

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号
特表2024-534302
(P2024-534302A)

(43)公表日 令和6年9月20日(2024.9.20)

(51)国際特許分類
G 0 6 N 3/0464(2023.01)

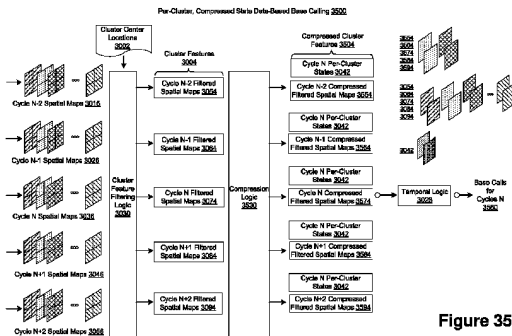
F I
G 0 6 N 3/0464

審査請求 未請求 予備審査請求 未請求 (全137頁)			
(21)出願番号	特願2023-579830(P2023-579830)	(71)出願人	500358711 イルミナ インコーポレイテッド アメリカ合衆国 カリフォルニア州 9 2 1 2 2 サンディエゴ イルミナ ウェイ 5 2 0 0
(86)(22)出願日	令和4年9月21日(2022.9.21)	(74)代理人	100108453 弁理士 村山 靖彦
(85)翻訳文提出日	令和6年2月27日(2024.2.27)	(74)代理人	100110364 弁理士 実広 信哉
(86)国際出願番号	PCT/US2022/044293	(74)代理人	100133400 弁理士 阿部 達彦
(87)国際公開番号	WO2023/049215	(72)発明者	ギャヴィン・デレク・パーナビー アメリカ合衆国・カリフォルニア・9 2 1 2 2・サン・ディエゴ・イルミナ・ウ エイ・5 2 0 0
(87)国際公開日	令和5年3月30日(2023.3.30)		
(31)優先権主張番号	63/247,296		
(32)優先日	令和3年9月22日(2021.9.22)		
(33)優先権主張国・地域又は機関	米国(US)		
(31)優先権主張番号	63/247,301		
(32)優先日	令和3年9月22日(2021.9.22)		
(33)優先権主張国・地域又は機関	米国(US)		
(31)優先権主張番号	17/944,809		
	最終頁に続く		最終頁に続く

(54)【発明の名称】 圧縮された状態ベースのベースコール

(57)【要約】

本技術は、システムを開示し、含む。本システムは、クラスターの配列決定画像を処理し、空間畳み込み特徴を生成するように構成された空間畳み込みニューラルネットワークと、空間畳み込み特徴から、クラスターの中心を含む空間畳み込み特徴のサブセットを選択するように構成されたフィルタリング論理と、空間畳み込み特徴のサブセットを圧縮特徴のセットに圧縮するように構成された圧縮論理と、圧縮特徴のセット内の圧縮特徴についての状態情報にアクセスするように構成されたコンテキスト化論理と、ステートフルな圧縮特徴のセットを処理し、時間畳み込みステートフル特徴を生成するように構成された時間畳み込みニューラルネットワークと、時間畳み込みステートフル特徴に基づいてクラスターについてのベースコールを生成するように構成されたベースコール論理と、を備える。



【特許請求の範囲】

【請求項 1】

システムであって、

クラスターの配列決定画像を処理し、空間畳み込み特徴を生成するように構成された空間畳み込みニューラルネットワークと、

前記空間畳み込み特徴から、前記クラスターの中心を含む空間畳み込み特徴のサブセットを選択するように構成されたフィルタリング論理と、

空間畳み込み特徴の前記サブセットを圧縮特徴のセットに圧縮するように構成された圧縮論理であって、空間畳み込み特徴の前記サブセットが M 個のチャンネルを有し、圧縮特徴の前記セットが N 個のチャンネルを有し、 $M > N$ である、圧縮論理と、

圧縮特徴の前記セット内の圧縮特徴についての状態情報にアクセスし、前記状態情報を前記圧縮特徴に付加してステートフルな圧縮特徴のセットを生成するように構成されたコンテキスト化論理と、

ステートフルな圧縮特徴の前記セットを処理し、時間畳み込みステートフル特徴を生成するように構成された時間畳み込みニューラルネットワークと、

前記時間畳み込みステートフル特徴に基づいて前記クラスターについてのベースコールを生成するように構成されたベースコール論理と、を備える、システム。

【請求項 2】

チャンネルごとの状態が、前記配列決定画像を独立して処理することに対応して線形モデルによってクラスターごとに生成される強度変動補正係数である、請求項 1 に記載のシステム。

【請求項 3】

前記強度変動補正係数が、前記クラスターの強度プロファイル間のスケール変動を補償する増幅係数を含む、請求項 2 に記載のシステム。

【請求項 4】

前記強度変動補正係数が、複数のチャンネルに沿った前記クラスターの前記強度プロファイル間のシフト変動を補償するチャンネル固有オフセット係数を含む、請求項 2 又は 3 に記載のシステム。

【請求項 5】

前記強度変動補正係数が、前記シフト変動を補償する共通オフセット係数を含む、請求項 2 ～ 4 のいずれか一項に記載のシステム。

【請求項 6】

前記強度変動補正係数が、配列決定ランの現在の配列決定サイクルにおいて決定された現在の強度統計値の分析と、前記現在の配列決定サイクルに先行する前記配列決定ランの 1 つ以上の以前の配列決定サイクルにおいて決定された過去の強度統計値の分析とを組み合わせることに基づいて、クラスターごとに生成される、請求項 2 ～ 5 のいずれか一項に記載のシステム。

【請求項 7】

前記クラスターの前記強度プロファイルが、前記複数のチャンネル内のチャンネルに対するチャンネル固有強度値によって特徴付けられる、請求項 4 ～ 6 のいずれか一項に記載のシステム。

【請求項 8】

前記チャンネルごとの状態が、前記チャンネルに関するチャンネル固有状態値によって特徴付けられる、請求項 1 ～ 7 のいずれか一項に記載のシステム。

【請求項 9】

前記チャンネルごとの状態が、クラスターごとのチャンネル固有ビットを用いてビットごとに符号化される、請求項 1 ～ 8 のいずれか一項に記載のシステム。

【請求項 10】

前記チャンネルのサブセットに関するチャンネル固有状態値が、前記クラスターごとのチャンネル固有ビットを用いてビットごとに符号化される、請求項 9 に記載のシステム。

10

20

30

40

50

【請求項 1 1】

前記チャンネル固有状態値が、前記チャンネルにわたって平均化されて、パンチャネル状態値を生成し、前記パンチャネル状態値が、前記クラスターごとのチャンネル固有ビットを用いてビットごとに符号化される、請求項 9 又は 1 0 に記載のシステム。

【請求項 1 2】

前記チャンネルごとの状態が、前記クラスターごとのチャンネル固有ビットとビットごとに連結される、請求項 9 ～ 1 1 のいずれか一項に記載のシステム。

【請求項 1 3】

前記チャンネルごとの状態が、前記クラスターごとのチャンネル固有ビットとビットごとに合計される、請求項 9 ～ 1 2 のいずれか一項に記載のシステム。

10

【請求項 1 4】

前記チャンネルごとの状態が、前記現在の配列決定サイクル及び前記以前の配列決定サイクルにおいて空間処理論理によって生成されたクラスターごとのチャンネル固有ビットの圧縮値に基づく、請求項 1 ～ 1 3 のいずれか一項に記載のシステム。

【請求項 1 5】

前記チャンネルごとの状態が、前記圧縮値の平均である、請求項 1 4 に記載のシステム。

【請求項 1 6】

前記チャンネルごとの状態が、前記圧縮値のうちの最大のものである、請求項 1 4 に記載のシステム。

【請求項 1 7】

前記チャンネルごとの状態が、前記圧縮値のうちの最小のものである、請求項 1 4 に記載のシステム。

20

【請求項 1 8】

前記チャンネルごとの状態が、前記圧縮値の指数加重平均である、請求項 1 4 に記載のシステム。

【請求項 1 9】

前記指数加重平均が、前の配列決定サイクルよりも最近の配列決定サイクルを加重することに基づいて決定される、請求項 1 8 に記載のシステム。

【請求項 2 0】

前記チャンネルごとの状態が、前記圧縮値の移動平均である、請求項 1 4 又は 1 5 に記載のシステム。

30

【請求項 2 1】

前記移動平均が、以前の配列決定サイクルからの圧縮値のローリングサブセットを使用する、請求項 2 0 に記載のシステム。

【請求項 2 2】

前記チャンネルごとの状態が、アクティブ状態値及び非アクティブ状態値を含む、請求項 1 4 ～ 2 1 のいずれか一項に記載のシステム。

【請求項 2 3】

前記アクティブ状態値及び前記非アクティブ状態値が、チャンネル固有である、請求項 2 2 に記載のシステム。

40

【請求項 2 4】

前記アクティブ状態値及び前記非アクティブ状態値が、先行するベースコールから決定される、請求項 2 2 又は 2 3 に記載のシステム。

【請求項 2 5】

前記アクティブ状態値及び前記非アクティブ状態値が、前記圧縮値の大域的最大値及び大域的最小値に基づいて決定される、請求項 2 2 ～ 2 4 のいずれか一項に記載のシステム。

【請求項 2 6】

前記アクティブ状態値及び前記非アクティブ状態値が、前記圧縮値の指数加重平均に基づいて決定される、請求項 2 2 ～ 2 5 のいずれか一項に記載のシステム。

50

【請求項 27】

前記空間畳み込みニューラルネットワークへの入力、現在の配列決定サイクル、1つ以上の以前の配列決定サイクル、及び1つ以上の次の配列決定サイクルについての配列決定画像のスライディングウィンドウを含む、請求項1～26のいずれか一項に記載のシステム。

【請求項 28】

特徴セットが、前記現在の配列決定サイクル、前記以前の配列決定サイクル、及び前記次の隣接配列決定サイクルについての配列決定画像の前記スライディングウィンドウ内の配列決定画像のチャンネル固有ピクセルから抽出され、クラスターごとの中心の特徴セットが、前記現在の配列決定サイクル、前記以前の配列決定サイクル、及び前記次の配列決定サイクルについての前記特徴セットから選別される、請求項27に記載のシステム。

10

【請求項 29】

チャンネル固有ビットセットが、前記現在の配列決定サイクル、前記以前の配列決定サイクル、及び前記次の配列決定サイクルについての前記クラスターごとの中心の特徴セットから圧縮される、請求項28に記載のシステム。

【請求項 30】

システムであって、

クラスターの強度プロファイルが配列決定画像のチャンネル固有ピクセルにわたって分散される入力処理し、前記クラスターの前記強度プロファイルが前記チャンネル固有ピクセルから抽出された特徴に集約される空間的にコンパクトな出力を生成するように構成された空間処理論理と、

20

前記クラスターの中心で検出されたピーク強度を特徴付ける前記特徴から、前記クラスターごとの中心の特徴を選別するように構成されたクラスター集束論理と、

前記クラスターごとの中心の特徴をクラスターごとのチャンネル固有ビットに蒸留するように構成された圧縮論理であって、前記クラスターごとの中心の特徴がM個の特徴チャンネルを有し、前記クラスターごとのチャンネル固有ビットがN個のビットチャンネルを有し、 $M > N$ である、圧縮論理と、

前記クラスターごとのチャンネル固有ビットについてチャンネルごとの状態を生成するように構成された状態生成論理と、

前記クラスターごとのチャンネル固有ビット及び前記チャンネルごとの状態を処理し、時間的にコンパクトな出力を生成するように構成された時間的処理論理と、

30

前記時間的にコンパクトな出力に基づいてベースコールを生成するように構成されたベースコール論理と、を備える、システム。

【発明の詳細な説明】

【技術分野】

【0001】

(優先権出願)

本出願は、2021年9月22日に出願された米国特許仮出願第63/247,301号(代理人整理番号ILLM 1043-2/IP-2208-PRV)の利益を主張する、2022年9月14日に出願された「COMPRESSED STATE-BASED BASE CALLING」と題する米国特許非仮出願第17/944,948号(代理人整理番号ILLM 1043-4/IP-2208-US)の優先権及びその利益を主張する。

40

【0002】

本出願は、2021年9月22日に出願された米国特許仮出願第63/247,296号(代理人整理番号ILLM 1043-1/IP-2073-PRV)の利益を主張する、2022年9月14日に出願された「STATE-BASED BASE CALLING」と題する米国特許非仮出願第17/944,809号(代理人整理番号ILLM 1043-3/IP-2073-US)の優先権及び利益を主張する。

【0003】

50

優先権出願は、本明細書に完全に記載されているかのように、全ての目的のために参照により本明細書に組み込まれる。

【0004】

(発明の分野)

開示される技術は、人工知能型コンピュータ及びデジタルデータ処理システム、並びに知能（すなわち、知識ベースのシステム、推論システム、及び知識取得システム）を模倣するための対応するデータ処理方法及び製品に関し、不確実性を伴う推論のためのシステム（例えば、ファジー論理システム）、適応システム、機械学習システム、及び人工ニューラルネットワークを含む。具体的には、開示される技術は、データを分析するための深層畳み込みニューラルネットワークなどの深層ニューラルネットワークを使用することに関する。

10

【0005】

(組み込み)

以下は、本明細書に完全に記載されているかのように、全ての目的のために参照により組み込まれる。

【0006】

2021年5月4日出願された「EQUALIZATION-BASED IMAGE PROCESSING AND SPATIAL CROSSTALK ATTENUATOR」と題する米国特許非仮出願第17/308,035号（代理人整理番号ILLM 1032-2/IP-1991-US）、

20

2020年10月27日出願された「SYSTEMS AND METHODS FOR PER-CLUSTER INTENSITY CORRECTION AND BASE CALLING」と題する米国特許非仮出願第63/106,256号（代理人整理番号ILLM 1034-1/IP-2026-PRV）、

2018年3月1日出願された「OPTICAL DISTORTION CORRECTION FOR IMAGED SAMPLES」と題する米国特許非仮出願第15/909,437号、

2020年3月20日出願された「TRAINING DATA GENERATION FOR ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する米国特許非仮出願第16/825,987号（代理人整理番号ILLM 1008-16/IP-1693-US）、

30

2020年3月20日出願された「ARTIFICIAL INTELLIGENCE-BASED GENERATION OF SEQUENCING METADATA」と題する米国特許非仮出願第16/825,991号（代理人整理番号ILLM 1008-17/IP-1741-US）、

2020年3月20日出願された「ARTIFICIAL INTELLIGENCE-BASED BASE CALLING」と題する米国特許非仮出願第16/826,126号（代理人整理番号ILLM 1008-18/IP-1744-US）、

2020年3月20日出願された「ARTIFICIAL INTELLIGENCE-BASED QUALITY SCORING」と題する米国特許非仮出願第16/826,134号（代理人整理番号ILLM 1008-19/IP-1747-US）、

40

2020年3月21日出願された「ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する米国特許非仮出願第16/826,168号（代理人整理番号ILLM 1008-20/IP-1752-US）、

2021年2月12日出願された「ARTIFICIAL INTELLIGENCE-BASED BASE CALLING OF INDEX SEQUENCES」と題する米国特許非仮出願第17/175,546号（代理人整理番号ILLM 1015-2/IP-1857-US）、

2021年2月19日出願された「ARTIFICIAL INTELLIGENCE-BASED MANY-TO-MANY BASE CALLING」と題する米国

50

特許非仮出願第17/180,542号(代理人整理番号ILLM 1016-2/IP-1858-US)、

2021年2月15日に出願された「KNOWLEDGE DISTILLATION-BASED COMPRESSION OF ARTIFICIAL INTELLIGENCE-BASED BASE CALLER」と題する米国特許非仮出願第17/176,151号(代理人整理番号ILLM 1017-2/IP-1859-US)、

2020年8月28日に出願された「DETECTING AND FILTERING CLUSTERS BASED ON ARTIFICIAL INTELLIGENCE-PREDICTED BASE CALLS」と題する米国特許仮出願第63/072,032号(代理人整理番号ILLM1018-1/IP-1860-PRV)、

2021年3月16日に出願された「TILE LOCATION AND/OR CYCLE BASED WEIGHT SET SELECTION FOR BASE CALLING」と題する米国特許仮出願第63/161,880号(代理人整理番号ILLM 1019-1/IP-1861-PRV)、

2021年3月16日に出願された「NEURAL NETWORK PARAMETER QUANTIZATION FOR BASE CALLING」と題する米国特許仮出願第63/161,896号(代理人整理番号ILLM 1019-2/IP-2049-PRV)、

2021年2月15日に出願された「HARDWARE EXECUTION AND ACCELERATION OF ARTIFICIAL INTELLIGENCE-BASED BASE CALLER」と題する米国特許非仮出願第17/176,147号(代理人整理番号ILLM 1020-2/IP-1866-US)、

2021年8月3日に出願された「BASE CALLING USING MULTIPLE BASE CALLER MODELS」と題する米国特許仮出願第63/228,954号(代理人整理番号ILLM 1021-1/IP-1856-PRV)、

2021年2月18日に出願された「DATA COMPRESSION FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLING」と題する米国特許非仮出願第17/179,395号(代理人整理番号ILLM 1029-2/IP-1964-US)、

2021年2月19日に出願された「SPLIT ARCHITECTURE FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLER」と題する米国特許非仮出願第17/180,480号(代理人整理番号ILLM 1030-2/IP-1982-US)、

2021年2月19日に出願された「BUS NETWORK FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLER」と題する米国特許非仮出願第17/180,513号(代理人整理番号ILLM 1031-2/IP-1965-US)、

2021年3月31日に出願された「ARTIFICIAL INTELLIGENCE-BASED BASE CALLER WITH CONTEXTUAL AWARENESS」と題する米国特許仮出願第63/169,163号(代理人整理番号ILLM 1033-1/IP-2007-PRV)、

2021年6月29日に出願された「SELF-LEARNED BASE CALLER, TRAINED USING OLIGO SEQUENCES」と題する米国特許仮出願第63/216,419号(代理人整理番号ILLM 1038-1/IP-2050-PRV)、

2021年6月29日に出願された「SELF-LEARNED BASE CALLER, TRAINED USING ORGANISM SEQUENCE」と題する米国特許仮出願第63/216,404号(代理人整理番号ILLM 1038-2/IP-2094-PRV)、

2021年7月19日に出願された「SPECIALIST SIGNAL PROF

10

20

30

40

50

ILERS FOR BASE CALLING」と題する米国特許仮出願第63／223,408号（代理人整理番号ILLM 1041-1／IP-2063-PRV）、

2021年7月28日に出願された「QUALITY SCORE CALIBRATION OF BASECALLING SYSTEMS」と題する米国特許仮出願第63／226,707号（代理人整理番号ILLM 1045-1／IP-2093-PRV）、

2021年7月1日に出願された「EFFICIENT ARTIFICIAL INTELLIGENCE-BASED BASE CALLING OF INDEX SEQUENCES」と題する米国特許仮出願第63／217,644号（代理人整理番号ILLM 1046-1／IP-2135-PRV）、

2014年10月31日に出願された「IMAGE ANALYSIS USEFUL FOR PATTERNED OBJECTS」と題する米国特許非仮出願第14／530,299号、

2014年12月3日に出願された「METHODS AND SYSTEMS FOR ANALYZING IMAGE DATA」と題する米国特許非仮出願第15／153,953号、

2018年1月5日に出願された「PHASING CORRECTION」と題する米国特許非仮出願第15／863,241号、

2013年9月6日に出願された「CENTROID MARKERS FOR IMAGE ANALYSIS OF HIGH DENSITY CLUSTERS IN COMPLEX POLYNUCLEOTIDE SEQUENCING」と題する米国特許非仮出願第14／020,570号、

2009年9月23日に出願された「METHOD AND SYSTEM FOR DETERMINING THE ACCURACY OF DNA BASE IDENTIFICATIONS」と題する米国特許非仮出願第12／565,341号、

2007年3月30日に出願された「SYSTEMS AND DEVICES FOR SEQUENCE BY SYNTHESIS ANALYSIS」と題する米国特許非仮出願第12／295,337号、

2008年1月28日に出願された「IMAGE DATA EFFICIENT GENETIC SEQUENCING METHOD AND SYSTEM」と題する米国特許非仮出願第12／020,739号、

2013年3月15日に出願された「BIOSENSORS FOR BIOLOGICAL OR CHEMICAL ANALYSIS AND SYSTEMS AND METHODS FOR SAME」と題する米国特許非仮出願第13／833,619号（代理人整理番号IP-0626-US）、

2016年6月7日に出願された「BIOSENSORS FOR BIOLOGICAL OR CHEMICAL ANALYSIS AND METHODS OF MANUFACTURING THE SAME」と題する米国特許非仮出願第15／175,489号（代理人整理番号IP-0689-US）、

2013年4月26日に出願された「MICRODEVICES AND BIOSENSOR CARTRIDGES FOR BIOLOGICAL OR CHEMICAL ANALYSIS AND SYSTEMS AND METHODS FOR THE SAME」と題する米国特許非仮出願第13／882,088号（代理人整理番号IP-0462-US）、

2012年9月21日に出願された「METHODS AND COMPOSITIONS FOR NUCLEIC ACID SEQUENCING」と題する米国特許非仮出願第13／624,200号（代理人整理番号IP-0538-US）、

2011年1月13日に出願された「DATA PROCESSING SYSTEM AND METHODS」と題する米国特許非仮出願第13／006,206号、

2018年3月26日に出願された「DETECTION APPARATUS HA

10

20

30

40

50

VING A MICROFLUOROMETER, A FLUIDIC SYSTEM, AND A FLOW CELL LATCH CLAMP MODULE」と題する米国特許非仮出願第15/936,365号、

2019年9月11日に出願された「FLOW CELLS AND METHODS RELATED TO SAME」と題する米国特許非仮出願第16/567,224号、

2019年6月12日に出願された「DEVICE FOR LUMINESCENT IMAGING」と題する米国特許非仮出願第16/439,635号、

2017年5月12日に出願された「INTEGRATED OPTOELECTRONIC READ HEAD AND FLUIDIC CARTRIDGE USEFUL FOR NUCLEIC ACID SEQUENCING」と題する米国特許非仮出願第15/594,413号、

2019年3月12日に出願された「ILLUMINATION FOR FLUORESCENCE IMAGING USING OBJECTIVE LENS」と題する米国特許非仮出願第16/351,193号、

2009年12月15日に出願された「DYNAMIC AUTOFOCUS METHOD AND SYSTEM FOR ASSAY IMAGER」と題する米国特許非仮出願第12/638,770号、及び

2013年3月1日に出願された「KINETIC EXCLUSION AMPLIFICATION OF NUCLEIC ACID LIBRARIES」と題する米国特許非仮出願第13/783,043号。

【背景技術】

【0007】

本セクションで考察される主題は、単に本セクションにおける言及の結果として、先行技術であると想定されるべきではない。同様に、本セクションで言及した問題、又は背景として提供された主題と関連付けられた問題は、先行技術において以前に認識されていると想定されるべきではない。本セクションの主題は、単に異なるアプローチを表し、それ自体はまた、特許請求される技術の実装形態に対応し得る。

