

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7558243号
(P7558243)

(45)発行日 令和6年9月30日(2024.9.30)

(24)登録日 令和6年9月19日(2024.9.19)

(51)国際特許分類

G 0 6 T 7/00 (2017.01)

F I

G 0 6 T 7/00 3 5 0 C

請求項の数 17 (全43頁)

(21)出願番号	特願2022-502320(P2022-502320)	(73)特許権者	521427427
(86)(22)出願日	令和2年3月12日(2020.3.12)		レチンエイアイ メディカル アーゲー
(65)公表番号	特表2022-525264(P2022-525264 A)		スイス国, 3 0 1 0 ベルン, 3, フライブルクシュトラーセ
(43)公表日	令和4年5月11日(2022.5.11)	(74)代理人	110001416
(86)国際出願番号	PCT/EP2020/056710		弁理士法人信栄事務所
(87)国際公開番号	WO2020/187705	(72)発明者	トロン, プルーン ソランジュ ガランス
(87)国際公開日	令和2年9月24日(2020.9.24)		スイス国, 3 0 1 0 ベルン, 3, フライブルクシュトラーセ
審査請求日	令和5年3月8日(2023.3.8)	(72)発明者	デ ザネット, サンドロ イヴォ セバス
(31)優先権主張番号	19163191.0		チャーノ
(32)優先日	平成31年3月15日(2019.3.15)		スイス国, 3 0 1 0 ベルン, 3, フライブルクシュトラーセ
(33)優先権主張国・地域又は機関	欧州特許庁(EP)	(72)発明者	アバストロブーロス, ステファノス
(31)優先権主張番号	19206339.4		スイス国, 3 0 1 0 ベルン, 3, フライブルクシュトラーセ
(32)優先日	令和1年10月30日(2019.10.30)		最終頁に続く
	最終頁に続く		

(54)【発明の名称】 特徴点検出

(57)【特許請求の範囲】

【請求項1】

特徴点検出のためのモデルを訓練する方法であって、
第1の画像および第2の画像を取得するステップと、
前記モデルを用いて、前記第1の画像に対する第1のスコアマップおよび前記第2の画像に対する第2のスコアマップを生成するステップと、
前記第1のスコアマップに基づいて、前記第1の画像内の第1の複数の関心点を選択するステップと、
前記第2のスコアマップに基づいて、前記第2の画像内の第2の複数の関心点を選択するステップと、
前記第1の複数の関心点のうちの1つの第1の関心点と、前記第2の複数の関心点のうちの1つの第2の関心点とのペアワイズマッチングを行うステップと、
前記第1の画像と前記第2の画像との間のグラウンドトゥルス変換に基づいて、前記ペアワイズマッチの正しさをチェックして、リワードマップを生成するステップと、
前記第1のスコアマップまたは前記第2のスコアマップと、前記リワードマップとを結合するステップと、
前記結合の結果に基づいて前記モデルを更新するステップとを含む方法。

【請求項2】

前記複数の関心点を選択するステップは、前記画像内の任意の点から最近傍関心点までの距離に最大制限を課すことを含む、請求項1に記載の方法。

【請求項 3】

前記ペアワイズマッチングは、前記第 1 の関心点において前記第 1 の画像で検出された特徴と、前記第 2 の関心点において前記第 2 の画像で検出された特徴との間の類似性に基づいて実行される、請求項 1 または請求項 2 に記載の方法。

【請求項 4】

前記マッチングは、前記第 1 の関心点を、複数の第 2 の関心点のうちの 1 つの第 2 の関心点にマッチングすることによって第 1 の方向に実行され、前記第 2 の関心点は、前記第 1 の関心点における特徴に最も類似した特徴を有する、請求項 1 から請求項 3 のいずれか一項に記載の方法。

【請求項 5】

前記マッチングはさらに、前記第 2 の関心点を複数の第 1 の関心点のうちの 1 つの第 1 の関心点にマッチングすることによって第 2 の方向に実行され、前記第 1 の関心点は、前記第 2 の関心点における特徴に最も類似した特徴を有する、請求項 4 に記載の方法。

【請求項 6】

前記リワードマップが、グラウンドトゥルスデータに従って、マッチングに成功した関心点に対するリワードを示し、前記グラウンドトゥルスデータに従って、マッチングに失敗した関心点に対するリワードがないことを示す、請求項 1 から請求項 5 のいずれか一項に記載の方法。

【請求項 7】

前記結合が、前記第 1 のスコアマップまたは前記第 2 のスコアマップ、および前記リワードマップを前記関心点についてのみ結合することを含む、請求項 1 から請求項 6 のいずれか一項に記載の方法。

【請求項 8】

前記結合が、偽陽性マッチをおそらくランダムに選択し、真陽性マッチおよび前記偽陽性マッチの選択分についてのみ、前記第 1 のスコアマップまたは前記第 2 のスコアマップ、および前記リワードマップを結合することによって前記真陽性マッチの数と前記偽陽性マッチの数とを均衡させることを含む、

前記真陽性マッチとは、正しさのチェックに合格した関心点であり、前記偽陽性マッチとは、正しさのチェックに合格しなかった関心点である、請求項 1 から請求項 7 のいずれか一項に記載の方法。

【請求項 9】

前記結合が、前記第 1 のスコアマップまたは前記第 2 のスコアマップと、前記リワードマップとの間の二乗差の合計を計算することを含む、請求項 1 から請求項 8 のいずれか一項に記載の方法。

【請求項 10】

特徴点検出のためのモデルを訓練する装置であって、

制御ユニットと、

メモリを備え、前記メモリは、前記制御ユニットに、

第 1 の画像と第 2 の画像を取得するステップと、

前記モデルを用いて、前記第 1 の画像に対する第 1 のスコアマップおよび前記第 2 の画像に対する第 2 のスコアマップを生成するステップと、

前記第 1 のスコアマップに基づいて、前記第 1 の画像内の第 1 の複数の関心点を選択するステップと、

前記第 2 のスコアマップに基づいて、前記第 2 の画像内の第 2 の複数の関心点を選択するステップと、

前記第 1 の複数の関心点のうちの 1 つの第 1 の関心点と、前記第 2 の複数の関心点のうちの 1 つの第 2 の関心点とのペアワイズマッチングを行うステップと、

前記第 1 の画像と前記第 2 の画像との間のグラウンドトゥルス変換に基づいて、前記ペアワイズマッチの正しさをチェックして、リワードマップを生成するステップと、

前記第 1 のスコアマップまたは前記第 2 のスコアマップと、前記リワードマップとを結合

10

20

30

40

50

するステップと、

前記結合の結果に基づいて、前記モデルを更新するステップと、
を実行させる命令を含む、装置。

【請求項 11】

第1の画像を第2の画像にレジストレーションする方法であって、

第1の画像と第2の画像を取得するステップと、

請求項1から10のいずれかに記載の方法または装置によって訓練されたモデルを用いて、前記第1の画像に対する第1のスコアマップおよび前記第2の画像に対する第2のスコアマップを生成するステップと、

前記第1のスコアマップに基づいて、前記第1の画像内の第1の複数の関心点を選択するステップと、

前記第2のスコアマップに基づいて、前記第2の画像内の第2の複数の関心点を選択するステップと、

前記第1の複数の関心点のうちの1つの第1の関心点と、前記第2の複数の関心点のうちの1つの第2の関心点とのペアワイズマッチングを行うステップと、を含む方法。

【請求項 12】

前記複数の関心点を選択するステップは、前記画像内の任意の点から最近傍関心点までの距離に最大制限を課すことを含む、請求項11に記載の方法。

【請求項 13】

前記ペアワイズマッチングは、前記第1の関心点において前記第1の画像で検出された特徴と、前記第2の関心点において前記第2の画像で検出された特徴との間の類似性に基づいて実行される、請求項11または請求項12に記載の方法。

【請求項 14】

前記マッチングは、前記第1の関心点を、複数の第2の関心点のうちの1つの第2の関心点にマッチングすることによって第1の方向に実行され、前記第2の関心点は、前記第1の関心点における特徴に最も類似した特徴を有する、請求項11から請求項13のいずれか一項に記載の方法。

【請求項 15】

前記マッチングはさらに、前記第2の関心点を、複数の第1の関心点のうちの1つの第1の関心点にマッチングすることによって第2の方向に実行され、前記第1の関心点は、前記第2の関心点における特徴に最も類似した特徴を有する、請求項14に記載の方法。

【請求項 16】

第1の画像を第2の画像にレジストレーションする装置であって、

制御ユニットと、

メモリを備え、前記メモリは、前記制御ユニットに、

第1の画像と第2の画像を取得するステップと、

請求項1から請求項10のいずれか一項に記載の方法または装置によって訓練されたモデルを用いて、前記第1の画像に対する第1のスコアマップおよび前記第2の画像に対する第2のスコアマップを生成するステップと、

前記第1のスコアマップに基づいて、前記第1の画像内の第1の複数の関心点を選択するステップと、

前記第2のスコアマップに基づいて、前記第2の画像内の第2の複数の関心点を選択するステップと、

前記第1の複数の関心点のうちの1つの第1の関心点と、前記第2の複数の関心点のうちの1つの第2の関心点とのペアワイズマッチングを行うステップと、

を実行させる命令を含む、装置。

【請求項 17】

第1の画像を第2の画像にレジストレーションするコンピュータプログラムであって、前記コンピュータプログラムは、制御ユニットによって実行されたとき、前記制御ユニットに、

10

20

30

40

50

第 1 の画像と第 2 の画像を取得するステップと、

請求項 1 から請求項 10 のいずれか一項に記載の方法または装置によって訓練されたモデルを用いて、前記第 1 の画像に対する第 1 のスコアマップおよび前記第 2 の画像に対する第 2 のスコアマップを生成するステップと、

前記第 1 のスコアマップに基づいて、前記第 1 の画像内の第 1 の複数の関心点を選択するステップと、

前記第 2 のスコアマップに基づいて、前記第 2 の画像内の第 2 の複数の関心点を選択するステップと、

前記第 1 の複数の関心点のうちの 1 つの第 1 の関心点と、前記第 2 の複数の関心点のうちの 1 つの第 2 の関心点とのペアワイズマッチングを行うステップと、

を実行させる命令を含む、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、特徴検出に関するものである。本発明は、より詳細には、画像レジストレーションのための特徴検出に関するものである。本発明はさらに、検出された特徴点を用いた画像レジストレーションに関するものである。

【背景技術】

【0002】

2 つ以上の画像を同じ大域的空間参照に整列させるプロセスである画像レジストレーションは、コンピュータビジョン、パターン認識、および医用画像解析の分野において重要な要素である。医療分野でのイメージングは困難な作業である。そのため、通常の写真撮影などに比べて、画質が悪くなることがある。ノイズやブラーなどの画像アーティファクトの存在や、画像化された組織の性質と相まって、自然画像に最適化された古典的な先行技術の特徴検出器では対応しきれない。さらに、新しい特徴検出器をドメインごとに手動で作成することは、時間のかかる作業であり、成功の保証はない。

【0003】

既存のレジストレーションアルゴリズムは、領域ベースのアプローチと特徴ベースのアプローチに分類され得る。領域ベースのアプローチでは、一般的に、相互相関 [15]、相互情報量 [33、4]、位相相関 [24] などの類似性指標を用いて、画像ペアの強度パターンを比較し、変換を推定する。しかし、照明が変化した場合やオーバーラップする領域が小さい場合には、領域ベースのアプローチは困難または不可能になる。逆に、特徴ベースのアプローチでは、画像ペアの対応点を抽出し、検出された特徴点間の距離を最小化する変換を探索する。特徴ベースの手法は、領域ベースのレジストレーション手法に比べて、強度、スケール、回転などの変化に強く、そのため、医用画像のレジストレーションなどの問題に適していると考えられる。一般的に、2 つの画像の特徴抽出とマッチングは、注目点の検出、各特徴の記述、対応する特徴のマッチング、マッチングを用いた画像間の変換の推定という 4 つのステップを含む。理解されるように、検出ステップはそれ以降のすべてのステップに影響を与えるため、レジストレーションを成功させるためには非常に重要である。高い画像のカバレッジと、コントラストの低い画像での安定したキーポイントが必要となる。

【0004】

文献では、局所的な関心点検出器が徹底的に評価されている。SIFT [29] は、コンピュータビジョンにおいて、おそらく最もよく知られた検出器／記述子である。SIFT は、異なるスケールでコーナーやプロブを計算してスケール不変性を付加し、局所勾配を用いて記述子を抽出する。Root SIFT [9] は、SIFTと比較して結果が向上することが示された。高速堅牢特徴量 (Speeded-Up Robust Features、SURF) [12] は、Haar フィルタと積分画像を用いたより高速の代替手法であり、それに対し、KAZE [7] は、非線形スケール空間を利用して、より正確なキーポイント検出を実現している。

10

20

30

40

50

【0005】

眼底画像の分野では、血管樹と分岐点解析に頼る手法が広く用いられている[28、21]。しかし、血管樹を正確にセグメンテーションすることは難しく、また、血管が少ない画像ではレジストレーションに失敗することが多い。代替的レジストレーション技術は、再現可能な局所的特徴のマッチングに基づいている。Chenらは、低品質のマルチモーダル網膜画像からHarrisコーナーを検出し[22]、部分的強度不変特徴(Harris-PIIFD)記述子を割り当てた[14]。重複領域が30%を超えるような低品質の画像に対しても良好な結果を達成したが、再現性が低いという特徴がある。Wangら[41]は、SURF特徴量を用いて再現性を高め、大量の外れ値を排除するポイントマッチングの新技术を導入したが、重複領域が50%未満になると成功率が大きく低下する。

10

【0006】

また、Cattinら[13]は、SURF法を用いて、血管がはっきりしない場合でも、網膜画像のモザイクを効率的に作成できることを示した。しかし、この手法は、自己相似性の高い画像の場合にしか成功しないようである。Dーサドル検出器／記述子[34]は、眼底画像レジストレーション(FIRE) データセット において、レジストレーション成功率の点で従来の手法を上回ることが示されており[23]、低品質な領域でも関心点を検出することができる。

【0007】

最近では、深層学習の登場により、CNNアーキテクチャに基づく学習済み検出器が、最先端のコンピュータビジョン検出器を凌駕することが示された[19、17、44、32、10]。学習不変特徴変換(LIFT)[32]は、パッチを用いて完全に微分可能な深層CNNを学習し、古典的な運動からの構造復元(Structure from Motion, SfM)システムからの監視に基づいて、関心点の検出、方向推定、記述子の計算を行う。SuperPoint[17]は、関心点検出器と記述子を訓練するための自己教師付きフレームワークを導入した。これは、LIFT、SIFT、有向Fastと回転Brief(Oriented Fast and Rotated Brief, ORB)と比較して、HPatches[11]上でのホモグラフィ推定結果は最先端のものとなっている。しかし、訓練手順は複雑であり、自己教師型であるため、ネットワークはコーナー上の点しか見つけることができないという問題がある。局所特徴ネットワーク(LF-NET)[32]は、我々の手法に最も近い。Onoらは、キーポイント検出器および記述子を、一方の枝が微分可能であって他方の非微分可能な枝の出力にフィードする2枝の設定で、エンド・ツー・エンドで訓練した。彼らは、検出器を、画像ペア間の再現性向けに最適化した。

20

30

【0008】

Truongら[39]は、Root-SIFT、SURF、KAZE、ORB[36]、二元ロバスト不変スケーラブルなキーポイント(Binary Robust Invariant Scalable Keypoints, BRISK)[27]、高速網膜キーポイント(FREAK)[6]、LIFT、SuperPoint[17]、LF-NET[32]を用いて、網膜眼底画像の画像マッチングとレジストレーション品質の両面から評価を行った。その結果、マッチング性能ではSuperPointが他の製品よりも優れているが、レジストレーション品質ではLIFTが最も高い結果を示し、KAZEとSIFTが僅差で続くことがわかった。彼らは、これらの検出器の問題点として、互いに密集した位置にある特徴点を検出し、同様の記述子で特徴付けられる可能性があることを強調した。これは、偽のマッチにつながる可能性があり、したがって、不正確なまたは失敗したレジストレーションにつながる。

