 6 】

図 6 で、複数のステップサイズは、たとえば上述のように 4、2、1 を備え得る。いくつかの実施形態では、ステップサイズは 8、4、2、1 を備える。いくつかの実施形態では、他のステップサイズが使用され得る。いくつかの実施形態では、4、2、1 の検索ステップサイズを利用することによって、同じ検出精度の従来の方法と比較して約 3 0 % スピードアップする。

40

【 0 0 5 7 】

本方法では、カスケード分類器上で開いた手のひらまたはオブジェクトがあり得るという確信が高まるにつれてステップサイズが徐々に減少するため、検出精度が保持または改善され得る。たとえば、真の陰性を返す最初の 2 段階は、たとえば 4 - 2 - 1 検索の場合の 4 などの、最大ステップサイズを保持することができる。すなわち、最初の 2 段階に位置が戻った時に、さらなる検索のために 3 つの位置をスキップする。カスケード分類器が段階 3 ~ 8 に行った場合、段階 3 ~ 8 に行った位置の周囲に $+ / - 2$ の検索が実行され得る。最後に、検出器からジェスチャが検出されると、検索位置が、オブジェクトが検出された位置の周囲で $+ / - 1$ 改良され得る。したがって、検索ステップサイズは、オブジェ

50

クトまたは開いた手のひらの外の領域では4でよく、それによって余分な計算を回避することができる。さらに、検出されているジェスチャを含む領域内の、すべてのピクセル位置が検索され得る点に留意されたい。このように、図6に示される処理は、従来のジェスチャ検出方法と比較してサイクルを30%節約しながら、検出精度を保持することができる。

【0058】

いくつかの実施形態では、8個のカスケード接続された段階がある。前のカスケード段階が、すべての開いた手のひらの矩形および他の矩形を次の段階で処理されるように渡す。段階が進むにつれて、偽陰性の量が減少する。また、最後の段階が完了すると、検出されたジェスチャがある場合はそれが返される。任意のピクセル位置で、すべての段階が終了でき、それによって画像がジェスチャを有していないことを示す。

10

【0059】

さらに、ステップサイズは、ビデオフレームの4つの境界のフレーム境界内に収まるように切り取られてよい。たとえば、4のステップサイズを作ることが有効なピクセル領域を越える場合、ステップサイズは、検索ピクセル位置を有効領域内の最後の可能なピクセルに保持するために切り取られてよい。これによって、境界ボックスエッジが有効なピクセル領域の境界に近いオブジェクトの検出精度を向上させることができる。

【0060】

図7を参照すると、グラフ700は、各カスケード接続された段階で発生し得る処理の例示的な量を示している。この例では、いくつかの実施形態では、ピクセル位置の86%が、開いた手のひらを検出することなしに段階0で返されている。この場合、スキャンされた画像は、単一のジェスチャである開いた手のひらを含んでいる。ピクセル位置の別の10%は段階1で返されており、その位置が開いた手のひらを有していないことを示している。残りの段階2から段階7は、開いた手のひらのないより少数の位置を返している。すべての段階が開いた手のひらを通過する場合、開いた手のひらの位置を返す位置は、x軸ラベル8に表示される。図7が示すように、段階0ではピクセルの大部分が所望のポーズの兆候を有していなかったため、それらのピクセルの周囲を検索する必要はないと思われる。処理および時間を節約することができる。図7によって、時間の大部分が、ピクセル位置(40×40ピクセルの大きさ)の大多数を処理する際に段階0および段階1で費やされているということがわかる。検出精度に影響を与えることなく、段階0および段階1によって処理されるピクセルの量を低減して、それによってサイクルを節約することが望ましい場合がある。いくつかの実施形態は、ステップサイズを狭めて検索する、またはモデルマッチングするというアイデアを実装する。

20

30

【0061】

図8を参照すると、図表800は、ビデオフレーム上で水平および垂直にスライドする40×40ピクセルのモデルディメンションブロック810を示している。いくつかの実施形態では、ブロック(40×40)は、カスケード接続された段階における検出の第1の段階で使用される増分領域を表し得る。モデルディメンションブロック(以下、「ブロック」と省略する。)810は、画像にわたって垂直および水平にスライドさせることによってジェスチャの画像を徐々にスキャンすることができる。一旦ブロック810がそのディメンション内にジェスチャを検出すると、ブロック810がジェスチャを見つけた位置内のより細かい解像度のスキャンが実行され得る。このスキャンは、上述のカスケード接続された段階と一致する場合がある。たとえば、ブロック810は、ブロック810がジェスチャ820を見つけるまで、段階0のスキャンだけを実行することができる。ブロック810が、開いた手のひらのジェスチャ820が位置する場所まで垂直および水平にスライドすると、実際にはジェスチャがそこに位置していることをより正確に決定するために、より高い段階のスキャンが実行され得る。いくつかの実施形態では、モデルディメンションに一致するために、ビデオ内の開いた手のひらのサイズが、モデルによって表される開いた手のひらと一致するようにビデオフレームが正規化され得る。

40

【0062】

50

様々な実施形態による方法および装置は、以下の利点を提供することができる。

【 0 0 6 3 】

1 . カスケード分類器の応答を使用して、4、2、1のステップサイズの適合および改良で検出精度を落とさずに、サイクルおよび帯域幅を30%以上節約する。

【 0 0 6 4 】

2 . 係合が一定の時間（たとえば、500ミリ秒）の間静止しているオブジェクトを含む、第1のオブジェクト検出の後、サイクル、帯域幅、および電気を90%以上節約することができる。