【0008】

計算能力の急速な改善により、近年、多くのコンピュータビジョンタスクにおいて、深層畳み込みニューラルネットワーク (convolution neural network、CNN) が、著しく改善された精度で、大成功を収めることが可能となった。推論段階の間、多くのアプリケーションは、厳密な電力消費要件を伴う、1つの画像の低遅延の処理を必要とし、これにより、グラフィックス処理ユニット (graphics processing unit、GPU) 及び他の汎用プラットフォームの効率が低下し、そのことは、特定のアクセレーションハードウェア、例えば、フィールドプログラマブルゲートアレイ (Field Programmable Gate Array、FPGA) にとっては、深層学習アルゴリズムの推論に特に効果的となるようにデジタル回路をカスタマイズすることによって、好機をもたらすこととなる。しかしながら、携帯型及び埋め込み式システムにCNNを配備することは、大きいデータ量、集中的な計算、様々なアルゴリズム構造、及び頻繁なメモリアクセスのために依然として困難である。

【0009】

畳み込みが、CNNにおけるほとんどの演算を提供するので、畳み込みアクセレーションスキームが、ハードウェアCNNアクセラレータの効率及び性能に大きく影響することになる。畳み込みは、カーネル及び特徴マップに沿ってスライドする4つのレベルのループを伴う、積和 (multiply and accumulate、MAC) 演算を含む。第1のループレベルは、1つのカーネルウィンドウ内のピクセルのMACを計算する。第2のループレベルは、様々な異なる入力特徴マップにわたるMACの積の和を累積する。第1及び第2のループレベルを完了した後、バイアスを追加することにより、最終出力ピクセルが得られる。第3のループレベルは、入力特徴マップ内で、カーネルウィンドウをスライドさせる。第4のループレベルは、様々な異なる出力特徴マップを発生させる。

【0010】

FPGAは、特に、推論タスクを加速化するために、より多くの関心を集め、より広く普及してきた。それは、FPGAが、(1)再構成可能性が高く、(2)CNNの急速な進化にキャッチアップするために必要な開発時間の速さという点で、特定用途向け集積回路(application specific integrated circuit、ASIC)と比較して優れており、(3)良好な性能を有し、(4)GPUと比較して、エネルギー効率が優れている、ということに起因する。FPGAの高い性能及び高い効率性は、特定の計算のためにカスタマイズされた回路を合成して、カスタマイズされたメモリシステムで数十億回の演算を直接処理することによって実現することができる。例えば、現代のFPGAにおける数百から数千のデジタル信号処理(digital signal processing、DSP)ブロックは、コア

10

量込み込み演算、例えば、高度の並列処理を伴う積和演算をサポートする。外部オンチップメモリとオンチッププロセッシングエンジン(processing engine、PE)と間の専用データバッファは、フィールドプログラマブルゲートアレイ(FPGA)チップ上に、数十メガバイトのオンチップブロックランダムアクセスメモリ(block random access memory、BRAM)を構成することにより、優先データフローを実現するように設計することができる。

【0011】

高性能を達成するためにリソース利用を最大化しながら、データ通信を最小限に抑えるために、CNNアクセレーションの効率的なデータフロー及びハードウェアアーキテクチャが望まれている。アクセレーションハードウェア上の様々なCNNアルゴリズムの推論プロセスを加速化し、高い性能、高い効率、及び高い柔軟性を実現するための方法論及びフレームワークを設計するための好機が生じることとなる。

20

【0012】

次世代配列決定(NGS)技術の重要な特徴は並列化であり、いくつかの配列決定プラットフォームの基礎となる主な機構は、合成による配列決定(SBS)である。簡単に説明すると、一本鎖DNA鋳型の相補的塩基を順次構築し、蛍光の一連の生画像における合成情報を捕捉することによって、数千万から数億のランダムDNA断片が同時に配列決定される。

【0013】

画像データから実際のシーケンス情報(すなわち、{A, C, G, T}中の文字列)を抽出することは、2つの計算タスク、すなわち、画像分析及びベースコールを含む。画像分析の主な機能は、画像データを各DNAフラグメントについての蛍光強度データに翻訳することであり、一方、ベースコールの目的は、得られた強度データからシーケンス情報を推測することである。

30

【0014】

ベースコール精度を低下させる可能性がある変動の確率的及び状況的原因がいくつかある。例えば、ベースコールにおけるk-merバイアスは、配列決定されたゲノムのGC含量によって影響を受ける。ベースコーラーは、DNAのGCリッチ領域に適用された場合に、主にシーケンス複雑性の減少に起因するが、増幅工程中のポリメラーゼ連鎖反応(PCR)バイアスの結果としても、バイアスを示し得る。

40

【0015】

ベースコールの正確さは、シーケンスアセンブリ、SNPコール、及び遺伝子型コールを含む様々な下流の適用にとって本質的に重要である。ベースコール精度を改善することは、より小さい配列決定カバレッジで下流アプリケーションの所望の性能を達成することを可能にすることができ、これは配列決定コストの低減につながる。

【0016】

ベースコールのためにニューラルネットワークを訓練することは、大量のコンピュータメモリを必要とし、これは、画像サイズ及び数の増加とともに指数関数的に増加する。ディープニューラルネットワークを最適化するためのバックプロパゲーションアルゴリズムは中間活性化の記憶を必要とするので、コンピュータメモリは制限要因となる。これらの

50

中間活性化のサイズ及び数は、入力サイズ及び数に比例して増加するので、メモリは、より大きくより多くの画像で急速に満たされる。

【0017】

ニューラルネットワークを使用するベースコーラー、例えば、共同所有の特許出願第16/826,126号、同第16/826,134号、同第16/826,168号、同第17/175,546号、同第17/180,542号、同第17/176,151号、同第63/072,032号、同第63/161,880号、同第63/161,896号、同第17/176,147号、同第63/228,954号、同第17/179,395号、同第17/180,480号、同第17/180,513号、同第63/169,163号及び同第63/217,644号に開示されるものは、一実施態様に従って、配列決定サイクルのスライディングウィンドウについての画像データを使用してベースコール予測を行う。より多くの配列決定サイクルからの画像データを含むようにスライディングウィンドウのサイズを増加させることは、ニューラルネットワークの複雑さを増加させ、利用可能な計算及びメモリに更なる負担を加えることにもなる。

10

【0018】

多数の過去の配列決定サイクルからの文脈情報を組み込むようにベースコール操作を構成する機会が生じる。特にk-merバイアスの減衰に対して、誤差率が低減されたより正確なベースコールがもたらされ得る。

【図面の簡単な説明】

【0019】

20

図面では、同様の参照文字は、概して、異なる図全体を通して同様の部分を指す。また、図面は必ずしも縮尺通りではなく、その代わりに、開示された技術の原理を例示することを強調している。以下の説明において、開示された技術の様々な実施態様は、以下の図面を参照して説明される。

【0020】

【図1】開示された状態ベースのベースコールの高レベル図である。

【図2】配列決定画像内のピクセルの過去の強度値に基づいて生成された状態情報を使用してベースコール動作を実行する一実施態様を示す。

【図3】ベースコールについての入力として処理される配列決定画像内のピクセルについてピクセルごとの状態を生成する一実施態様を示す。

30

【図4】ピクセルごとの状態に基づいてベースコールを生成する一実施態様を示す。

【図5】ベースコールについての入力として処理される配列決定画像内のピクセルのチャンネルごとの状態を生成する一実施態様を示す。

【図6】ベースコールについての入力として処理される配列決定画像内のピクセルのチャンネルごとのMIN（最小）状態を生成する一実施態様を示す。

【図7】ベースコールについての入力として処理される配列決定画像内のピクセルについてのパンチャネルMIN（最小）状態を生成する一実施態様を示す。

【図8】ベースコールについての入力として処理される配列決定画像内のピクセルのチャンネルごとのMAX（最大）状態を生成する一実施態様を示す。

【図9】ベースコールについての入力として処理される配列決定画像内のピクセルについてのパンチャネルMAX（最大）状態を生成する一実施態様を示す。

40

【図10】ベースコールについての入力として処理される配列決定画像内のピクセルのチャンネルごとのAVG（平均）状態を生成する一実施態様を示す。

【図11】ベースコールについての入力として処理される配列決定画像内のピクセルについてのパンチャネルAVG（平均）状態を生成する一実施態様を示す。

【図12】ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとのMIN/MAX（最小及び最大）状態1200を生成する一実施態様を示す。

【図13】ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとのMIN AVG（最小及び平均）状態1300を生成する一実施態様

50

を示す。

【図 1 4】ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとの MAX AVG（最大及び平均）状態 1 4 0 0 を生成する一実施態様を示す。

【図 1 5】ベースコールについての入力として処理された配列決定画像内のピクセルのチャンネルごとの MIN MAX AVG（最小、最大、及び平均）状態を生成する一実施態様を示す。

【図 1 6】将来のベースコールにおいて使用するために、以前にコールされた塩基に依存して状態情報を生成する一実施態様を示す。

【図 1 7 A】現在の配列決定サイクル 1 1（サイクル 1 1）のためのベースコール指向の ON 及び OFF 状態生成を示す。

【図 1 7 B】現在の配列決定サイクル 1 1（サイクル 1 1）のためのベースコール指向の ON 及び OFF 状態生成を示す。

【図 1 7 C】現在の配列決定サイクル 1 1（サイクル 1 1）のためのベースコール指向の ON 及び OFF 状態生成を示す。

【図 1 8 A】図 1 7 A、1 7 B、及び 1 7 C を拡張し、次の配列決定サイクル 1 2（サイクル 1 2）において OFF 状態がどのように更新されるかを示す。

【図 1 8 B】図 1 7 A、1 7 B、及び 1 7 C を拡張し、次の配列決定サイクル 1 2（サイクル 1 2）において OFF 状態がどのように更新されるかを示す。

【図 1 8 C】図 1 7 A、1 7 B、及び 1 7 C を拡張し、次の配列決定サイクル 1 2（サイクル 1 2）において OFF 状態がどのように更新されるかを示す。

【図 1 9 A】図 1 8 A、1 8 B、及び 1 8 C を更に拡張し、更に次の配列決定サイクル 1 3（サイクル 1 3）において ON 状態がどのように更新されるかを示す。

【図 1 9 B】図 1 8 A、1 8 B、及び 1 8 C を更に拡張し、更に次の配列決定サイクル 1 3（サイクル 1 3）において ON 状態がどのように更新されるかを示す。

【図 1 9 C】図 1 8 A、1 8 B、及び 1 8 C を更に拡張し、更に次の配列決定サイクル 1 3（サイクル 1 3）において ON 状態がどのように更新されるかを示す。

【図 2 0 A】図 1 9 A、図 1 9 B、及び図 1 9 C を更に拡大し、ON 状態及び OFF 状態の両方が、更に次の配列決定サイクル 1 4（サイクル 1 4）においてどのように更新されるかを示す。

【図 2 0 B】図 1 9 A、図 1 9 B、及び図 1 9 C を更に拡大し、ON 状態及び OFF 状態の両方が、更に次の配列決定サイクル 1 4（サイクル 1 4）においてどのように更新されるかを示す。

【図 2 0 C】図 1 9 A、図 1 9 B、及び図 1 9 C を更に拡大し、ON 状態及び OFF 状態の両方が、更に次の配列決定サイクル 1 4（サイクル 1 4）においてどのように更新されるかを示す。

【図 2 1】指数加重平均化を使用してベースコールについての状態情報を生成する一実施態様を示す。

【図 2 2】指数加重平均化を使用してベースコールについてのチャンネルごとの状態情報を生成する一実施態様を示す。

【図 2 3】指数加重平均化を使用してベースコールについてのチャンネルごとの状態情報を生成する例を示す。

【図 2 4】以前にコールされた塩基によって指向される指数加重平均化を使用して、ベースコールについてのチャンネルごとの状態情報を生成する一実施態様を示す。

【図 2 5】以前にコールされた塩基によって指向される指数加重平均化を使用して、ベースコールについてのチャンネルごとの状態情報を生成する例を示す。

【図 2 6】以前にコールされた塩基を使用して非クラスターピクセルについての状態情報を生成する一実施態様を示す。

【図 2 7】ベースコールについての状態データを生成する異なる実施態様を示す。

【図 2 8】ベースコール動作の処理の様々な態様における供給状態データの異なる実施態

10

20

30

40

50

様を示す。

【図 29】状態ベース及びニューラルネットワークベースのベースコールの一実施態様を示す。

【図 30】ベースコールクラスターについてのクラスターごとの状態データを使用する一実施態様を示す。

【図 31】対応するクラスターピクセルの過去の強度値を使用してクラスターごとの状態を生成する一実施態様を示す。

【図 32】クラスターピクセルに対応する空間量み込み特徴の過去の特徴値を使用してクラスターごとの状態を生成する一実施態様を示す。

【図 33】ピクセル強度を補間することによってクラスター強度を生成し、補間されたクラスター強度を使用して、ベースコールについてのクラスターごとの状態を生成する一実施態様を示す。

【図 34】ニューラルネットワークベースのベースコーラーとは別個のリアルタイム分析 (RTA) ベースコーラーを使用してクラスターごとの状態を生成する一実施態様を示す。

【図 35】圧縮特徴を生成する圧縮論理の一実施態様を示す。

【図 36】クラスターピクセルに対応する圧縮空間量み込み特徴の過去の特徴値を使用してクラスターごとの状態を生成する一実施態様を示す。

【図 37】圧縮特徴を有するクラスターごとの状態の組み込みのために、ニューラルネットワークベースのベースコーラーとは別個の RTA ベースコーラーを使用してクラスターごとの状態を生成する一実施態様を示す。

【図 38】クラスターごとの状態を圧縮特徴に組み込むために、対応するクラスターピクセルの過去の強度値を使用してクラスターごとの状態を生成する一実施態様を示す。

【図 39】疎なウェルごとの状態から密なピクセルごとの状態を生成する一実施態様を示す。

【図 40】ベースコール動作を実行するためのベースコーラーへの入力として、疎なウェルごとの状態、密なピクセルごとの状態、及びピクセルごとの強度値を提供する一実施態様を示す。

【図 41】ベースコール動作を実行するためのニューラルネットワークベースのベースコーラーへの入力として、疎なウェルごとの状態、密なピクセルごとの状態、及びピクセルごとの強度値を使用する一実施態様を示す。

【図 42 A】構成可能プロセッサを含む配列決定システムの一実施態様を示す。

【図 42 B】構成可能プロセッサを含む配列決定システムの一実施態様を示す。

【図 42 C】ベースコールセンサ出力などの、配列決定システムからのセンサデータの分析のためのシステムの簡略ブロック図である。

【図 43 A】ホストプロセッサによって実行されるランタイムプログラムの機能を含む、ベースコール動作の態様を示す簡略図である。

【図 43 B】構成可能プロセッサの構成の簡略図である。

【図 44】CPU 上で状態データを決定し、ベースコールのために CPU から FPGA に状態データをロードする一実施態様を示す。

【図 45】非ニューラルネットワークベースのベースコーラー RTA、状態情報を伴わないニューラルネットワークベースのベースコーラー Deep RTA (5ci_k14)、並びに追加／補足入力としてピクセルごとの MIN 状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep RTA with State (5ci_k14_DC) のベースコールパフォーマンスを比較する。

【図 46】異なる状態チャンネルにわたって、開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep RTA with State (5ci_k14_DC) のベースコールパフォーマンスを比較する。

【図 47】k-mer、特に 5mer に対する、開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep RTA with State (DC) のベ

10

20

30

40

50

ースコールパフォーマンスを示す。5merは、5つの塩基位置（例えば、ACGCG、GGGGG、TCGCG）についての反復塩基パターンを指す。

【図48】非ニューラルネットワークベースのベースコーラーRTA、状態情報を伴わないニューラルネットワークベースのベースコーラーDeepRTA（5ci_k14）、追加／補足入力としてピクセルごとのMIN状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（5ci_k14_DC_MIN）、並びに追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（5ci_k14_DC_MIN_MAX）のベースコールパフォーマンスを比較する。

10

【図49】非ニューラルネットワークベースのベースコーラーRTA、状態情報及び5サイクル隣接入力ウィンドウを伴わないニューラルネットワークベースのベースコーラーDeepRTA（5ci_k14）、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN状態情報を使用する開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（5ci_k14_DC_MIN）、3サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN状態情報及びMAX状態情報を使用する開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（3ci_k14_DC_MIN_MAX）、並びに5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN状態情報及びMAX状態情報を使用する開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（5ci_k14_DC_MIN_MAX）のベースコールパフォーマンスを比較する。

20

【図50】k-mer（特に5mer）上の5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（5ci_k14_DC_MIN_MAX）のベースコールパフォーマンスを、状態情報及び5サイクル隣接入力ウィンドウなしのニューラルネットワークベースのベースコーラーDeepRTA（5ci_k14）のベースコールパフォーマンスと比較する。5merは、5つの塩基位置（例えば、ACGCG、GGGGG、TCGCG）についての反復塩基パターンを指す。

30

【図51】等化器実装を伴う非ニューラルネットワークベースのベースコーラーRTA、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（5ci_k14_DC_MIN_MAX）、並びに5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのAVG状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（5ci_k14_DC_AVG）のベースコールパフォーマンスを比較する。

【図52】等化器実装を伴う非ニューラルネットワークベースのベースコーラーRTAと、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN、MAX、及びAVG状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（5ci_k14_DC_MIN_MAX_AVG）とのベースコールパフォーマンスを比較する。

40

【図53】等化器実装を伴う非ニューラルネットワークベースのベースコーラーRTA、5サイクル隣接入力ウィンドウ及びサイズ14（K=14）のフィルタバンクとともに追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State（5ci_k14_DC_MIN_MAX）、並びに5サイクル隣接入力ウィンドウ及びサイズ32（K=32）のフィルタバンクとともに追加／補足入力としてピク

50

セルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5ci_k14_DC_MIN_MAX)のベースコールパフォーマンスを比較する。

【図54】上記で議論されるように、指数加重平均化実施の異なるアルファパラメータ構成（例えば、0.05、0.07、0.10、0.12）について、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with Stateのベースコールパフォーマンスを比較する。

【図55】開示された技術の一実施態様による、指数加重平均によって決定される状態値を追跡するグラフである。

【図56】開示された技術の一実施態様による、指数加重平均によって決定される状態値を追跡するグラフである。

【図57】コールされたリードの数、リード1及びリード1についてのリードミスマッチ誤差率及びそれらの平均、単一ヌクレオチド多型（SNP）再現性、SNO精度、SNPコール精度（F1スコア）、挿入／欠失（Indel）再現性、Indel精度、及びIndelコール精度（F1スコア）のような二次解析タスク及びメトリクスの観点から、等化器実装を有するRTA（RTA+Eq）、状態情報を有しないDeepRTA（DeepRTA（k14_1m））、及び開示されたDeepRTA with State（DeepRTA_extrachannels（k14_1, _onoff_expavg））のベースコールパフォーマンスを比較する。

【発明を実施するための形態】

【0021】

以下の考察は、開示される技術を当業者が作製及び使用することを可能にするために提示され、特定の用途及びその要件に関連して提供される。開示される実施態様に対する種々の修正は、当業者には容易に明らかとなり、本明細書で定義される一般原理は、開示された技術の趣旨及び範囲から逸脱することなく、その他の実施態様及び用途に適用され得る。したがって、開示される技術は、示される実施態様に限定されることを意図するものではなく、本明細書に開示される原理及び特徴と一致する最も広い範囲を与えられるものである。

【0022】

様々な実施態様の詳細な説明は、添付の図面と併せて読むと、より良く理解することができる。図が様々な実施態様の機能ブロックの図を示す限りにおいて、機能ブロックは、必ずしもハードウェア回路間の分割を示すものではない。したがって、例えば、機能ブロック（例えば、モジュール、プロセッサ、又はメモリ）のうちの1つ以上は、単一のハードウェア（例えば、汎用信号プロセッサ又はランダムアクセスメモリのブロック、ハードディスクなど）又は複数のハードウェアに実装されてもよい。同様に、プログラムは、スタンドアロンプログラムであってもよく、オペレーティングシステム内のサブルーチンとして組み込まれてもよく、インストールされたソフトウェアパッケージ内の機能である等でもよい。様々な実施態様は、図面に示された配置及び手段に限定されないことを理解されたい。

【0023】

モジュールとして指定された図の処理エンジン及びデータベースは、ハードウェア又はソフトウェアで実装することができ、図に示されるように、正確に同じブロックで分割される必要はない。いくつかのモジュールは、異なるプロセッサ、コンピュータ若しくはサーバ上に実装されてもよく、又は多数の異なるプロセッサ、コンピュータ若しくはサーバの中で広がることもできる。加えて、モジュールの一部は、達成される機能に影響を及ぼすことなく、図に示されるものとは並行して、又は異なる順序で操作され得ることが理解されるであろう。図のモジュールはまた、方法におけるフローチャートステップと考えることができる。また、モジュールは、必ずしもメモリ内に隣接して配置された全てのコードを有する必要はない。コードのいくつかの部分は、他のモジュール又は他の機能からのコードが間に配置された状態で、コードの他の部分から分離することができる。

【0024】

序論

開示された技術は、状態ベースのベースコールに関する。特に、開示された技術は、現在の配列決定サイクルについてのベースコールを生成するときに、以前の配列決定サイクルからのデータに関する状態情報を現在の配列決定サイクルからのデータの分析に組み込むことに関する。例えば、N番目の配列決定サイクルについてのベースコールを生成する場合、開示された技術は、配列決定サイクル1～N-1からのデータに関するベースコール論理状態情報に組み込むことができる。

【0025】

以下の説明は、開示される状態ベースのベースコールの様々な実施態様を説明する。実施態様は、異なるデータ及び処理態様によって異なる。例えば、状態情報を生成する異なる「方法」又は「論理」は、異なるタイプの状態をもたらす。また、状態生成のために過去の値が追跡される「物」は、実施態様ごとに異なり得る。更に、いったん生成されると、状態情報がベースコールのために「いつ」、「どこで」、及び「どのように」処理されるかは、開示された技術の様々な実施態様を生じさせる。

【0026】

状態ベースのベースコール

図1は、開示された状態ベースのベースコール100の高レベル図である。ベースコールは、配列のヌクレオチド組成を決定するプロセスである。一実施態様では、ベースコールは、IlluminaのiSeq、HiSeqX、HiSeq3000、HiSeq4000、HiSeq2500、NovaSeq6000、NextSeq550、NextSeq1000、NextSeq2000、NextSeqDx、MiSeq、及びMiSeqDxなどの配列決定システムによって行われる、配列決定ラン（又は配列決定反応）中に生成された画像データ、すなわち配列決定画像の分析を伴う。他の実施態様において、ベースコールは、非画像配列決定データからシーケンスリードを推測することを含み得る。