40

【発明の概要】

【発明が解決しようとする課題】

【0009】

本発明の一態様は、上記で概説した問題の少なくとも1つを解決すること、または本明

50

細書で説明した利点の少なくとも1つを提供することである。

【課題を解決するための手段】

【0010】

本発明の第1の態様によれば、特徴点検出のためのモデルを訓練する方法は、
第1の画像および第2の画像を取得するステップと、
前記モデルを用いて、前記第1の画像に対する第1のスコアマップおよび前記第2の画像に対する第2のスコアマップを生成するステップと、
前記第1のスコアマップに基づいて、前記第1の画像内の第1の複数の関心点を選択するステップと、
前記第2のスコアマップに基づいて、前記第2の画像内の第2の複数の関心点を選択するステップと、
前記第1の複数の関心点のうちの1つの第1の関心点と、前記第2の複数の関心点のうちの1つの第2の関心点とのペアワイズマッチングを行うステップと、
前記第1の画像と前記第2の画像との間のグランドトゥールース変換に基づいて、前記ペアワイズマッチの正しさをチェックして、リワードマップを生成するステップと、
前記スコアマップと前記リワードマップを結合または比較するステップと、
前記結合または比較の結果に基づいて前記モデルを更新するステップとを含む。

10

【0011】

第1の画像と第2の画像との間のグランドトゥールース変換に基づくスコアマップとリワードマップの結合または比較の結果に基づいてモデルを更新することにより、2つの画像をレジストレーションするという目的にさらに的を絞った、高度に直接的なリワードが得られる。したがって、モデルの訓練が改善される。本明細書に開示される任意の実施形態において、モデルは、例えば、学習された関数、人工ニューラルネットワーク、または分類器であってもよい。

20

【0012】

前記複数の関心点を選択するステップは、前記画像内の任意の点から最近傍関心点までの距離に最大制限を課すことを含み得る。これにより、ほとんどの関心点が画像の小さい領域に集まってしまう状況を回避できる。この機能は、全体的に優れた画像レジストレーションをもたらすマッチングポイントを得ることを助長する。

【0013】

前記ペアワイズマッチングは、前記第1の関心点において前記第1の画像で検出された特徴と、前記第2の関心点において前記第2の画像で検出された特徴との間の類似性に基づいて実行されてもよい。これにより、他方の画像において類似した特徴記述子を有する関心点に、関心点をマッチングさせることができる。

30

【0014】

前記マッチングはさらに、前記第1の関心点を、複数の第2の関心点のうちの1つの第2の関心点にマッチングすることによって第1の方向に実行されてもよく、前記第2の関心点は、前記第1の関心点における特徴に最も類似した特徴を有する。これにより、マッチする関心点の適切な選択をもたらす。

【0015】

前記マッチングはさらに、前記第2の関心点を複数の第1の関心点のうちの1つの第1の関心点にマッチングすることによって第2の方向に実行されてもよく、前記第1の関心点は、前記第2の関心点における特徴に最も類似した特徴を有する。これは、マッチングする関心点の選択を改善するのに役立つ。

40

【0016】

前記リワードマップは、グランドトゥールースデータに従って、マッチングに成功した関心点に対するリワードを示してもよく、前記グランドトゥールースデータに従って、マッチングに失敗した関心点に対するリワードがないことを示してもよい。これは、よりの的を絞ったリワードマップを提供することで、訓練手順を改善するのに役立つ。

【0017】

50

前記結合または比較は、前記スコアマップおよびリワードマップを前記関心点についてのみ結合または比較することを含む。関心点ではない他の点は、数が多く、訓練手順を助けるための十分な情報を追加しない可能性がある。

【0018】

前記結合または比較が、偽陽性マッチを（おそらくランダムに）選択し、真陽性マッチおよび前記偽陽性マッチの選択分についてのみ前記スコアマップおよびリワードマップを結合または比較することによって前記真陽性マッチの数と前記偽陽性マッチの数を均衡させることを含んでもよく、前記真陽性マッチとは、正しさのチェックに合格した関心点であり、前記偽陽性マッチとは、正しさのチェックに合格しなかった関心点である。これは、「マッチしない」に対する訓練へのバイアスをさらに減らすのに役立つ。

10

【0019】

前記結合または比較は、前記スコアマップと前記リワードマップの間の二乗差の合計を計算することを含んでもよい。これは、訓練手順の適切な成分を構成してもよい。

【0020】

本発明の別の態様によれば、特徴点検出のためのモデルを訓練する装置は、制御ユニットと、メモリを備え、前記メモリは、前記制御ユニットに、第1の画像と第2の画像を取得するステップと、前記モデルを用いて、前記第1の画像に対する第1のスコアマップおよび前記第2の画像に対する第2のスコアマップを生成するステップと、前記第1のスコアマップに基づいて、前記第1の画像内の第1の複数の関心点を選択するステップと、前記第2のスコアマップに基づいて、前記第2の画像内の第2の複数の関心点を選択するステップと、前記第1の複数の関心点のうちの1つの第1の関心点と、前記第2の複数の関心点のうちの1つの第2の関心点とのペアワイズマッチングを行うステップと、前記第1の画像と前記第2の画像との間のグラントゥールス変換に基づいて、前記ペアワイズマッチの正しさをチェックして、リワードマップを生成するステップと、前記スコアマップと前記リワードマップを結合または比較するステップと、前記結合または比較の結果に基づいて、前記モデルを更新するステップと、を実行させる命令を含む。

20

【0021】

本発明の別の態様によれば、第1の画像を第2の画像にレジストレーションする方法が提供され、前記方法は、第1の画像と第2の画像を取得するステップと、請求項1から10のいずれかに記載の方法または装置によって生成されたモデルを用いて、前記第1の画像に対する第1のスコアマップおよび前記第2の画像に対する第2のスコアマップを生成するステップと、前記第1のスコアマップに基づいて、前記第1の画像内の第1の複数の関心点を選択するステップと、前記第2のスコアマップに基づいて、前記第2の画像内の第2の複数の関心点を選択するステップと、前記第1の複数の関心点のうちの1つの第1の関心点と、前記第2の複数の関心点のうちの1つの第2の関心点とのペアワイズマッチングを行うステップと、を含む。

30

【0022】

前記複数の関心点を選択するステップは、前記画像内の任意の点から最近傍関心点までの距離に最大制限を課すことを含んでもよい。

【0023】

前記ペアワイズマッチングは、前記第1の関心点において前記第1の画像で検出された特徴と、前記第2の関心点において前記第2の画像で検出された特徴との間の類似性に基づいて実行されてもよい。

40

【0024】

前記マッチングは、前記第1の関心点を、複数の第2の関心点のうちの1つの第2の関心点にマッチングすることによって第1の方向に実行されてもよく、前記第2の関心点は、前記第1の関心点における特徴に最も類似した特徴を有する。

【0025】

前記マッチングはさらに、前記第2の関心点を、複数の第1の関心点のうちの1つの第1の関心点にマッチングすることによって第2の方向に実行されてもよく、前記第1の関

50

心点は、前記第2の関心点における特徴に最も類似した特徴を有する。

【0026】

前記複数の関心点を選択するステップは、前記画像内の任意の点から最近傍関心点までの距離に最大制限を課すことを含んでもよい。

【0027】

前記ペアワイズマッチングは、前記第1の関心点において前記第1の画像で検出された特徴と、前記第2の関心点において前記第2の画像で検出された特徴との間の類似性に基づいて実行されてもよい。

【0028】

前記マッチングは、前記第1の関心点を、複数の第2の関心点のうちの1つの第2の関心点にマッチングすることによって第1の方向に実行されてもよく、前記第2の関心点は、前記第1の関心点における特徴に最も類似した特徴を有する。

【0029】

前記マッチングはさらに、前記第2の関心点を、複数の第1の関心点のうちの1つの第1の関心点にマッチングすることによって第2の方向に実行されてもよく、前記第1の関心点は、前記第2の関心点における特徴に最も類似した特徴を有する。

【0030】

本発明の別の態様によれば、第1の画像を第2の画像にレジストレーションする装置が提供され、装置は、制御ユニット、例えば、少なくとも1つのコンピュータプロセッサと、メモリを備え、前記メモリは、前記制御ユニットに、第1の画像と第2の画像を取得するステップと、記載された方法または装置によって生成されたモデルを用いて、前記第1の画像に対する第1のスコアマップおよび前記第2の画像に対する第2のスコアマップを生成するステップと、前記第1のスコアマップに基づいて、前記第1の画像内の第1の複数の関心点を選択するステップと、前記第2のスコアマップに基づいて、前記第2の画像内の第2の複数の関心点を選択するステップと、前記第1の複数の関心点のうちの1つの第1の関心点と、前記第2の複数の関心点のうちの1つの第2の関心点とのペアワイズマッチングを行うステップと、を実行させる命令を含む。

【0031】

本発明の別の態様によれば、記載された方法または装置によって生成されたモデルが提供される。

【0032】

本発明の一態様は、キーポイント検出のための半教師付き学習方法である。検出器は、多くの場合、再現性（L F - N E Tなど）に対して最適化されており、画像ペア間の関連するマッチの品質に対しては最適化されていない。本発明の一態様は、強化学習を使用して、再現性のある安定した関心点を密にカバーして抽出し、特定のドメイン上で正しいマッチングを最大化するように特別に設計された訓練手順である。このような特定のドメインの例は、困難な網膜細隙灯画像である。

【0033】

当業者であれば、上述の特徴は、有用と思われる任意の方法で組み合わせ得ることを理解できるだろう。さらに、システムおよび装置に関して記載された修正および変形は、同様に、方法およびコンピュータプログラム製品に適用されてもよく、また、方法に関して記載された修正および変形は、同様に、システムおよび装置に適用されてもよい。

【0034】

以下、本発明の態様を、図面を参照しながら、例を用いて説明する。図面は図式化されており、縮尺通りに描かれていない場合がある。

【図面の簡単な説明】

【0035】

【図1】 実用例でポイントがどのようにマッチングされ得るかを示す図である。

【図2A】 画像ペアの訓練のステップの一例を示す図である。

【図2B】 損失計算の一例を示す図である。

【図 2 C】 U n e t - 4 の一例の模式図である。

【図 3】 細隙灯データセットの画像の例を示す図である。

【図 4 A】 非前処理済みデータを用いて、206 ペアの細隙灯データセットについて評価された検出器／記述子の性能指標の概要を示す図である。

【図 4 B】 前処理されたデータを用いて、206 ペアの細隙灯データセットについて評価された検出器／記述子の性能指標の概要を示す図である。

【図 5】 別の実用例でポイントがどのようにマッチングされ得るかを示す図である。

【図 6】 連続した画像をレジストレーションして作成したモザイク図である。

【図 7】 特徴点検出のためのモデルを訓練するシステムのブロック図である。

【図 8】 特徴点検出用モデルを学習する方法のフローチャートである。

10

【図 9】 第 1 の画像を第 2 の画像にレジストレーションするシステムのブロック図である。

【図 10】 第 1 の画像を第 2 の画像にレジストレーションする方法のフローチャートである。

【発明を実施するための形態】

【0036】

特定の例示的な実施形態について、添付の図面および紙面を参照しながら、より詳細に説明する。

【0037】

詳細な構造や要素など、本明細書に開示される事項は、例示的な実施形態の包括的な理解を助けるために提供される。したがって、例示的な実施形態は、これらの具体的に定義された事項がなくても実行され得る。また、よく知られている操作や構造は、不必要な詳細によって説明が不明瞭になり得るため、詳細には記載されていない。

20

【0038】

本明細書で開示された技術は、任意のアプリケーション領域の画像レジストレーションに適用され得る。

【0039】

この問題に対する、既知の完全教師付き機械学習ソリューションは、2つの異なる視点からの画像の位置を関連付ける人手による注釈付きのグラウンドトゥールスを必要とする。自然画像では、グラウンドトゥールスは静的な設定で作成できるが、医療データは非常に動的であり、患者が関与している。このため、グラウンドトゥールスの取得は非常に困難であり、実現不可能となっている。

30

【0040】

多くの特徴検出器やキーポイント検出器で検出された特徴の分布は、エッジやコーナーに集中している。医療分野では通常、広い領域には特徴がないため、特徴点が凝集し、その結果、不正確な変換につながるマッチングが発生する。

【0041】

例として以下のような側面が挙げられる。

1. 従来技術では、再現性などの間接的な指標を用いていたが、本技術ではターゲット領域での最終的なマッチング成功率を用いて訓練を行う。

2. アルゴリズムは、合成増強のみで訓練され、グラウンドトゥールスデータの問題を解決する。これにより、ターゲットデータのみで検出器を訓練できる。

40

3. 特徴点が画像全体に均等に分布している。

【0042】

次のような利点が、部分的または完全に実現され得る。

1. 本発明の特定の実施形態を使用することにより、特徴点検出器を学習し、特定の、他の方法では実行不可能なイメージングドメイン向けに最適化できる。さらに、優れた特徴記述子が再利用され得る。

2. グラウンドトゥールスが必要なく、ターゲットドメインからのサンプルのみが必要である。

3. より良いマッチング率を得るために前処理や特徴の手動による加工は必要ない。

50

4. 特徴点（およびマッチング）が均等に分布しているため、2つの画像間の推定変換の精度が大幅に向上する。

【0043】

次のようなさらなる利点も部分的または完全に実現され得る。

1. 対象領域からより多くのデータが得られた場合、それを使用して検出器をさらに改善できる。

2. 検出器を記述子アルゴリズムに最適に適合させ得る。より優れた記述子が見つかった場合、実装コストやデータを追加することなく、新しい検出器が訓練され得る。

【0044】

本発明の一態様は、画像中の特徴点を検出するモデルを訓練するための訓練手順である。この訓練手順は、あらゆる種類の画像に適用され得る。これは、例えば、医療分野、衛星画像や屋外の画像に対しても可能である。また、画像は、携帯電話や携帯型カメラ、超音波診断装置、眼科用細隙灯撮像装置などによって作成されてもよい。

10

【0045】

例えば、画像は2D画像であっても3D画像であってもよい。2D画像の場合、2D（または1D）の特徴が検出されてもよく、点の座標は2次元であってもよい。3D画像の場合、1D、2Dまたは3Dの特徴が検出されてもよく、点の座標は3次元であってもよい。

【0046】

例えば、画像は1D画像であってもよい。1D画像の場合、1Dの特徴が検出され、点の座標は1次元であってもよい。一般に、画像は、任意の正の整数Nに対してN次元であり得る。

20

【0047】

例えば、画像は、写真やX線画像などである。

【0048】

例えば、モデルによって検出された特徴点は、画像のレジストレーションに使用されてもよい。

【0049】

本発明の一態様は、特定の特徴点記述子に対する最適な特徴点検出を求める方法または装置である。その意味で、このアルゴリズムは、ターゲット記述子アルゴリズムのマッチング能力を最適化する画像内の特徴点を見つける。このターゲット記述子アルゴリズムは、例えば、添付論文に記載されているように、root SIFTであり得るが、ORB、BRISK、BRISFなどの他の記述子であってもよい。

30

【0050】

以下の説明では、関心点という用語とキーポイントという用語は互換的に使用される。

【0051】

モデルの訓練には、次のようなステップが使用され得る。

【0052】

1. グラウンドトゥールズホモグラフィ $H_I = H_I$, I に関連する一対の画像 $I \in \mathbb{R}^{H \times W}$, $I' \in \mathbb{R}^{H' \times W}$ が与えられた場合、モデルは各画像に対する以下のスコアマップを提供してもよい。