【 0 0 6 5 】

3 . 最初のオブジェクト検出性能が、以前の利用 / 係合統計から改善する。ユーザ特性に応じて、処理時間を最大50%減少することができる。

【 0 0 6 6 】

4 . 実施形態は、ハードウェア開発および並行処理にとって使いやすい。

【 0 0 6 7 】

5 . 実施形態が高速であり、モバイル電池式デバイス上の手のジェスチャ認識を実現可能にする。

【 0 0 6 8 】

図9Aを参照すると、流れ図900は、いくつかの実施形態による例示的な方法論を示している。いくつかの実施形態では、ブロック902から開始して、複数のジェスチャモデルから1つまたは複数のジェスチャモデルを使用して、1つまたは複数の参照画像のそれぞれの中でジェスチャが検出され得る。検出は、たとえば、上述の処理のうちのいずれかによって実行され得る。1つまたは複数のジェスチャモデルは、図2または図3で説明した例示的なジェスチャのうちのいずれでもよく、又は任意の数のジェスチャに基づく他の任意のタイプのジェスチャモデルでもよい。ブロック902における検出は、1つまたは複数の参照画像内のスキャンされたジェスチャと複数のジェスチャモデルを比較することによって、1つまたは複数のジェスチャモデルを使用することができる。たとえば、右の開いた手のひらが参照画像内に記録されると、次いで、スキャンされたジェスチャが、最も一致するジェスチャモデルが見つかるまで、まず左の開いた手のひらのモデル、次いで左の握った拳のモデル、次いで右の開いた手のひらのモデル、以下同様と比較される。いくつかの実施形態では、適切な比較を実行するために、まず参照画像内のスキャンされたジェスチャが、まずジェスチャモデルの比較可能なサイズにスケーリングされ得る。ブロック902は、たとえば、デバイス100またはカメラ200によって実装され得る。

【 0 0 6 9 】

ブロック904で、第1のジェスチャモデルが、ブロック902で検出されたジェスチャに最も密接に一致すると選択され得る。第1のジェスチャモデルは、ブロック902で比較された1つまたは複数のジェスチャモデルの中のジェスチャモデルでよい。たとえば、右の開いた手のひらのジェスチャモデルが、右の開いた手のひらの検出されたジェスチャの第1のジェスチャモデルとして選択され得る。ブロック904は、たとえば、デバイス100またはカメラ200によって実装され得る。

【 0 0 7 0 】

ブロック906で、画像のシーケンス内のジェスチャを検索するために、第1のジェスチャモデルが他のジェスチャモデルよりも優先され得る。一旦1つまたは複数の参照画像内のジェスチャが検出されると、ブロック904で選択された第1のジェスチャモデルは、ソース画像内のジェスチャを検索する際に比較される第1のモデルでよい。いくつかの実施形態では、1つまたは複数の参照画像は、ソース画像のシーケンスの以前の画像である。たとえば、1つまたは複数の参照画像はビデオの最初のフレームでよく、ビデオの残りのフレームはジェスチャ検出を対象とする後続の画像でよい。したがって、参照画像内で検出されたジェスチャは、後続のソース画像内に存在するジェスチャである可能性が高く、第1のジェスチャモデルの優先順位付けを非常に効率的にする。ブロック906は、たとえば、デバイス100またはカメラ200によって実装され得る。

10

20

30

40

50

【 0 0 7 1 】

ブロック 9 0 8 で、ソース画像内でジェスチャが実行されているかどうかを決定するために、第 1 のジェスチャモデルを使用して、ソース画像のシーケンスがスキャンされ得る。言い換えれば、第 1 のジェスチャモデルが他のジェスチャモデルよりも優先されるので、ソース画像のシーケンスをスキャンするために第 1 のジェスチャモデルが使用される。ブロック 9 0 8 は、たとえば、デバイス 1 0 0 またはカメラ 2 0 0 によって実装され得る。

【 0 0 7 2 】

ブロック 9 1 0 で、ソース画像のシーケンス内でジェスチャが実行されていると決定されると、次いで、ジェスチャ検出のために別のジェスチャモデルを使用する前にスキャンが終了してよい。いくつかの実施形態では、第 1 のジェスチャモデルによってジェスチャが検出されない場合、スキャンを継続することができる。ブロック 9 1 0 は、たとえば、デバイス 1 0 0 またはカメラ 2 0 0 によって実装され得る。

【 0 0 7 3 】

図 9 B を参照すると、流れ図 9 5 0 は、いくつかの実施形態による、別の例示的な方法論を示している。ブロック 9 5 2、9 5 4、および 9 5 6 は、それぞれブロック 9 0 2、9 0 4、および 9 0 6 に関連して論じた説明と一致し得る。ブロック 9 5 2、9 5 4、および 9 5 6 は、たとえば、デバイス 1 0 0 またはカメラ 2 0 0 によって実装され得る。

【 0 0 7 4 】

いくつかの実施形態では、1 つまたは複数の参照画像内で検出されたジェスチャの他の特性も同様に優先され得る。実施形態に応じて、これらの優先順位付けのうちのいくつか、またはすべてが実装されてもよく、またいずれも実装されなくてもよく、実施形態はそれに限定されない。たとえば、ブロック 9 5 8 で、1 つまたは複数の参照画像内で検出されたジェスチャの位置は、画像内の他の位置よりも優先され得る。たとえば、ジェスチャのソース画像のスキャンを開始するとき、1 つまたは複数の参照画像内でジェスチャが検出された位置が、他の位置よりも先にスキャンされ得る。ブロック 9 5 8 は、たとえば、デバイス 1 0 0 またはカメラ 2 0 0 によって実装され得る。