【0027】

配列決定システム104は、核酸の配列決定に使用され得る。適用可能な技術としては、核酸がアレイ中の固定場所（例えば、フローセルのウェル）に付着され、アレイが繰り返し撮像されるものが挙げられる。そのような実施態様では、配列決定システム104は、特定のヌクレオチド塩基タイプを別のヌクレオチド塩基と区別するために使用され得る、2つの異なる色チャンネルの画像を取得し得る。より具体的には、配列決定システム104は、「ベースコール」と称されるプロセスを実施してもよく、これは一般に、撮像サイクルでの画像の所与のスポットの場所について、ベースコール（例えば、アデニン（A）、シトシン（C）、グアニン（G）、又はチミン（T））を決定するプロセスを指す。2チャンネルベースコール中、例えば、2つの画像から抽出された画像データを使用して、塩基同一性を2つの画像の強度の組み合わせとして符号化することによって、4つの塩基タイプのうちの1つの存在が決定され得る。2つの画像のそれぞれにおける所与のスポット又は場所について、塩基同一性は、信号同一性の組み合わせが[オン、オン]、[オン、オフ]、[オフ、オン]、又は[オフ、オフ]であるかどうかに基づいて決定され得る。

【0028】

配列決定システム104からの出力データは、リアルタイム分析モジュール（図示せず）に通信することができる。リアルタイム分析モジュールは、様々な実施態様では、画像データを分析する（例えば、画質スコアリング、ベースコールなど）、ビームの特性（例えば、焦点、形状、強度、電力、輝度、位置）をグラフィカルユーザインターフェース（GUI）に報告又は表示する、などのためのコンピュータ可読命令を実行する。これらの動作は、ダウンストリーム分析時間を最小限に抑え、撮像ラン中にリアルタイムフィードバック及びトラブルシューティングを提供するために、撮像サイクル中にリアルタイムで実行され得る。実施態様において、リアルタイム分析モジュールは、配列決定システム104の撮像サブシステムに通信可能に結合され、それを制御するコンピューティングデバ

イスであり得る。

【0029】

以下の説明は、一実施態様に従って、配列決定画像がどのように生成され、それらが何を描写するのかを概説する。

【0030】

いくつかの実施態様では、ベースコールは、配列決定画像に符号化された強度データを、ヌクレオチドシーケンスに復号化する。一実施態様では、Illuminaのシーケンシングプラットフォームは、ベースコールのための循環可逆的ターミネーション（Cyclic Reversible Termination、CRT）化学を採用する。このプロセスは、新たに添加された各ヌクレオチドの放出信号を追跡しながら、蛍光標識されたヌクレオチドを有するテンプレート鎖に相補的な新生鎖を伸長させることに依存する。蛍光標識されたヌクレオチドは、ヌクレオチド型のフルオロフォアシグナルをアンカーする、3'の取り外し可能なブロックを有する。

10

【0031】

配列決定は、反復サイクルで行われ、各々は3つの工程、つまり、（a）蛍光標識されたヌクレオチドを添加することによる出現鎖の伸長と、（b）配列決定システム104の光学サブシステムの1以上のレーザを使用してフルオロフォアを励起させ、光学サブシステムの異なるフィルタを通した撮像によって配列決定画像を生成することと、（c）次の配列決定サイクルに備えてフルオロフォアを切断し、3'ブロックを除去することと、の3つの工程を各々含む反復サイクルで行われる。取り込み及び撮像サイクルは、指定された数の配列決定サイクルまで繰り返され、リード長を定義する。このアプローチを使用して、各サイクルはテンプレート鎖に沿って新しい位置を照合する。

20

【0032】

Illuminaシーケンサの途方もないパワーは、CRT反応を受ける数百万のクラスター又は更には数十億のクラスター（「クラスター」とも呼ばれる）を同時に実行及び感知する能力に起因する。クラスターは、テンプレート鎖の約1000個の同一のコピーを含むが、クラスターのサイズ及び形状は様々である。クラスターは、配列決定ラン前に、入力ライブラリのブリッジ増幅又は排除増幅によって、テンプレートストランドから成長させる。増幅及びクラスター伸長の目的は、撮像デバイスが一本鎖のフルオロフォア信号を確実に感知することができないため、放出される信号の強度を増加させることである。しかしながら、クラスター内の鎖の物理的距離は小さいため、撮像デバイスは鎖のクラスターを単一のスポットとして知覚する。

30

【0033】

シーケンシングは、フローセル（又はバイオセンサ）、すなわち、入力ストランドを保持する小さなガラススライドにおいて行われる。フローセルは、顕微鏡撮像、励起レーザ、及び蛍光フィルタを含む光学システムに接続される。フローセルは、レーンと呼ばれる複数のチャンバを含む。レーンは互いに物理的に分離されており、試料の交差汚染なしに区別可能な異なるタグ付き配列決定ライブラリを含むことができる。いくつかの実施態様では、フローセルは、パターン化された表面を含む。「パターン化された表面」は、固体支持体の露出層内又はその上における、異なる領域の配置を指す。

40

【0034】

配列決定システム104の撮像デバイス（例えば、電荷結合素子（Charge-Coupled Device、CCD）又は相補的金属酸化物半導体（Complementary Metal-Oxide-Semiconductor、CMOS）センサなどの固体撮像素子）は、タイルと呼ばれる一連の非重複領域において、レーンに沿った複数の場所でスナップショットを取る。例えば、レーン当たり64個又は96個のタイルが存在し得る。タイルは数十万～数百万個のクラスターを保持する。

【0035】

配列決定ランの出力は、配列決定画像である。配列決定画像は、ピクセル化ユニット（例えば、ピクセル、スーパーピクセル、サブピクセル）のグリッド（又はアレイ）を使用

50

するクラスター及びそれらの周囲の背景の強度放射を示す。強度放射は、ピクセル化ユニットの強度値として記憶される。配列決定画像は、ピクセル化ユニットのグリッドの、 w 個 $\times$ h 個の寸法を有し、ここで、 w (幅) 及び h (高さ) は、 $1\sim 100,000$ の範囲の任意の数 (例えば、 115×115 、 200×200 、 1800×2000 、 2200×25000 、 2800×3600 、 4000×400) である。いくつかの実施態様では、 w 及び h は同じである。他の実施態様では、 w 及び h は異なる。配列決定画像は、配列決定ラン中にヌクレオチドシーケンスにヌクレオチドを組み込む結果として生成される強度放射を示す。強度放射は、関連するクラスター及びそれらの周囲の背景に由来するものである。

【0036】

一実施態様によれば、データフロー論理 (図示せず) は、ベースコールについてのベースコーラー 144 に配列決定画像を提供する。一実施態様によれば、ベースコーラー 144 は、パッチごとに (又はタイルごとに) 配列決定画像にアクセスする。パッチのそれぞれは、配列決定画像を形成するピクセル化されたユニットのグリッド内のピクセル化ユニットのサブグリッド (又はサブアレイ) である。パッチは、ピクセル化ユニットのサブグリッドの q 個 $\times$ r 個の寸法を有し、 q (幅) 及び r (高さ) は、 $1\sim 10000$ の範囲の任意の数である (例えば、 3×3 、 5×5 、 7×7 、 10×10 、 15×15 、 25×25 、 64×64 、 78×78 、 115×115)。いくつかの実施態様では、 q 及び r は同じである。他の実施態様では、 q 及び r は互いに異なる。いくつかの実施態様では、1つの配列決定画像から抽出されたパッチは、同じサイズである。他の実施態様では、パッチは、異なるサイズのものである。いくつかの実施態様では、パッチどうしは、(例えば、エッジの上で) 重なり合うピクセル化ユニットを有することができる。

【0037】

いくつかの実施態様によれば、配列決定により、対応する m 個の画像チャンネルに対して、配列決定サイクルごとに、 m 個の配列決定画像を生成する。すなわち、配列決定画像の各々は、1つ以上の画像 (又は強度) チャンネル (カラー画像の赤、緑、青 (RGB) のチャンネルと類似) を有する。一実施態様では、各画像チャンネルは、複数のフィルタ波長帯域のうちの1つに対応する。別の実施態様では、各画像チャンネルは、配列決定サイクルにおける複数の撮像事象のうちの1つに対応する。更に別の実施態様では、各画像チャンネルは、特定のレーザによる照射と特定の光学フィルタを通した撮像との組み合わせに対応する。画像パッチは、特定の配列決定サイクルについて、 m 個の画像チャンネルの各々からタイリング (又はアクセス) される。4-、2-、及び1-チャンネル化学などの異なる実施態様では、 m は、4又は2である。他の実施態様では、 m は、1、3、又は4よりも大きい。他の実施態様では、画像は、赤色チャンネル及び緑色チャンネルの代わりに、又はそれに加えて、青色チャンネル及び紫色チャンネル内にあり得る。

【0038】

例えば、配列決定ランが、2つの異なる画像化チャンネル、すなわち、青色チャンネルと緑色チャンネルとを使用して実施されると考えてみる。すると、各配列決定サイクルにおいて、配列決定ランは、青色画像及び緑色画像を生成する。このようにして、配列決定ランの k 個の配列決定サイクルのシリーズについて、 k 対の青色画像及び緑色画像のシーケンスが出力として生成され、配列決定画像として記憶される。したがって、 k 対の青色及び緑色画像パッチのシーケンスが、ベースコーラー 144 によるパッチレベル処理に対して生成される。

【0039】

配列決定システム 104 はまた、他の実施態様に従って、非画像配列決定データを生成することができる。一実施態様では、配列決定データは、分子伸長中の水素イオンの放出によって誘発される pH 変化に基づき得る。pH 変化は検出され、組み込まれた塩基の数に比例する電圧変化に変換される。更に別の実施態様では、配列決定データは、例えば、塩基の同一性を決定すると同時に、クラスターがナノ細孔を通過する際、又はその開口部付近を通過する際に現在の破壊を測定するためにバイオセンサを使用するナノ細孔検知か

10

20

30

40

50

ら構築され得る。一実施態様では、ナノ細孔ベースの配列決定は、以下の概念に基づき得る：ナノ細孔を介して膜を通してDNA（又はRNA）の一本鎖を通過させ、膜にわたって電位差を印加する。細孔内に存在するヌクレオチドは、細孔の電気抵抗に影響を及ぼし得、そのため、経時的な電流測定は、細孔を通過するDNA塩基のシーケンスを示すことができる。この電流信号（プロットされたときにその外観に起因する「押しつぶし」）は、シーケンサによって収集された生データである。これらの測定値は、4 kHz 周波数（例えば）で取られた16ビットの整数データ取得（Data Acquisition、DAC）値として記憶され得る。1秒当たり～450塩基対のDNA鎖速度を用いて、これは、平均して、塩基当たり約9つの生観察を与え得る。次いで、この信号を処理して、個々の読み取りに対応する開孔信号の破断を識別し得る。これらの生信号の伸長は、ベースコールされ得、DAC値をDNA塩基のシーケンスに変換するプロセスである。いくつかの実施態様では、配列決定データは、正規化又はスケーリングされたDAC値を含み得る。

10

【0040】

現在の配列決定データ112は、配列決定ランの現在の配列決定サイクルについて配列決定データ104によって生成された配列決定データを含む。図1において、現在の配列決定サイクルは、「N」配列決定サイクルとして識別される。

【0041】

以前の配列決定データ116は、配列決定ランの1つ以上の以前の配列決定サイクルについて配列決定データ104によって生成された配列決定データを含む。以前の配列決定サイクルは、現在の配列決定サイクルに先行する。図1において、以前の配列決定サイクルは、「1～N-1」配列決定サイクルとして識別される。他の実施態様では、以前の配列決定データ116は、1～N-1配列決定サイクルのサブセットについての配列決定データを含む。

20

【0042】

状態生成器126は、現在の配列決定データ112及び以前の配列決定データ116を使用して、現在の配列決定サイクルの現在の状態データ136を生成する。状態生成器126は、所望の結果を生成するために配列決定データに適用される値又は関数であり得る。状態生成器126は、加算、減算、除算、乗算、又はそれらの組み合わせを含むが、これらに限定されない、種々の数学的操作のいずれかによって、配列決定データに適用され得る。状態生成器126は、数式、論理関数、コンピュータ実装アルゴリズムなどとする

30

【0043】

一実施態様では、状態生成器126は、現在の配列決定データ112及び以前の配列決定データ116の要約統計値を蓄積することによって現在の状態データ136を生成する。要約統計値の例は、最大値、最小値、平均（average）（平均（mean））、指数加重平均、移動（ランニング）平均、指数移動平均、モード、標準偏差、分散、歪度、尖度、パーセンタイル、及びエントロピーを含む。他の実施態様では、状態生成器126は、要約統計値に基づいて二次統計値を決定する。二次統計値の例は、デルタ、和、一連の最大値、一連の最小値、一連の最大値の最小値、及び一連の最小値の最大値を含む。

40

【0044】

ベースコーラー144は、現在の配列決定データ112及び現在の状態データ136の処理に応答して、現在の配列決定サイクルについての現在のベースコールデータ154を生成する。現在のベースコールデータ154は、1つ以上のクラスターについてのベースコールを含み得る。いくつかの実施態様では、現在の配列決定データ112及び現在の状態データ136は、ベースコーラー144による処理の前に組み合わせられる。結合は、例えば、合計演算、要素単位の乗算演算、要素単位の乗算及び合計（畳み込み）演算、並びに連結演算によってもたらしことができる。

【0045】

ベースコーラー144の例は、リアルタイム分析（RTA）、Blind Call、f

50

reeIbis、Softy、AYB、OnlineCall、BM-BC、ParticleCall、TotalRecaller、naiveBayesCall、Srfim、BayesCall、Ibis、Rolexa、Alta-Cyclic、及びBustardなどの、Illuminaプラットフォームに利用可能な異なるベースコール手順を含む。ベースコーラー144の例には、共同所有の特許出願第16/825,987号、同第16/825,991号、同第16/826,126号、同第16/826,134号、同第16/826,134号、同第16/826,168号、同第17/175,546号、同第17/180,542号、同第17/176,151号、同第63/072,032号、同第63/161,880号、同第63/161,896号、同第17/176,147号、同第63/228,954号、同第17/179,395号、同第17/180,480号、同第17/180,513号、同第63/169,163号及び同第63/217,644号に開示されているものなどの、Illuminaのニューラルネットワークベースの提供物も含まれる。これらは、本明細書では集合的に「DeepRTA」又は「深層学習一次解析 (Deep Learning Primary Analysis)」と呼ばれる。ベースコーラー144の更に他の例は、Metricor、Nanocal、DeepNano、Nanonet、Scrappie、Albacore、Guppy、Basecrawller、Chiron、Halcyon、MinCall、SACall、Causalcall、及びWaveNanoなどのOxford Nanopore Technologies (ONT) に利用可能な異なるベースコール手順を含む。

10

20

【0046】

ピクセル化された状態データ

図2は、配列決定画像内のピクセルの過去の強度値に基づいて生成された状態情報200を使用してベースコール動作を実行する一実施態様を示す。

【0047】

アクション202において、メモリは、配列決定ランの現在の配列決定サイクルについて、複数のピクセル内のそれぞれのピクセルについてのそれぞれの現在の強度値を記憶する（例えば、データストア112内に）。アクション212において、メモリは、現在の配列決定サイクルに先行する配列決定ランの1つ以上の以前の配列決定サイクルについて、それぞれのピクセルについてのそれぞれの以前の強度値のシーケンスを記憶する（例えば、データストア116内に）。一実施態様では、それぞれのピクセルについてのそれぞれの強度値は、複数のチャンネルについてのチャンネル固有強度値によってそれぞれ特徴付けられる。一実施態様では、チャンネルは、特定のレーザによる照射と特定の光学フィルタを通した撮像との組み合わせに対応する。別の実施態様では、チャンネルはフィルタ波長帯域に対応する。更に別の実施態様では、チャンネルは、配列決定サイクルにおける撮像イベントに対応する。

30

【0048】

動作222において、メモリへのアクセスを有する状態生成器126は、(i) それぞれの現在の強度値、及び(ii) それぞれの以前の強度値のシーケンス内のそれぞれの以前の強度値に依存して、それぞれのピクセルに対するそれぞれの現在の状態値を生成する。また、状態生成器126は、それぞれの現在状態値をメモリに記憶する。いくつかの実施態様では、それぞれのピクセルについてのそれぞれの状態値は、複数のチャンネルのチャンネル固有状態値によってそれぞれ特徴付けられる。一実施態様では、それぞれの現在の状態値は、それぞれの現在の強度値を用いてピクセル単位で符号化される。いくつかの実施態様では、複数のチャンネル内のチャンネルのサブセットについてのチャンネル固有状態値は、それぞれの現在の強度値を用いてピクセル単位で符号化される。いくつかの実施態様では、チャンネル固有状態値は、それぞれのパンチャンネル現在状態値を生成するために、複数のチャンネル内のチャンネルにわたって平均化される。一実施態様では、それぞれのパンチャンネル現在状態値は、それぞれの現在の強度値を用いてピクセル単位で符号化される。

40

【0049】

50

いくつかの実施態様では、ピクセル単位符号化は、ピクセル単位の連結を含む。他の実施態様では、ピクセル単位符号化はピクセル単位の合計を含む。更に他の実施態様では、ピクセル単位符号化は、要素単位の乗算を含む。更に他の実施態様では、ピクセル単位符号化は、要素単位の乗算及び合計（畳み込み）を含む。

【0050】

一実施態様では、それぞれの状態値は、それぞれのピクセルの過去の強度パターンを特徴付けるように構成される。別の実施態様では、過去の強度パターンは、ベースコーラー144のベースコール精度の損失を補償するように構成される。更に別の実施態様では、過去の強度パターンは、k-merの塩基をベースコールするときのベースコーラー144のベースコール精度の損失を補償するように構成される。更に別の実施態様では、それぞれの状態値は、クラスターのそれぞれの信号プロファイルを識別するように構成される。

10

【0051】

一実施態様では、それぞれの現在の状態値は、(i)それぞれの以前の強度値、及び(ii)それぞれの現在の強度値から、現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されたそれぞれの現在の平均強度である。いくつかの実施態様では、それぞれの現在の平均強度は各々、チャンネル固有の現在の平均強度によって特徴付けられる。

【0052】

別の実施態様では、それぞれの現在の状態値は、(i)それぞれの以前の強度値、及び(ii)それぞれの現在の強度値から、現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されたそれぞれの現在の最大強度である。いくつかの実施態様では、それぞれの現在の最大強度は各々、チャンネル固有の現在の最大強度によって特徴付けられる。

20

【0053】

更に別の実施態様では、それぞれの現在の状態値は、(i)それぞれの以前の強度値、及び(ii)それぞれの現在の強度値から、現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されたそれぞれの現在の最大強度である。いくつかの実施態様では、それぞれの現在の最大強度は各々、チャンネル固有の現在の最大強度によって特徴付けられる。

【0054】

更に別の実施態様では、それぞれの現在の状態値は、(i)それぞれの以前の強度値、及び(ii)それぞれの現在の強度値から、現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されたそれぞれの現在の最小強度である。いくつかの実施態様では、それぞれの現在の最小強度は各々、チャンネル固有の現在の最小強度によって特徴付けられる。

30

【0055】

更に別の実施態様では、それぞれの現在の状態値は、(i)それぞれの以前の強度値、及び(ii)それぞれの現在の強度値から、現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されたそれぞれの現在の指数加重平均強度である。いくつかの実施態様では、それぞれの現在の指数加重平均強度は、前の配列決定サイクルよりも最近の配列決定サイクルを加重することに基づいて決定される。他の実施態様では、それぞれの現在の指数加重平均強度は各々、チャンネル固有の現在の指数加重平均強度によって特徴付けられる。

40

【0056】

更に別の実施態様では、それぞれの現在の状態値は、(i)それぞれの現在の強度値、及び(ii)それぞれの以前の強度値のローリングサブセットから、現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されたそれぞれの現在の移動平均強度である。いくつかの実施態様では、それぞれの現在の移動平均強度は各々、チャンネル固有の現在の移動平均強度によって特徴付けられる。

【0057】

50

いくつかの実施態様では、それぞれの現在の強度値の各々は、それぞれの現在の強度値とそれぞれの現在のアクティブ状態強度及びそれぞれの現在の非アクティブ状態強度との比較に基づいて、アクティブ状態バケット又は非アクティブ状態バケットに帰属される。一実施態様では、それぞれの現在のアクティブ状態強度は、(i) それぞれの以前の強度値から、現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されたそれぞれの現在の大域的 maximum 強度である。いくつかの実施態様では、それぞれの現在の大域的 maximum 強度は各々、チャンネル固有の現在の大域的 maximum 強度によって特徴付けられる。

【0058】

一実施態様では、それぞれの現在の非アクティブ状態強度は、(i) それぞれの以前の強度値から、現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されたそれぞれの現在の大域的 minimum 強度である。いくつかの実施態様では、それぞれの現在の大域的 minimum 強度は各々、チャンネル固有の現在の大域的 minimum 強度によって特徴付けられる。

10

【0059】

一実施態様では、それぞれの現在の状態値は、現在の配列決定サイクルにおいてそれぞれのピクセルについて生成されたそれぞれの現在のアクティブ状態値及びそれぞれの現在の非アクティブ状態値を更に含む。いくつかの実施態様では、それぞれの現在アクティブ状態値は各々、チャンネル固有の現在アクティブ状態値によって特徴付けられる。他の実施態様では、それぞれの現在の非アクティブ状態値は各々、チャンネル固有の現在の非アクティブ状態値によって特徴付けられる。

【0060】

一実施態様では、対象チャンネル内のピクセルの現在のアクティブ状態値は、(i) 現在の配列決定サイクルにおいて検出され、アクティブ状態バケットに帰属する対象チャンネル内のピクセルについての現在の強度値、及び (ii) 以前の配列決定サイクルにおいて検出された対象チャンネル内のピクセルの以前の強度値から、現在の配列決定サイクルにおいて決定される。

20

【0061】

いくつかの実施態様では、現在の強度値は、現在の強度値と、対象チャンネル内の以前の強度値から決定された大域的 maximum 値及び大域的 minimum 値との比較に基づいて、アクティブ状態バケットに帰属する。他の実施態様では、現在の強度値は、現在の強度値と、対象チャンネル内の先行する配列決定サイクルで決定された以前のアクティブ状態値及び以前の非アクティブ状態値との比較に基づいて、アクティブ状態バケットに帰属する。一実施態様では、現在のアクティブ状態値は、現在の強度値及び以前の強度値から決定された指数加重平均である。別の実施態様では、現在のアクティブ状態値は、現在の強度値及び以前の強度値から決定される平均である。更に別の実施態様では、現在のアクティブ状態値は、現在の強度値及び以前の強度値のローリングサブセットから決定される移動平均である。更に別の実施態様では、現在のアクティブ状態値は、現在の強度値及び以前の強度値から決定される minimum 値である。更に別の実施態様では、現在のアクティブ状態値は、現在の強度値及び以前の強度値から決定される maximum 値である。