40

$$S = f_{\theta}(I), S' = f_{\theta}(I').$$

このステップでは、2つのホモグラフィ変換 H_I および $H_{I'}$ をそれぞれ使用して、原画像から2つの画像 I および I' を生成してもよい。すると、それら2つの画像間のホモグラフィは、 $H = H_I^* H_{I'}$ となる。例えば、ホモグラフィ変換 H_I および $H_{I'}$ は、ランダムジェネレータを用いてランダムに生成されてもよい。2つのキーポイント確率マップ S と S' は、モデルを使用して以下のように生成される。

$$S' = f_{\theta}(I), S' = f_{\theta}(I')$$

【0053】

2. ウィンドウサイズ w の標準的な微分不可能ノンマキシマムサプレッション (NMS

50

）を用いて、両方のスコアマップ上で関心点の位置を抽出してもよい。このステップでは、関心点の位置を、ウィンドウサイズ w のノンマキシマムサプレッションを用いて両スコアマップ S 、 S' 上で抽出してもよい。これは、すべてのマス目 $w \times w$ において、最大値のみを局所的に保持し、他のすべての値を0に設定することを意味する。これにより、画像全体ではっきりとした疎な点が得られる。つまり、キーポイント確率マップ S 、 S' の各マス目 $w \times w$ に最大値のキーポイントのみが保持されるため、キーポイントの数が減ることになる。

【0054】

ウィンドウサイズ w は試行錯誤により選択してもよい。この幅 w は、予め与えられていても（固定のアルゴリズムパラメータ）、入手した画像 I に基づいて動的に決定されてもよい。

10

【0055】

これは、この手順における任意のステップであることを理解されたい。さらに、標準的な微分不可能NMSの代わりに、代替アルゴリズムを使用してもよい。

【0056】

3. 検出された各キーポイントに対して、128のroot-SIFT特徴記述子が計算されてもよい。このステップでは、例えば、root-SIFT記述子アルゴリズムを使用して、ステップ2で見つかった各キーポイントに特徴記述子ベクトルを割り当てる。特徴記述子ベクトルは、例えば、128の長さを有していてもよい。他の長さを代わりに使用してもよい。本明細書の他の箇所で述べたように、別のタイプの特徴記述子を代替的に使用してもよい。

20

【0057】

例えば、特徴記述子は、画像 I または I' に存在する画像情報に基づいて計算される。例えば、SIFT特徴記述子は、キーポイント周辺の画像 I または I' における画像勾配を記述する。

【0058】

4. 画像 I からのキーポイントは、例えば[1]のようなブルートフォースマッチャーを用いて、画像 I' のキーポイントとマッチさせてもよいし、逆に、画像 I' のキーポイントを画像 I からのキーポイントとマッチさせてもよい。例えば、両方向で発見されたマッチのみが保持される。このステップでは、画像 I からのキーポイントが、他方の画像 I' の、ステップ3)で計算された最も近い記述子を用いて、画像 I' のキーポイントとマッチングされてもよい。このため、画像 I のキーポイントに対して計算された記述子出力を、画像 I' のキーポイントに対して計算された記述子出力と比較してもよい。ここで、「最も近い記述子」とは、事前に定義された類似性尺度に従って最も類似した記述子出力を意味する。特定の実装では、2番目に近いポイントが非常に類似した記述子を有する場合、マッチングは破棄される。

30

【0059】

マッチングは、両方向（画像 I から画像 I' へ、および画像 I' から画像 I へ）に行い得る。両方向で発見されたマッチのみが保持される。代替の実装では、一方の方向のみで見つかったマッチの一部またはすべてが、マッチのセットに保持されてもよい。

40

【0060】

任意の適切な種類のマッチャーが使用され得る。このステップは、ブルートフォースマッチャーに限定されない。

【0061】

特定の実装では、ステップ4のマッチングは、グラウンドトゥルスホモグラフィ H （または H_I または $H_{I'}$ ）を考慮することなく、画像 I および I' の両方における記述子の出力に基づいて実行される。

【0062】

5. マッチは、グラウンドトゥルスホモグラフィ H に従ってチェックされる。画像 I の対応するキーポイント x が、 H を適用した後の I^0 のポイント x^0 の近傍に入る場合

50

、マッチは真陽性と定義される。これは次のように定式化され得る。

【数 1】

$$\|H * x - x'\| \leq \varepsilon$$

式中、 ε は、例えば、3 画素として選択され得る。T を真のマッチングの集合とする。

【0063】

このステップでは、ステップ 4) で見つかったすべてのマッチングポイントについて、画像 I のマッチングポイントがホモグラフィ H を用いて画像 I' の点に変換される。例えば、画像 I のマッチングポイントの座標は、ホモグラフィ H を用いて画像 I' の点の対応する座標に変換される。特定の実施形態では、ステップ 4) で見つかったマッチングポイントのサブセットのみが変換される。

10

【0064】

画像 I' 内のグラウンドトゥルス対応点を得るためにホモグラフィ H を使用して変換した後の画像 I 内のキーポイント x が、ステップ 4 で見つかったマッチングポイント x' に十分に近い場合、マッチングは真陽性と定義されてよい。

【0065】

例えば、「十分に近い」とは、キーポイント x のグラウンドトゥルス対応点が、マッチングポイント x' から特定の閾値イプシロンよりも小さい距離を有することを意味すると定義され得る。この距離は、例えば、ユークリッド距離であってもよい。特定の実装では、イプシロンは 3 画素であってもよい。しかし、これは限定ではない。

20

【0066】

モデル f_θ は、ニューラルネットワーク、例えば、畳み込みニューラルネットワークであってもよい。しかしながら、これは限定ではない。代替的な実施形態では、統計モデルなど、他のタイプのモデルが使用され得る。

【0067】

本発明の一態様によれば、ステップ 2、3、4、および／または 5 は、モデルを訓練するために使用されてもよい。この訓練は、モデルまたはニューラルネットワークのバックプロパゲーションを含んでもよい。

30

【0068】

例えば、訓練手順によって最適化される関数（すなわち、コスト関数または損失）は、ステップ 5 で見つかったマッチに基づいてもよい。

【0069】

したがって、キーポイントのマッチング能力が最適化され得る。

【0070】

損失関数は、以下のように提供されてもよい。

【数 2】

$$L_{\text{simple}} = (f_\theta(I) - R)^2$$

40

ここで、リワード行列 R は、以下のように定義され得る。

【数 3】

$$R_{x,y} \in \mathbb{R}^{H \times W} = \begin{cases} 1, & \text{for } (x,y) \in T \\ 0, & \text{otherwise} \end{cases}$$

【0071】

しかし、上記文献の L_{simple} の定式化の欠点は、場合によっては、陽性リワードを与えられたポイントと陰性リワードを与えられたポイントとの間に比較的大きなクラス

50

不均衡が生じ、特に訓練の最初の段階では、後者が圧倒的に優勢になることであると考えられる。ほとんどがゼロの値であるリワード R が与えられた場合、モデル f_{θ} は、レジストレーションのための画像特徴に対応するキーポイントの望ましい表示ではなく、ゼロの出力に収束する可能性がある。

【0072】

好ましくは、この不均衡を解消するために、サンプルマイニングを使用する。つまり、すべての n 個の真陽性点を選択し、すべての偽陽性ではなく、偽陽性のセットから追加の n 個の点をランダムにサンプリングする。 $2n$ 個の真陽性特徴点と、マイニングされた偽陽性キーポイントを通してのみ、バックプロパゲートを行い得る。

【0073】

代替的に、バックプロパゲーションに用いる真陽性点と偽陽性点の数は同一でなくてもよい。真陽性点の数に対して、偽陽性点の数がある程度減っていれば十分である。

【0074】

このようにして、真陽性キーポイントと偽陽性キーポイントの数が（ほぼ）同じになる。これで、真陽性サンプル数と偽陽性サンプル数における不均衡の回避を助長できる。

【0075】

偽陽性よりも真陽性の方が多い場合には、見つかったすべてのマッチに対して勾配をバックプロパゲーションしてもよい。

【0076】

このマイニングは、数学的には、真陽性キーポイントの位置と、マイニングされた（偽陽性）特徴点の（ランダムな）サブセットの位置で 1 に等しく、それ以外では 0 に等しいマスク M として定式化され得る。

【0077】

したがって、損失は例えば以下のように定式化され得る。

【数 4】

$$L = \sum (f_{\theta}(I) - R)^2 \cdot M$$

【0078】

好ましくは、損失は、マスク M による関与する特徴点のそれぞれの損失の平均値として定式化され得る。

【数 5】

$$L = \frac{\sum (f_{\theta}(I) - R)^2 \cdot M}{\sum M}$$

式中、ドットは要素ごとの乗算を意味し、上付きの 2 は要素ごとの二乗を意味し、マイナスは要素ごとの減算を意味し、総和は行列のすべての要素に対して行われる。モデル f_{θ} の出力は、その点がキーポイントである確率または可能性を示し得る。

【0079】

好ましくは、損失関数は、特徴記述子に従ってマッチすると判断された点のみに基づいて計算される。特定のアプリケーションでは、真陰性に対して訓練を行わないことで、より良い結果および／またはより安定した結果が達成され得る。

【0080】

ステップ 2 は、画像上にキーポイントをより規則的に分散させるのに役立ち得る。例えば、キーポイント確率マップ上に幅 w 、高さ h の固定された非最大サプレッションウィンドウを設けることで、2 つのキーポイント間の最大距離が x 方向に $2 * w$ 、 y 方向に $2 * h$ となることを確実にする。例えば、ウィンドウの幅 w は、ウィンドウの高さ h と等しく

10

20

30

40

50

てもよい。または、ウィンドウの大きさを画像に依存させ、例えば画像の情報量に依存させることも可能である。

【0081】

ステップ3で生成された記述子は、それぞれの画像のキーポイントに存在する特徴の記述子の間の類似性に基づいて、画像Iのキーポイントを画像I'のキーポイントにマッチングするために使用されてもよい。

【0082】

異なるタイプの記述子を使用してもよい。例えば、SIFTは（検出器と記述子の両方のコンポーネントを有する）適切な記述子であってよい。

【0083】

一般的に、マッチングは3つのステップで行われてもよい。

1) 2つの画像から関心のある特徴点を検出する。

2) 検出された各特徴点について、一意の記述（特徴ベクトル）を生成する。

3) 特徴ベクトル間の類似性または距離尺度（ユークリッド距離など）を用いて、一方の画像の特徴点を他方の画像の特徴点にマッチングする。

【0084】

本方法の利点は、キーポイント確率マップを生成する前に、画像の前処理を行う必要がないことであると考えられる。モデルまたはニューラルネットワークを使用してステップ1でキーポイント確率マップを生成する場合、モデルまたはニューラルネットワークは、より良いキーポイントのための最適化プロセスにおいて、暗黙的に前処理を行うことを学習してもよい。

【0085】

説明した例では、画像を合成的に変換して（ランダムホモグラフィを使用して）訓練セットを作成している。しかし、この手法は、実際のデータペアを訓練するためにも使用され得る。その場合例えば、グラウンドトゥルスホモグラフィHは、人間の観察者または別の方法で決定され得る。

【0086】

本明細書で開示の技術を用いて訓練されたネットワークは、フルサイズのグレースケール画像上のGLAmpointsと呼ばれる安定した関心点の位置を予測できる。以下では、使用する訓練セットの生成と訓練手順の例を開示する。標準的な畳み込みネットワークアーキテクチャが採用される可能性があるため、これらのアーキテクチャについては最後に簡単に説明するのみである。

【0087】

本発明者らは、糖尿病性網膜症（DR）、緑内障、加齢黄斑変性症（AMD）などの様々な眼疾患の診断に広く用いられているヒトの網膜のデジタル眼底画像に着目した研究を行っている[37、47]。同じセッションで取得された、オーバーラップが小さい網膜画像の場合、レジストレーションを使用して、網膜のより広い領域を描写するモザイクを作成できる。画像のモザイク処理により、眼科医は網膜を1枚の大きな画像で表示することができ、診断や治療計画を立てる際に役立つ。また、異なる時間に撮影された網膜画像のモザイク処理は、眼疾患の進行を監視したり、識別したりするのに重要であることが示されている。さらに重要なことに、DRの眼科レーザー治療において、レジストレーションの応用が模索されている。レジストレーションは、手術中に血管をリアルタイムで追跡し、網膜へのレーザー照射を正確に行い、健康な組織へのダメージを最小限に抑えることを確実にする。

【0088】

モザイク処理は、通常、画像から再現性のある関心点を抽出し、それらを関連付ける変換を探索することに依存する。その結果、キーポイントの検出は、このパイプラインの中で最も重要な段階であり、それ以降のすべての段階、つまりレジストレーションの成功を左右する条件となる。

【0089】

同時に、従来の特徴検出器は汎用的なものであり、屋外で、合焦し、エッジやコーナーがシャープでノイズの少ない画像に対して手動で最適化されている。しかし、拡大され、歪んでおり、ノイズが多く、合焦の保証がなく、エッジがシャープでない軟部組織が描かれている可能性がある医用画像では、通常、機能しない（図3参照）。

【0090】

図3は、細隙灯データセットからの画像の例で、レジストレーションにとって困難な条件を示している。A) 血管が少なく、露出オーバーのため、コントラストやコーナーが弱く、B) モーションブラー、C) フォーカスブラー、D) 撮影時のアーティファクトと反射。

【0091】

このような画像では、従来の手法では最適な結果が得られず、レジストレーションの後の段階で、ランダムサンプリングコンセンサス (RanSAC) [20]、バンドル調整 [38]、同時定位とマッピング (SLAM) [18] など、より高度な最適化が必要となる。これらのケースでは、特徴点のグラウンドトゥールズが欠落しているため、教師付き学習法は失敗するか、適用できない。

【0092】

本開示では、半教師付きの方法で特徴点を学習する方法が開示されている。学習済み特徴検出器は、ヒューリスティックに基づく手法よりも優れた性能を呈することが示されたが、それらは通常、再現性向けに最適化されており、その結果、最終的なマッチングの際に性能が低下する可能性がある。一方、本開示によれば、GLAMPointsと呼ばれるキーポイントは、最終的なマッチング能力のために訓練されてもよく、例えば、スケール不変特徴変換 (SIFT) 記述子と関連付けられた場合、マッチング性能とレジストレーション品質において、先行技術を凌駕する。

【0093】

図1は、一対の細隙灯画像の間で検出されたキーポイントとそのマッチングの例を示す。第1列は原画像であり、第2列は前処理済みデータを示す。検出されたキーポイントは白で示されている。白線は真陽性マッチを、黒線は偽陽性マッチを示す。

【0094】

図1に示すように、GLAMPoints (A行で示す) は、SIFT (B行で示す) よりも多くの正しいマッチを生み出している。特徴点に基づいたレジストレーションは、ポイントマッチングと変換の推定により、本質的に微分不可能である。強化学習 (RL) を用いると、検出されたポイントを古典的なRLの意味での決定と仮定することで、この問題を回避できる。これにより、特定の画像モダリティに特化した畳み込みニューラルネットワーク (CNN) の学習に、重要な性能指標であるマッチング能力を直接使用することが可能になる。

【0095】

訓練されたネットワークは、フルサイズのグレースケール画像上の安定した関心点の位置を予測できる。これらの関心点を、本明細書では「GLAMPoints」と呼ぶ。以下では、訓練セットの生成方法および訓練手順を開示する。標準的な畳み込みネットワークアーキテクチャを使用することが可能であるため、このアーキテクチャについては最後に簡単に説明するにとどめる。

【0096】

一例として、眼科分野の訓練セット、すなわち、レーザー治療に使用される細隙灯眼底映像を選択した。このアプリケーションでは、網膜組織を正確に切除するためにライブレジストレーションが行われる。例となる訓練データセットは、300~700画素×150~400画素の範囲の異なる解像度を有する1336画像のベースセットからの画像で構成されている。これらの画像は、複数のカメラやデバイスで撮影されており、画像モダリティの多様性をカバーしている。これらの画像は、健康な人や糖尿病性網膜症の人など、10人の異なる患者の眼の検査から得られたものである。フル解像度のサンプルは、大きな画像にゼロでパディング、またはランダムに切り抜く等で256×256画素に縮小