【 0 0 7 5 】

ブロック 9 6 0 で、1 つまたは複数の参照画像内で検出されたジェスチャのスケールが、ソース画像内の他のスケールよりも優先され得る。たとえば、1 つまたは複数の参照画像内で検出されたジェスチャのサイズが約 40×40 ピクセルだった場合、ソース画像内のジェスチャをスキャンする際に 40×40 スケールがまず使用され得る。ブロック 9 6 0 は、たとえば、デバイス 1 0 0 またはカメラ 2 0 0 によって実装され得る。

【 0 0 7 6 】

ブロック 9 6 2 で、第 1 のジェスチャモデルの後ではあるが、第 2 のジェスチャモデルが他のジェスチャモデルよりも優先され得る。たとえば、第 1 のジェスチャモデルが右の開いた手のひらである場合、第 2 のジェスチャモデルは右の握った手として指定されてよい。第 2 のジェスチャモデルの選択は、何らかの最初のジェスチャの後で、次に現れる最も可能性が高いジェスチャに基づいて、統計的方法または分析に基づいてよい。他の場合、第 2 のジェスチャは、たとえば、参照画像の異なる場所で見つかった、またはユーザ入力に基づく、以前に記録または検出されたジェスチャとして選択され得る。第 1 のジェスチャモデルの後ではあるが、第 2 のジェスチャモデルを、他のジェスチャモデルよりも優先することによって、第 1 のジェスチャモデルがジェスチャモデルの検出に失敗した場合、ジェスチャを検索するために第 2 のジェスチャモデルが使用されることを示し得る。ブロック 9 6 2 は、たとえば、デバイス 1 0 0 またはカメラ 2 0 0 によって実装され得る。

【 0 0 7 7 】

ブロック 9 6 4 で、ジェスチャがソース画像のシーケンス内で実行されているかどうかを決定するために、たとえばブロック 9 5 6、9 5 8、9 6 0、または 9 6 2 内で、優先順位付けのうちのいずれかを使用して、ソース画像のシーケンスがスキャンされ得る。やはり、これらの優先順位付けのうちのいくつか、すべてが使用されてもよく、いずれも使

10

20

30

40

50

用されなくてもよい。これらは単にいくつかの例にすぎないので、これらの開示において論じた説明のうちのいずれかと一致する他の優先順位付けが使用され得る。ブロック 964 は、たとえば、デバイス 100 またはカメラ 200 によって実装され得る。

【0078】

決定ブロック 966 で、方法は、ソース画像のシーケンス内でジェスチャが検出されるかどうかを決定することができる。検出は、ブロック 964 で論じた優先順位付け、およびこれらの開示で論じた検出方法のいずれか（たとえば、カスケード接続された段階等）に基づいてよい。いくつかの実施形態では、本明細書で論じた優先順位付けとともに、当分野で知られている他の検出方法が使用され得る。決定ブロック 966 は、たとえば、デバイス 100 またはカメラ 200 によって実装され得る。

10

【0079】

決定ブロック 966 での決定がイエスの場合、次いでブロック 968 で、ジェスチャ検出のために別のジェスチャモデルを使用する前にスキャンが終了してよい。ブロック 970 で、決定がノーの場合、次いで、ジェスチャが実行されているかどうかを決定するために、第 2 のジェスチャモデルを使用してソース画像のシーケンスがスキャンされ得る。たとえば、右の握った手が、ソース画像のシーケンス内のジェスチャをスキャンするために使用される第 2 のジェスチャモデルでよい。さらに、第 2 のジェスチャモデルとともに他の優先順位付けも採用され得る。他の場合、他の優先順位付け、たとえば、位置優先順位付け、スケール優先順位付け等は使用することができず、あるいは異なる位置、またはスケール等を優先するためにそれらが修正され得る。当業者は、本明細書の開示による優先順位付けの多くの組合せが可能であること、および実施形態はそれに限定されないことが理解できるであろう。

20

【0080】

上で論じた方法、システム、およびデバイスは例である。様々な実施形態は、様々な手順または構成要素を、適宜省略、置換、または追加することができる。たとえば、代替構成では、説明された方法は、説明された順序とは異なる順序で実行されてよく、および/または様々な段階が追加、省略、および/または結合されてよい。また、特定の実施形態に関して説明された特徴は、他の様々な実施形態と結合されてよい。本実施形態の異なる態様および要素が、同様の方法で結合されてよい。また、技術は進化するので、要素の多くは、本開示の範囲をこれらの具体例に限定しない例である。

30

【0081】

実施形態の完全な理解を提供するために、具体的な詳細が本明細書に記載されている。しかしながら、実施形態はこれらの具体的な詳細なしに実践され得る。たとえば、よく知られている回路、処理、アルゴリズム、構造、および技法は、実施形態を曖昧にするのを回避するために、不要な詳細なしに示されている。この説明は、例示的な実施形態を提供するにすぎず、本発明の範囲、適用性、または構成を限定することを意図するものではない。むしろ、本実施形態の上記の説明は、本発明の実施形態を実装するための可能な説明を当業者に提供するであろう。本発明の趣旨および範囲から逸脱することなしに、要素の機能および構成に様々な変更が行われてよい。

【0082】

また、いくつかの実施形態は、流れ図またはブロック図として示される処理として説明された。それぞれは操作を連続的な処理として説明し得るが、操作の多くは並行して、または同時に実行され得る。加えて、操作の順序は並べ替えられてよい。処理は、図面には含まれていない追加のステップを有してよい。さらに、本方法の実施形態は、ハードウェア、ソフトウェア、ファームウェア、ミドルウェア、マイクロコード、ハードウェア記述言語、またはそれらの任意の組合せによって実装され得る。ソフトウェア、ファームウェア、ミドルウェア、またはマイクロコードに実装されると、関連するタスクを実行するためのプログラムコードまたはコードセグメントが、記憶媒体などのコンピュータ可読媒体に格納され得る。プロセッサは、関連するタスクを実行することができる。