30

【0062】

一実施態様では、現在のアクティブ状態値は、先行する配列決定サイクルから搬送され、現在の強度値が現在の配列決定サイクルにおける非アクティブ状態バケットに帰属する場合、現在の配列決定サイクルにおいて再決定されない。

40

【0063】

一実施態様では、対象チャンネル内のピクセルの現在の非アクティブ状態値は、(i) 現在の配列決定サイクルにおいて検出され、非アクティブ状態バケットに帰属する対象チャンネル内のピクセルについての現在の強度値と、(ii) 以前の配列決定サイクルにおいて検出された対象チャンネル内のピクセルの以前の強度値とから、現在の配列決定サイクルにおいて決定される。

【0064】

いくつかの実施態様では、現在の強度値は、対象チャンネル内の以前の強度値から決定さ

50

れる大域的 maximum 値及び大域的 minimum 値との現在の強度値の比較に基づいて、非アクティブ状態バケットに帰属する。他の実施態様では、現在の強度値は、現在の強度値と、以前の非アクティブ状態値及び対象チャネル内の先行する配列決定サイクルにおいて決定された以前の非アクティブ状態値との比較に基づいて、非アクティブ状態バケットに帰属する。

【0065】

いくつかの実施態様では、現在の強度値が大域的 maximum 値よりも大域的 minimum 値に近いとき、現在の強度値は非アクティブ状態バケットに帰属する。いくつかの実施態様では、現在の強度値が大域的 minimum 値よりも大域的 maximum 値に近いとき、現在の強度値はアクティブ状態バケットに帰属する。

【0066】

一実施態様では、現在の非アクティブ状態値は、現在の強度値及び以前の強度値から決定された指数加重平均である。別の実施態様では、現在の非アクティブ状態値は、現在の強度値及び以前の強度値から決定される平均である。更に別の実施態様では、現在の非アクティブ状態値は、現在の強度値及び以前の強度値のローリングサブセットから決定された移動平均である。更に別の実施態様では、現在の非アクティブ状態値は、現在の強度値及び以前の強度値から決定される最小値である。更に別の実施態様では、現在の非アクティブ状態値は、現在の強度値及び以前の強度値から決定される maximum 値である。

【0067】

いくつかの実施態様では、現在の非アクティブ状態値は、先行する配列決定サイクルから搬送され、現在の強度値が現在の配列決定サイクルにおけるアクティブ状態バケットに帰属する場合、現在の配列決定サイクルにおいて再決定されない。

【0068】

アクション 232 において、メモリへのアクセスを有するベースコーラー 144 は、(i) それぞれの現在の強度値、及び (ii) それぞれの現在の状態値の処理に応答して、現在の配列決定サイクルについてのベースコールを生成する。いくつかの実施態様では、現在の配列決定サイクルについてのベースコールは、それぞれの現在の信号値及びそれぞれの以前の信号値が検出される 1 つ以上のクラスターについてのベースコールを含む。

【0069】

一実施態様では、ベースコーラー 144 は、ニューラルネットワークである。いくつかの実施態様では、ニューラルネットワークは、畳み込みニューラルネットワークである。一実施態様では、畳み込みニューラルネットワークは、複数の空間畳み込み層と複数の時間畳み込み層とを含む。

【0070】

いくつかの実施態様では、ベースコーラー 144 は、既に実行された配列決定ランについてオフラインで生成された配列決定画像を使用して訓練される。一実施態様では、それぞれのピクセルについてのそれぞれの状態値は、ベースコーラー 144 によるベースコールに先立って、既に実行された配列決定ランの各配列決定サイクルについてオフラインで計算される。一実施態様では、ベースコーラー 144 は、それぞれの状態値を使用してベースコール精度の損失を補償するように訓練される。

【0071】

いくつかの実施態様では、それぞれの現在状態値は、配列決定ランの各配列決定サイクルにおいて反復的に生成される。いくつかの実施態様では、メモリは、配列決定ランの次の配列決定サイクルについて、それぞれのピクセルについてのそれぞれの次の強度値のシーケンスを記憶するように更に構成される。いくつかの実施態様では、ベースコーラー 144 は、(i) それぞれの現在の強度値、(ii) 以前の配列決定サイクルのうちの 1 つ以上についてのそれぞれの以前の強度値のシーケンスにおけるそれぞれの以前の強度値、(iii) 次の配列決定サイクルのうちの 1 つ以上についてのそれぞれの次の強度値のシーケンスにおけるそれぞれの次の強度値、及び (iv) それぞれの状態値を処理することに応答して、現在の配列決定サイクルについてのベースコールを生成するように更に構成される。いくつかの実施態様では、それぞれの状態値は、それぞれの以前の強度値及びそ

10

20

30

40

50

れぞれの次の強度値を用いてピクセル単位で符号化される。

【0072】

ピクセルごとの状態

図3は、ベースコールについての入力として処理される配列決定画像内のピクセルについてピクセルごとの状態300を生成する一実施態様を示す。図3は、サイズ3×3の例示的なピクセルグリッド302N、302N-1、...、302を示す。図3はまた、ピクセルごとの状態グリッド304を示す。ピクセルグリッド302N、302N-1、...、302は、ベースコールについてのベースコーラー144への入力として提供される配列決定画像を表す。ピクセルグリッド302N、302N-1、...、302内のピクセルは、複数の画像チャンネル、例えば、赤色チャンネル及び青色チャンネルに対する強度値を有することができる。ピクセルグリッド302Nは、配列決定ランの現在の配列決定サイクルNについて生成され、現在の配列決定サイクルNにおけるクラスターのセットにおけるヌクレオチド組み込みの結果として生成された強度発光を示す。同様に、ピクセルグリッド302N-1は、配列決定ランの以前の配列決定サイクルN-1について生成され、以前の配列決定サイクルN-1におけるクラスターのセットにおけるヌクレオチド組み込みの結果として生成された強度発光を示す。同様に、ピクセルグリッド302は、配列決定ランの第1の配列決定サイクル1に対して生成され、第1の配列決定サイクル1におけるクラスターのセットへのヌクレオチド組み込みの結果として生成される強度発光を示す。

10

【0073】

空間的な観点から、ピクセルグリッド302N、302N-1、...、302は、9個のピクセル1~9を共有すると考えることができる。図3では、9つのピクセル1~9は、ピクセルグリッド302N、302N-1、...、302にわたって同じピクセルごとの上付き文字によってインデックス付けされている。時間的観点から、9個のピクセル1~9は、現在の時間ステップNにおける強度値の第1のセットと、以前の時間ステップN-1における強度値の第2のセットと、第1の時間ステップ1における強度値の第3のセットと、を示すと考えることができる。強度値の第1、第2、及び第3のセット内の強度値は、時間的順序で配置されていると考えることができる。図3では、時変強度値は、ピクセルグリッド302N、302N-1、...、302にわたる異なるピクセルごとの上付き文字によってインデックス付けされる。

20

30

【0074】

所与のピクセルについての現在のピクセル状態318は、現在の配列決定Nにおける所与のピクセルについての現在の強度値312と、それぞれの以前の配列決定サイクルN-1~1における所与のピクセルの以前の強度値314、...、316とに基づいて、現在の配列決定Nにおいて計算され得る。所与のピクセルの以前の強度値314、...、316は、以前の強度位置N-1から始まり、以前の強度位置1で終わる以前の強度シーケンスに配置されていると考えることができる。

【0075】

一実施態様では、現在のピクセル状態318は、現在の強度値312及び以前の強度値314、...、316を入力として取る状態生成論理（又は関数）の出力である。一実施態様では、状態生成論理は、現在の強度値312及び以前の強度値314、...、316の中から最大値を選択し、その最大値を現在のピクセル状態318として使用することができる。別の実施態様では、状態生成論理は、現在の強度値312及び以前の強度値314、...、316の中から最小値を選択し、その最小値を現在のピクセル状態318として使用することができる。更に別の実施態様では、状態生成論理は、現在の強度値312と以前の強度値314、...、316との平均（average）（平均（mean））を決定し、その平均を現在のピクセル状態318として使用することができる。更に別の実施態様では、状態生成論理は、現在の強度値312と以前の強度値31

40

50

4、．．． 3 1 6との移動（ランニング）平均を決定し、移動平均を現在のピクセル状態 3 1 8として使用することができる。更に別の実施態様では、状態生成論理は、現在の強度値 3 1 2及び以前の強度値 3 1 4、．．． 3 1 6の指数移動平均を決定し、指数移動平均を現在のピクセル状態 3 1 8として使用することができる。更に別の実施態様では、状態生成論理は、現在の強度値 3 1 2及び以前の強度値 3 1 4、．．． 3 1 6の標準偏差を決定し、標準偏差を現在のピクセル状態 3 1 8として使用することができる。このプロセスは、9個のピクセル1～9の各々について実施され、それぞれの現在のピクセル状態が9個のピクセル1～9の各々について決定されるようにする。図3に示す例によれば、9個のピクセル1～9について状態生成論理によって生成された現在のピクセル状態は、ピクセルごとの状態グリッド3 0 4を形成する。一実施態様では、状態生成論理は、状態生成器1 2 6によって実施される。

10

【0 0 7 6】

ピクセル単位の状態符号化

図4は、ピクセルごとの状態3 0 4に基づいてベースコールを生成する一実施態様を示す。一実施態様では、符号化論理4 0 2は、例えば、合計、連結、要素単位の乗算、又は要素単位の乗算及び合計（畳み込み）によって、ピクセルグリッド3 0 2 Nとピクセル単位の状態グリッド3 0 4とをピクセルごとに組み合わせる（4 0 0）。ピクセルグリッド3 0 2 Nとピクセルごとの状態グリッド3 0 4の組み合わせは、現在のベースコールデータ1 5 4を生成するために、ベースコーラー1 4 4によって処理される。

【0 0 7 7】

チャンネルごとの状態

図5は、ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとの状態5 0 0を生成する一実施態様を示す。図5は、例示的なピクセルPに関する。図5において、ピクセルPは、2つの例示的なチャンネル（又は強度チャンネル）、すなわちチャンネル1 5 0 2及びチャンネル2 5 0 6にわたる強度値を有する。他の実施態様では、ピクセルPは、より少ない又はより多い数のチャンネルを有することができる。図5に示す例では、第1のチャンネル5 0 2は、5つの配列決定サイクル1～5に対して強度値5 1 2、5 1 3、．．． 5 1 6の第1のセットを有する。同様に、第2のチャンネル5 0 6は、5つの配列決定サイクル1～5について第2の組の強度値5 2 2、5 2 3、．．． 5 2 6を有する。

20

30

【0 0 7 8】

次いで、強度値の第1のセットに基づいて、第1のチャンネル5 0 2について第1のチャンネル状態5 3 2が決定される。同様に、第2のチャンネル状態5 3 6は、強度値の第2のセットに基づいて第2のチャンネル5 0 6について決定される。第1及び第2のチャンネル状態5 3 2及び5 3 6は、例えば、最小値選択関数、最大値選択関数、平均関数、指数加重平均関数、移動平均関数、指数移動平均関数、又は標準偏差関数を実施することによって、状態生成論理によって決定されることができる。

【0 0 7 9】

次いで、少なくとも1つのベースコールが、第1のチャンネル状態5 3 2及び第2のチャンネル状態5 3 6に基づいて、第5の配列決定サイクル（サイクル5）のために生成される。これは、一実施態様では、サイクル5のチャンネル1強度値5 1 6を第1のチャンネル状態5 3 2と組み合わせてチャンネル1 5 0 2の第1の組み合わせを生成することと、サイクル5のチャンネル2強度値5 2 6を第2のチャンネル状態5 3 6と組み合わせてチャンネル2 5 0 6の第2の組み合わせを生成することと、を含む。次いで、ベースコーラー1 4 4は、チャンネル1 5 0 2についての第1の組み合わせ及びチャンネル2 5 0 6についての第2の組み合わせを処理して、サイクル5についての1つ以上のベースコール（例えば、1つ以上のクラスターについて）を生成する。

40

【0 0 8 0】

別の実施態様では、ベースコーラー1 4 4は、サイクル5のチャンネル1強度値5 1 6、サイクル5のチャンネル2強度値5 2 6、サイクル5の第1のチャンネル状態5 3 2、及びサ

50

イクル5の第2のチャンネル状態536を、4つの別個の入力チャンネル（例えば、RGBスタイル画像チャンネル）として処理し、サイクル5のベースコールを生成する。当業者は、データ／チャンネルを組み合わせ、それらを組み合わせとして処理する他の現代又は将来の方法が、本開示の範囲内で同様に適用され、同等に使用され得ることを理解するであろう。

【0081】

更に別の実施態様では、第1及び第2のチャンネル状態532及び536は、例えば、第1及び第2のチャンネル状態532及び536に平均化関数を適用することによって、いわゆる「パンチャンネル状態」を生成するために組み合わせられる。次いで、サイクル5のこのパンチャンネル状態は、サイクル5のチャンネル1強度値516及びサイクル5のチャンネル2強度値526とともに処理されて、サイクル5のベースコール、すなわち、チャンネルごとの実施態様では4つであるのに対して合計3つの別個の入力チャンネルを生成する。

【0082】

図5は、例示及び簡略化のために1つのピクセルPのみに関するが、状態生成一及び状態符号化一応答性ベースコールのステップは、入力配列決定画像中の複数のピクセルに対してピクセル単位及びチャンネル単位で実施されることが理解され、これは部分的には、クラスター強度プロファイル及びそれらの状態が、隣接するピクセルの群によって定義され、したがって、クラスターに対するベースコールを生成するために群として分析されるからである。

【0083】

チャンネルごとのMIN状態

図6は、ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとのMIN（最小）状態600を生成する一実施態様を示す。図6において、列601は、ベースコールについてのベースコーラー144によって処理される入力配列決定画像の複数のピクセルにインデックスを付ける。入力配列決定画像は、配列決定ランの現在の配列決定サイクルNにおけるクラスターのセットについて捕捉された強度プロファイルを表す。図6において、列602は、列601においてインデックス付けされたピクセル及び現在の配列決定サイクルNについての第1のチャンネル（画像チャンネル）におけるピクセル単位の強度値を表す。図6において、列603は、列601においてインデックス付けされたピクセル及び現在の配列決定サイクルNについての第2のチャンネル（画像チャンネル）におけるピクセル単位の強度値を表す。

【0084】

図6において、列604は、列601においてインデックス付けされたピクセルについて、第1の配列決定サイクル1と現在の配列決定サイクルNとの間の第1のチャンネルにおいて観察された強度値から選択された第1のチャンネルにおけるピクセル単位の最小強度値を表す。図6において、列605は、列601においてインデックス付けされたピクセルについて、第1の配列決定サイクル1と現在の配列決定サイクルNとの間の第2のチャンネルにおいて観察された強度値から選択された、第2のチャンネルにおけるピクセル単位の最小強度値を表す。

【0085】

一実施態様では、ベースコーラー144は、列602、603、604、及び605を別個の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

【0086】

パンチャンネルMIN状態

図7は、ベースコールについての入力として処理される配列決定画像内のピクセルについてパンチャンネルMIN（最小）状態700を生成する一実施態様を示す。図7において、列708は、図6の列604及び605をピクセル単位で平均化する。一実施態様では、ベースコーラー144は、列602、603、及び708を別個の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

【0087】

チャンネルごとのMAX状態

図8は、ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとのMAX（最大）状態800を生成する一実施態様を示す。図8において、列804は、列601においてインデックス付けされたピクセルについて、第1の配列決定サイクル1と現在の配列決定サイクルNとの間の第1のチャンネルにおいて観察された強度値から選択された第1のチャンネルにおけるピクセル単位の最大強度値を表す。図8において、列805は、列601においてインデックス付けされたピクセルについて、第1の配列決定サイクル1と現在の配列決定サイクルNとの間の第2のチャンネルにおいて観察された強度値から選択された第2のチャンネルにおけるピクセル単位の最大強度値を表す。

10

【0088】

一実施態様では、ベースコーラー144は、列602、603、804、及び805を別個の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

【0089】

パンチャネルMAX状態

図9は、ベースコールについての入力として処理される配列決定画像内のピクセルについてのパンチャネルMAX（最大）状態900を生成する一実施態様を示す。図9において、列908は、図8の列804及び805をピクセル単位で平均化する。一実施態様では、ベースコーラー144は、列602、603、及び908を別個の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

20

【0090】

チャンネルごとのAVG状態

図10は、ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとのAVG（平均）状態1000を生成する一実施態様を示す。図10において、列1004は、列601においてインデックス付けされたピクセルについて、第1の配列決定サイクル1と現在の配列決定サイクルNとの間の第1のチャンネルにおいて観察された強度値のそれぞれのピクセル単位の平均として計算された第1のチャンネルにおけるピクセル単位の平均強度値を表す。図10において、列1005は、列601においてインデックス付けされたピクセルについて第1の配列決定サイクル1と現在の配列決定サイクルNとの間で第2のチャンネルにおいて観察された強度値のそれぞれのピクセル単位の平均として計算された第2のチャンネルにおけるピクセル単位の平均強度値を表す。

30

【0091】

一実施態様では、ベースコーラー144は、列602、603、1004、及び1005を別個の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

【0092】

パンチャネルAVG状態

図11は、ベースコールについての入力として処理される配列決定画像内のピクセルについてのパンチャネルAVG（平均）状態1100を生成する一実施態様を示す。図11において、列1108は、図10の列1004及び1005をピクセル単位で平均化する。一実施態様では、ベースコーラー144は、列602、603、及び1108を別個の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

40

【0093】

チャンネルごとのMIN MAX状態

図12は、ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとのMIN/MAX（最小及び最大）状態1200を生成する一実施態様を示す。一実施態様では、ベースコーラー144は、列602、603、604、60

50

5、804、及び805を別々の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

【0094】

チャンネルごとのMIN AVG状態

図13は、ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとのMIN AVG（最小及び平均）状態1300を生成する一実施態様を示す。一実施態様では、ベースコーラー144は、列602、603、604、605、1004、及び1005を別々の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

【0095】

チャンネルごとのMAX AVG状態

図14は、ベースコールについての入力として処理される配列決定画像内のピクセルについてチャンネルごとのMAX AVG（最大及び平均）状態1400を生成する一実施態様を示す。一実施態様では、ベースコーラー144は、列602、603、804、805、1004、及び1005を別々の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

【0096】

チャンネルごとのMIN MAX AVG状態

図15は、ベースコールについての入力として処理される配列決定画像内のピクセルについて、チャンネルごとのMIN MAX AVG（最小、最大、及び平均）状態1500を生成する一実施態様を示す。一実施態様では、ベースコーラー144は、列602、603、604、605、804、805、1004、及び1005を別々の入力チャンネルとして処理して、現在の配列決定サイクルNについて1つ以上のベースコールを生成する。

【0097】

当業者であれば、上記で論じたチャンネル及び状態の任意の組み合わせ又は配列決定又は配置が、ベースコールについてのベースコーラー144への入力として提供され得ることを理解するであろう。上記で説明したチャンネル及び状態を組み合わせ、配列決定、配置する他の同時及び将来の方法は、本開示の範囲内である。例えば、パンチャンネル状態は、指数加重平均、移動平均、標準偏差、分散などの他の集約／累積関数を使用してチャンネルごとの状態を組み合わせることによって生成することができる。別の例では、チャンネルごとのMIN状態は、パンチャンネルのAVG MAX状態などと連結され得る。

【0098】

ベースコール指向状態生成

図16は、将来のベースコールにおいて使用するために、以前にコールされた塩基に依存して状態情報を生成する（1600）一実施態様を示す。図17A、17B、及び17Cは、現在の配列決定サイクル11（サイクル11）のためのベースコール指向ON及びOFF状態生成1700を示す。図18A、18B、及び18Cは、図17A、17B、及び17Cを拡張し、次の配列決定サイクル12（サイクル12）においてOFF状態がどのように更新されるか（1800）を示す。図19A、19B、及び19Cは、図18A、18B、及び18Cを更に拡大し、更に次の配列決定サイクル13（サイクル13）においてON状態がどのように更新されるか（1900）を示す。図20A、図20B、及び図20Cは、図19A、図19B、及び図19Cを更に拡大し、更に次の配列決定サイクル14（サイクル14）においてON状態及びOFF状態の両方がどのように更新されるか（2000）を示す。

【0099】

1つの蛍光色素（又は同じ若しくは同様の励起／発光スペクトルの2つ以上の色素）を使用する配列決定反応におけるヌクレオチド取り込みを検出するための異なる戦略を区別する1つの手段は、配列決定サイクル中に発生する蛍光遷移の存在又は相対的な欠如、又はその間のレベルに関して組み込みを特徴付けることによるものである。したがって、配

10

20

30

40

50

列決定戦略は、配列決定サイクルに対するそれらの蛍光プロファイルによって例示することができる。本明細書に開示される戦略の場合、「1」又は「0」は、ヌクレオチドが（例えば、蛍光によって検出可能な）信号状態にある蛍光状態（1／ON）、あるいはヌクレオチドが（例えば、撮像ステップで検出されないか、又は最小限にしか検出されない）暗状態にある蛍光状態（0／OFF）を示す。「0」、「OFF」、又は「暗」状態は、必ずしも信号の完全な欠如又は不在を指すとは限らない。いくつかの実施態様では、信号（例えば、蛍光）の全体的な欠如又は不在があり得るが、他の実施態様では、OFF状態であっても何らかの検出可能な信号が存在し得る。最小の又は減少した蛍光信号（例えば、背景信号）もまた、第1の画像から第2の画像へ（又はその逆）の蛍光の変化を確実に区別することができる限り、「0」、「OFF」、又は「暗」状態の範囲に含まれると考えられる。

10

【0100】

本明細書で使用される場合、用語「暗」又は「OFF」は、検出器によって検出されるバックグラウンド信号と比較して有意でない、検出器によって検出される所望の信号の量を言うことが意図される。例えば、物体の特徴は、その特徴の信号対雑音比が実質的に低い、例えば、1未満であるとき、暗又はOFFであると見なされ得る。いくつかの実施態様では、暗特徴又はOFF特徴は、いかなる量の所望の信号も生成しないことがある（すなわち、信号が生成又は検出されない）。他の実施態様では、バックグラウンドに対して非常に低い量の信号は、暗又はOFFと見なされ得る。

20

【0101】

1つの実施態様において、1つの蛍光色素（又は同じ若しくは類似の励起／発光スペクトルの2つの色素）及び2つの画像化事象を使用して、配列決定反応におけるヌクレオチド取り込みを検出及び決定するための例示的な戦略は、以下の検出表によって例示される。