10

20

30

40

50

した。このサイズ縮小は、学習プロセスを高速化し、画像の数を増やすために行われた。しかし、上述の寸法と訓練セットの内容は、単に非限定的な例として本明細書に提供されていることに留意されたい。

【0097】

BをサイズH×Wのベース画像のセットとする。ステップiごとに、2つの別々の、ランダムにサンプリングされたホモグラフィ変換 g_i 、 g'_i を適用することで、原画像 B_i から画像ペア I_i 、 I'_i が生成される。したがって画像 I_i と画像 I'_i は、ホモグラフィ

【数6】

$$H_{I_i I'_i} = (g'_i * g_i)^{-1}$$

10

に基づいて関連づけられている。ホモグラフィの生成方法については、本明細書の別の箇所で詳しく説明している。幾何学的変換に加えて、ガウスノイズ、コントラストの変化、照明、ガンマ、モーションブラー、画像の逆数など、標準的なデータ拡張を使用してもよい。これらの外観変換のサブセットは、各画像 I および I' に対してランダムに選択されてもよい。ある実施形態では、ネットワークが同じ画像ペアを2度見ることがないように、各ステップでベース画像に異なる幾何学的変換と外観変換が適用される。

【0098】

特徴点検出器を訓練するには、古典的な強化学習（RL）の機能を利用してもよい。RLは、ある環境下での行動の確率を推定し、いくつかのステップでリワードを最大化することを目的としている。特徴点検出は、画像のすべての位置で1つの行動をとること、すなわち、特徴点として選択するか、背景として選択することとみなすことができる。学習した関数は次のように定義され得る。

$$f_{\theta}(I) \rightarrow -S$$

20

ここでSは、画素単位の特徴点確率のマップを表す。キーポイントの位置の直接的なグラウンドトゥールズがない場合、代わりにリワードを計算できる。このリワードは、マッチングの成功に基づくものであり得る。このマッチングの成功は、コンピュータビジョンの古典的なマッチングステップの後に計算されてもよい。

【0099】

訓練は以下のように行われてもよい。

30

1. グラウンドトゥールズホモグラフィ $H = H_I$ 、 I に関連する一対の画像 $I \in \mathbb{R}^{H \times W}$ 、 $I' \in \mathbb{R}^{H \times W}$ が与えられた場合、モデルは、各画像に対する以下のスコアマップを提供してもよい。

$$S = f_{\theta}(I) \text{ および } S' = f_{\theta}(I').$$

2. 関心点の位置は、標準的な微分不可能ノンマキシマムサプレッション（NMS）を用いて、ウィンドウサイズwで両方のスコアマップ上に抽出されてもよい。

3. 検出された各キーポイントに対して、128個のroot SIFT特徴記述子が計算される。

4. 例えば[1]のようなブルートフォースマッチャーを用いて、画像 I からのキーポイントが画像 I' のキーポイントとマッチングされ得る、または、画像 I' のキーポイントが画像 I からのキーポイントとマッチングされ得る。例えば、両方向で発見されたマッチのみが保持される。

40

5. マッチは、グラウンドトゥールズホモグラフィ H に従ってチェックされる。画像 I の対応するキーポイント x が、 H を適用した後の I' のポイント x' の近傍に入る場合、マッチは真陽性と定義される。これは次のように定式化される。

【数7】

$$\|H * x - x'\| \leq \varepsilon$$

50

ここで、 ε は、例えば、3 画素として選択され得る。T は真のマッチのセットを表す。

【0100】

古典的な RL の枠組みでは、ある行動が取られ、つまり特徴点が選択され、真陽性点の集合に入った場合、陽性リワードが得られる。すべてのその他の点／画素には 0 のリワードが与えられる。そして、リワード行列 R は、例えば以下のように定義され得る。

【数 8】

$$R_{x,y} \in \mathbb{R}^{H \times W} = \begin{cases} 1, & \text{for } (x,y) \in T \\ 0, & \text{otherwise} \end{cases}$$

10

これにより、次のような損失関数が得られる。

【数 9】

$$L_{\text{simple}} = (f_{\theta}(I) - R)^2$$

【0101】

しかし、この方式の大きな欠点は、陽性リワードを与えられた点と陰性リワードを与えられた点との間のクラスの不均衡が大きく、特に訓練の最初の段階では、後者が圧倒的に優勢になる可能性があることである。ほとんどがゼロの値であるリワード R が与えられた場合、ゼロの出力に収束する可能性がある。ハードマイニングは記述子の訓練を促進することが示されている [37]。偽陽性マッチに対してネガティブハードマイニングを行うことで、本手法のパフォーマンスも向上する可能性があるが、本研究では検討していない。代わりに、不均衡を解消するために、サンプルマイニングを使用してもよく、すべての n 個の真陽性点と追加の n 個を、偽陽性のセットからランダムにサンプリングすることができる。バックプロパゲーションは、2 n 個の真陽性特徴点とマイニングされた偽陽性キーポイントを介して実行されてもよい。特定の実施形態では、バックプロパゲーションは、2 n 個の真陽性特徴点とマイニングされた偽陽性キーポイントを通じてのみ実行される。偽陽性よりも真陽性の方が多い場合は、見つかったすべてのマッチに対して勾配をバックプロパゲーションしてもよい。このマイニングは、マスク M として数学的に定式化することができ、真陽性キーポイントの位置とマイニングされた特徴点のサブセットの位置では 1 に等しく、それ以外では 0 に等しくなる。したがって損失は次のように定式化され得る。

20

30

【数 10】

$$L(\theta, I) = \frac{\sum (f_{\theta}(I) - R)^2 \cdot M}{\sum M}$$

40

ここで、記号 $\cdot$ は要素ごとの乗算を表す。

【0102】

図 2 A は、ある基本画像 B のエポック i における画像ペア I と I' の訓練手順を示す。図 2 B は、損失計算の例を示す。図 2 C は、U n e t - 4 の模式図である。

【0103】

訓練ステップの概要を図 2 に示す。重要なことに、ステップ 1 のみが損失に関して微分可能であることが観察されよう。学習は、微分不可能なアクションの結果であるリワードに対して、監督なしで直接行われる。また、使用されている記述子は、回転不変性のない S I F T バージョンであることにも注目されたい。その理由は、細隙灯画像では、回転不

50

変性 S I F T 検出器／記述子よりも、回転依存性 S I F T 検出器／記述子の方が性能が良いためである。評価の目的は、検出器のみを調査することであり、そのため、一貫性を保つために回転依存型 S I F T 記述子を使用した。

【0104】

f_{θ} の学習には、最終的にシグモイド活性化を行う標準的な 4 レベルの深層 U n e t [34] を用いた。これは、バッチ正規化と正規化線形ユニット (R e L U) アクティベーションを用いた 3×3 の畳み込みブロックを含む (図 2 C 参照)。キーポイント検出のタスクは、画素単位のバイナリセグメンテーション (クラス関心点か否か) に類似しているため、バイナリセグメンテーションやセマンティックセグメンテーションのタスクで過去に成功を収めている U n e t モデルが有望であると考えられた。

【0105】

以下では、テストデータセットと評価プロトコルについて説明する。既存の検出器を、G L A M p o i n t s のような本明細書で開示する技術と定量的・定性的に比較する。

【0106】

この研究では、訓練済みモデルを、いくつかの眼底画像データセットと自然画像でテストした。医療画像については、2 つのデータセットを使用した。

【0107】

まず、「細隙灯」データセット。3 人の患者の網膜動画から、サイズが 338 画素 (p x) ~ 660 p x $\times 190 \sim 350$ p x の 206 フレームペアをランダムに選択し、テストサンプルとした。図 3 に例を示すが、見ての通り、これらのフレームには複数のアーティファクトがあり、特に難しいデータセットとなっている。ペアは、 $20 \sim 100\%$ の範囲でオーバーラップするように選択されている。これらのペアは、アフィン変換と 15 度までの回転によって関連付けられている。専用のソフトウェアツール (O m n i v i e w e r 1) を用いて、すべてのペアの画像に、少なくとも 5 つの対応点に関して手動で注釈を付けた。これらのランドマークを用いて、ペア間のグラウンドトゥルスホモグラフィを推定した。正しいマッチングを確認するために、推定されたホモグラフィを専門家がすべて検証し、誤ったマッチングを修正した。

【0108】

次に、F I R E データセット [22]。F I R E データセットは、一般に公開されている網膜画像レジストレーションデータセットであり、グラウンドトゥルス注釈付きである。 129 枚の網膜画像から構成されており、 134 組の画像ペアを形成している。 2912×2912 画素の原画像は、学習セットと同様の解像度になるように、元のサイズの 15% に縮小されている。このような画像の例を図 5 に示す。

【0109】

眼底画像のテストでは、前処理として、緑チャンネルを分離し、適応型ヒストグラムイコライゼーションとバイラテラルフィルタを適用してノイズを低減し、エッジの外観を向上させた。前処理の効果は図 1 で確認できる。このプロセスは、D e Z a n e t t a [46] によると、検出と描写の間に改善された結果をもたらす。しかし、G L A M p o i n t s は、原画像と前処理済み画像の両方でうまく動作する。そこで、性能を比較するために、両方のケースでの評価結果を提示する。

【0110】

さらに、キーポイントの検出とマッチングを自然画像に対しても行った。これには、O x f o r d データセット [30]、E F データセット [48]、W e b c a m データセット [40、24]、V i e w P o i n t データセット [45] を用い、合計で 195 組のペアを作成した。これらのデータセットで合計 195 組になる可能性がある。より詳細な情報は、この説明の他の部分で述べられている。

【0111】

評価基準としては、再現性、キーポイントの平均検出数、成功率などを考慮した。これらについては、以下で詳しく説明する。

【0112】

10

20

30

40

50

1. 再現性は、両方の画像に現れる対応点の割合を表す。

【数 1 1】

$$\frac{|\{x \in P, x' \in P' \mid \|H_{I,I'} * x - x'\| < \varepsilon\}|}{|P| + |P'|}$$

【0 1 1 3】

検出された点のセットは、対応するキーポイント x と x' を伴う画像 I と I' に対して、それぞれ P 、 P' と表される。 H 、 I は、基準画像と変換後の画像とを関連付けるグラウンドトゥールズホモグラフィである。 ε は、2つの点間の距離カットオフ（3画素に設定）である。

【0 1 1 4】

2. 画像ごとに検出されたキーポイントの平均数。マッチングは、[28] で提案されている最近傍距離比（NNDR）戦略を用いて検出された。第1の最近傍と第2の最近傍の間の記述子距離比が、ある閾値 t 未満であれば、2つのキーポイントはマッチングする。マッチング性能に関しては、以下の指標が評価された。

（A）AUC、これは、（ t の値を変化させて作成した）ROC（受信動作特性）曲線の下面積である。これにより、[15、43、42]と同様に、各手法の識別力を評価できる。

（b）マッチングスコア、これは、共有視点領域において検出器が抽出した特徴量の総数に対する正しいマッチングの比率として定義される[29]。この指標により、総特徴量パイプラインの性能を評価できる。

（c）カバレッジ率、これは、正しくマッチしたキーポイントによる画像のカバレッジを測定する。この計算のため、[7]で提案された手法を採用して、正しくマッチしたキーポイントからカバレッジマスクを生成し、それぞれのキーポイントに一定の半径（25px）の円盤を追加した。

【0 1 1 5】

RanSACアルゴリズムを適用して、基準画像と変換画像を関連付けるホモグラフィ

【数 1 2】

$$\hat{H}$$

を計算することで、検出済みマッチから外れ値を取り除いた。上記の指標はすべて、マッチング性能を示すものである。

【0 1 1 6】

3. 成功率：[13、41]と同様に、各検出器が達成したレジストレーション品質と精度を評価した。これを行うために、以下に示すように、基準画像の6つの固定点の再投影誤差を互いに比較した。

【数 1 3】

$$c_i, i = \{1, \dots, 6\}$$

ホモグラフィが見つかった各画像ペアについて、対応関係の中央値誤差MEE、最大値誤差MAE、およびルート平均二乗誤差RMSEを用いてレジストレーションの品質を評価した。これらの指標を用いて、MEEとMAEの異なる閾値を定義し、「許容できる」レジストレーションと「不正確な」レジストレーションを定義した。細隙灯データセットでは、（ $MEE < 10$ 、 $MAE < 30$ ）の場合には、画像ペアを「許容できる」レジストレーションとし、それ以外の場合には「不正確な」レジストレーションとした。一方、F

10

20

30

40

50

IREデータセット [22] の網膜全体の画像では、($MEE < 1 : 50$, $MAE < 10$) が「許容できる」レジストレーションとなる。閾値の値は、結果を後から見て経験的に求めたものである。

【0117】

最後に、各クラスの成功率を算出し、レジストレーションが各カテゴリに該当する画像ペアの割合とした。これらの指標を、実世界での全体的な性能の最も重要な定量的評価基準と考えてもよい。

【0118】

root-SIFT記述子に関連付けられた検出器の性能を評価するために、マッチング能力とレジストレーション品質を、よく知られた検出器および記述子と比較した。その中でも、SIFT [2]、root-SIFT [28, 8]、KAZE [6]、LIFT [44] は、Truongら [39] によって、眼底画像に対して良好な性能を示すことが示された。さらに、この手法は、他のCNNベースの検出器・記述子と比較され、LF-**NET** [31] やSuperPoint [16] と比較された。本発明者らは、LIFT (Picadillyで事前訓練済み)、SuperPoint、LF-**NET** (屋内データで事前訓練済み、屋外データで事前訓練したものよりも眼底画像ではかなり良い結果が得られる可能性がある) の実装と、SIFTとKAZEのOpenCVの実装を使用した。Root-SIFTの回転依存型バージョンは、回転不変型バージョンに比べてテストセットでのパフォーマンスが高いため、使用されてもよい。

【0119】

GLAMPointsの訓練は、Tensorflow [4] を用いて、ミニバッチサイズを5に設定し、Adamオプティマイザ [25] を用いて、デフォルトのパラメータである学習率 = $0 : 001$ および $\beta = (0.9, 0.999)$ で行った。このモデルを4回のクロスバリデーションで検証したところ、成功率の標準偏差が1%という類似した結果が得られた。その後、8936枚のベース画像からなる全データセットを対象に、35回のエポックで再訓練を行った。GLAMPoints (NMS10) は、10pxのNMSウィンドウを用いて訓練・テストされた。他のNMSウィンドウも適用可能であり、同様の性能が得られることに留意されたい。

【0120】

表1：細隙灯データセットの206枚の画像に対する各検出器のレジストレーションクラスごとの成功率 (%)。許容できるレジストレーションは、($MEE < 10$, $MAE < 30$) を有すると定義されている。最良の結果を太字で示す。

10

20

30

40

50

【表 1】

a) 非前処理データ

	失敗 [%]	不正確 [%]	許容できる [%]
SIFT	14.56	63.11	22.33
KAZE	24.27	61.65	14.08
SuperPoint	17.48	48.54	33.98
LIFT	0.0	43.69	56.31
LF-NET	0.0	39.81	60.19
GLAMpoints (NMS10)	0.0	36.41	63.59

b) 前処理済みデータ

	失敗 [%]	不正確 [%]	許容できる [%]
SIFT	1.94	47.57	50.49
KAZE	1.46	54.85	43.69
SuperPoint	7.77	51.46	40.78
LIFT	0.0	39.81	60.19
LF-NET	0.0	36.89	63.11
GLAMpoints (NMS10)	0.0	31.55	68.45