40

【0083】

50

本開示の趣旨から逸脱することなしに、様々な修正形態、代替構造、および均等物が使用され得る。たとえば、上記の要素は、より大きなシステムの構成要素にすぎない場合があり、他のルールが本発明の適用よりも優先されてもよく、あるいは本発明の適用を修正してもよい。また、いくつかのステップが、上記の要素が考慮される前に、その間に、またはその後に行われてよい。したがって、上記の説明は本開示の範囲を限定しない。

【0084】

複数の態様を上記で説明してきたが、ここでは、そのような態様が実装され得るコンピューティングシステムの例を図10に関連して説明する。1つまたは複数の態様によれば、図10に示されるようなコンピュータシステムは、コンピューティングデバイスの一部として組み込まれてよく、本明細書に記載の特徴、方法、および/または方法ステップのうちの、いずれか、および/またはすべてを、実装、実行、および/または遂行することができる。たとえば、図4、5、6、9A、および9Bに示されるようなブロックのうちのいずれか、またはすべてを実装するために、プロセッサ1010、非一時的ワーキングメモリ1035、および通信サブシステム1030のうちの1つまたは複数が使用され得る。たとえば、コンピュータシステム1000は、ハンドヘルドデバイスの構成要素のうちのいくつかを表すことができる。ハンドヘルドデバイスは、カメラおよび/またはディスプレイユニットなどの入力センサユニットを備えた任意のコンピューティングデバイスでよい。ハンドヘルドデバイスの例には、これに限定されないが、ビデオゲーム機、タブレット、スマートフォン、およびモバイルデバイスがある。いくつかの実施形態では、コンピュータシステム1000は、上述のデバイス100または200を実装するように構成される。図10は、本明細書に記載されるように、他の様々な実施形態によって提供される方法を実行することができる、および/またはホストコンピュータシステム、遠隔キオスク/端末、販売時点情報管理システムデバイス、モバイルデバイス、セットトップボックス、および/またはコンピュータシステムとして機能することができる、コンピュータシステム1000の一実施形態の概略的な説明を提供する。図10は、様々な構成要素の一般的な説明を提供することを意図するにすぎず、それらの構成要素のうちのいずれか、および/またはすべては適宜利用され得る。したがって、図10は、個々のシステム要素がどのように比較的別々に、または比較的より統合された方法で実装され得るかを広範に示している。

【0085】

コンピュータシステム1000は、バス1005を介して電氣的に結合され得る（または、適宜通信することができる）ハードウェア要素を備えるように示されている。ハードウェア要素は、これに限定されないが、1つまたは複数の汎用プロセッサ、および/または1つまたは複数の専用プロセッサ（デジタル信号処理チップ、グラフィックアクセラレーションプロセッサ、および/または同類のものなど）を含む1つまたは複数のプロセッサ1010、これに限定されないが、カメラ、マウス、キーボード、および/または同類のものを組み得る1つまたは複数の入力デバイス1015、ならびに、これに限定されないが、ディスプレイユニット、プリンタ、および/または同類のものを組み得る1つまたは複数の出力デバイス1020を含む得る。

【0086】

コンピュータシステム1000は、これに限定されないが、ローカルおよび/またはネットワークアクセス可能記憶装置を備え得る1つまたは複数の非一時的記憶装置1025をさらに含んでもよく（および/または、それと通信してもよく）、ならびに/あるいは、これに限定されないが、プログラム可能な、フラッシュで更新可能な、および/または同等の、ディスクドライブ、ドライブアレイ、光記憶装置、ランダムアクセスメモリ（「RAM」）および/または読み出し専用メモリ（「ROM」）などのソリッドステート記憶装置を含んでもよい。そのような記憶装置は、これに限定されないが、様々なファイルシステム、データベース構造、および/または同類のものを組み、任意の適切なデータ記憶装置を実装するように構成され得る。

【0087】

コンピュータシステム 1000 は、これに限定されないが、モデム、ネットワークカード（ワイヤレスまたはワイヤード）、赤外線通信デバイス、ワイヤレス通信デバイスおよび/またはチップセット（ブルートゥース（登録商標）デバイス、802.11 デバイス、Wi-Fi（登録商標）デバイス、WiMAX（登録商標）のデバイス、セルラー通信設備など）および/または同類のものを含み得る、通信サブシステム 1030 も含み得る。通信サブシステム 1030 は、ネットワーク（一例を挙げると、以下で説明するネットワークなど）、他のコンピュータシステム、および/または本明細書に記載の他の任意のデバイスとのデータの交換を可能にすることができる。多くの実施形態では、コンピュータシステム 1000 は、上述のように、RAM または ROM デバイスを含み得る非一時的ワーキングメモリ 1035 をさらに備え得る。

10

【0088】

コンピュータシステム 1000 は、オペレーティングシステム 1040、デバイスドライバ、実行可能ライブラリ、および/または 1 つまたは複数のアプリケーションプログラム 1045 などの他のコードを含む、非一時的ワーキングメモリ 1035 内に位置するように示される、ソフトウェア要素も備えてよく、様々な実施形態によって提供される、ならびに/あるいは方法を実装するように、および/または本明細書に記載されるように他の実施形態によって提供されるシステムを構成するように設計され得る、コンピュータプログラムを備えてよい。単に例にすぎないが、上述の方法に関して説明した、たとえば図 4、5、6、9A、および 9B に関して説明した 1 つまたは複数の手順は、コンピュータ（および/またはコンピュータ内のプロセッサ）によって実行可能なコードおよび/または命令として実装されてよく、ある態様では、そのようなコードおよび/または命令は、説明した方法によって 1 つまたは複数の操作を実行するように汎用コンピュータ（または他のデバイス）を構成および/または適合するために使用され得る。