【0102】

【表1】

	画像1	画像2
A	ON	OFF
C	OFF	ON
G	OFF	OFF
T	ON	ON

30

【0103】

ヌクレオチド取り込みを検出するための他の戦略（例えば、4チャンネル化学及び1チャンネル化学）は、本開示の範囲内であり、別々に議論されない。

【0104】

アクション1602において、開示された技術は、所与のピクセルについて、以前の配列決定サイクルにおいてコールされた、すなわち、以前にコールされた塩基の基礎となるチャンネル単位のアクティブ（ON）及び非アクティブ（OFF）分類にアクセスする。これは、図17Aの例によって示される。図17Aにおいて、ピクセルPは、サイクル1～10について塩基がコールされる対応するクラスターの強度発光を示す。いくつかの実施態様では、ピクセルPは、対応するクラスター中心の場所座標によって決定される、対応するクラスターの中心を含む。コールされた塩基は、第1のチャンネル（チャンネル1）内の強度値及び第2のチャンネル（チャンネル2）内の強度値に対するチャンネル単位のアクティブ（ON）及び非アクティブ（OFF）分類1754及び1756から生じる。

40

【0105】

アクション1612において、開示された技術は、所与のピクセル及び現在の配列決定サイクルについて、チャンネルごとのアクティブ（ON）及び非アクティブ（OFF）分類1754及び1756に基づいて、チャンネルごとのアクティブ（ON）及び非アクティブ（OFF）状態についてのチャンネル単位の要約統計値を蓄積する。これは、図17Bの例

50

によって示される。図 17B では、第 1 のチャンネル 1702 の非アクティブ (OFF) 状態は、非アクティブ (OFF) 状態に分類された列 1754 内の強度値に累積関数を適用することによって、配列決定サイクル 11 (サイクル 11) について決定される。累積関数の例には、最大値選択、最小値選択、平均 (average) (平均 (mean))、指数加重平均、移動 (ランニング) 平均、指数移動平均、モード、標準偏差、分散、歪度、尖度、パーセンタイル、及びエントロピーが含まれる。図 17B では、第 1 のチャンネル 1712 のアクティブ (ON) 状態は、アクティブ (ON) 状態に分類される列 1754 内の強度値に累積関数を適用することによって、サイクル 11 について決定される。図 17B では、第 2 のチャンネル 1722 の非アクティブ (OFF) 状態は、非アクティブ (OFF) 状態に分類される列 1756 内の強度値に累積関数を適用することによって、サイクル 11 について決定される。図 17B では、第 2 のチャンネル 1732 のアクティブ (ON) 状態は、アクティブ (ON) 状態に分類された列 1756 内の強度値に累積関数を適用することによって、サイクル 11 について決定される。

10

20

30

40

50

【0106】

アクション 1622 において、開示された技術は、所与のピクセル及び現在の配列決定サイクルについて、チャンネルごとのアクティブ (ON) 状態及び非アクティブ (OFF) 状態についての累積されたチャンネル単位の要約統計値を、所与のピクセルの現在の強度チャンネルと組み合わせる。これは、図 17C の例によって示される。図 17C では、1764 及び 1766 は、サイクル 11 においてピクセル P に対して登録された第 1 及び第 2 のチャンネルに対するそれぞれの強度値である。図 17C において、ベースコーラー 144 は、1764、1766、1702、1712、1722、及び 1732 を別個の入力チャンネルとして処理して、配列決定サイクル 11 においてピクセル P の対応するクラスターについてのベースコールを生成する。

【0107】

アクション 1632 において、開示された技術は、現在の配列決定サイクルについて、所与のピクセルのチャンネルごとのアクティブ (ON) 状態及び非アクティブ (OFF) 状態並びに強度チャンネルについての累積されたチャンネル単位の要約統計値の組み合わせに基づいて、少なくとも 1 つのベースコールを生成する。これは、図 18A の例によって示される。図 18A において、ベースコール「G」は、サイクル 11 において、「OFF」及び「OFF」チャンネル単位の分類 1854 及び 1856 に基づいてピクセル P の対応するクラスターに対して生成され、これらは、図 18C における処理 1764、1766、1702、1712、1722、及び 1732 に応答してベースコーラー 144 によって行われる。

【0108】

図 18B では、サイクル 11 における「OFF」及び「OFF」チャンネル単位の分類 1854 及び 1856 (及び「ON」分類の欠如) に応答して、第 1 及び第 2 のチャンネル 1802 及び 1822 の OFF チャンネル状態のみがサイクル 12 において更新され、その結果、累算関数が再適用されて、両方が「OFF」分類 1854 及び 1856 を有するサイクル 11 のチャンネル単位の強度値も考慮される。いくつかの実施態様では、チャンネル単位の ON 状態及び OFF 状態は、単一変数としてメモリ内に維持され、例えば、累積関数が平均化関数であったときに平均を計算するために使用された除数を乗算することによって、更新インスタンスごとに逆計算される。第 1 及び第 2 のチャンネル 1812 及び 1832 についての ON チャンネル状態は、サイクル 12 において不変に保たれ、サイクル 11 からの同じ値すなわち、1712 及び 1732 をそれぞれ継承する。

【0109】

図 18C において、1864 及び 1866 は、サイクル 12 でピクセル P に対して登録された第 1 及び第 2 のチャンネルに対するそれぞれの強度値である。図 18C において、ベースコーラー 144 は、1864、1866、1802、1812、1822、及び 1832 を別個の入力チャンネルとして処理して、配列決定サイクル 12 におけるピクセル P の対応するクラスターについてのベースコールを生成する。

【0110】

図19Aにおいて、ベースコール「T」は、「ON」及び「ON」チャンネル単位の分類1954及び1956に基づいて、サイクル12におけるピクセルPの対応するクラスターについて生成され、これらは、図18Cにおける処理1864、1866、1802、1812、1822、及び1832に回答して、ベースコーラー144によって行われる。

【0111】

図19Bでは、サイクル12における「ON」及び「ON」チャンネル単位の分類1954及び1956（及び「OFF」分類の欠如）に回答して、第1及び第2のチャンネル1912及び1932についてのONチャンネル状態のみがサイクル13において更新され、その結果、累算関数が再適用されて、両方が「ON」分類1954及び1956を有するサイクル12のチャンネル単位の強度値も考慮される。第1及び第2のチャンネル1902及び1922のOFFチャンネル状態は、サイクル13において不変に保たれ、サイクル12からの同じ値すなわち、1802及び1822をそれぞれ継承する。

【0112】

図19Cにおいて、1964及び1966は、サイクル13でピクセルPに対して登録された第1及び第2のチャンネルに対するそれぞれの強度値である。図19Cにおいて、ベースコーラー144は、1964、1966、1902、1912、1922、及び1932を別個の入力チャンネルとして処理して、配列決定サイクル13においてピクセルPの対応するクラスターについてのベースコールを生成する。

【0113】

図20Aにおいて、ベースコール「A」は、「ON」及び「OF」チャンネル単位の分類2054及び2056に基づいて、サイクル13においてピクセルPの対応するクラスターに対して生成され、これらは、図19Cにおける処理1964、1966、1902、1912、1922、及び1932に回答してベースコーラー144によって行われる。

【0114】

図20Bでは、サイクル13における「ON」及び「OFF」チャンネル単位の分類2054及び2056に回答して、第1及び第2のチャンネル2002、2012、2022、及び2032それぞれのON及びOFFチャンネル状態の両方がサイクル14において更新され、その結果、累積関数が再適用されて、「ON」及び「OFF」分類2054及び2056を有するサイクル13のチャンネル単位の強度値も考慮される。

【0115】

図20Cにおいて、2064及び2066は、サイクル14でピクセルPに対して登録された第1及び第2のチャンネルに対するそれぞれの強度値である。図20Cにおいて、ベースコーラー144は、2064、2066、2002、2012、2022、及び2032を別々の入力チャンネルとして処理して、配列決定サイクル14におけるピクセルPの対応するクラスターについてのベースコールを生成する。

【0116】

ベースコール指向状態生成では、状態情報が1配列決定サイクルだけ遅れることに留意されたい。すなわち、サイクル12をベースコールするために使用される状態情報1802、1812、1822、及び1832は、サイクル1～11に基づき、サイクル12を含まない。同様に、サイクル13をベースコールするために使用される状態情報1902、1912、1922、及び1932は、サイクル1～12に基づき、サイクル13を含まない。

【0117】

指数加重平均化

図21は、指数加重平均化を使用してベースコールについての状態情報を生成する（2100）一実施態様を示す。一実施態様では、状態情報を計算するときに、より最近の配列決定サイクルからの入力信号（例えば、強度値）により大きな重みを与えるために、指数加重平均化が使用される。その結果、指数加重平均化によって生成された状態情報推定

10

20

30

40

50

値は、信号の変化に追従し、外れ値に対してよりロバストである。

【0118】

アクション2102において、開示された技術は、所与のピクセルについて、以前の配列決定サイクル（例えば、最初の10回又は20回の配列決定サイクル）からの過去の強度値についての要約統計値を蓄積する。要約統計値を累積する累積関数の例には、最大値選択、最小値選択、平均（average）（平均（mean））、指数加重平均、移動（ランニング）平均、指数移動平均、モード、標準偏差、分散、歪度、尖度、パーセンタイル、及びエントロピーが含まれる。

【0119】

アクション2112において、開示された技術は、所与のピクセルについて、累積された要約統計値に基づいて開始指数加重平均を初期化する。一実施態様では、開始指数加重平均は、過去の強度値の平均（例えば、全ての他のピクセル又は全ての他のクラスターピクセルの過去の強度値の平均）、過去の強度値から選択された大域的 maximum（大域的 MAX）（例えば、全ての他のピクセル又は全ての他のクラスターピクセルの過去の強度値の大域的 MAX）、又は過去の強度値から選択された大域的 minimum（大域的 MIN）（例えば、他の全ての画素又は他の全てのクラスターピクセルの過去の強度値の大域的 MIN）であり得る。

【0120】

アクション2122において、開示された技術は、所与のピクセル及び現在の配列決定サイクルについて、開始指数加重平均と現在のピクセル強度値との加重組み合わせに基づいて現在の指数加重平均を決定する。

【0121】

アクション2132において、開示された技術は、現在の指数加重平均を、現在の配列決定サイクルにおけるベースコールについての現在の状態値として使用する例えば、所与のピクセルに対応するクラスターをベースコールする。一実施態様では、これは、ベースコーラー144が現在の状態値及び現在のピクセル強度値を処理し、現在の配列決定サイクルについて少なくとも1つのベースコールを生成することを含む。

【0122】

アクション2142において、開示された技術は、所与のピクセル及び次の配列決定サイクルについて、現在の指数加重平均と次のピクセル強度値との加重組み合わせに基づいて、次の指数加重平均を決定する。

【0123】

アクション2152において、開示された技術は、次の配列決定サイクルにおけるベースコール（例えば、所与のピクセルに対応するクラスターのベースコール）についての次の状態値として次の指数加重平均を使用する。一実施態様では、これは、ベースコーラー144が次の状態値及び次のピクセル強度値を処理し、次の配列決定サイクルについての少なくとも1つのベースコールを生成することを含む。

【0124】

一実施態様では、指数加重平均論理は、以下のように表される。

$$y[k] = (1 - \alpha) * y[k-1] + \alpha * x[k]$$

式中、

$y[k]$ は、現在の配列決定サイクルの指数加重平均であり、

$y[k-1]$ は、以前の配列決定サイクルの指数加重平均であり、

$x[k]$ は、現在の配列決定サイクルの入力信号（例えば、強度値）であり、

アルファは、加重パラメータであり、

【0125】

加重パラメータ「アルファ」は、0と1との間の値であり得る $\{0 \leq \alpha \leq 1\}$ 。アルファが0に等しいとき $\{\alpha = 0\}$ 出力は、 $y[k-1]$ （平均化なし）である。アルファが1に等しいとき $\{\alpha = 1\}$ 、出力は、 $x[k]$ である。他のアルファ値では、出力は強度の指数加重平均である。アルファは、フィルタが更新に対してどれだ

10

20

30

40

50

け迅速に反応するかを表す。より小さい値はより遅く反応し、より多くの平均化を行うが、より大きい値はより最近の入力値により多くの重みを置く。

【0126】

なお、 $y[k]$ は、上述の式を更新するときを除いて使用されないで、単一の変数として記憶され、適所で更新され得る。 $y[k-1]$ は、例えば、予想される強度（例えば、全てのクラスター／ウェルにわたる平均初期強度）に初期化され得る。一実施態様では、最初の20回の配列決定サイクル（サイクル1～20）を使用して、オンライン推定なしで初期平均強度を推定することができる。初期平均強度は、 $y[k-1]$ についての初期推定値としてプラグインすることができ、次いで、オンライン指数加重平均化を、配列決定サイクル20（サイクル20）後に使用することができる。

10

【0127】

CPU上で実施する場合、上記の式を次のように書き換えることにより、1回の乗算で効率的に実現できる。

$$y[k] = y[k-1] * \text{アルファ} (x[k] - y[k-1])$$

【0128】

いくつかの実施態様では、アルファは0に近い小さい値であり得、したがって、より多くの平均化が実行される。他の実施態様では、特定の用途に最も適した v/s 応答時間を平均化することをトレードオフする値を見つけるために、アルファを変更することができる。

【0129】

20

チャンネルごとの指数加重平均化

図22は、指数加重平均を使用してベースコールについてのチャンネルごとの状態情報を生成する(2200)一実施態様を示す。図23は、指数加重平均を使用してベースコールについてのチャンネルごとの状態情報を生成する(2300)例を示す。

【0130】

アクション2202において、開示された技術は、所与のピクセルについて、以前の配列決定サイクル例えば、最初の10回又は20回の配列決定サイクルからのチャンネルごとの過去の強度値についてのチャンネル単位の要約統計値を蓄積する。チャンネル単位の要約統計値を累積するチャンネルごとの累積関数の例には、最大／大域的 maximum 値選択、最小／大域的最小値選択、平均 (average) (平均 (mean))、指数加重平均、移動 (ランニング) 平均、指数移動平均、モード、標準偏差、分散、歪度、尖度、パーセンタイル、及びエントロピーが含まれる。これは、図23の例によって示される。

30

【0131】

図23において、第1のチャンネル (チャンネル1) は、最初の5回の配列決定サイクル (サイクル1～5) について過去の強度値2374を有する。図23では、第2のチャンネル (チャンネル2) は、サイクル1～5について過去の強度値2376を有する。更に、大域的 maximum (大域的 MAX) 値2312が、過去の強度値2374からチャンネル1に対して選択され、大域的最小 (大域的 MIN) 値2314が、過去の強度値2374からチャンネル1に対して選択される。同様に、大域的 MAX 値2316は、過去の強度値2376からチャンネル2に対して選択され、大域的 MIN 値2318は、過去の強度値2376からチャンネル2に対して選択される。

40

【0132】

動作2212において、開示された技術は、チャンネルごとに、蓄積されたチャンネル単位の要約統計値に基づいて、開始アクティブ (ON) 及び非アクティブ (OFF) 状態の対を初期化する。これは、図23の例によって示される。図23において、チャンネル1及びサイクル5の後、大域的 MAX 2312は、チャンネル1のアクティブ (ON) 状態として初期化され、大域的 MIN 2314は、チャンネル1の非アクティブ (OFF) 状態として初期化される。図23では、チャンネル2及びサイクル5の後、大域的 MAX 2316は、チャンネル2のアクティブ (ON) 状態として初期化され、大域的 MIN 2318は、チャンネル2の非アクティブ (OFF) 状態として初期化される。

50

【0133】

動作2222において、開示された技術は、所与のピクセル及び現在の配列決定サイクルについて、チャンネル単位で、現在のチャンネル強度値を、開始アクティブ（ON）及び非アクティブ（OFF）状態の対との比較に基づいて、開始アクティブ（ON）状態又は開始非アクティブ（OFF）状態に帰属させる。これは、図23の例によって示される。図23では、チャンネル1及びサイクル6について、強度値2322が登録される。強度値2322は大域的MAX 2312に量的により近いので、大域的MAX 2312はチャンネル1のアクティブ（ON）状態のプロキシとして機能し、アクティブ（ON）状態におけるチャンネル1の開始指数加重平均として使用されるので、強度値2322はアクティブ（ON）状態に帰属する。図23では、チャンネル2及びサイクル6について、強度値2326が登録される。

10

【0134】

強度値2326は大域的MIN 2318に量的により近いので、大域的MIN 2318はチャンネル2の非アクティブ（OFF）状態のプロキシとして機能し、非アクティブ（OFF）状態におけるチャンネル2の開始指数加重平均として使用されるので、強度値2326は非アクティブ（OFF）状態に帰属する。

【0135】

アクション2232において、開示された技術は、所与のピクセルについて、所与のチャンネル内で、指数加重平均を使用して、現在のチャンネル強度値と帰属状態の以前の状態値との加重組み合わせに基づいて帰属状態を更新する。

20

【0136】

アクション2242において、開示された技術は、所与のピクセルについて、所与のチャンネル内で、未帰属状態の以前の状態値からの未帰属状態を維持する。

【0137】

アクション2232及び2242は、図23に例として示されている。図23において、サイクル6について、状態情報は、チャンネル1及び2について以下のように決定される。サイクル6及びチャンネル1のアクティブ（ON）状態2342は、指数加重平均2332を使用して決定される。なぜなら、サイクル6及びチャンネル1の強度値2322は、チャンネル1のアクティブ（ON）状態に帰属するからである。指数加重平均2332は、上述の式に関して説明したように、アルファ値として「0.12」を使用する。チャンネル1の非アクティブ（OFF）状態2344は変更されずに維持され、チャンネル1の大域的MIN 2314としてサイクル5から継承される。

30

【0138】

サイクル6及びチャンネル2の非アクティブ（OFF）状態2348は、指数加重平均2338を使用して決定される。なぜなら、サイクル6及びチャンネル2の強度値2326は、チャンネル2の非アクティブ（OFF）状態に帰属するからである。指数加重平均2338も、上述の式に関して説明したように、アルファ値として「0.12」を使用する。チャンネル2のアクティブ（ON）状態2346は、変更されずに維持され、チャンネル2の大域的MAX 2316としてサイクル5から継承される。

【0139】

次に、サイクル6のチャンネル単位のアクティブ（ON）及び非アクティブ（OFF）状態2342、2344、2346、及び2348が、サイクル6について登録された強度値2322及び2326とともに使用されて、サイクル6について少なくとも1つのベースコールが生成される。

40

【0140】

図23において、サイクル7について、状態情報は、チャンネル1及び2について以下のように決定される。サイクル7及びチャンネル1の非アクティブ（OFF）状態2374は、指数加重平均2364を使用して決定される。なぜなら、サイクル7及びチャンネル1の強度値2352は、チャンネル1の非アクティブ（OFF）状態に帰属するからである。指数加重平均2364も、上述の式に関して説明したように、アルファ値として「0.12

50

」を使用する。チャンネル 1 のアクティブ（ON）状態 2 3 7 2 は、不変に保たれ、チャンネル 1 の指数加重平均 2 3 4 2 としてサイクル 6 から継承される。

【0 1 4 1】

サイクル 7 及びチャンネル 2 のアクティブ（ON）状態 2 3 7 6 は、指数加重平均 2 3 6 6 を使用して決定される。なぜなら、サイクル 7 及びチャンネル 2 の強度値 2 3 5 6 は、チャンネル 2 のアクティブ（ON）状態に帰属するからである。指数加重平均 2 3 6 6 も、上述の式に関して説明したように、アルファ値として「0.12」を使用する。チャンネル 2 の非アクティブ（OFF）状態 2 3 7 8 は、変更されずに保たれ、チャンネル 2 の指数加重平均 2 3 4 8 としてサイクル 6 から継承される。

【0 1 4 2】

次に、サイクル 7 のチャンネル単位のアクティブ（ON）及び非アクティブ（OFF）状態 2 3 7 2、2 3 7 4、2 3 7 6、及び 2 3 7 8 が、サイクル 7 について登録された強度値 2 3 5 2 及び 2 3 5 6 とともに使用されて、サイクル 7 について少なくとも 1 つのベースコールが生成される。

【0 1 4 3】

図 2 3 において、所与の配列決定サイクルにおけるアクティブ（ON）状態及び非アクティブ（OFF）状態へのチャンネル単位の属性は、所与の配列決定サイクル（例えば、図 2 3 のサイクル 6 及び 7）について登録されたチャンネルごとの強度値に基づくことに留意されたい。次に、所与の配列決定サイクルにおけるアクティブ（ON）及び非アクティブ（OFF）状態へのチャンネル単位の属性が、以前の配列決定サイクルで行われたベースコーラに基づく、ベースコール指向のチャンネルごとの指数加重平均化について説明する。

【0 1 4 4】

ベースコール指向のチャンネルごとの指数加重平均化

図 2 4 は、以前にコールされた塩基によって指向される指数加重平均化を使用して、ベースコールについてのチャンネルごとの状態情報を生成する（2 4 0 0）一実施態様を示す。図 2 5 は、以前にコールされた塩基によって指向される指数加重平均化を使用して、ベースコールについてのチャンネルごとの状態情報を生成する（2 5 0 0）例を描写する。

【0 1 4 5】

図 2 4 において、動作 2 4 0 2 及び 2 4 1 2 は、図 2 2 の動作 2 2 0 2 及び 2 2 1 2 と同様である。アクション 2 4 2 2 において、アクティブ（ON）及び非アクティブ（OFF）状態へのチャンネル単位の属性は、1 配列決定サイクルだけ遅れ、以前の配列決定サイクルにおいてコールされた塩基に基づいて作成される。これは、図 2 5 の例によって示される。図 2 5 では、サイクル 6 において、ベースコール「A」2 5 0 2 が、それぞれチャンネル 1 及び 2 の「ON」分類 2 5 0 4 及び「OFF」分類 2 5 0 6 に基づいて行われる。

【0 1 4 6】

アクション 2 4 3 2 において、サイクル 6 における「ON」及び「OFF」分類 2 5 0 4 及び 2 5 0 6 並びにベースコール 2 5 0 2 は、指数加重平均化を適用することによってサイクル 7 の状態情報 2 3 4 2、2 3 4 4、2 3 4 6、及び 2 3 4 8 を決定するために使用される。アクション 2 4 4 2 において、サイクル 7 の未帰属状態も、サイクル 6 におけるベースコール 2 5 0 2 に基づいて識別され、変更されないまま維持される。

【0 1 4 7】

サイクル 7 では、サイクル 6 で決定された状態情報 2 3 4 2、2 3 4 4、2 3 4 6、及び 2 3 4 8、並びにサイクル 7 で登録されたチャンネルごとの強度値 2 3 7 2 及び 2 3 7 6 に基づいて、「OFF」及び「ON」分類 2 5 1 4 及び 2 5 1 6、並びに次にベースコール「C」2 5 1 2 が生成されることに留意されたい。

【0 1 4 8】

続いて、サイクル 7 における「OFF」及び「ON」分類 2 5 1 4 及び 2 5 1 6 並びにベースコール 2 5 1 2 は、指数加重平均化を適用することによってサイクル 8 についての状態情報 2 3 7 2、2 3 7 4、2 3 7 6、及び 2 3 7 8 を決定するために使用される。サイクル 8 の未帰属状態も、サイクル 7 におけるベースコール 2 5 1 2 に基づいて識別され