【0121】

表 1 は、細隙灯データセットで評価したレジストレーションの成功率を示している。前処理を行わない場合、比較に用いたほとんどの検出器は、前処理を行った画像に比べて低い性能を示した。それに比べて、提示するモデルは、前処理なしの画像で良好な性能を示した。表 1 に示すように、SIFT、KAZE、SuperPoint の許容できるレジストレーション成功率（A）は、前処理を行った画像と行っていない画像の間で 20～30% 低下しているのに対し、GLAMpoints、LIFT、LF-NET の許容できるレジストレーション成功率（A）は、3～6% の低下のみを提示している。また、LF-NET、LIFT、GLAMpoints は、前処理によらず、安定した平均キーポイント数（LF-NET が約 485、IFT、GLAMpoints が約 350）を検出しているが、他の検出器では 2 倍の減少となっている。

【0122】

テストした実施形態では、GLAMpoints は、KAZE、SIFT、SuperPoint よりも、許容できるレジストレーションの成功率において、少なくとも 18% 上回り得る。同じカテゴリにおいて、LF-NET と LIFT は、元データと前処理済みデータにおいて、それぞれ 3% と 5%、7% と 8% 上回っている。また、LF-NET は眼底画像で訓練した場合、この特定の指標とデータセットで同様の結果が得られる可能性があるが、LF-NET の学習手順では、相対ポーズと対応する深度マップを伴う画像ペアを使用することに注目することも重要であり、深度マップは、眼底画像に関して得るには、不可能ではないにせよ非常に困難であり得る。

【0123】

さらに、前処理とは無関係に、GLAMpoints モデルは、不正確なレジストレーション（I）においても、成功したレジストレーションにおいても全体的に MEE と RM

S E が最小となった。すべての検出器において、許容できるレジストレーションの M E E と R M S E は 1 画素以内で類似している。クラスごとに対応する M E E と R M S E の詳細は、本明細書の別箇所にある。G L A M p o i n t s の前処理に依存しない安定した結果は、この検出器が高品質の画像で S I F T 以上とは言わないまでも同程度の性能を発揮する一方で、低品質でコントラストの弱い画像でも性能が低下しないことを示している。

【0124】

S I F T は多数のキーポイントを抽出するが（未処理画像では平均 205.69 個、前処理済みでは 431.03 個）、それらのキーポイントはクラスターで現れる。その結果、再現性が比較的高くても（多くのおそらく有効なマッチ）、関心点の位置が近いため、最近傍距離比（N N D R）故に多くのマッチが拒否され、真陽性マッチは少なくなり、これは図 4 から明らかなようにカバレッジが低いことからわかる。その結果、M : s c o r e と A U C が小さくなる。再現性の値が同様の場合、我々のアプローチは、広く拡散しそのマッチング能力が訓練された（カバレッジ率が最も高い）関心点を抽出し、その結果、図 4 に見られるように、より多くの真陽性マッチング（2 番目に高い M : s c o r e と A U C）が得られる。

【0125】

図 4 は、206 ペアの細隙灯データセットで評価した検出器／記述子の性能指標の概要を示したものである。図 4 A は、非前処理済みのデータの結果を示しており、図 4 B は前処理済みデータの結果を示している。

【0126】

また、S u p e r P o i n t は、カバレッジが比較的小さいにもかかわらず、最も高い M : s c o r e と A U C を獲得したことにも注目されよう（図 4 参照）。しかし、この場合、M : s c o r e と A U C は人為的に高くなっており、これは、S u p e r P o i n t が検出するキーポイントの数が非常に少なく（非前処理画像と前処理画像の平均でそれぞれ 35, 88 と 59, 21）、再現性が最も低いため、正しいマッチングの可能性が少ないためである。そのため、真陽性マッチ率が低いにもかかわらず、高いマッチング性能を示している。このことは、不正確なレジストレーションや失敗したレジストレーションが多いことからわかる（未処理データでは 48.54% と 17.48%、前処理済み画像では 51.46% と 7.77%、表 1）。

【0127】

最後に、L F - N E T は非常に高い再現性（元データで最高、前処理済み画像で 2 番目に大きい）を示したが、その M : s c o r e と A U C はランキングの下位に位置していることが注目される（図 4）。これは、L F - N E T 検出器の訓練が、マッチングの目的よりも再現性を優先していたためと考えられる。

【0128】

表 2：F I R E データセットの非前処理画像に対する各検出器の成功率（%）。許容できるレジストレーションを、（M E E < 1 : 5, M A E < 10）を有するものとして定義している。最良の結果を太字で示す。

【表 2】

	失敗 [%]	不正確 [%]	許容できる [%]
SIFT	2.24	91.79	5.97
KAZE	14.10	84.33	1.49
SuperPoint	0.0	53.73	46.27
LIFT	0.0	57.46	42.54
LF-NET	0.0	73.88	26.12
GLAMpoints (NMS10)	0.0	52.99	47.01

【0129】

GLAMpointsはFIREデータセットでも評価した。すべての画像は血管のコントラストが高い良質なものであるため、前処理を施す必要はない。表2は、レジストレーションの成功率を示している。クラスごとのMEEとRMSEの平均値と標準偏差は、この説明の別箇所にある。提示する手法は、成功率と、失敗していないレジストレーションの全体的な精度の両方で、優れた結果を示した。興味深いことに、GLAMpointsとSIFTの間で、許容できるレジストレーションの成功率に41.04%の差があることに気付く。GLAMpointsとSIFTは同じ記述子(SIFT)を使用しているため、この差は検出器によってのみ説明される。実際、図5に見られるように、SIFTでは、血管樹や画像の境界にのみ高密度に配置された限られた数のキーポイントが検出されるのに対し、GLAMpoints(NMS10)では、焦点や血管外帯などの困難な領域を含む網膜全体の関心点が抽出され、正しいマッチングの数が大幅に増加している。

10

【0130】

図5は、FIREデータセットの一对の画像について、前処理を行わずに検出された関心点とそれに対応するマッチング結果を示したものである。黒い点は、検出された関心点を示している。白線は真陽性マッチを、黒線は偽陽性マッチを示す。行A)は、GLAMpointsを用いて得られた関心点とマッチを示し、行B)は、SIFTを用いて得られた関心点とマッチを示している。

【0131】

GLAMpoints(NMS10)が他のすべての検出器よりも優れているとはいえ、LIFTとSuperPointもFIREデータセットで非常によく動作している。実際に、このデータセットでは、はっきりとしたコントラストの血管樹上の明確なコーナーを示している。LIFTは、画像全体に渡ってキーポイントを抽出することができ、SuperPointは、合成されたプリミティブな形状のコーナーを検出するように訓練されている。しかし、細隙灯データセットで明らかになったように、SuperPointの性能は、特徴がはっきりしていない画像では強く悪化する。

20

【0132】

図5は、提示されたGLAMpoints(NMS10)とSIFTの方法に関して、FIREデータセットの画像ペア間のマッチを示している。すべての検出器のマッチのさらなる例は、この説明の他箇所にある。

30

【0133】

特定の実施形態において、本明細書に開示された特徴検出器は、眼底細隙灯ビデオからモザイクを作成できるシステムまたは方法において使用され得る。これを行うために、各フレーム上でキーポイントおよび記述子が抽出され、RanSaCを用いて連続する画像間のホモグラフィが推定される。次に、計算されたホモグラフィに基づいて、画像をワープする。25枚から558枚の画像を含む10本のビデオを用いて、連続するフレームをレジストレーションすることにより、モザイクを生成した。レジストレーションが失敗するまでの平均フレーム数(キーポイントが抽出されなかったり、一对の画像間でマッチしなかったりするため)を計算した。これら10本のビデオにわたり、失敗までのレジスター済み平均フレーム数は、GLAMpoints(NMS15)では9.98、SIFTでは1.04であった。このようなモザイクの例を図6に示す。

40

【0134】

図6は、失敗するまで連続した画像をレジストレーションして作成したモザイクを示す。A) GLAMpoints、非前処理済みデータ、53フレーム、B) SIFT、前処理済みデータ、34フレーム、C) SIFT、非前処理済み画像、11フレーム。

【0135】

同じビデオから、SIFTは前処理済みデータでは34フレーム後に失敗し、元データでは11フレーム後にのみ失敗したが、GLAMpointsは53枚の連続した画像のレジストレーションに成功し、視覚的なエラーもなかった。このモザイクは、フレーム間のマッチングで作成され、バンドル調整は行われていないことに留意されたい。[46]

50

に記載のものと同じブレンド方法が使用された。

【0136】

検出の実行時間は、 660×350 pxの解像度の84ペアの画像に対して計算された。GLAMpointsアーキテクチャはGeForce GTX GPUで実行され、NMSとSIFTはCPUを使用した。GLAMpoints (NMS10)とSIFTの実行時間の平均値と標準偏差を表3に示す。

【0137】

表3：本検出器とSIFT検出器の画像検出平均実行時間 [ms]

【表3】

	GLAMpoints (NMS10)	SIFT
前処理	0.0	16.64 ± 0.93
検出画像 I	CNN: 16.28 ± 96.86 NMS: 11.2 ± 1.05	28.94 ± 1.88
計	27.48 ± 98.74	45.58 ± 4.69

【0138】

細隙灯画像で訓練したGLAMpointsを用いて、自然画像での結果を算出した。自然画像では大域的に、GLAMpointsは、許容できるレジストレーションに関して75.38%の成功率に達し、これに対し、最も性能の高かった回転不変型SIFT検出器は85.13%、SuperPointは83.59%であった。また、AUC、M：score、カバレッジ率においても先行技術の結果まで上がり、それぞれ第2位、第2位、第1位のスコアであった。また、再現性において、GLAMpointsは、SIFT、KAZE、LF-NETよりも多くの画像のレジストレーションに成功したにもかかわらず、SIFT、KAZE、LF-NETに次いで最下位から二番目となっており、再現性は検出器の性能を測るための最適な指標ではないことが改めて示された。この評価基準の詳細については、本明細書の別の箇所で説明する。最後に、このデータセットの屋外画像は、GLAMpointsが訓練された医療用眼底画像とは全く異なるものであり、これは実質的な汎化特性を示していることに留意する必要があると得る。

【0139】

提示した手法は、この説明ではGLAMpointsと呼ぶ深層RLを用いて、学習可能な検出器を訓練する。この検出器は、例えば、医療用眼底画像の画像マッチングやレジストレーションにおいて、先行技術を凌駕する可能性がある。実験の結果、以下のことが実証された。(1) 検出器は、特定の記述子に関連するマッチング能力を直接訓練したものであり、パイプラインの一部のみが微分可能である。他の多くの検出器は再現性を重視して設計されているが、キーポイントは再現性があってもマッチングには適していないため、ミスリードする恐れがある。(2) 学習は合成データのみで行い得る。これにより、手動で注釈を付ける手間が省け、学習データの使用量における融通性を提供する。(3) 訓練方法は領域に関して柔軟に対応しており、医療用眼底画像で成功するように最適化されているが、他の種類の画像にも適用可能である。(4) 訓練済みCNNは、他の最先端の検出器と比較して、コーナー／特徴が少ないテクスチャの低い画像において、正しいマッチングにつながる良好なキーポイントを有意に多く検出する。その結果、画像の明示的な前処理は必要ないことがわかった。(5) 既に存在するあらゆる特徴記述子は、対応する優れた検出器を学習することで改善できる可能性がある。

【0140】

別の実施形態では、キーポイントの位置とともに、回転不変の記述子を計算することが可能である。検出と記述の両方は、同様の方法でエンドツーエンドで訓練され得る。さらに、本実験はU-Net CNNアーキテクチャを用いて行われたが、場合によってはU-Net (UNet) よりも優れた性能をもたらす可能性のある他のCNNアーキテクチャも適用可能である。

【0141】

以下では、訓練方法に関する補足的な詳細を開示する。これらの詳細は、非限定的な例とみなされることが理解されるであろう。

【0142】

回転不変性のあるSIFT記述子と回転不変性のないSIFT記述子の性能比較を行った。良く学習された正確なマッチポイント (GLAMPoints) 検出器は、眼底画像において、回転不変性のないSIFT記述子の方が回転不変性のあるバージョンよりも性能が良かったため、回転依存性のあるSIFT (スケール不変特徴変換) 記述子に関連付けて学習・テストされた。両バージョンのSIFT記述子について、前処理済みの細隙灯データセットで評価した指標の詳細を表4に示す。

【0143】

表4：206ペアの細隙灯データセットの前処理済みペアに関して計算されたメトリクス。各カテゴリの最良の結果を太字で示す。

【表4】

	回転不変性SIFT	回転依存性SIFT
許容できるレジストレーションの成功率 [%]	49.03	50.49
不正確なレジストレーションの成功率 [%]	50.49	47.57
失敗したレジストレーションの成功率 [%]	0.49	1.94
Mスコア	0.0470 ± 0.003	0.056 ± 0.003
カバレッジ率	0.1348 ± 0.010	0.15 ± 0.012
AUC	0.1274 ± 0.008	0.143 ± 0.007

【0144】

以下に、ホモグラフィ生成の例示的な方法を概説する。このホモグラフィ生成は、単一の画像を含むデータから、画像のペアを含む訓練データを生成するために行われてもよい。学習に用いるサイズH×Wのベース画像の集合をBとする。ステップiごとに、ランダムにサンプリングされた2つの別個のホモグラフィ変換 g_i , g'_i を適用して、原画像 B_i から画像ペア I_i , I'_i を生成する。これらのホモグラフィ変換はそれぞれ、回転、剪断、透視、スケール、並進から構成されている。他の変換タイプの組み合わせも代替的に使用できる。パラメータの例示的な最小値と最大値を表5に示す。

【0145】

表5：訓練中のランダムなホモグラフィ生成のための例示的なパラメータ

【表5】

スケール	透視	並進	剪断	角度
最小スケール 0.7	最小透視パラメータ 0.000001	最大水平変位 100	最大/最小水平剪断 -0.2 / 0.2	最大角度 25
最大スケール 1.3	最大透視パラメータ 0.0008	最大垂直変位 100	最大/最小垂直剪断 -0.2 / 0.2	

【0146】

図7は、特徴点検出用モデルを訓練するシステム701を示す。システム701は、制御ユニット705と、通信ユニット704と、メモリ706とを含んでもよい。制御ユニット705は、任意のプロセッサまたは複数の協働するプロセッサを含んでもよい。制御ユニット705は、代替的に、専用の電子回路によって実装されてもよい。通信ユニット704は、カメラ702やディスプレイ703などの周辺機器を接続するための任意の種類のインターフェースを含んでもよく、また、データの交換および／または外部機器の制

御のために、例えば、ネットワーク接続を含んでもよい。代替実施形態では、カメラ 702 および／またはディスプレイは、単一の装置として、システム 701 に組み込まれてもよい。代替実施形態では、カメラ 702 によって撮影された画像は、外部データベース（図示せず）に保存され、その後、通信ユニット 704 に送信されてもよい。同様に、システム 701 によって生成されたデータは、ディスプレイに表示される前に、外部データベースに保存されてもよい。そのために、通信ユニット 704 は、例えばネットワークを介して、データサーバに接続されていてもよい。

【0147】

制御ユニット 705 は、システム 701 の動作を制御する。例えば、制御ユニット 705 は、メモリ 706 に格納されたコードを実行する。メモリ 706 は、RAM、ROM、フラッシュ、磁気ディスク、またはその他の揮発性もしくは不揮発性のコンピュータ可読媒体などの任意の記憶手段、またはそれらの組み合わせを含んでもよい。例えば、コンピュータ命令は、不揮発性のコンピュータ可読媒体に格納されてもよい。メモリ 706 は、画像 709、モデル 708 および任意の他のデータなどのデータ 707 をさらに含んでもよい。プログラムコードは、機能単位またはモジュールに分割されていてもよい。ただし、これは限定されるものではない。

【0148】

動作時、制御ユニット 705 は、カメラ 702 から複数の画像を取得する、または、カメラ 702 で撮影された画像を外部記憶媒体から取得して、画像 709 をメモリ 706 に保存するように構成されていてもよい。