20

【0089】

これらの命令および/またはコードのセットは、上述の非一時的記憶装置 1025 などのコンピュータ可読記憶媒体に格納され得る。場合によっては、記憶媒体は、コンピュータシステム 1000 などのコンピュータシステム内に組み込まれ得る。他の実施形態では、記憶媒体は、コンピュータシステム（たとえば、コンパクトディスクなどのリムーバブル媒体）と分離していてもよく、および/またはインストールパッケージで提供されてもよいので、記憶媒体は、格納された命令および/またはコードで汎用コンピュータをプログラム、構成、および/または適合するために使用され得る。これらの命令は、コンピュータシステム 1000 によって実行可能な実行可能コードの形式をとってもよく、ならびに/あるいは、ソースおよび/またはインストール可能コードの形式をとってもよく、コンピュータシステム 1000 上でコンパイルおよび/またはインストールされると（たとえば、任意の多種多様な一般的に利用可能なコンパイラ、インストールプログラム、圧縮/解凍ユーティリティ等のいずれかを使用して）、実行可能コードの形式をとる。

30

【0090】

実質的な変形は、特定の要件に応じて行われてよい。たとえば、カスタマイズされたハードウェアが使用されてもよく、および/あるいは特定の要素がハードウェア、ソフトウェア（アプレット等のポータブルソフトウェアを含む）、またはその両方に実装されてもよい。さらに、ネットワーク入力/出力デバイスなどの他のコンピューティングデバイスへの接続が採用されてよい。

40

【0091】

いくつかの実施形態は、本開示による方法を実行するために、コンピュータシステム（コンピュータシステム 1000 など）を採用することができる。たとえば、説明された方法の手順のうちのいくつかまたはすべては、非一時的ワーキングメモリ 1035 に含まれる 1 つまたは複数の命令の 1 つまたは複数のシーケンス（オペレーティングシステム 1040、および/またはアプリケーションプログラム 1045 などの他のコードに組み込まれ得る）を実行するプロセッサ 1010 に応答してコンピュータシステム 1000 によって実行され得る。そのような命令は、1 つまたは複数の非一時的記憶装置 1025 などの

50

別のコンピュータ可読媒体から、非一時的ワーキングメモリ 1035 に読み込まれてよい。単に例示にすぎないが、非一時的ワーキングメモリ 1035 に含まれる命令のシーケンスの実行は、プロセッサ 1010 に、本明細書に記載の方法の 1 つまたは複数の手順、たとえば、図 4、5、6、9A、および 9B のいずれかに関して説明された方法の要素のうちの 1 つまたは複数を実行させることができる。

【0092】

本明細書で使用される場合、「機械可読媒体 (machine-readable medium)」および「コンピュータ可読媒体 (computer-readable medium)」という用語は、機械に特定の方法で操作させるデータの提供に關与する任意の媒体を指す。コンピュータシステム 1000 を使用して実装されたある実施形態では、様々なコンピュータ可読媒体は、実行するためにプロセッサ 1010 に命令/コードを提供することに関与してもよく、そのような命令/コードを (たとえば、信号として) 格納および/または搬送するために使用されてもよい。多くの実装形態では、コンピュータ可読媒体は、物理的および/または有形の記憶媒体である。そのような媒体は、これに限定されないが、不揮発性媒体、揮発性媒体、および伝送媒体を含む、多くの形式をとることができる。不揮発性媒体は、たとえば、非一時的記憶装置 1025 などの光および/または磁気ディスクを含む。揮発性媒体は、これに限定されないが、非一時的ワーキングメモリ 1035 などのダイナミックメモリを含む。伝送媒体は、これに限定されないが、同軸ケーブル、バス 1005 を備えるワイヤを含む銅線および光ファイバ、ならびに通信サブシステム 1030 の様々なコンポーネント (および/または通信サブシステム 1030 が他のデバイスとの通信を提供する媒体) を含む。したがって、伝送媒体は、波形 (これに限定されないが、電波および赤外線データ通信中に生成されるような、無線、音響、および/または光波形を含む) の形式をとることもできる。

【0093】

1 つまたは複数の例では、説明された機能は、ハードウェア、ソフトウェア、ファームウェア、またはそれらの任意の組合せに実装され得る。ソフトウェアに実装されると、機能は、1 つまたは複数の命令またはコードとしてコンピュータ可読媒体に格納されてもよく、それを介して伝送されてもよい。コンピュータ可読媒体は、コンピュータデータ記憶媒体を含み得る。データ記憶媒体は、本開示に記載の技法を実装するための命令、コード、および/またはデータ構造を取り出すために、1 つまたは複数のコンピュータ、あるいは 1 つまたは複数のプロセッサによってアクセス可能な任意の利用可能な媒体でよい。本明細書で使用される場合、「データ記憶媒体 (data storage media)」は製品を指し、一時的な伝搬信号を指すものではない。例を挙げると、これに限定されないが、そのようなコンピュータ可読媒体は、RAM、ROM、EEPROM、CD-ROM、または他の光ディスクストレージ、磁気ディスクストレージ、または他の磁気ストレージデバイス、フラッシュメモリ、あるいは所望のプログラムコードを命令またはデータ構造の形式で格納するために使用され得る、およびコンピュータによってアクセス可能な他の任意の媒体を備え得る。本明細書で使用されるディスク (disk) およびディスク (disc) は、コンパクトディスク (CD)、レーザーディスク (登録商標)、光ディスク、デジタル多用途ディスク (DVD)、フロッピーディスク (登録商標)、およびブルーレイディスクを含み、ディスク (disk) は通常データを磁氣的に再生し、ディスク (disc) はデータをレーザーで光学的に再生する。上記の組合せも、コンピュータ可読媒体の範囲に含まれるべきである。