10

20

30

40

50

、変更されないまま維持される。

【0149】

図示されていないが、サイクル8についてのベースコールは、サイクル8について登録されたチャンネル単位の強度値と、サイクル8についてサイクル7で決定された状態情報2372、2374、2376、及び2378とに基づいて生成される。

【0150】

異なる実施態様では、指数加重平均（EWA）を使用して決定された状態情報は、最小値選択論理、最大値選択論理、平均化論理など、何らかの他の論理を使用して決定された状態情報の代わりに、それに加えて、又はそれと組み合わせて使用され得る。

【0151】

非クラスターピクセル

次に、クラスターピクセルと非クラスターピクセルとの区別について説明する。クラスターピクセルは、クラスター中心の場所座標によって決定されるクラスターの中心を含むピクセルである。非クラスターピクセルはクラスター中心を含まない。非クラスターピクセルはクラスター強度（又は背景強度）を示すことに留意されたい。それらはクラスター中心と一致しない。

【0152】

クラスターピクセルーベースコール関係の簡単な説明もここでは有用である。クラスターは、複数のピクセルによって、例えば、3×3ピクセルのピクセルグリッドによって示される。いくつかの実施態様では、ベースコーラー144は、異なる配列決定サイクルにおける所与のクラスターのそれぞれの強度プロファイルの特徴付けるために、所与のピクセルグリッド内のピクセルについて異なる強度値を有する配列決定画像を処理する。処理に応答して、一実施態様では、ベースコーラー144は、所与のクラスターの中心を含む所与のピクセルグリッド内の中心ピクセルのみを参照することによって、異なる配列決定サイクルにおける所与のクラスターについてのそれぞれのベースコールを指定する出力を生成する。すなわち、所与のピクセルグリッド全体が所与のクラスターの強度プロファイルの特徴付けるとしても、それぞれのベースコールは、中心ピクセルのみに関して行われる。所与のピクセルグリッドの非中心ピクセルは、ベースコールを生成するためにベースコーラー144によって分析される。しかしながら、中心ピクセルのみがベースコールを表現するために使用される。

【0153】

ベースコールを表現するために中心ピクセルのみが使用される実施態様では、非クラスターピクセルの状態を決定するためにベースコールが利用可能でないので、ベースコール指向状態生成は困難であり得る。この制限は、以下に説明する技術によって補償することができる。

【0154】

図26は、以前にコールされた塩基を使用して非クラスターピクセルの状態情報を生成する（2600）一実施態様を示す。クラスター中心を含むピクセル、すなわち、クラスターピクセル2602については、例えば、図16～図20C並びに図24及び図25に関して上述したように、クラスターの以前のベースコールを使用して状態データを生成する（2612）。

【0155】

クラスター中心を含まないピクセル、すなわち、非クラスターピクセル2608については、一実施態様によれば、最も近いクラスターピクセルについての状態データが、状態データを生成するために使用される（2618）。別の実施態様では、全ての他のクラスターピクセルについての状態データの平均が、非クラスターピクセル2608の状態データを生成するために使用される（2628）。

【0156】

状態入力

図27は、ベースコールについての状態データ2700を生成する異なる実施態様を示

10

20

30

40

50

す。開示された技術は、「エンティティ」2702についての状態情報2722を、その現在の構成*i*及び過去の構成*i*−1、*i*−2、...、1に基づいて生成する。エンティティ2702の例は、任意のタイプの信号測定値、ピクセル測定値、電圧測定値、現在の測定値、pHスケール測定値、処理の中間出力、畳み込み特徴例えば、特徴マップ、圧縮特徴例えば、圧縮特徴マップ）、異なるタイプの出力（例えば、例えば、softmaxスコア、シグモイドスコア、回帰スコア）、ベースコール、及びリードを含む。次に、状態情報2722及びエンティティ2702の現在の構成*i*に応じて、現在のベースコールが行われる。

【0157】

状態情報2722をエンティティ2702と組み合わせるために、いくつかの実施態様では、状態情報2722の次元数は、エンティティ2702の次元数と調整される必要があり得ることに留意されたい（例えば、マッチングによって互換性をもたせる）、又はその逆も同様である。異なる実施態様では、これは、クローニング、パディング（例えば、0パディング）、連結、畳み込み、合計、転置畳み込みなどのような次元数変更動作によって達成され得る。例えば、状態情報2722が1×1の次元数を有し、エンティティ2702が3×3の次元数を有する場合、状態情報2722の9つのクローンがエンティティ2702と連結され得る。

【0158】

供給状態入力

図28は、ベースコール動作の処理の様々な態様における供給状態データ2800の異なる実施態様を示す。一実施態様では、ベースコーラー144は、様々な処理モジュール1〜*n*（例えば、前処理層、ニューラルネットワーク層、後処理層、出力層）を有することができる。状態データ2802は、ベースコール2808を生成するために、ベースコーラー144の処理モジュール1〜*n*のいずれかに提供され得る。例えば、状態データ2802は、ベースコーラー144の第1の処理モジュール（例えば、第1の畳み込み層）による処理のために入力データ2802（例えば、入力画像）と組み合わせることができる。別の例では、状態データ2802は、ベースコーラー144の中間出力2812（例えば、先行する畳み込み層によって生成された中間特徴マップ）と組み合わせることができる。更に別の例では（図示せず）、状態データ2802は、ベースコーラー144の最終出力（例えば、最終畳み込み層によって生成された最終特徴マップ）と組み合わせることができる。そのような実施態様では、状態データ2802と最終出力との組み合わせは、ベースコール2808を生成するために、出力層／ベースコーラー層（例えば、ソフトマックス層、回帰層、シグモイド層）によって処理され得る。

【0159】

状態ベース及びニューラルネットワークベースのベースコーラー

以下の考察は、本明細書に記載のニューラルネットワークベースのベースコーラー2900に焦点を当てている。ニューラルネットワークベースのベースコーラー2900は、ベースコーラー144の一実施態様であり、本明細書ではまとめて「DeepRTA」と呼ばれる。まず、一実施態様に従って、ニューラルネットワークベースのベースコーラー2900への入力を説明する。次いで、ニューラルネットワークベースのベースコーラー2900の構造及び形態の例を示す。最後に、一実施態様による、ニューラルネットワークベースのベースコーラー2900の出力を説明する。

【0160】

データフロー論理は、ベースコールについての、ニューラルネットワークベースのベースコーラー2900に配列決定画像を提供する。ニューラルネットワークベースのベースコーラー2900は、パッチごとに（又はタイルごとに）配列決定画像にアクセスする。パッチのそれぞれは、配列決定画像を形成するピクセル化されたユニットのグリッド内のピクセル化ユニットのサブグリッド（又はサブアレイ）である。パッチは、ピクセル化ユニットのサブグリッドの*q*個×*r*個の寸法を有し、*q*（幅）及び*r*（高さ）は、1〜10000の範囲の任意の数である（例えば、3×3、5×5、7×7、10×10、15×

10

20

30

40

50

15、 25×25 、 64×64 、 78×78 、 115×115 ）。いくつかの実施態様では、 q 及び r は同じである。他の実施態様では、 q 及び r は互いに異なる。いくつかの実施態様では、1つの配列決定画像から抽出されたパッチは、同じサイズである。他の実施態様では、パッチは、異なるサイズのものである。いくつかの実施態様では、パッチどうしは、（例えば、エッジの上で）重なり合うピクセル化ユニットを有することができる。

【0161】

シーケンシングにより、対応する m 個の画像チャンネルに対して、配列決定サイクルごとに、 m 個の配列決定画像を生成する。すなわち、配列決定画像の各々は、1つ以上の画像（又は強度）チャンネル（カラー画像の赤、緑、青（RGB）のチャンネルと類似）を有する。一実施態様では、各画像チャンネルは、複数のフィルタ波長帯域のうちの1つに対応する。別の実施態様では、各画像チャンネルは、配列決定サイクルにおける複数の撮像事象のうちの1つに対応する。更に別の実施態様では、各画像チャンネルは、特定のレーザによる照射と特定の光学フィルタを通した撮像との組み合わせに対応する。画像パッチは、特定の配列決定サイクルについて、 m 個の画像チャンネルの各々からタイリング（又はアクセス）される。4-、2-、及び1-チャンネル化学などの異なる実施態様では、 m は、4又は2である。他の実施態様では、 m は、1、3、又は4よりも大きい。他の実施態様では、画像は、赤色チャンネル及び緑色チャンネルの代わりに、又はそれに加えて、青色チャンネル及び紫色チャンネル内にあり得る。

【0162】

例えば、配列決定ランが、2つの異なる画像化チャンネル、すなわち、青色チャンネルと緑色チャンネルとを使用して実施されると考えてみる。すると、各配列決定サイクルにおいて、配列決定ランは、青色画像及び緑色画像を生成する。このようにして、配列決定ランの k 個の配列決定サイクルのシリーズについて、 k 対の青色画像及び緑色画像のシーケンスが出力として生成され、配列決定画像として記憶される。したがって、 k 対の青色及び緑色画像パッチのシーケンスが、ニューラルネットワークベースのベースコーラー2900によるパッチレベル処理に対して生成される。

【0163】

ベースコールの反復1回（又はフォワードパス又は単一の順方向横断の1つのインスタンス）に対する、ニューラルネットワークベースのベースコーラー2900への入力画像データは、複数の配列決定サイクルを含むスライディングウィンドウに対するデータを含む。スライディングウィンドウは、例えば、現在の配列決定サイクル、1つ以上の先行する配列決定サイクル、及び1つ以上の後続の配列決定サイクルを含むことができる。

【0164】

一実施態様では、入力画像データは、ベースコールされる現在の（時点 t ）配列決定サイクルのデータが、（ i ）左隣接／コンテキスト／以前の／先行する／前の（時点 $t-1$ ）配列決定サイクルのデータ、及び（ i ）右隣接／コンテキスト／次の／連続する／後続の（時点 $t+1$ ）配列決定サイクルのデータを伴うように、3回の配列決定サイクルのデータを含む。

【0165】

別の実施態様では、入力画像データは、5回の配列決定サイクルのデータを含み、ベースコールされる現在の（時間 t ）配列決定サイクルのデータは、（ i ）第1の左隣接／コンテキスト／以前の／先行する／前の（時間 $t-1$ ）配列決定サイクルのデータと、（ i ）第2の左隣接／コンテキスト／以前の／先行する／前の（時間 $t-2$ ）配列決定サイクルのデータと、（ i ）第1の右隣接／コンテキスト／次の／連続する／後続の（時間 $t+1$ ）配列決定サイクルのデータと、（ i ）第2の右隣接／コンテキスト／次の／連続する／後続の（時間 $t+2$ ）配列決定サイクルのデータと、を伴う。

【0166】

更に別の実施態様では、入力画像データは、7回の配列決定サイクルのデータを含み、ベースコールされる現在の（時間 t ）配列決定サイクルのデータは、（ i ）第1の左隣接／コンテキスト／以前の／先行する／前の（時間 $t-1$ ）配列決定サイクルのデータと、

(i i) 第2の左隣接／コンテキスト／以前の／先行する／前の(時間 $t-2$)配列決定サイクルのデータと、(i i i) 第3の左隣接／コンテキスト／以前の／先行する／前の(時間 $t-3$)配列決定サイクルのデータと、(i v) 第1の右隣接／コンテキスト／次の／連続する／後続の(時間 $t+1$)配列決定サイクルのデータと、(v) 第2の右隣接／コンテキスト／次の／連続する／後続の(時間 $t+2$)配列決定サイクルのデータと、(v i) 第3の右隣接／コンテキスト／次の／連続する／後続の(時間 $t+3$)配列決定サイクルのデータと、を伴う。他の実施態様では、入力画像データは、単一の配列決定サイクルのデータを含む。更に他の実施態様では、入力画像データは、10回、15回、20回、30回、58回、75回、92回、130回、168回、175回、209回、225回、230回、275回、318回、325回、330回、525回、又は625回の配列決定サイクルのデータを含む。

10

【0167】

1つの実施態様によれば、ニューラルネットワークベースのベースコーラー2900は、その畳み込み層を介して画像パッチを処理し、代替表現を生成する。次いで、その代替表現は、出力層(例えば、ソフトマックス層)によって使用され、現在の配列決定サイクル(時間 t)、又は配列決定サイクルの各々(すなわち、現在の配列決定サイクル(時間 t)、第1及び第2の先行する配列決定サイクル(時間 $t-1$ 、時間 $t-2$)、及び第1及び第2の後続の配列決定サイクル(時間 $t+1$ 、時間 $t+2$))のいずれかに対するベースコールを生成する。得られたベースコールは、配列決定リードを形成する。

【0168】

一実施態様では、ニューラルネットワークベースのベースコーラー2900は、特定の配列決定サイクルに対する単一のターゲットクラスターのためのベースコールを出力する。別の実施態様では、ニューラルネットワークベースのベースコーラー2900は、特定の配列決定サイクルで複数のターゲットクラスター内の各ターゲットクラスターに対するベースコールを出力する。更に別の実施態様では、ニューラルネットワークベースのベースコーラー2900は、複数の配列決定サイクル内の各配列決定サイクルで、複数のターゲットクラスター内の各ターゲットクラスターに対するベースコールを出力し、それによって、各ターゲットクラスターに対するベースコールシーケンスを生成する。

20

【0169】

一実施態様では、ニューラルネットワークベースのベースコーラー2900は、多層パーセプトロン(Multilayer Perceptron、MLP)である。別の実施態様では、ニューラルネットワークベースのベースコーラー2900は、フィードフォワードニューラルネットワークである。更に別の実施態様では、ニューラルネットワークベースのベースコーラー2900は、完全に接続されたニューラルネットワークである。更なる実施態様では、ニューラルネットワークベースのベースコーラー2900は、完全畳み込みニューラルネットワークである。更に別の実施態様では、ニューラルネットワークベースのベースコーラー2900は、セマンティックセグメンテーションニューラルネットワークである。なお別の更なる実施態様では、ニューラルネットワークベースのベースコーラー2900は、敵対的生成ネットワーク(generative adversarial network、GAN)である。更に別の実施態様では、ニューラルネットワークベースのベースコーラー2900は、Transformer、BERT、及びDETRのようなマルチヘッドアテンション機構を含む。

30

40

【0170】

一実施態様では、ニューラルネットワークベースのベースコーラー2900は、複数の畳み込み層を有する畳み込みニューラルネットワーク(CNN)である。別の実施態様では、ニューラルネットワークベースのベースコーラー2900は、長い短期メモリネットワーク(long short-term memory network、LSTM)、双方向LSTM(bi-directional LSTM、Bi-LSTM)、又はゲートされた反復単位(gated recurrent unit、GRU)などの反復ニューラルネットワーク(recurrent neural network、RNN)である。更に別の実施態様では、ニューラルネットワークベースのベースコーラ

50

ー 2 9 0 0 は、C N N 及び R N N の両方を含む。

【 0 1 7 1 】

更に他の実施態様では、ニューラルネットワークベースのベースコーラー 2 9 0 0 は、1 D 畳み込み、2 D 畳み込み、3 D 畳み込み、4 D 畳み込み、5 D 畳み込み、拡張又は膨張畳み込み、転置畳み込み、深さ単位分離可能畳み込み、点単位畳み込み、 1×1 畳み込み、グループ畳み込み、平坦化 (flattened) 畳み込み、空間及びクロスチャネル (spatial and cross-channel) 畳み込み、シャッフルグループ化 (shuffled grouped) 畳み込み、Transformer、BERT、空間的分離可能畳み込み、並びに逆畳み込みを使用することができる。ニューラルネットワークベースのベースコーラー 2 9 0 0 は、ロジスティック回帰／対数損失、多クラスクロスエントロピー／ソフトマックス損失、二値クロスエントロピー損失、平均二乗誤差損失、L1 損失、L2 損失、平滑 L1 損失、及び Huber 損失などの、1 つ以上の損失機能を使用することができる。ニューラルネットワークベースのベースコーラー 2 9 0 0 は、TFRecord、圧縮符号化 (例えば、PNG)、シャープ化、マップ変換に対する平行コール、バッチング、プリフェッチ、モデル並列、データ並列、及び同期／非同期確率的勾配降下法 (stochastic gradient descent、SGD) のような、任意の並列、効率、及び圧縮方式を使用することができる。ニューラルネットワークベースのベースコーラー 2 9 0 0 は、アップサンプリング層、ダウンサンプリング層、反復接続、ゲート及びゲートされたメモリユニット (LSTM 又は GRU など)、残留ブロック、残留接続、ハイウェイ接続、スキップ接続、ピープホール接続、活性化関数 (例えば、非線形変換関数 (正規化線形ユニット (rectifying linear unit、ReLU)、リーキー ReLU、指数関数的ライナーユニット (exponential liner unit、ELU)、シグモイド、及び双曲線正接 (tanh)) など)、バッチ正規化層、規則化層、ドロップアウト、プール層 (例えば、最大又は平均プール)、グローバル平均プール層、及び注意機構のような非線形変換機能を含み得る。

【 0 1 7 2 】

ニューラルネットワークベースのベースコーラー 2 9 0 0 は、逆伝播ベースの勾配更新技術を使用して学習する。ニューラルネットワークベースのベースコーラー 2 9 0 0 が学習するために使用され得る例示的な勾配降下技術としては、確率的勾配降下、バッチ勾配降下、及びミニバッチ勾配降下が挙げられる。ニューラルネットワークベースのベースコーラー 2 9 0 0 が学習するために使用され得る勾配降下最適化アルゴリズムのいくつかの例としては、Momentum、Nesterov 加速勾配、Adagrad、Adadelata、RMSprop、Adam、AdaMax、Nadam、及び AMSGrad が挙げられる。

【 0 1 7 3 】

一実施態様では、ニューラルネットワークベースのベースコーラー 2 9 0 0 は、異なる配列決定サイクルに対するデータの処理を分離するために、専用のアーキテクチャを使用する。上記の専用アーキテクチャを使用する動機をまず説明する。上述のように、ニューラルネットワークベースのベースコーラー 2 9 0 0 は、現在の配列決定サイクル、1 つ以上の先行する配列決定サイクル、及び 1 つ以上のその後の配列決定サイクルに対する画像パッチを処理する。追加の配列決定サイクルのデータは、配列固有のコンテキストを提供する。ニューラルネットワークベースのベースコーラー 2 9 0 0 は、学習中にシーケンスごとに固有のコンテキストを学習し、それらをベースコールする。更に、事前及び事後配列決定サイクルのデータは、プレフェージング及びフェージング信号の二次の寄与を現在の配列決定サイクルに提供する。

【 0 1 7 4 】

しかしながら、異なる配列決定サイクルで、かつ異なる画像チャネル内に捕捉される画像は、位置合わせ不良であり、互いに残留位置合わせ誤差を有する。この位置合わせ不良を考慮するために、専用アーキテクチャは、配列決定サイクル間では情報を混合せず、同一の配列決定サイクル内でのみ情報を混合する、空間畳み込み層を含む。

【 0 1 7 5 】

10

20

30

40

50

空間畳み込み層（又は空間ロジック）は、畳み込みの「専用の非共有」シーケンスを介して複数の配列決定サイクルの各々に対して独立してデータを処理することによって分離を操作する、いわゆる「分離された畳み込み」を使用する。分離された畳み込みは、任意の他の配列決定サイクルのデータ及び得られた特徴マップ上で畳み込むことなく、所与の配列決定サイクル、すなわち、サイクル内のみのデータ及び得られた特徴マップ上で畳み込む。

【0176】

例えば、入力画像データが、(i) ベースコールされる現在の（時間 t ）配列決定サイクルに対する現在の画像パッチ、(ii) 以前の（時間 $t-1$ ）配列決定サイクルに対する以前の画像パッチ、及び (iii) 次の（時間 $t+1$ ）配列決定サイクルに対する次の画像パッチと、を含むと考える。次いで、専用アーキテクチャは、3つの別個の畳み込みパイプライン、すなわち、現在の畳み込みパイプライン、以前の畳み込みパイプライン、及び次の畳み込みパイプラインを開始する。現在のデータ処理パイプラインは、現在の（時間 t ）配列決定サイクルに対する現在の画像パッチを入力として受信し、複数の空間畳み込み層を介して独立してそれを処理して、最終空間畳み込み層の出力としていわゆる「現在の空間畳み込み表現」を生成する。以前の畳み込みパイプラインは、以前の（時間 $t-1$ ）配列決定サイクルに対する以前の画像パッチを入力として受信し、複数の空間畳み込み層を介して独立してそれを処理して、最終空間畳み込み層の出力としていわゆる「以前の空間畳み込み表現」を生成する。次の畳み込みパイプラインは、次の（時間 $t+1$ ）配列決定サイクルに対する次のデータを入力として受信し、複数の空間畳み込み層を介して独立してそれを処理して、最終空間畳み込み層の出力としていわゆる「次の空間畳み込み表現」を生成する。

【0177】

いくつかの実施態様では、現在、以前、及び次の畳み込みパイプラインは、並行して実行される。いくつかの実施態様では、空間畳み込み層は、専用アーキテクチャ内の空間畳み込みネットワーク（又はサブネットワーク）の一部である。

【0178】

ニューラルネットワークベースのベースコーラー 2900 は、配列決定サイクル間、すなわち、サイクル間で情報を混合する時間畳み込み層（又は時相論理）を更に含む。時間畳み込み層は、空間畳み込みネットワークからそれらの入力を受信し、それぞれのデータ処理パイプラインに対して最終空間畳み込み層によって生成される空間畳み込み表現で動作する。

【0179】

時間畳み込み層のサイクル間動作性自由度は、空間畳み込みネットワークへの入力として供給される画像データ内に存在する位置合わせ不良特性が、空間畳み込み層のシーケンスによって実行される、分離された畳み込みのスタック又はカスケードによって空間畳み込み表現からパージされるという事実から生じる。

【0180】

時間畳み込み層は、スライディングウィンドウベースでの後続の入力で入力チャンネル上でグループごとに畳み込む、いわゆる「組み合わせ畳み込み」を使用する。一実施形態では、後続の入力は、以前の空間畳み込み層又は以前の時間畳み込み層によって生成される後続の出力である。

【0181】

いくつかの実施態様では、時間畳み込み層は、専用アーキテクチャ内の時間畳み込みネットワーク（又はサブネットワーク）の一部である。時間畳み込みネットワークは、空間畳み込みネットワークからその入力を受信する。一実施態様では、時間畳み込みネットワークの第1の時間畳み込み層は、配列決定サイクル間の空間畳み込み表現をグループごとに組み合わせる。別の実施態様では、時間畳み込みネットワークの後続の時間畳み込み層は、以前の時間畳み込み層の後続する出力を組み合わせる。最終時間畳み込み層の出力は、出力を生成する出力層に供給される。出力は、1つ以上の配列決定サイクルで1つ以上