【0149】

システムは、モデル 708 を訓練する訓練モジュール 711 を含んでもよい。モデルは、ニューラルネットワーク、例えば、畳み込みニューラルネットワーク、または、統計モデルなどの他のモデルを含んでもよく、そのモデルパラメータは、訓練モジュール 711 によって実行される訓練プロセスによって調整されてもよい。

【0150】

訓練モジュール 711 は、入力値をモデル 708 に供給すること、入力値に応答してモデル 708 が出力する出力値を評価すること、および評価の結果に基づいてモデル 708 のモデルパラメータを調整することを含む訓練手順を実行するように構成されてもよい。

【0151】

通信ユニット 704 は、制御ユニット 705 の制御下で、複数の画像を受信するように構成されていてもよい。これらの画像 709 は、メモリ 707 に保存されてもよい。

【0152】

任意で、制御ユニット 705 は、カメラ 702（内部カメラまたは外部カメラ）を制御して、画像を生成し、それらを通信ユニット 704 に送信し、メモリ 707 に保存する。

【0153】

システム 701 は、単一の画像の画像ペアを生成する前処理モジュール 717 を含んでもよい。例えば、前処理モジュール 717 は、ランダムな変換を生成し、その変換を第 1 の画像に適用して第 2 の画像を生成するように構成されている。あるいは、前処理モジュール 717 は、2 つのランダムな変換を生成し、第 1 のランダムな変換を適用して特定の画像から第 1 の画像を生成し、第 2 のランダムな変換を特定の画像に適用して同じ特定の画像から第 2 の画像を生成するように構成されてもよい。生成するランダム変換の種類は、特定のアプリケーション領域で発生する典型的な動きに対応するように慎重に構成され得る。代替的な実施形態では、カメラによって生成された 2 つの画像が手動でレジストレーションされ、2 つの画像間の変換が利用可能になるようにする。

【0154】

プロセッサは、第 1 の画像および第 2 の画像を含むそのような一対の画像を、メモリ 709 から取得してもよい。

【0155】

システム 701 は、モデルを用いて、第 1 の画像に対する第 1 のスコアマップおよび第

10

20

30

40

50

2の画像に対する第2のスコアマップを生成するように構成されたスコアマップ生成器713を含んでもよい。すなわち、スコアマップ生成器713は、任意の前処理動作（正規化など、または他の種類の前処理）を行うように構成されていてもよい。しかし、前処理なしでも良好な結果が得られることが確認されている。結果として得られた画像は、モデル708への入力として提供されてもよい。入力に応答してモデル708によって生成された対応する出力は、別の画像（スコアマップ）を含んでもよく、その各画素は、画像を別の画像にレジストレーションする目的で、その点が適切な関心点であるという可能性に関連している。スコアマップ生成器713は、第1の画像および第2の画像に対する処理を、一方が他方から独立して別々に、すなわち、他方の画像の内容に関する知識を用いずに実行するように構成されていてもよいことが観察される。

10

【0156】

システム701は、さらに関心点セクタ702を備えていてもよい。関心点セクタ712は、第1のスコアマップに基づいて第1の画像内の第1の複数の関心点を選択し、第2のスコアマップに基づいて第2の画像内の第2の複数の関心点を選択するように構成されてもよい。この場合も、両画像の処理は、別々の独立した処理であってもよい。例えば、スコアマップ上で最大のスコアを有する点を関心点として選択してもよい。特定の実施形態では、隣り合う関心点間の最大および／または最小の距離が、アルゴリズム手段によって課されてもよい。例えば、画素の各 $N \times M$ ブロックにおいて、最高スコアの画素のみが選択される。隣接する点の間の最大および／または最小の距離に影響を与えるための他のアルゴリズムが想定されてもよい。

20

【0157】

システムは、マッチングモジュール716を含んでもよい。マッチングモジュール716は、画像をペアで処理するように構成されていてもよい。具体的には、マッチングモジュール716は、第1の複数の関心点のうちの第1の関心点と、第2の複数の関心点のうちの第2の関心点とのペアワイズマッチングを行う。言い換えると、第1の画像内の関心点と第2の画像内の関心点とのマッチングを行う。例えば、両画像の関心点について特徴記述子を算出し、第1の画像の関心点の特徴記述子と第2の画像の関心点の特徴記述子との間で類似度を算出する。最も高い類似度を有するペアが、マッチングペアとして選択されてもよい。他のマッチングの方法が想定されてもよい。

【0158】

システムは、マッチングモジュール714によって生成されたペアワイズマッチの正しさをチェックするように構成された検証モジュール715を含んでもよい。このため、検証モジュール715は、グラウンドトゥルス情報にアクセスしてもよい。例えば、画像ペアが、同じ画像の差分（アフィン）変換を用いて人工的に生成された場合、この変換には、この画像ペアの点のグラウンドトゥルスマッチングが含まれる。したがって、第1の画像の点に変換を適用すると、第2の画像の対応するマッチングポイントが得られるはずである。第2の画像内のマッチングポイントと第1の画像内のポイントのグラウンドトゥルス変換との間の距離（例えばユークリッド距離）は、マッチングポイントの誤差とみなすことができる。リワードは、このような誤差の指標に基づいていてもよく、誤差が小さいほどリワードが高くなり、逆に誤差が大きいほどリワードが低くなる。このようにして、各マッチングポイントについて、リワードを計算することができる。これにより、リワードマップ、またはリワードマトリクスを得ることができる。このようにして、第1の画像内の関心点のリワードは、マッチングモジュール714によって見つけられた、第2の画像内のマッチングポイントとのマッチングの成功に関連している。

30

40

【0159】

システムは、スコアマップとリワードマップを結合または比較するための結合モジュール716を含んでもよい。すなわち、スコアマップ生成器713によって生成されたポイントに対するスコアが高く、検証モジュール715によって生成されたそのポイントに対するリワードも高かった場合（「真陽性」）、結合モジュール716は、将来的に類似の関心点を識別するために、モデル708を強化するための値を決定してもよい。一方

50

、スコアマップ生成器 7 1 3 によって生成された点のスコアは高かったが、検証モジュール 7 1 5 によって生成されたその点のリワードが低かった場合（「偽陽性」）、結合モジュール 7 1 6 は、将来的に類似の関心点を識別することを控えるようにモデル 7 0 8 を強化する値を決定してもよい。特定の実施形態では、結合モジュール 7 1 6 は、偽陽性のサブセットに対してのみリワードを決定するように構成される。例えば、画像内の真陽性の数が M 個であった場合、最大で M 個の偽陽性が考慮される。すべての値を合計して、例えば、総リワード関数を計算してもよい。

【0 1 6 0】

訓練モジュール 7 1 1 は、結合または比較の結果に基づいて、モデルを更新するように構成されてもよい。これは、それ自体が当技術分野で知られているモデルの訓練ステップである。更新する正確なパラメータは、使用されるモデルのタイプ、例えば、最近傍、ニューラルネットワーク、畳み込みニューラルネットワーク、U - n e t、または他のタイプのモデルに依存する。

10

【0 1 6 1】

関心点セクタ 7 1 2 は、画像内の任意の点から関心点のうちの最も近い 1 つまでの距離に最大制限を課すように構成されてもよい。

【0 1 6 2】

マッチングモジュール 7 1 4 は、第 1 の関心点において第 1 の画像で検出された特徴と、第 2 の関心点において第 2 の画像で検出された特徴ととの間の類似性に基づいて、ペアワイズマッチングを実行するように構成されてもよい。

20

【0 1 6 3】

マッチングモジュール 7 1 4 は、第 1 の関心点を、複数の第 2 の関心点のうちの 1 つの第 2 の関心点にマッチングすることによって、第 1 の方向のマッチングを実行するように構成されてもよく、第 2 の関心点は、第 1 の関心点における特徴に最も類似する特徴を有する。マッチングモジュール 7 1 4 は、第 2 の関心点を、複数の第 1 の関心点のうちの 1 つの第 1 の関心点にマッチングすることによって、さらなるマッチングを第 2 の方向に実行するように構成されてもよく、第 1 の関心点は、第 2 の関心点における特徴に最も類似した特徴を有している。マッチングモジュール 7 1 4 は、両方向でマッチしないあらゆる全てのマッチを破棄するように構成されてもよい。

【0 1 6 4】

30

検証モジュール 7 1 5 は、リワードマップによって、グラウンドトゥールースデータに従ってマッチに成功した関心点に対するリワードを示し、グラウンドトゥールースデータに従ってマッチに失敗した関心点に対するリワードがないことを示すように構成されてもよい。

【0 1 6 5】

結合モジュール 7 1 6 は、関心点に対するスコアマップおよびリワードマップのみを結合または比較するように構成されてもよい。

【0 1 6 6】

結合モジュール 7 1 6 は、偽陽性マッチの（おそらくランダムな）選択を行い、真陽性マッチおよび偽陽性マッチの選択分についてのみスコアマップおよびリワードマップを結合または比較することによって、真陽性マッチの数と偽陽性マッチの数とを均衡させるように構成されてもよく、真陽性マッチは、正しさのチェックに合格した関心点であり、偽陽性マッチは、正しさのチェックに合格しなかった関心点である。ランダムな選択を行う代わりに、結合モジュール 7 1 6 は、例えば、最も低いリワードマップ値を有する偽陽性マッチを選択するように構成されてもよい。

40

【0 1 6 7】

結合モジュール 7 1 6 は、スコアマップとリワードマップとの間の二乗差の合計を計算するように構成されてもよい。

【0 1 6 8】

通信ユニット 7 0 4 は、ディスプレイ 7 0 3 によって、進捗情報や結合モジュールの最新出力の値などの関心情報を出力するように構成されてもよい。

50

【0169】

図8は、特徴点検出のモデルを学習する方法を示す図である。ステップ801で、前処理ステップは、例えばランダム変換を用いて原画像から画像のペアを生成することと、こうして生成された画像のペアのグラウンドトゥールズ情報としてランダム変換を保存することを含んでもよい。

【0170】

ステップ802は、画像のペアの第1の画像および第2の画像を取得することを含んでもよい。ステップ803は、モデルを使用して、第1の画像に対する第1のスコアマップおよび第2の画像に対する第2のスコアマップを生成することを含んでもよい。ステップ804は、第1のスコアマップに基づいて、第1の画像における第1の複数の関心点を選択し、第2のスコアマップに基づいて、第2の画像における第2の複数の関心点を選択することを含んでもよい。

【0171】

ステップ805は、第1の複数の関心点のうちの第1の関心点と、第2の複数の関心点のうちの第2の関心点とのペアワイズマッチングを行うことを含んでもよい。ステップ806は、第1の画像と第2の画像との間のグラウンドトゥールズ変換に基づいて、ペアワイズマッチの正しさをチェックして、リワードマップを生成することを含んでもよい。ステップ807は、スコアマップとリワードマップを結合または比較することを含んでもよい。ステップ808は、結合または比較の結果に基づいて、モデルを更新することを含んでもよい。ステップ809で、プロセスは、より多くの訓練が望まれるかどうかをチェックしてもよい。より多くの訓練が望まれる場合、プロセスは、次の画像ペアを取得することによって、ステップ802から進んでもよい。

【0172】

図9は、第1の画像を第2の画像にレジストレーションするためのシステム901を示す。この装置は、制御ユニット905と、通信ユニット904と、データ907およびプログラムコード910を格納するためのメモリ906とを備える。特徴点検出のためのモデルを訓練するためのシステム701に関して本明細書で説明したハードウェアおよび代替的な実装オプションに関連する考慮事項は、第1の画像を第2の画像にレジストレーションするためのシステム901にも適用される。例えば、プログラムコード910は、代替的に、専用の電子回路によって実装されてもよい。さらに、カメラ902およびディスプレイ903は、任意の外部装置であってもよいし、システム901に組み込まれて一体型の装置を形成してもよい。システム901は、制御ユニット905、例えば、少なくとも1つのコンピュータプロセッサと、例えば、カメラ902および／またはディスプレイ903と通信するための通信ユニット904と、プログラムコード910または制御ユニットに特定のステップを実行させるための命令を含むメモリ906（任意の種類の記憶媒体）とを備える。メモリ906は、学習済みモデル908および画像909を格納できるように構成されている。例えば、カメラ902から受信した画像を、メモリ909に格納してもよい。

【0173】

システムは、例えば、格納された画像909の中から、例えば、カメラ902によって撮影され、カメラ902から受信された2つの画像である第1の画像および第2の画像を取得するように構成された取得モジュール917を含んでもよい。システムは、学習済みモデル908を使用して、第1の画像に対する第1のスコアマップおよび第2の画像に対する第2のスコアマップを生成するように構成されたスコアマップ生成器913をさらに含んでもよい。例えば、正規化などの任意の前処理を行った後の両方の画像をモデル908に入力してもよく、各画像に応答して生成された出力はスコアマップである。

【0174】

システムは、第1のスコアマップに基づいて第1の画像内の第1の複数の関心点を選択し、第2のスコアマップに基づいて第2の画像内の第2の複数の関心点を選択するように構成された関心点セレクタを含んでもよい。この選択は、特徴記述子と、両画像における

関心点の特徴記述間の類似性とを用いて、関心点セクタ 712 によって実行される選択と同様に行われてもよい。システムは、第 1 の複数の関心点のうちの第 1 の関心点を、第 2 の複数の関心点のうちの第 2 の関心点とペアワイズマッチングするように構成されたマッチングモジュール 914 を含んでもよい。

【0175】

システムはさらに、レジストレーションモジュール 918 を含んでもよい。レジストレーションモジュール 918 は、マッチした関心点に基づいて形態学的変換を決定するように構成されてもよい。レジストレーションモジュール 918 は、マッチした関心点に基づいて、第 1 の画像内の各点を第 2 の画像内の対応する点にマッピングするように構成されてもよい。例えば、マッチした点に基づいて、アフィン変換または非アフィン変換が決定されてもよい。これは、別のパラメータフィッティング手順を含んでいてもよい。しかし、マッチングポイントのセットからそのような変換を生成する方式は、それ自体が当技術分野で知られており、本明細書ではさらに詳細には説明しない。例えば、変換は、第 1 の画像に適用されてもよいし、第 2 の画像に適用されてもよい。通信ユニット 904 は、変換された画像をディスプレイ 903 によって出力するように構成されてもよい。

【0176】

関心点セクタ 912 は、画像内の任意の点から関心点のうちの最も近い 1 つまでの距離に最大制限を課すように構成されてもよい。

【0177】

マッチングモジュール 914 は、第 1 の関心点において第 1 の画像で検出された特徴と、第 2 の関心点において第 2 の画像で検出された特徴との間の類似性に基づいて、ペアワイズマッチングを実行するように構成されてもよい。

【0178】

マッチングモジュール 914 は、第 1 の関心点を複数の第 2 の関心点のうちの 1 つの第 2 の関心点にマッチングすることによって、第 1 の方向のマッチングを実行するように構成されてもよく、第 2 の関心点は、第 1 の関心点における特徴に最も類似した特徴を有する。マッチングモジュール 914 はさらに、第 2 の関心点を複数の第 1 の関心点のうちの 1 つの第 1 の関心点にマッチングすることによって第 2 の方向のマッチングを実行するように構成されてもよく、第 1 の関心点は、第 2 の関心点における特徴に最も類似した特徴を有している。例えば、マッチングモジュール 914 は、両方向ではマッチしないマッチを破棄するように構成されてもよい。