【0094】

コードは、1 つまたは複数のデジタル信号プロセッサ (DSP)、汎用マイクロプロセッサ、特定用途向け集積回路 (ASIC)、フィールドプログラマブル論理アレイ (FPGA)、または他の同等の集積またはディスクリート論理回路などの、1 つまたは複数のプロセッサによって実行され得る。したがって、本明細書で使用される場合、「プロセッサ (processor)」という用語は、上記の構造、または本明細書に記載の技法の実装に適した他の任意の構造のうちのいずれかを指すことができる。さらに、いくつかの

態様では、本明細書に記載の機能は、符号化および復号化するように構成された専用ハードウェアおよび/またはソフトウェアモジュール内に提供されてもよく、結合されたコーデックに組み込まれてもよい。また、本技法は、1つまたは複数の回路あるいは論理要素に完全に実装され得る。

【0095】

本開示の技法は、ワイヤレスハンドセット、集積回路（IC）、またはICのセット（たとえば、チップセット）を含む、広範なデバイスまたは装置に実装され得る。本開示において、様々なコンポーネント、モジュール、またはユニットは、開示された技法を実行するように構成されたデバイスの機能的な側面を強調するために説明されているが、必ずしも異なるハードウェアユニットによる実現を必要としない。むしろ、上述のように、様々なユニットがコーデックハードウェアユニットに組み合わせられてもよく、コンピュータ可読媒体に格納された適切なソフトウェアおよび/またはファームウェアとともに、上述のような1つまたは複数のプロセッサを含む、相互運用可能なハードウェアユニットの集合によって提供されてもよい。

10

【0096】

様々な例を説明してきた。これらおよび他の例は、以下の特許請求の範囲内である。

【符号の説明】

【0097】

- 100 コンピューティングデバイス
- 110 カメラ
- 120 タッチスクリーン
- 200 カメラ
- 210 ジェスチャ
- 220 異なるジェスチャ
- 300 データベース
- 302 例示的なジェスチャ
- 304 例示的なジェスチャ
- 306 例示的なジェスチャ
- 308 例示的なジェスチャ
- 310 例示的なジェスチャ
- 312 例示的なジェスチャ
- 314 例示的なジェスチャ
- 316 例示的なジェスチャ
- 318 例示的なジェスチャ
- 320 例示的なジェスチャ
- 400 ブロック図
- 500 処理フロー
- 600 図
- 700 グラフ
- 800 図表
- 810 モデルディメンションブロック
- 820 ジェスチャ
- 900 流れ図
- 950 流れ図
- 1000 コンピュータシステム
- 1005 バス
- 1010 プロセッサ
- 1015 入力デバイス
- 1020 出力デバイス
- 1025 非一時的記憶装置