のクラスターをベースコールするために使用される。

【0182】

ニューラルネットワークベースのベースコーラー2900に関する更なる詳細は、参照により本明細書に組み込まれる、2019年3月21日出願の「ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する米国特許仮出願第62/821,766号（代理人整理番号ILLM1008-9/IP-1752-PRV）に見出すことができる。

【0183】

ピクセルごとの状態入力

図29は、状態ベース及びニューラルネットワークベースのベースコールの一実施態様を示す。図29に示す例では、配列決定サイクルの第1のウィンドウは、配列決定サイクル $N-2$ 、 $N-1$ 、 N 、 $N+1$ 、及び $N+2$ を含む。一実施態様では、それぞれの配列決定サイクル $N-2$ 、 $N-1$ 、 N 、 $N+1$ 、及び $N+2$ についてのそれぞれの画像パッチ（配列決定画像／クラスター画像）2914、2924、2934、2944、及び2954は、現在の配列決定サイクル N についての状態データ2912とピクセル単位で組み合わせられる（例えば、連結される）。状態データ2912は、上述の技術のいずれかを使用して決定することができる。

【0184】

画像パッチ2914、2924、2934、2944、2954をまとめてクラスター画像（配列決定画像）2950と呼ぶ。ピクセルごと、チャンネルごとの状態データ2940は、5つの配列決定サイクル $N-2$ 、 $N-1$ 、 N 、 $N+1$ 、及び $N+2$ があるので、状態データ2912の5つのインスタンスを含む。

【0185】

状態データ2912は、図3及び図4に関して上述した実施態様と同様に、クラスター画像2950内の対応するピクセルに対するピクセルごとの状態値を有する。また、一実施態様では、クラスター画像2950が2つのチャンネル（強度チャンネル）を有するので、状態データ2912も2つの状態チャンネルを有する。他の実施態様では、状態データ2912は、異なる実施態様において上記で説明したように、1つのチャンネルのみを有するか、又は3つ以上のチャンネルを有することができる。

【0186】

高レベルでは、ピクセルごと、チャンネルごとの状態データ2940が、空間論理106（又は空間ネットワーク若しくは空間サブネットワーク若しくは空間畳み込みニューラルネットワーク）による処理のためにクラスター画像2950と組み合わせられる。ピクセルごと、チャンネルごとの状態データ2940及びクラスター画像2950の組み合わせは、空間論理106を介して別々に処理されて、それぞれの配列決定サイクル $N-2$ 、 $N-1$ 、 N 、 $N+1$ 、及び $N+2$ に対するそれぞれの空間マップ2916、2926、2936、2946、及び2956（又は中間結果又は空間出力セット又は空間特徴マップセット）を生成する。空間畳み込みネットワーク106は、1D、2D、又は3D畳み込みを使用することができる。

【0187】

空間論理106は、空間畳み込み層のシーケンス（又はカスケード）を含む。各空間畳み込み層は、分離された畳み込みを実装する複数の空間畳み込みフィルタを有するフィルタバンクを有する。したがって、各空間畳み込み層は、複数の空間特徴マップを出力として生成する。ある対象のサブジェクト空間畳み込み層によって生成される空間特徴マップの数は、その対象の空間畳み込み層に構成された空間畳み込みフィルタの数の関数である。例えば、対象の空間畳み込み層が14個の空間畳み込みフィルタを有する場合、対象の空間畳み込み層は14個の空間特徴マップを生成する。集合的観点から、14個の空間特徴マップは、14個のチャンネル（又は深さ次元=14）との空間特徴マップの体積（又はテンソル）と見なすことができる。

【0188】

更に、対象の空間畳み込み層に続く次の空間畳み込み層も、14個の空間畳み込みフィルタで構成することができる。そのような場合、次の空間畳み込み層は、入力として、14個の空間特徴マップを処理して、対象の空間畳み込み層を生成し、その対象の空間畳み込み層自体が、14個の新しい空間特徴マップを出力として生成する。図29は、それぞれの配列決定サイクル $N-2$ 、 $N-1$ 、 N 、 $N+1$ 、及び $N+2$ について空間ネットワーク106の最終空間畳み込み層によって生成された5つの空間特徴マップセット2916、2926、2936、2946、及び2956を示す。例えば、5つの空間特徴マップセット2916、2926、2936、2946、及び2956のそれぞれは、14個の特徴マップを有する。

【0189】

10

別の例では、空間ネットワーク106の7つの空間畳み込み層のカスケードによって、7つの空間特徴マップセットのシーケンスを生成することができる。対象配列決定サイクル i についてのサイクルごとの入力パッチデータ及び状態データの組み合わせは、例えば、 115×115 の空間次元数及び2（元の配列決定画像の2つの画像チャンネルに起因して）の深度次元数を有することができる。一実施態様では、7つの空間畳み込み層の各々は、後続の空間特徴マップボリュームの空間次元数を2だけ、例えば、 10×10 から 8×8 に低減する、 3×3 の畳み込みを使用する。

【0190】

第1の空間特徴マップボリュームは、 113×113 の空間次元数（すなわち、第1の空間畳み込み層の 3×3 の畳み込みによって 115×115 から減少したもの）と、14の深さ次元数（すなわち、第1の空間畳み込み層内の14個の空間畳み込みフィルタによる、14個の特徴マップ又は14個のチャンネル）と、を有し得る。第2の空間特徴マップボリュームは、 111×111 の空間次元数（すなわち、第2の空間畳み込み層の 3×3 の畳み込みによって 113×113 から減少したもの）と、14の深さ次元数（すなわち、第2の空間畳み込み層内の14個の空間畳み込みフィルタによる14個の特徴マップ又は14個のチャンネル）と、を有し得る。第3の空間特徴マップボリュームは、 109×109 の空間次元数（すなわち、第3の空間畳み込み層の 3×3 の畳み込みによって 111×111 から減少したもの）と、14の深さ次元数（すなわち、第3の空間畳み込み層内の14個の空間畳み込みフィルタによる14個の特徴マップ又は14個のチャンネル）と、を有し得る。第4の空間特徴マップボリュームは、 107×107 の空間次元数（すなわち、第4の空間畳み込み層の 3×3 の畳み込みによって 109×109 から減少したもの）と、14の深さ次元数（すなわち、第4の空間畳み込み層内の14個の空間畳み込みフィルタによる14個の特徴マップ又は14個のチャンネル）と、を有し得る。第5の空間特徴マップボリュームは、 105×105 の空間次元数（すなわち、第5の空間畳み込み層の 3×3 の畳み込みによって 107×107 から減少したもの）と、14の深さ次元数（すなわち、第5の空間畳み込み層内の14個の空間畳み込みフィルタによる14個の特徴マップ又は14個のチャンネル）と、を有し得る。第6の空間特徴マップボリュームは、 103×103 の空間次元数（すなわち、第6の空間畳み込み層の 3×3 の畳み込みによって 105×105 から減少したもの）と、14の深さ次元数（すなわち、第6の空間畳み込み層内の14個の空間畳み込みフィルタによる14個の特徴マップ又は14個のチャンネル）と、を有し得る。第7の空間特徴マップボリュームは、 101×101 の空間次元数（すなわち、第7の空間畳み込み層の 3×3 の畳み込みによって 103×103 から減少したもの）と、14の深さ次元数（すなわち、第7の空間畳み込み層内の14個の空間畳み込みフィルタによる14個の特徴マップ又は14個のチャンネル）と、を有し得る。

【0191】

図29に示されるマルチサイクルの例に類似して、5つの配列決定サイクル $N-2$ 、 $N-1$ 、 N 、 $N+1$ 、及び $N+2$ 、並びに5つのサイクルごとの画像パッチ2914、2924、2934、2944、及び2954、並びに状態データ2912の5つのインスタンスについて、空間論理106は、7つの空間特徴マップボリュームの5つのそれぞれのシーケンスを別々に生成し、図29の空間マップ2916、2926、2936、294

50

6、及び2956は、最終的な空間特徴マップボリュームの5つの別々のインスタンスと等価である。

【0192】

空間マップ2916、2926、2936、2946、及び2956は、時相論理2928によって処理されて、現在の配列決定サイクルNについてのベースコール2930を生成する。

【0193】

クラスターごと／ウェルごとの状態

図30は、ベースコールクラスターについてのクラスターごとの状態データを使用する一実施態様を示す。図29とは異なり、図30では、状態データは、入力クラスター画像2914、2924、2934、2944、及び2954と組み合わせられない。代わりに、状態データは、空間論理106の出力と組み合わせられる。クラスター画像2914、2924、2934、2944、及び2954は、クラスターピクセル及び非クラスターピクセルの両方を含むことに留意されたい。結果として、空間マップ3016、3026、3036、3046、及び3056は、クラスターピクセル及び非クラスターピクセルの両方から発する空間畳み込み特徴を含む。

【0194】

いくつかの実施態様では、クラスター特徴フィルタリング論理3030は、クラスター中心場所3002を使用して、非クラスターピクセルに対応する空間マップ3016、3026、3036、3046、及び3056からそれらの特徴をフィルタ除去する。結果として得られるフィルタリングされたクラスターごとの空間マップ3054、3064、3074、3084、及び3094は、クラスター特徴3004と総称されるクラスターピクセルに対応する特徴のみを含む。一実施態様では、空間マップ3016、3026、3036、3046、及び3056の次元数は 101×101 であり、フィルタリングされたクラスターごとの空間マップ3054、3064、3074、3084、及び3094の次元数は 25×25 である。そのような実施態様では、 $25 \times 25 \times k$ テンソルは、元のクラスターピクセルと、クラスターピクセルの連続的に生成された畳み込み特徴とを含む、クラスターピクセルに対応するデータに連続畳み込み演算を適用した結果として生成される空間特徴を含む。

【0195】

ニューラルネットワークベースのベースコーラー2900への開始入力において状態データ2940がクラスター画像2950と組み合わせられる図29とは異なり、図30では、現在の配列決定サイクルNのクラスターごとの状態3042が、ニューラルネットワークベースのベースコーラー2900の中間出力、すなわち、フィルタリングされたクラスターごとの空間マップ3054、3064、3074、3084、及び3094と組み合わせられる。

【0196】

いくつかの実施態様では、クラスターごとの状態3042の次元数は、フィルタリングされたクラスターごとの空間マップ3054、3064、3074、3084、及び3094の次元数と一致するように修正される（例えば、次元数をトリミングすること、次元数を追加すること、パディングすること（例えば、0パディングすること）、クローニングすることなどによって）。異なる実施態様では、クラスター単位の状態3042並びにフィルタリングされたクラスター単位の空間マップ3054、3064、3074、3084、及び3094のそれぞれの5つのインスタンスは、連結、合計、要素単位の乗算、要素単位の乗算及び合計（畳み込み）など、上記で説明した技法を使用してそれぞれ組み合わせることができる。

【0197】

次いで、図30に示される例によれば、それぞれのフィルタリングされたクラスターごとの空間マップ3054、3064、3074、3084、及び3094のそれぞれの5つの組み合わせ、並びにクラスターごとの状態3042のそれぞれの5つのインスタンス

10

20

30

40

50

が、時相論理 2 9 2 8 によって処理されて、現在の配列決定サイクル N についてのベースコール 3 0 6 0 が生成される。ベースコール 3 0 6 0 は、クラスターピクセル、したがってクラスター特徴 3 0 0 4 に対応するクラスターについて作成される。

【 0 1 9 8 】

次に、クラスターごとの状態 3 0 4 2 の異なる変形例、及びこれらの変形例が開示された技術の異なる実施態様に従ってどのように決定されるかについて説明する。

【 0 1 9 9 】

図 3 1 は、対応するクラスターピクセルの過去の強度値を使用してクラスターごとの状態を生成する (3 1 0 0) 一実施態様を示す。図 3 1 において、クラスターごとの状態 3 0 4 2 は、上述したように、例えば、最小値選択関数、最大値選択関数、平均化関数、指数加重平均化関数などを使用して、以前の配列決定サイクル及び現在の配列決定サイクルからのクラスターピクセルの過去の強度値の要約統計値を蓄積することによって決定される。次いで、クラスターごとの状態 3 0 4 2 は、クラスターピクセル及び現在のベースコール反復／演算の空間畳み込み特徴と組み合わせられて、対応するクラスターについてのベースコールを生成する。これらのステップは、図 3 1 のアクション 3 1 0 2、3 1 1 2、3 1 2 2、及び 3 1 3 2 によって示される。

【 0 2 0 0 】

図 3 2 は、クラスターピクセルに対応する空間畳み込み特徴の過去の特徴値を使用してクラスターごとの状態を生成する (3 2 0 0) 一実施態様を示す。図 3 1 において、クラスターごとの状態 3 0 4 2 は、上述したように、例えば、最小値選択関数、最大値選択関数、平均化関数、指数加重平均化関数などを使用して、以前の配列決定サイクル及び現在の配列決定サイクルからのクラスターピクセルに対応する空間畳み込み特徴の過去の特徴値の要約統計値を蓄積することによって決定される。次いで、クラスターごとの状態 3 0 4 2 は、クラスターピクセル及び現在のベースコール反復／演算の空間畳み込み特徴と組み合わせられて、対応するクラスターについてのベースコールを生成する。これらのステップは、図 3 2 のアクション 3 2 0 2、3 2 1 2、3 2 2 2、及び 3 2 3 2 によって示される。

【 0 2 0 1 】

図 3 3 は、ピクセル強度を補間することによってクラスター強度を生成し (3 3 0 0)、補間されたクラスター強度を使用して、ベースコールについてのクラスターごとの状態を生成する一実施態様を示す。一実施態様では、配列決定画像内のピクセルのグループの強度値を補間して (例えば、線形補間、双線形補間、三次補間を使用することによって)、クラスター強度を推定することができる。いくつかの実施態様では、クラスター中心場所及びクラスター形状データを使用して、ピクセルの所与のグループを選択し、対応するクラスターのクラスター強度を補間することができる。他の実施態様では、クラスター形状データを使用して、部分ピクセルの加重された強度を使用してクラスター強度を決定することができる。

【 0 2 0 2 】

例えば、最小値選択関数、最大値選択関数、平均化関数、指数加重平均化関数等を使用して、前述のように、以前の配列決定サイクル及び現在の配列決定サイクルにおいて補間されたクラスター強度の要約統計値を累積して、現在の配列決定サイクルのクラスターごとの状態 3 0 4 2 を生成することができる。次いで、クラスターごとの状態 3 0 4 2 は、クラスターピクセル及び現在のベースコール反復／演算の空間畳み込み特徴と組み合わせられて、対応するクラスターについてのベースコールを生成する。これらのステップは、図 3 2 のアクション 3 3 0 2、3 3 1 2、3 3 2 2、及び 3 3 3 2 によって示される。

【 0 2 0 3 】

クラスターごとの状態としての R T A 係数

図 3 4 は、ニューラルネットワークベースのベースコーラー 2 9 0 0 とは別個のリアルタイム分析 (R T A) ベースコーラーを使用して、クラスターごとの状態を生成する (3 4 0 0) 一実施態様を示す。

【0204】

R T A ベースコーラーに関して、R T A は、線形強度抽出器を使用して、ベースコールについての配列決定画像から特徴を抽出することができるベースコーラーである。以下の議論は、R T A による強度抽出及びベースコールの1つの実施態様を説明する。この実施態様では、R T A は、テンプレートサイクルと呼ばれるいくつかの数の初期配列決定サイクルからの配列決定画像を使用してタイル上のクラスターの場所を識別するテンプレート画像を生成するためにテンプレート生成ステップを実行することができる。テンプレート画像は、その後の位置合わせ及び強度抽出ステップのための基準として使用することができる。テンプレート画像は、テンプレートサイクルの各配列決定画像内の輝点を検出してマージすることによって生成することができ、これは、配列決定画像を鮮鋭化し（例えば、ラプラシアン畳み込みを使用して）、空間的に分離された O t s u 手法によって「ON」閾値を決定し、その後、サブピクセル場所補間を用いて5ピクセル局所最大検出を行うことを含む。別の例では、タイル上のクラスターの場所は、基準マーカーを使用して識別することができる。生物学的試料が画像化される固体支持体は、固体支持体に付着されたプローブに対する試料又はその画像の配向の決定を容易にするために、そのような基準マーカーを含むことができる。例示的な基準は、限定されないが、ビーズ（蛍光部分又は標識プローブが結合され得る核酸などの部分を有するか又は有しない）、既知の又は決定可能な特徴に結合された蛍光分子、又は形態学的形状を蛍光部分と組み合わせる構造を含む。例示的な基準は、参照により本明細書に組み込まれる米国特許出願公開第2002/0150909号に記載されている。

10

20

【0205】

次いで、R T A は、テンプレート画像に対して現在の配列決定画像を登録することができる。これは、画像相関を使用して現在の配列決定画像をサブ領域上のテンプレート画像に整列させることによって、又は非線形変換（例えば、完全6パラメータ線形アフィン変換）を使用することによって達成することができる。

【0206】

R T A は、配列決定画像のカラーチャネル間のクロストークを補正するためにカラーマトリックスを生成することができる。R T A は、経験的な位相補正を実施して、位相誤差によって引き起こされる配列決定画像中のノイズを補償することができる。

【0207】

異なる補正が配列決定画像に適用された後、R T A は、配列決定画像における各スポット場所についての信号強度を抽出することができる。例えば、所与のスポット場所について、信号強度は、スポット場所内のピクセルの強度の加重平均を決定することによって抽出され得る。例えば、中心ピクセル及び隣接ピクセルの加重平均は、バイリニア又はバイキュービック補間を用いて実行されてもよい。いくつかの実施態様では、画像内の各スポット場所は、数ピクセル（例えば、1～5ピクセル）を備えてもよい。

30

【0208】

次に、R T A は、抽出された信号強度を空間的に正規化して、サンプリングされた画像全体にわたる照明の変動を説明することができる。例えば、強度値は、5番目及び95番目のパーセンタイルがそれぞれ0及び1の値を有するように正規化され得る。画像の正規化された信号強度（例えば、各チャネルの正規化された強度）を使用して、画像内の複数のスポットの平均鮮明度を計算することができる。

40

【0209】

いくつかの実施態様では、R T A は、抽出された信号強度の信号対雑音比を最大化するために等化器を使用することができる。等化器は、配列決定画像におけるクラスター強度データの信号対雑音比を最大化するように、（例えば、最小二乗推定、適応等化アルゴリズムを使用して）訓練することができる。一実施態様では、等化器は、空間クロストーク並びに他の形態のクロストーク及びノイズを補正するための訓練された係数を含むことができる。いくつかの実施態様では、等化器は単一のルックアップテーブル（LUT）であり得る。他の実施態様では、等化器は、「等化器フィルタ」又は「畳み込みカーネル」と

50

も呼ばれる、サブピクセル解像度をもつ複数の L U T をもつ L U T バンクであり得る。一実施態様では、等化器内の L U T の数は、配列決定画像のピクセルが分割され得るサブピクセルの数に依存し得る。例えば、ピクセルが $n \times n$ サブピクセル（例えば、 5×5 サブピクセル）に分割可能である場合、等化器は n^2 個の L U T（例えば、 25 個の L U T）を生成することができる。更なる実施態様では、等化器は高速フーリエ変換（Fast Fourier Transform、FFT）に基づくことができる。更なる実施態様では、等化器は、Winograd 畳み込みに基づくことができる。

【0210】

訓練の一実施態様では、配列決定画像からのデータは、ウェルサブピクセル場所によってビン化される。例えば、 5×5 L U T の場合、ウェルの $1/25$ 番目の中心はビン（1，1）にあり（例えば、センサピクセルの左上隅）、ウェルの $1/25$ 番目はビン（1，2）にあり、以下同様である。一実施態様では、各ビンについての等化器係数は、それぞれのビンに対応するウェルからのデータのサブセットに対する最小二乗推定を使用して決定される。このようにして、結果として得られる推定等化器係数は、ビンごとに異なる。

【0211】

各 L U T / 等化器フィルタ / 畳み込みカーネルは、訓練から学習された複数の係数を有する。一実施態様では、L U T 内の係数の数は、クラスターをベースコールするために使用されるピクセルの数に対応する。例えば、クラスターをベースコールするために使用されるピクセル（画像又はピクセルパッチ）の局所グリッドがサイズ $p \times p$ （例えば、 9×9 ピクセルパッチ）の場合、各 L U T は p^2 の係数（例えば、 81 の係数）を有する。

【0212】

一実施態様では、訓練は、信号対ノイズ比を最大化するように、ベースコールされるターゲットクラスターからの強度放射及び 1 以上の隣接クラスターからの強度放射を表すピクセルの強度値を混合／結合するように構成されている等化器係数を生成する。信号対ノイズ比において最大化された信号は、ターゲットクラスターからの強度放射であり、信号対ノイズ比において最小化されたノイズは、隣接するクラスターからの強度放射、すなわち、空間クロストークにいくつかのランダムノイズを（例えば、背景強度放射を説明するために）加えたものである。等化器係数は重みとして使用され、混合／結合は、等化器係数とピクセルの強度値との間で要素単位の乗算を実行して、ピクセルの強度値の重み付き和を計算すること、すなわち畳み込み演算を含む。

【0213】

RTA は、最適化された強度データに数学的モデルを適合させることによってベースコールを行うことができる。使用され得る適切な数学的モデルは、例えば、 k 平均クラスターリングアルゴリズム、 k 平均様クラスターリングアルゴリズム、期待値最大化クラスターリングアルゴリズム、ヒストグラムベースの方法などを含む。4 つのガウス分布は、1 つの分布がデータセットにおいて表される 4 つのヌクレオチドの各々に適用されるように、2 チャネル強度データのセットに適合され得る。1 つの特定の実施態様では、期待値最大化（expectation maximization、EM）アルゴリズムが適用され得る。EM アルゴリズムの結果として、各 X 、 Y 値（それぞれ 2 つのチャネル強度の各々を参照する）について、ある X 、 Y 強度値が、データが適合される 4 つのガウス分布のうちの 1 つに属する尤度を表す値を生成することができる。4 つの塩基が 4 つの別個の分布を与える場合、各 X 、 Y 強度値はまた、4 つの塩基の各々に対して 1 つずつ、4 つの関連する尤度値を有する。4 つの尤度値の最大値は、ベースコールを示す。例えば、クラスターが両方のチャネルにおいて「OFF」である場合、ベースコールは G である。クラスターが 1 つのチャネルにおいて「OFF」であり、別のチャネルにおいて「ON」である場合、ベースコールは C 又は T のいずれかであり（どちらのチャネルが ON であるかに応じて）、クラスターが両方のチャネルにおいて「ON」である場合、ベースコールは A である。

【0214】

RTA についての更なる詳細は、2018 年 3 月 1 日に出願された「OPTICAL DISTORTION CORRECTION FOR IMAGED SAMPLES