【0179】

図 10 は、第 1 の画像を第 2 の画像にレジストレーションする例示的な方法を示す。ステップ 1002 は、第 1 の画像および第 2 の画像、例えば、カメラからの 2 つの撮像画像を取得することを含んでもよい。ステップ 1003 は、適切に訓練されたモデル、例えば、本明細書に開示されたシステムまたは方法によって訓練されたモデルを使用して、第 1 の画像に対する第 1 のスコアマップおよび第 2 の画像に対する第 2 のスコアマップを生成することを含んでもよい。ステップ 1004 は、第 1 のスコアマップに基づいて、第 1 の画像における第 1 の複数の関心点を選択することと、第 2 のスコアマップに基づいて、第 2 の画像における第 2 の複数の関心点を選択することを含んでもよい。ステップ 1005 は、例えば、関心点に関連する特徴記述子に基づいて、第 1 の複数の関心点のうちの第 1 の関心点と、第 2 の複数の関心点のうちの第 2 の関心点とのペアワイズマッチングを行うことを含んでもよい。任意で、ステップ 1006 は、マッチした関心点に応じて形態学的変換を生成することにより、その画像ペアをレジストレーションすることを含んでもよい。任意で、ステップ 1006 は、変換を画像に適用し、変換された画像をディスプレイを用いて出力すること、または変換された画像を保存することを含む。あるいは、ステップ 1006 は、マッチした関心点を使用するためにその画像ペアを継ぎ合わせ、継ぎ合わせた画像を保存または表示することを含む。

【0180】

本発明の一部またはすべての態様は、ソフトウェア、特にコンピュータプログラム製品

10

20

30

40

50

の形で実施されるのに適し得る。コンピュータプログラム製品は、非一時的コンピュータ可読媒体に格納されたコンピュータプログラムを含んでもよい。また、コンピュータプログラムは、光ファイバケーブルや空気などの伝送媒体によって伝播される光信号や電気磁気信号などの信号によって表されてもよい。コンピュータプログラムは、その一部または全部が、コンピュータシステムで実行するのに適したソースコード、オブジェクトコード、または疑似コードの形態を有していてもよい。例えば、コードは、1つ以上のプロセッサによって実行可能であってもよい。

【0181】

本明細書に記載されている例および実施形態は、本発明を限定するのではなく、説明する働きをする。当業者であれば、添付の特許請求の範囲およびその均等物によって定義される本開示の趣旨および範囲から逸脱することなく、代替の実施形態を設計することができるであろう。特許請求の範囲で括弧に入れた参照符号は、特許請求の範囲を限定するものと解釈しないものとする。特許請求の範囲または本明細書において別個の実体として記載されているアイテムは、記載されているアイテムの特徴を組み合わせた単一のハードウェアまたはソフトウェアアイテムとして実施され得る。

【0182】

以下の主題は、条項の形で開示されている。

【0183】

1. 特徴点検出のための分類器を訓練する方法であって、
第1の画像および第2の画像を取得するステップと、
前記分類器を用いて、前記第1の画像に対する第1のスコアマップおよび前記第2の画像に対する第2のスコアマップを生成するステップと、
前記第1のスコアマップに基づいて、前記第1の画像内の第1の複数の関心点を選択するステップと、
前記第2のスコアマップに基づいて、前記第2の画像内の第2の複数の関心点を選択するステップと、
前記第1の複数の関心点のうちの1つの第1の関心点と、前記第2の複数の関心点のうちの1つの第2の関心点とのペアワイズマッチングを行うステップと、
前記第1の画像と前記第2の画像との間のグラウンドトゥルス変換に基づいて、前記ペアワイズマッチの正しさをチェックして、リワードマップを生成するステップと、
前記スコアマップと前記リワードマップを結合または比較するステップと、
前記結合または比較の結果に基づいて前記分類器を更新するステップとを含む方法。

【0184】

2. 前記複数の関心点を選択するステップは、前記画像内の任意の点から最近傍関心点までの距離に最大制限を課すことを含む、条項1に記載の方法。

【0185】

3. 前記ペアワイズマッチングは、前記第1の関心点において前記第1の画像で検出された特徴と、前記第2の関心点において前記第2の画像で検出された特徴との間の類似性に基づいて実行される、条項1または2に記載の方法。

【0186】

4. 前記マッチングは、前記第1の関心点を、複数の第2の関心点のうちの1つの第2の関心点にマッチングすることによって第1の方向に実行され、前記第2の関心点は、前記第1の関心点における特徴に最も類似した特徴を有する、条項1から3のいずれかに記載の方法。

【0187】

5. 前記マッチングはさらに、前記第2の関心点を、複数の第1の関心点のうちの1つの第1の関心点にマッチングすることによって第2の方向に実行され、前記第1の関心点は、前記第2の関心点における特徴に最も類似した特徴を有する、条項4に記載の方法。

【0188】

6. 前記リワードマップが、マッチングに成功した関心点に対するリワードを示し、マ

ッチングに失敗した関心点に対するリワードがないことを示す、条項 1 から 5 のいずれかに記載の方法。

【0189】

7. 前記結合または比較が、前記スコアマップおよびリワードマップを前記関心点についてのみ結合または比較することを含む、条項 1 から 6 のいずれかに記載の方法。

【0190】

8. 前記結合または比較が、偽陽性マッチを（おそらくランダムに）選択し、真陽性マッチおよび前記偽陽性マッチの選択分についてのみ前記スコアマップおよびリワードマップを結合または比較することによって、前記真陽性マッチの数と前記偽陽性マッチの数とを均衡させることを含み、

前記真陽性マッチとは、正しさのチェックに合格した関心点であり、前記偽陽性マッチとは、正しさのチェックに合格しなかった関心点である、条項 1 から 7 のいずれかに記載の方法。

【0191】

8. 前記結合または比較が、前記スコアマップと前記リワードマップの間の二乗差の合計を計算することを含む、条項 1 から 7 のいずれかに記載の方法。

【0192】

9. 特徴点検出のための分類器を訓練する装置であって、
制御ユニット、例えば、少なくとも 1 つのコンピュータプロセッサと、
メモリを備え、前記メモリは、前記制御ユニットに、
第 1 の画像と第 2 の画像を取得するステップと、
前記分類器を用いて、前記第 1 の画像に対する第 1 のスコアマップおよび前記第 2 の画像に対する第 2 のスコアマップを生成するステップと、
前記第 1 のスコアマップに基づいて、前記第 1 の画像内の第 1 の複数の関心点を選択するステップと、
前記第 2 のスコアマップに基づいて、前記第 2 の画像内の第 2 の複数の関心点を選択するステップと、
前記第 1 の複数の関心点のうちの 1 つの第 1 の関心点と、前記第 2 の複数の関心点のうちの 1 つの第 2 の関心点とのペアワイズマッチングを行うステップと、
前記第 1 の画像と前記第 2 の画像との間のグラウンドトゥールズ変換に基づいて、前記ペアワイズマッチの正しさをチェックして、リワードマップを生成するステップと、
前記スコアマップと前記リワードマップを結合または比較するステップと、
前記結合または比較の結果に基づいて、前記分類器を更新するステップと、
を実行させる命令を含む、装置。

【0193】

10. 第 1 の画像を第 2 の画像にレジストレーションする方法であって、
第 1 の画像と第 2 の画像を取得するステップと、
条項 1 から 9 のいずれかに記載の方法または装置によって生成された分類器を用いて、前記第 1 の画像に対する第 1 のスコアマップおよび前記第 2 の画像に対する第 2 のスコアマップを生成するステップと、
前記第 1 のスコアマップに基づいて、前記第 1 の画像内の第 1 の複数の関心点を選択するステップと、
前記第 2 のスコアマップに基づいて、前記第 2 の画像内の第 2 の複数の関心点を選択するステップと、
前記第 1 の複数の関心点のうちの 1 つの第 1 の関心点と、前記第 2 の複数の関心点のうちの 1 つの第 2 の関心点とのペアワイズマッチングを行うステップと、を含む方法。

【0194】

11. 複数の関心点を選択するステップは、前記画像内の任意の点から最近傍関心点までの距離に最大制限を課すことを含む、条項 10 に記載の方法。

【0195】

12. 前記ペアワイズマッチングは、前記第1の関心点において前記第1の画像で検出された特徴と、前記第2の関心点において前記第2の画像で検出された特徴との間の類似性に基づいて実行される、条項10または11に記載の方法。

【0196】

13. 前記マッチングは、前記第1の関心点を、複数の第2の関心点のうちの1つの第2の関心点にマッチングすることによって第1の方向に実行され、前記第2の関心点は、前記第1の関心点における特徴に最も類似した特徴を有する、条項10から12のいずれかに記載の方法。

【0197】

14. 前記マッチングはさらに、前記第2の関心点を、複数の第1の関心点のうちの1つの第1の関心点にマッチングすることによって第2の方向に実行され、前記第1の関心点は、前記第2の関心点における特徴に最も類似した特徴を有する、条項13に記載の方法。

【0198】

15. 第1の画像を第2の画像にレジストレーションする装置であって、
制御ユニット、例えば、少なくとも1つのコンピュータプロセッサと、
メモリを備え、前記メモリは、前記制御ユニットに、
第1の画像と第2の画像を取得するステップと、
条項1から10のいずれかに記載の方法または装置によって生成された分類器を用いて、
前記第1の画像に対する第1のスコアマップおよび前記第2の画像に対する第2のスコアマップを生成するステップと、

前記第1のスコアマップに基づいて、前記第1の画像内の第1の複数の関心点を選択するステップと、

前記第2のスコアマップに基づいて、前記第2の画像内の第2の複数の関心点を選択するステップと、

前記第1の複数の関心点のうちの1つの第1の関心点と、前記第2の複数の関心点のうちの1つの第2の関心点とのペアワイズマッチングを行うステップと、を実行させる命令を含む、装置。

【0199】

参照

[1] OpenCV: cv::BFMatcher Class Reference

[2] OpenCV: cv::xfeatures2d::SIFT Class Reference

[3] 「Improving Accuracy and Efficiency of Mutual Information for Multi-modal Retinal Image Registration using Adaptive Probability Density Estimation」 Computerized Medical Imaging and Graphics, 37 (7-8): 597-606, 2013年

[4] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Man' e, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Vi'egas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. YuおよびX. Zheng著 「Tensor-Flow: Large-scale machine learning on heterogeneous syst

10

20

30

40

50

ems、2015年 tensorflow.orgから入手可能なソフトウェア

[5] A. Alahi、R. OrtizおよびP. Vandergheynst著 「FREAK: Fast Retina Keypoint」 Conference on Computer Vision and Pattern Recognition 予稿集 510 - 517ページ、2012年

[6] P. F. Alcantarilla、A. BartoliおよびA. J. Davison著 「KAZE features」 Computer Science 講義資料、7577 LNCS (6) : 214 - 227ページ、2012年

[7] J. Aldana-Iuit、D. Mishkin、O. ChumおよびJ. Matas著 「In the saddle: Chasing fast and repeatable features」 23rd International Conference on Pattern Recognition、675 - 680ページ、2016年

[8] R. ArandjeloviおよびA. Zisserman著 「Three things everyone should know to improve object retrieval」

[9] V. Balntas、E. Johns、L. TangおよびK. Mikolajczyk著 「PN-Net: Conjoined Triple Deep Network for Learning Local Image Descriptors」 CoRR、abs/1601.05030、2016年

[10] V. Balntas、K. Lenc、A. VedaldiおよびK. Mikolajczyk著 「HPatches: A Benchmark and Evaluation of Handcrafted and Learned Local Descriptors」 Conference on Computer Vision and Pattern Recognition、3852 - 3861ページ、2017年

[11] H. Bay、T. TuytelaarsおよびL. Van Gool著 「SURF: Speeded up Robust Features」 Computer Science 講義資料、3951 LNCS: 404 - 417ページ、2006年

[12] P. C. Cattin、H. Bay、L. J. V. GoolおよびG. Sz'ekely著 「Retina Mosaicing Using Local Features」 Medical Image Computing and Computer-Assisted Intervention、185 - 192ページ、2006年

[13] J. Chen、J. Tian、N. Lee、J. Zheng、T. R. SmithおよびA. F. Laine著 「A Partial Intensity Invariant Feature Descriptor for Multimodal Retinal Image Registration」 IEEE Transactions on Biomedical Engineering、57 (7) : 1707 - 1718ページ、2010年

[14] A. V. Cideciyan著 「Registration of Ocular Fundus Images」 IEEE Engineering in Medicine and Biology Magazine、14 (1) : 52 - 58、1995年

[15] A. L. Dahl、H. AanaesおよびK. S. Pedersen著 「Finding the Best Feature Detector-Descriptor Combination」 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission、318 - 325ページ、2011年

[16] D. DeTone、T. MalisiewiczおよびA. Rabinovich著 「SuperPoint: Self-Supervised Interest Point Detection and Description」 IEEE Conference on Computer Vision and Pattern Reco

10

20

30

40

50

gnition Workshops、224 - 236ページ、2018年

[17] H. Durrant-WhyteおよびT. Bailey著 「Simultaneous Localisation and Mapping (SLAM) : Part I The Essential Algorithms Technical report

[18] P. FischerおよびT. Brox著 「Descriptor Matching with Convolutional Neural Networks: a Comparison to SIFT」1 - 10ページ

[19] M. A. FischlerおよびR. C. Bolles著 「Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography」Communications of the ACM、24 (6) : 381 - 395、1981年6月

10

[20] Y. Hang、X. Zhang、Y. Shao、H. WuおよびW. Sun著 「Retinal Image Registration Based on the Feature of Bifurcation Point」第10回International Congress on Image and Signal Processing、BioMedical Engineering and Informatics、CISPBMEI、1 - 6ページ、2017年

[21] C. G. HarrisおよびM. Stephens著 「A Combined Corner and Edge Detector」the Alvey Vision Conference予稿集、1 - 6ページ、1988年

20

[22] C. Hernandez-Matas、X. Zabulis、A. TriantafyllouおよびP. Anyfanti著 「FIRE: Fundus Image Registration dataset」Journal for Modeling in Ophthalmology、4 : 16 - 28、2017年

[23] J. Z. Huang、T. N. Tan、L. MaおよびY. H. Wang著 「Phase Correlation-based Iris Image Registration Model」Journal of Computer Science and Technology、20 (3) : 419 - 425ページ、2005年

30

[24] N. Jacobs、N. RomanおよびR. Pless著 「Consistent Temporal Variations in Many Outdoor Scenes」the IEEE Conference on Computer Vision and Pattern Recognition予稿集、2007年

[25] D. P. Kingma及びJ. Ba著 「Adam: A Method for Stochastic Optimization」CoRR、abs/1412.6980、2014年

[26] S. Leutenegger、M. ChliおよびR. Y. Siegwart著 BRISK: Binary Robust Invariant Scalable Keypoints」the IEEE International Conference on Computer Vision予稿集、2548 - 2555ページ、2011年

40

[27] P. Li、Q. Chen、W. FanおよびS. Yuan著 「Registration of OCT Fundus Images with Color Fundus Images Based on Invariant Features」In Cloud Computing and Security - Third International Conference、471 - 482ページ、2017年

[28] D. G. Lowe著 「Distinctive Image Features from Scale Invariant keypoints」International Journal of Computer Vision、60ページ 2004年

[29] K. MikolajczykおよびC. Schmid著 「A Performa

50

nce Evaluation of Local Descriptors」 IEEE Transactions on Pattern Analysis and Machine Intelligence、27 (10) : 1615 - 1630ページ、2005年

[30] K. Mikolajczyk、T. Tuytelaars、C. Schmid、A. Zisserman、J. Matas、F. Schaffalitzky、T. KadirおよびL. Van Gool著 「A Comparison of Affine Region Detectors」 International Journal of Computer Vision

[31] Y. Ono、E. Trulls、P. FuaおよびK. M. Yi著 「LF-NEIT : Learning Local Features from Images」 In Advances in Neural Information Processing Systems、6237 - 6247ページ、2018年

[32] J. P. W. Pluim、J. B. A. MaintzおよびM. A. Viergever著 「Mutual Information Based Registration of Medical Images: A Survey」 IEEE Transactions on Medical Imaging、22 (8) : 986 - 1004ページ、2003年