20

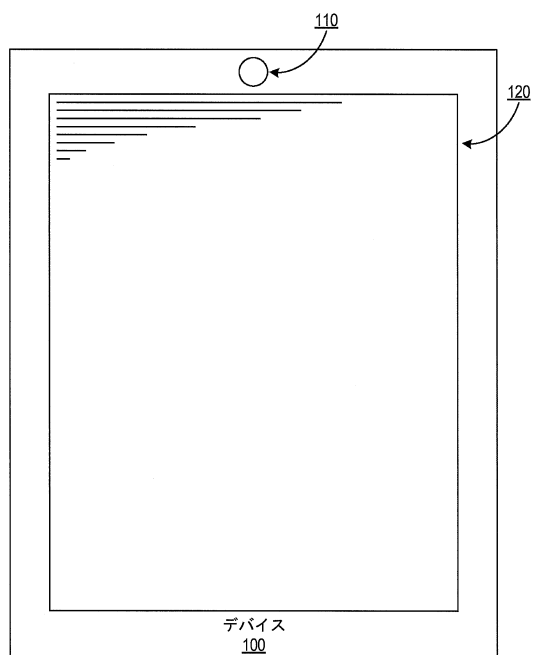
30

40

50

1 0 3 0 通信サブシステム
1 0 3 5 非一時的ワーキングメモリ
1 0 4 0 オペレーティングシステム
1 0 4 5 アプリケーションプログラム

【図 1】



【図 2 A】

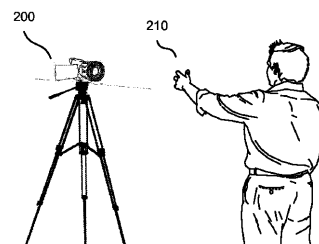


FIG. 2A

【図 2 B】

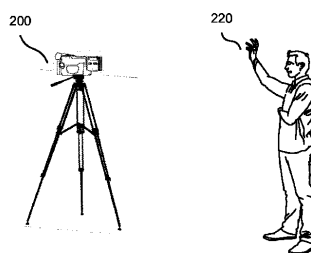


FIG. 2B

【図 3】

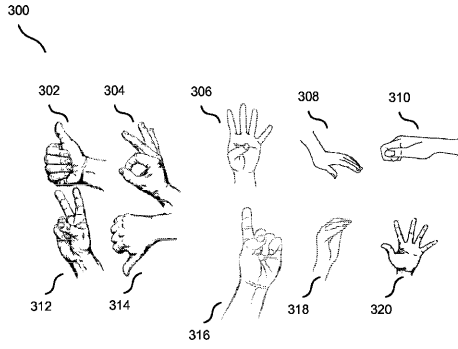
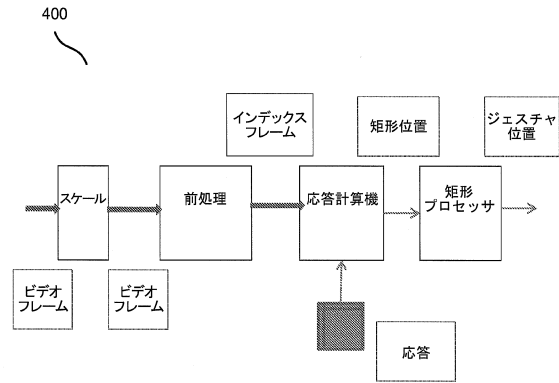
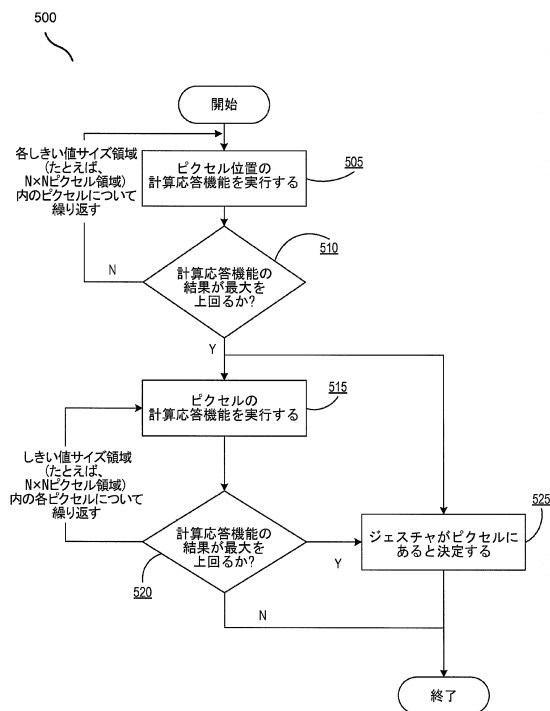