[33] R. Ramli、M. Yamani、I. Idris、K. Hasikin、N. K. A. Karim、A. Wahid、A. Wahab、I. Ahmady、F. Ahmady、N. A. KadriおよびH. Arof著 「Feature-Based Retinal Image Registration Using D-Saddle Feature」 2017年

[34] O. Ronneberger、P. FischerおよびT. Brox著 「U-Net: Convolutional Networks for Biomedical Image Segmentation」 Medical Image Computing and Computer-Assisted Intervention 掲載、234 - 241ページ、2015年

[35] E. Rublee、V. Rabaud、K. KonoligeおよびG. Bradski著 「ORB: An Efficient Alternative to SIFT or SURF」 the IEEE International Conference on Computer Vision 予稿集、2564 - 2571ページ、2011年

[36] C. Sanchez-galeana、C. Bowd、E. Z. Blumenthal、P. A. Gokhale、L. M. ZangwillおよびR. N. Weinreb著 「Using Optical Imaging Summary Data to Detect Glaucoma」。 Ophthalmology、1812 - 1818ページ、2001年

[37] E. Simo-Serra、E. Trulls、L. Ferraz、I. Kokkinos、P. FuaおよびF. Moreno-Noguer著 「Discriminative Learning of Deep Convolutional Feature Point Descriptors」 IEEE International Conference on Computer Vision、118 - 126ページ、2015. 4

[38] B. Triggs、P. F. McLauchlan、R. I. HartleyおよびA. W. Fitzgibbon著 「Bundle Adjustment - A Modern Synthesis」 Technical report

[39] P. Truong、S. De ZanetおよびS. Apostolopoulos著 「Comparison of Feature Detectors for Retinal Image Alignment」 ARVO、2019年

10

20

30

40

50

[40] Y. Verdie、K. M. Yi、P. FuaおよびV. Lepetit著 「TILDE: A Temporally Invariant Learned Detector」 the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 予稿集 5279 - 5288ページ、2015年

[41] G. Wang、Z. Wang、Y. ChenおよびW. Zhao著 「Biomedical Signal Processing and Control Robust Point Matching Method for Multimodal Retinal Image Registration」 Biomedical Signal Processing and Control、19:68 - 76ページ、2015年

10

[42] S. A. J. WinderおよびM. A. Brown著 「Learning Local Image Descriptors」 IEEE Conference on Computer Vision and Pattern Recognition、2007年

[43] S. A. J. Winder、G. HuaおよびM. A. Brown著 「Picking the Best DAISY」 IEEE Conference on Computer Vision and Pattern Recognition、178 - 185ページ、2009年

[44] K. M. Yi、E. Trulls、V. LepetitおよびP. Fua著 「LIFT: Learned Invariant Feature Transform」 於European Conference on Computer Vision - ECCV、467 - 483ページ、2016年

20

[45] K. M. Yi、Y. Verdie、P. FuaおよびV. Lepetit著 「Learning to Assign Orientations to Feature Points」 於IEEE Conference on Computer Vision and Pattern Recognition、107 - 116ページ、2016年

[46] S. D. Zanet、T. Rudolph、R. Richa、C. TappeinerおよびR. Sznitman著 「Retinal slit lamp video mosaicking」 International Journal of Computer Assisted Radiology and Surgery、11(6):1035 - 1041ページ、2016年

30

[47] L. Zhou、M. S. Rzeszutarski、L. J. SingermanおよびJ. M. Chokreff著 「The Detection and Quantification of Retinopathy Using Digital Angiograms」 IEEE Transactions on Medical Imaging、13(4):619 - 626ページ、1994年

[48] C. L. ZitnickおよびK. Ramnath著 「Edge Foci Interest Points」 於International Conference on Computer Vision、359 - 366ページ、2011年

40

【符号の説明】

【0200】

701、901 システム

702、902 カメラ

703、903 ディスプレイ

704、904 通信ユニット

705、905 制御ユニット

706、906 メモリ

707、907 データ

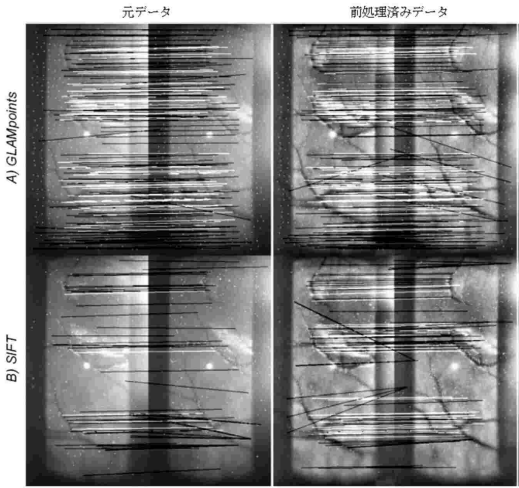
50

- 708、908 モデル
- 709、909 画像
- 711 訓練モジュール
- 712、912 関心点セクタ
- 713、913 スコアマップ生成器
- 714、914 マッチングモジュール
- 715 検証モジュール
- 716 結合モジュール
- 717 前処理モジュール
- 910 プログラムコード
- 917 取得モジュール
- 918 レジストレーションモジュール

【図面】

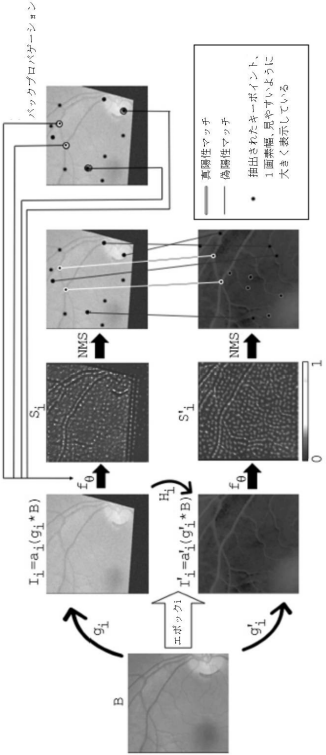
【図1】

Fig. 1



【図2A】

Fig. 2A



10

20

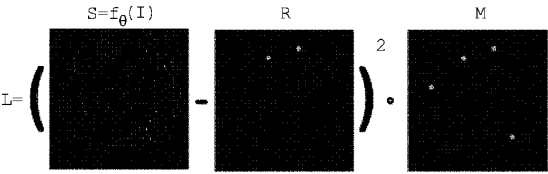
30

40

50

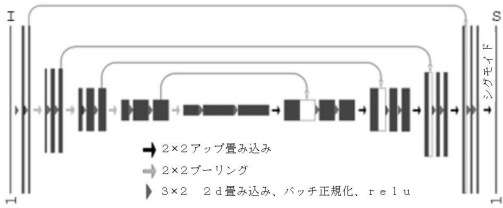
【図 2 B】

Fig. 2B



【図 2 C】

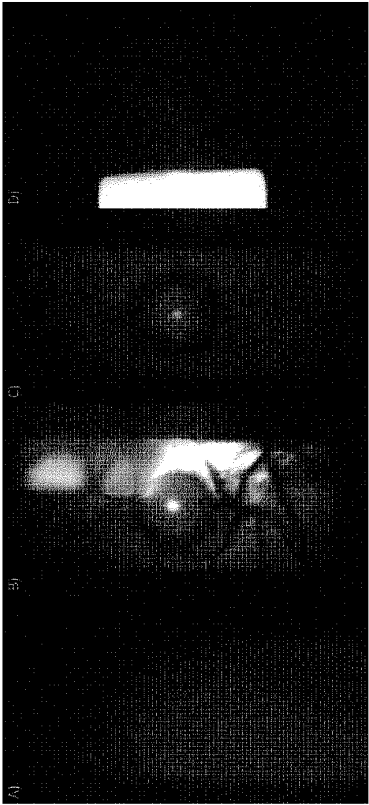
Fig. 2C



10

【図 3】

Fig. 3



【図 4 A】

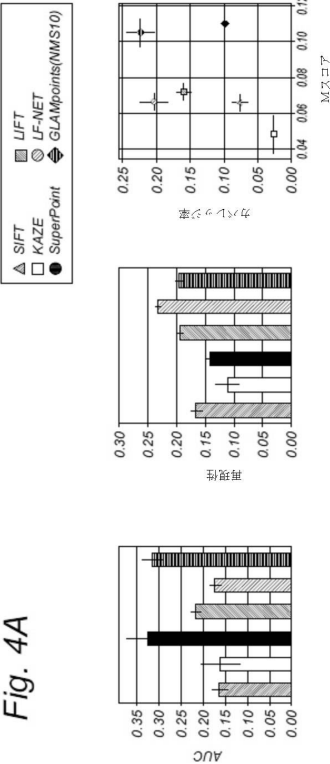


Fig. 4A

20

30

40

50

【図 4 B】

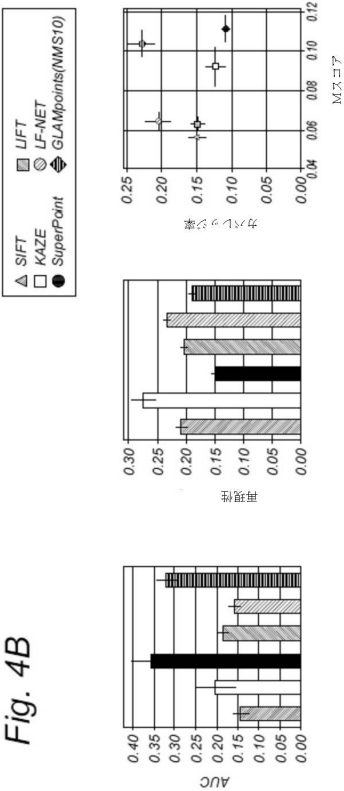
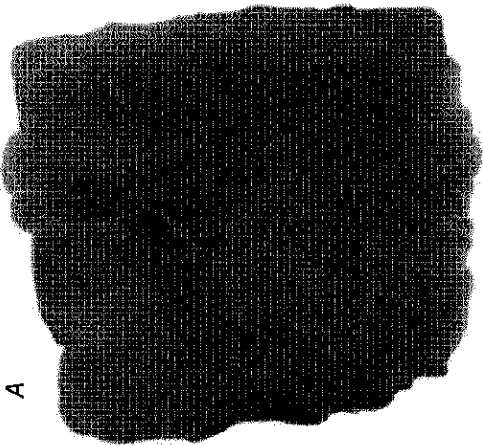


Fig. 4B

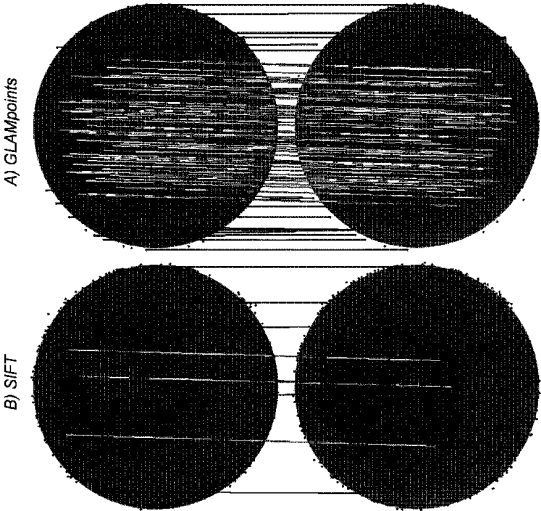
【図 6 A】



A

【図 5】

Fig. 5



A) GLAMpoints

B) SIFT

【図 6 B】



B

10

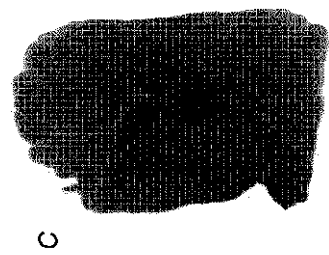
20

30

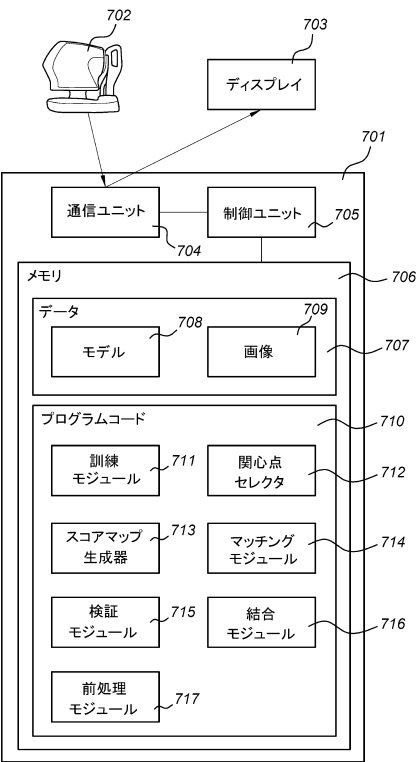
40

50

【図 6 C】



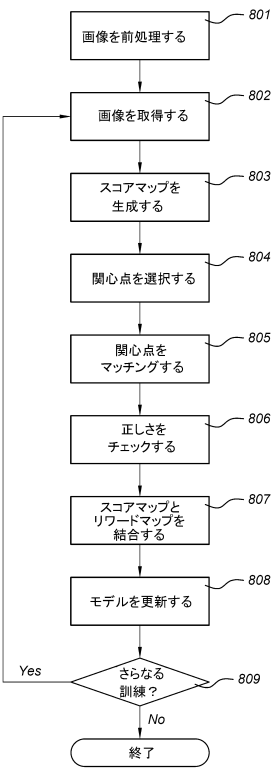
【図 7】
Fig. 7



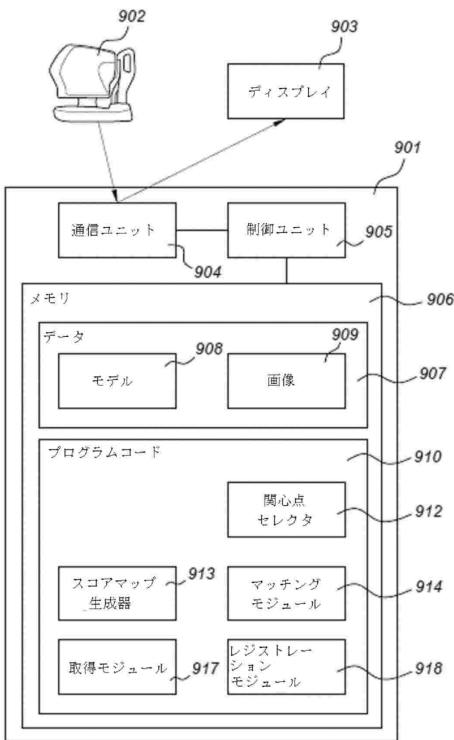
10

20

【図 8】
Fig. 8



【図 9】
Fig. 9



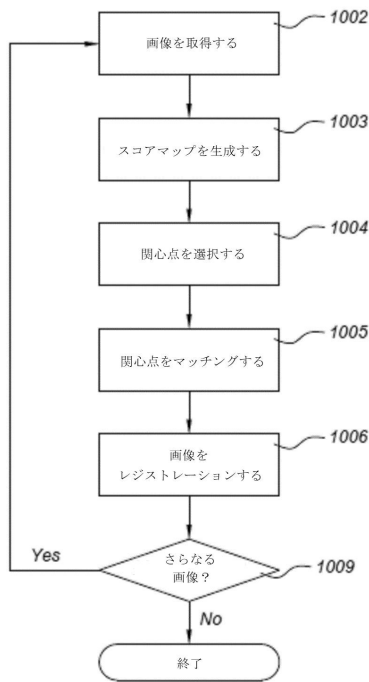
30

40

50

【図 10】

Fig. 10



10

20

30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関
欧州特許庁(EP)
イブルクシュトラーク

審査官 佐藤 実
(56)参考文献 Yuki Ono et al., LF-Net: Learning Local Features from Images, arxiv.org, 2018年11月22
日, <https://arxiv.org/abs/1805.09662>

(58)調査した分野 (Int.Cl., D B名)
G 0 6 T 7 / 0 0