FIG. 3

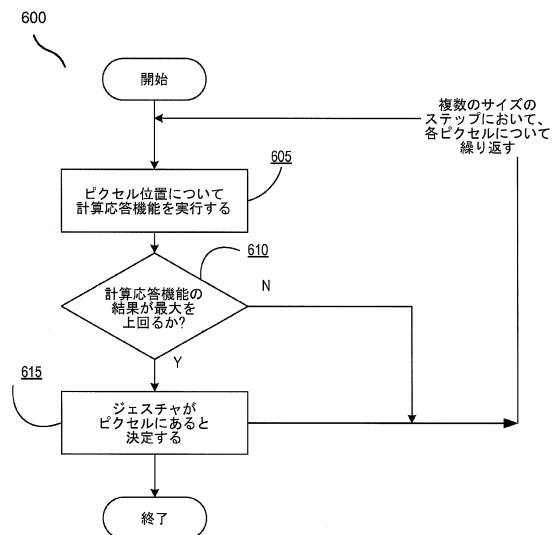
【図 4】



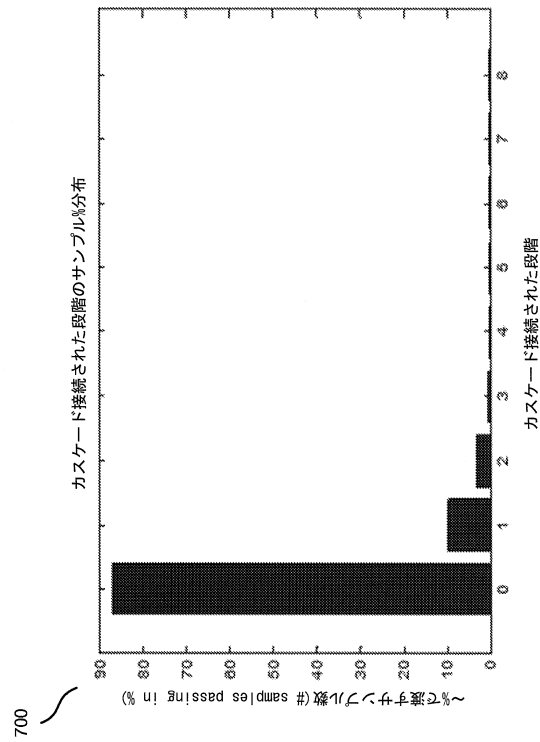
【図 5】



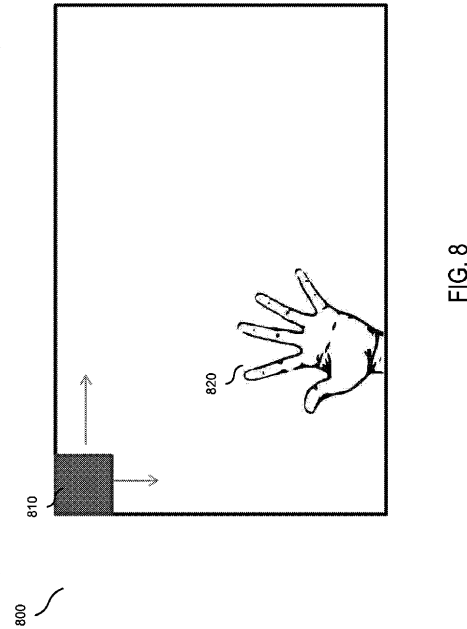
【図 6】



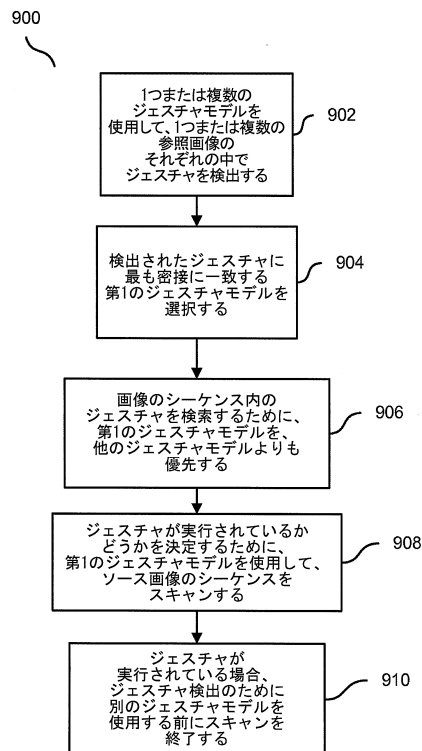
【図 7】



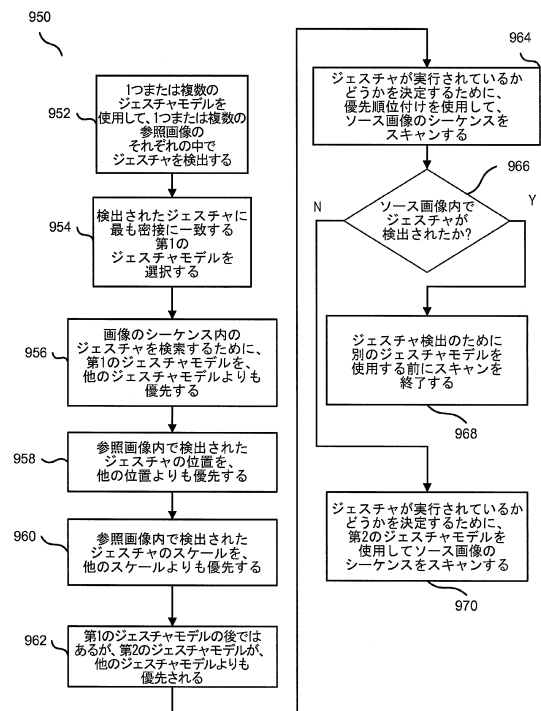
【図 8】



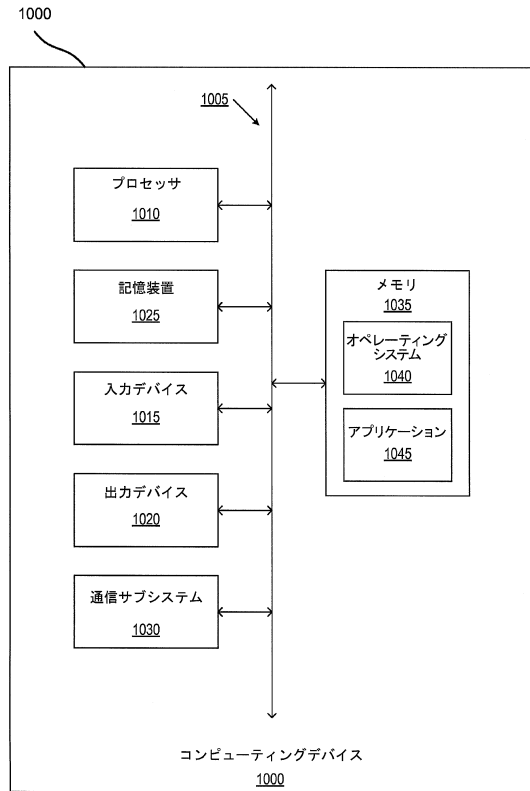
【図 9 A】



【図 9 B】



【図 10】



フロントページの続き

(72)発明者 フランシス・ビー・マクドゥーガル
アメリカ合衆国・カリフォルニア・９２１２１ - １７１４・サン・ディエゴ・モアハウス・ドライ
ヴ・５７７５

審査官 真木 健彦

(56)参考文献 米国特許出願公開第２０１０／０１９４７６２（ＵＳ，Ａ１）
国際公開第２００６／０８６５０８（ＷＯ，Ａ１）
米国特許出願公開第２０１１／０００７０７９（ＵＳ，Ａ１）
特開２００６ - １７２４３９（ＪＰ，Ａ）
特表２００９ - ５３０７２６（ＪＰ，Ａ）
特開２０１２ - ０９８９８８（ＪＰ，Ａ）
特開２０１１ - １９２０９０（ＪＰ，Ａ）
特開２０１２ - ０４８６０２（ＪＰ，Ａ）
特開２０１１ - ２２１６９９（ＪＰ，Ａ）
特開２００５ - １８４７２２（ＪＰ，Ａ）
特開２００５ - ０６２９１０（ＪＰ，Ａ）
浜田 康志，遷移ネットワークに基づく複雑背景下での手指ジェスチャの認識，情報処理学会研
究報告 Vol. 2005 No. 88，日本，社団法人情報処理学会，2005年 9月 6
日，2005-CVIM-150，P.9-16，ISSN 0919-6072

(58)調査した分野(Int.Cl.，ＤＢ名)

G 0 6 T	7 / 0 0	-	7 / 9 0
G 0 6 T	1 / 0 0		
G 0 6 F	3 / 0 1		
G 0 6 F	3 / 0 4 8	-	3 / 0 4 8 9