10

20

30

40

50

」という名称の米国特許非仮出願第15／909,437号に見出すことができる。2014年10月31日に出願された「IMAGE ANALYSIS USEFUL FOR PATTERNED OBJECTS」と題する米国特許非仮出願第14／530,299号、2014年12月3日に提出された「METHODS AND SYSTEMS FOR ANALYZING IMAGE DATA」と題する米国特許非仮出願第15／153,953号、2011年1月13日に提出された「DATA PROCESSING SYSTEM AND METHODS」と題する米国特許非仮出願第13／006,206号、及び2021年5月4日に提出された「EQUALIZATION-BASED IMAGE PROCESSING AND SPATIAL CROSS TALK ATTENUATOR」と題する米国特許非仮出願第17／308,035号（代理人整理番号ILLM 1032-2/IP-1991-US）であり、これらは全て、本明細書に完全に記載されているかのように参照により組み込まれる。

10

【0215】

クラスター集団内の多数（例えば、数千個、数百万個、数億個など）のクラスターの強度プロファイルにおけるクラスター間強度プロファイル変動は、ベースコールスループットの低下及びベースコール誤差率の増加を引き起こす。クラスター間強度プロファイル変動を補正するために、RTAは、クラスターごとにクラスターごとの変動補正係数を生成する。

【0216】

2チャンネルの実施態様では、クラスターごとの変動補正係数は、クラスター間強度プロファイル変動におけるスケール変動を考慮する増幅係数と、クラスター間強度プロファイル変動における第1及び第2の強度チャンネルに沿ったシフト変動をそれぞれ考慮する2チャンネル固有オフセット係数と、を含む。別の実装態様では、シフト変動は、異なる強度チャンネル（例えば、第1及び第2の強度チャンネル）に共通オフセット係数を使用することによって考慮される。

20

【0217】

RTAの変動補正論理は、1チャンネルの実施態様、4チャンネルの実施態様などを使用して生成された配列決定画像に同様に適用され得ることは、当業者には明らかであろう。例えば、4チャンネルの実施態様の場合、4チャンネル固有オフセット係数が、それぞれ4つの強度チャンネルのシフト変動を補正するために決定される。

30

【0218】

ターゲットクラスターのクラスターごとの変動補正係数は、配列決定ランの先行する配列決定サイクルでターゲットクラスターについて決定された過去の強度統計値の分析と、現在の配列決定サイクルでターゲットクラスターについて決定された現在の強度統計値の分析との組み合わせに基づいて、配列決定ランの現在の配列決定サイクルで生成される。クラスターごとの変動補正係数は、配列決定ランの次の配列決定サイクルでターゲットクラスターに登録された次の強度読み取り値を補正するために使用される。補正された次の強度読み取り値は、次の配列決定サイクルでターゲットクラスターをベースコールするために使用される。配列決定ランの連続する配列決定サイクルでそれぞれのクラスターのそれぞれの強度プロファイルにそれぞれのクラスターごとの変動補正係数を繰り返し適用した結果、強度プロファイルは一致するようになり、原点に（例えば、台形の下側隅にある）固定される。

40

【0219】

現在の配列決定サイクル

現在の配列決定サイクル*i*において、シーケンサは配列決定画像を生成する。配列決定画像は、クラスター集団内の複数のクラスターに登録された現在の強度データを含むとともに、現在の配列決定サイクル*i*でターゲットクラスターに登録された現在の強度データを含む。

【0220】

現在の強度データはRTAに提供される。RTAは、現在の強度データを処理し、現在

50

の配列決定サイクル i における標的クラスターについての現在のベースコールを生成する。

【0221】

現在の配列決定サイクル i では、ターゲットクラスターの強度プロファイルは、現在の強度データと、現在の配列決定サイクル i に先行する配列決定ランの配列決定サイクル、すなわち、先行する配列決定サイクル $1 \sim i-1$ でターゲットクラスターに登録された現在の過去の強度データと、を含む。現在の強度データ及び現在の過去の強度データを総称して、現在の利用可能な強度データとする。

【0222】

強度プロファイルでは、4つの強度分布は、4つの塩基 A、C、T、及び G に対応する。一実施態様では、現在のベースコールは、4つの強度分布のうち、現在の強度データが属する強度分布を決定することによって行われる。いくつかの実施態様では、これは、期待値最大化アルゴリズムを使用することによって達成される。期待値最大化アルゴリズムは、現在利用可能な強度データに最も適合する平均（重心）及び分布（共分散）を観測する尤度を繰り返し最大化する。

【0223】

期待値最大化アルゴリズムを使用することによって、現在の配列決定サイクル i で4つの強度分布が決定されると、4つの強度分布のそれぞれに属する現在の強度データの尤度が計算される。最も大きい尤度が現在のベースコールを与える。一例として、「m, n」が、それぞれ、第1及び第2の強度チャンネルにおける現在の強度データの強度値であるとする。期待値最大化アルゴリズムは、4つの強度分布のそれぞれに属する「m, n」の強度値の尤度を表す4つの値を生成する。4つの値の最大値は、呼び出された塩基を識別する。

【0224】

他の実施態様では、k-means クラスタリングアルゴリズム、k-means 様クラスタリングアルゴリズム、ヒストグラムベースの方法などがベースコールに使用され得る。

【0225】

次の配列決定サイクル

次の配列決定サイクル $i+1$ で、強度補正パラメータ決定器は、現在のベースコールに基づいて、ターゲットクラスターについての強度補正パラメータを決定する。2チャンネルの実施態様では、強度補正パラメータは、第1の強度チャンネルにおける分布強度、第2の強度チャンネルにおける分布強度、第1の強度チャンネルにおける強度誤差、第2の強度チャンネルにおける強度誤差、分布重心から原点の距離、及び分布強度対強度誤差の類似性測定を含む。

【0226】

強度補正パラメータのそれぞれを以下のように定義する。

1) 第1の強度チャンネルにおける分布強度は、現在の配列決定サイクル i でターゲットクラスターが属する塩基固有の強度分布の重心での第1の強度チャンネルにおける強度値である。塩基固有の強度分布は、現在のベースコールを呼び出すための基礎であることに留意されたい。

2) 第2の強度チャンネルにおける分布強度は、塩基固有の強度分布の重心での第2の強度チャンネルにおける強度値である。

3) 第1の強度チャンネルにおける強度誤差は、第1の強度チャンネルにおける現在の強度データの測定強度値と第1の強度チャンネルにおける分布強度との間の差である。

4) 第2の強度チャンネルにおける強度誤差は、第2の強度チャンネルにおける現在の強度データの測定強度値と第2の強度チャンネルにおける分布強度との間の差である。

5) 分布重心から原点の距離は、塩基固有の強度分布の重心と、塩基固有の強度分布が（例えば、期待値最大化アルゴリズムを使用することによって）適合された多次元空間の原点との間のユークリッド距離である。他の実施態様では、マハラノビス距離及び最小共

10

20

30

40

50

分散行列式 (minimum covariance determinant、MCD) 距離などの距離メトリック、並びにそれらの関連する重心推定値が使用され得る。

6) 分布強度対強度誤差の類似性測定は、第1及び第2の強度チャンネルにおける分布強度と強度誤差との間のチャンネル単位のドット積の総和である。

【0227】

累積強度補正パラメータ決定器は、先行する配列決定サイクル $i-1$ からの過去の累積強度補正パラメータとともに強度補正パラメータを累積して、累積強度補正パラメータを決定する。累積の例としては、合計及び平均化が挙げられる。

【0228】

変動補正係数決定器は、決定された累積強度補正パラメータに基づいて、変動補正係数を決定する。

【0229】

次の配列決定サイクル $i+1$ で、シーケンサは配列決定画像を生成する。配列決定画像は、クラスター集団内の複数のクラスターに登録された次の強度データを含むとともに、次の配列決定サイクル $i+1$ でターゲットクラスターに登録された次の強度データを含む。

【0230】

強度補正器は、変動補正係数を次の強度データに適用して、補正された次の強度データを生成する。

【0231】

次の配列決定サイクル $i+1$ では、ターゲットクラスターの強度プロファイルは、補正された次の強度データと、次の配列決定サイクル $i+1$ に先行する配列決定ランの配列決定サイクル、すなわち、先行する配列決定サイクル $1 \sim i$ でターゲットクラスターに登録された次の過去の強度データと、を含む。補正された次の強度データ及び次の過去の強度データを総称して、次の利用可能な強度データとする。

【0232】

補正された次の強度データは、RTAに提供される。RTAは、補正された次の強度データを処理することができ、次の配列決定サイクル $i+1$ でターゲットクラスターについて次のベースコールを生成することができる。次のベースコールを生成するために、期待値最大化アルゴリズムは、次の利用可能な強度データに最も適合するように、補正された次の強度データに基づいて、平均（重心）及び分布（共変量）を観察することができる。

【0233】

期待値最大化アルゴリズムを使用することによって、次の配列決定サイクル $i+1$ で4つの強度分布が決定されると、4つの強度分布のそれぞれに属する補正された次の強度データの尤度が計算される。最も大きい尤度が次のベースコールを与える。

【0234】

ベースコールパイプラインは、クラスター別に実行され、クラスター集団内の複数のクラスターについて並行して実行されることに留意されたい。また、ベースコールパイプラインは、配列決定ランの連続する配列決定サイクルについて（例えば、対エンドの配列決定ランにおいてリード1の連続する150の配列決定サイクル及びリード2の別の連続する150の配列決定サイクルについて）繰り返し実行される。

【0235】

最小二乗法

最小二乗法は、累積強度補正パラメータ及び変動補正係数の閉形式の式を決定する。最小二乗法決定器は、強度モデラー及び最小化器を含む。

【0236】

強度モデルは、以下の式に従って、ターゲットクラスターの測定強度と変動補正係数との間の関係をモデル化する。

$$y_{c,i} = a x_{c,i} + d_i + n_{c,i} \quad \text{等式 (1)}$$

式中、

10

20

30

40

50

a は、ターゲットクラスターの増幅係数である。

d_i は、強度チャンネル i のチャンネル固有オフセット係数である。

$x_{c,i}$ は、現在の配列決定サイクル C でのターゲットクラスターについての強度チャンネル i における分布強度である。

$y_{c,i}$ は、現在の配列決定サイクル C でのターゲットクラスターについての強度チャンネル i における測定強度である。

$n_{c,i}$ は、現在の配列決定サイクル C でのターゲットクラスターについての強度チャンネル i の加法性ノイズである。

【0237】

最小化器は、最小二乗法を使用して、以下の式を最小化する。

10

【0238】

【数1】

$$\text{errorf}(\hat{a}, \hat{d}_i) = \sum_{c=1}^C \sum_{i=1}^2 (\hat{a}x_{c,i} + \hat{d}_i - y_{c,i})^2 \quad \text{等式(2)}$$

式中、

errorf は誤差関数である。

【0239】

20

【数2】

$\hat{a}$

は、ターゲットクラスターについての増幅係数である。

【0240】

【数3】

$\hat{d}_i$

30

は、強度チャンネル i のチャンネル固有オフセット係数である。

C は、現在の配列決定サイクルである。

【0241】

チェーンルールを使用して、最小化器は、増幅係数

【数4】

$\hat{a}$

及びチャンネル固有オフセット係数

【数5】

40

$\hat{d}_i$

に対する誤差関数の2つの偏微分を計算する。誤差関数を最小限に抑えるために、偏微分は等式2をゼロに設定する。

【0242】

【数6】

$$\frac{\partial errorf}{\partial \hat{a}} = \frac{\partial}{\partial \hat{a}} \left[\sum_{c=1}^c \sum_{i=1}^2 (\hat{a}x_{c,i} + \hat{d}_i - y_{c,i})^2 \right] = 0 \quad \text{等式(3)}$$

$$\frac{\partial errorf}{\partial \hat{d}_i} = \frac{\partial}{\partial \hat{d}_i} \left[\sum_{c=1}^c \sum_{i=1}^2 (\hat{a}x_{c,i} + \hat{d}_i - y_{c,i})^2 \right] = 0 \quad \text{等式(4)}$$

【 0 2 4 3 】

チャネル固有の強度誤差 $e_{c,i}$ は、以下のように定義される。

$$e_{c,i} = y_{c,i} - x_{c,i} \quad \text{等式(5)}$$

【 0 2 4 4 】

閉形式の式

第 1 の偏微分は、以下のように増幅係数

【数 7】

$\hat{a}$

の閉形式の式を決定する。

【 0 2 4 5 】

【数 8】

$$\frac{\partial errorf}{\partial \hat{a}} = \sum_{c=1}^c \sum_{i=1}^2 2x_{c,i} (\hat{a}x_{c,i} + \hat{d}_i - y_{c,i}) = 0 \quad \text{等式(6)}$$

$$= \sum_{c=1}^c \sum_{i=1}^2 2x_{c,i} (\hat{a}x_{c,i} - x_{c,i} + \hat{d}_i - y_{c,i} + x_{c,i}) = 0 \quad \text{等式(7)}$$

$$= \sum_{c=1}^c \sum_{i=1}^2 x_{c,i}^2 (\hat{a} - 1) + \hat{d}_i x_{c,i} - e_{c,i} x_{c,i} = 0 \quad \text{等式(8)}$$

$$= (\hat{a} - 1) \sum_{c=1}^c (x_{c,1}^2 + x_{c,2}^2) + \hat{d}_1 \sum_{c=1}^c x_{c,1} + \hat{d}_2 \sum_{c=1}^c x_{c,2} - \sum_{c=1}^c e_{c,1} x_{c,1} - \sum_{c=1}^c e_{c,2} x_{c,2} = 0 \quad \text{等式(9)}$$

【 0 2 4 6 】

累積強度補正パラメータの閉形式の式

【数 9】

$\bar{x}_1, \bar{x}_2, \bar{e}_1, \bar{e}_2, \overline{xx}$, 及び $\overline{xe}$

は、以下のように等式 9 を再特性化する。

【 0 2 4 7 】

【数 10】

$$= C(\hat{a} - 1)\overline{xx} + C\bar{x}_1\hat{d}_1 + C\bar{x}_2\hat{d}_2 - C\overline{xe} = 0 \quad \text{等式(10)}$$

$$= (\hat{a} - 1)\overline{xx} + \bar{x}_1\hat{d}_1 + \bar{x}_2\hat{d}_2 - \overline{xe} = 0 \quad \text{等式(11)}$$

10

20

30

40

50

式中、
【数 1 1】

$$\bar{x}_1 = \sum_{c=1}^C x_{c,1}$$
$$\bar{x}_2 = \sum_{c=1}^C x_{c,2}$$
$$\bar{e}_1 = \sum_{c=1}^C e_{c,1}$$
$$\bar{e}_2 = \sum_{c=1}^C e_{c,2}$$
$$\overline{xx} = \sum_{c=1}^C (x_{c,1}^2 + x_{c,2}^2)$$
$$\overline{xe} = \sum_{c=1}^C (x_{c,1}e_{c,1} + x_{c,2}e_{c,2})$$

中間項(1)

中間項(2)

中間項(3)

中間項(4)

中間項(5)

中間項(6)

10

20

【0 2 4 8】
累積強度補正パラメータのそれぞれを以下のように定義する。
1) 第 1 の累積強度補正パラメータ

【数 1 2】
 $\bar{x}_1$

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて測定された第 1 の強度チャンネルにおける分布強度の和である。

30

2) 第 2 の累積強度補正パラメータ

【数 1 3】
 $\bar{x}_2$

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて測定された第 2 の強度チャンネルにおける分布強度の和である。

3) 第 3 の累積強度補正パラメータ

【数 1 4】
 $\bar{e}_1$

40

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて計算された第 1 の強度チャンネルにおける強度誤差の和である。

4) 第 4 の累積強度補正パラメータ

【数 1 5】
 $\bar{e}_2$

50

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて計算された第 2 の強度チャンネルにおける強度誤差の和である。

5) 第 5 の累積強度補正パラメータ

【数 1 6】

$$\overline{xx}$$

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて計算された分布重心から原点の距離の和である。

10

6) 第 6 の累積強度補正パラメータ

【数 1 7】

$$\overline{xe}$$

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて計算された分布強度対強度誤差の類似性測定の和である。

【0 2 4 9】

第 2 の偏微分は、オフセット係数

20

【数 1 8】

$$\hat{d}_i$$

の閉形式の式を以下のように決定する。

【0 2 5 0】

【数 1 9】

$$\frac{\partial errorf}{\partial \hat{d}_i} = \sum_{c=1}^c \sum_{i=1}^2 2 (\hat{a}x_{c,i} + \hat{d}_i - y_{c,i}) = 0 \quad \text{等式(12)}$$

30

$$= \sum_{c=1}^c \sum_{i=1}^2 (\hat{a}x_{c,i} - x_{c,i} + \hat{d}_i - y_{c,i} + x_{c,i}) = 0 \quad \text{等式(13)}$$

$$= \sum_{c=1}^c \sum_{i=1}^2 x_{c,i}^2 \left((\hat{a} - 1) + \sum_{i=1}^2 \hat{d}_i + \sum_{c=1}^c x_{c,i} \right) = 0 \quad \text{等式(14)}$$

【0 2 5 1】

次いで、それぞれの強度チャンネルについて：

【数 2 0】

40

$$= C(\hat{a} - 1)\bar{x}_i + C\hat{d}_i - C\bar{e}_i = 0 \quad \text{等式(15)}$$

$$= (\hat{a} - 1)\bar{x}_i + \hat{d}_i - \bar{e}_i = 0 \quad \text{等式(16)}$$

【0 2 5 2】

第 1 の強度チャンネル、すなわち、i = 1 について：

【数 2 1】

$$\hat{d}_1 = \bar{e}_1 + (1 - \hat{a})\bar{x}_1 \quad \text{等式(17)}$$

50

式中、
【数 2 2】

$$\hat{d}_1$$

は、第 1 の強度チャネルのオフセット係数である。
【0 2 5 3】

第 2 の強度チャネル、すなわち、 $i = 2$ について：
【数 2 3】

$$\hat{d}_2 = \bar{e}_2 + (1 - \hat{a})\bar{x}_2 \quad \text{等式(18)}$$

10

式中、
【数 2 4】

$$\hat{d}_2$$

は、第 2 の強度チャネルに対するオフセット係数である。
【0 2 5 4】

等式 1 7 及び 1 8 の等式 1 1 への代入：
【数 2 5】

20

$$\frac{\partial \text{errorf}}{\partial \hat{a}} = (\hat{a} - 1)\overline{xx} + \bar{x}_1[\bar{e}_1 + (1 - \hat{a})\bar{x}_1] + \bar{x}_2[\bar{e}_2 + (1 - \hat{a})\bar{x}_2] - \overline{xe} = 0 \quad \text{等式(19)}$$

$$= (\hat{a} - 1)\overline{xx} + (1 - \hat{a})\bar{x}_1^2 + (1 - \hat{a})\bar{x}_2^2 + \bar{x}_1\bar{e}_1 + \bar{x}_2\bar{e}_2 - \overline{xe} = 0 \quad \text{等式(20)}$$

$$= (\hat{a} - 1)(\overline{xx} - \bar{x}_1^2 - \bar{x}_2^2) + \bar{x}_1\bar{e}_1 + \bar{x}_2\bar{e}_2 - \overline{xe} = 0 \quad \text{等式(21)}$$

$$\hat{a} - 1 = \frac{\bar{x}_1\bar{e}_1 + \bar{x}_2\bar{e}_2 - \overline{xe}}{\overline{xx} - \bar{x}_1^2 - \bar{x}_2^2} \quad \text{等式(22)}$$

$$\hat{a} = 1 + \frac{\bar{x}_1\bar{e}_1 + \bar{x}_2\bar{e}_2 - \overline{xe}}{\overline{xx} - \bar{x}_1^2 - \bar{x}_2^2} \quad \text{等式(23)}$$

30

式中、
【数 2 6】

$$\hat{a}$$

は、ターゲットクラスターについての増幅係数である。
【0 2 5 5】

別の実施態様では、クラスターごとのメモリ要件を低減するために、異なる強度チャネル（例えば、第 1 及び第 2 の強度チャネル）に対する共通オフセット係数は、制約

40

【数 2 7】

$$\hat{d}_1 = \hat{d}_2$$

を導入することによって以下のように決定される。

【0 2 5 6】

【数 2 8】

50

$$\bar{x} = \frac{1}{2C} \sum_{c=1}^C (x_{c,1} + x_{c,2}) \quad \text{中間項(1. 1)}$$

$$\overline{xx} = \frac{1}{C} \sum_{c=1}^C (x_{c,1}^2 + x_{c,2}^2) \quad \text{中間項(2. 1)}$$

$$\bar{e} = \frac{1}{2C} \sum_{c=1}^C (e_{c,1} + e_{c,2}) \quad \text{中間項(3. 1)}$$

$$\overline{xe} = \frac{1}{C} \sum_{c=1}^C (x_{c,1}e_{c,1} + x_{c,2}e_{c,2}) \quad \text{中間項(4. 1)}$$

$$\hat{a} = 1 + \frac{\overline{xe} - 2\bar{x}\bar{e}}{\overline{xx} - 2\bar{x}^2} \quad \text{等式(24)}$$

$$\hat{d}_1 = \bar{e} + \bar{x}(1 - \hat{a}) \quad \text{等式(25)}$$

$$\hat{d}_2 = \bar{e} + \bar{x}(1 - \hat{a}) \quad \text{等式(26)}$$

10

【0257】

最小二乗法は、閉形式の式の決定するために配列決定ランの前に実行されることが当業者には明らかであろう。決定されると、閉形式の式は、配列決定ランのそれぞれの配列決定サイクルで反復的に、クラスターごとに配列決定ラン中に生成された強度値に適用される。

20

【0258】

クラスターごとの変動補正係数及びそれらがどのように決定されるかについての更なる詳細は、2021年10月27日に出願された「SYSTEMS AND METHODS FOR PER-CLUSTER INTENSITY CORRECTION AND BASE CALLING」と題する米国特許仮出願第63/106,256号（代理人整理番号ILLM 1034-1/IP-2026-PRV）に見出すことができ、その全体が参照により本明細書に組み込まれる。

30

【0259】

図34では、クラスターごとの状態3042は、一実施態様による、現在の配列決定サイクルについてRTAによって決定されたクラスターごとの変動補正係数である。次いで、クラスターごとの状態3042は、クラスターピクセル及び現在のベースコール反復／演算の空間畳み込み特徴と組み合わせられて、対応するクラスターについてのベースコールを生成する。これらのステップは、図34のアクション3402、3412、3422、及び3432によって示される。

【0260】

圧縮ネットワーク

上述のように、ニューラルネットワークベースのベースコーラー2900の専用アーキテクチャは、対応する配列決定サイクルの画像パッチのスライディングウィンドウを処理する。後続のスライディングウィンドウの配列決定サイクルどうしの間には重複が存在する。これにより、ニューラルネットワークベースのベースコーラー2900は、重なり合う配列決定サイクルの画像パッチを冗長に処理することになる。これにより、計算リソースの無駄が生じる。例えば、一実施態様では、ニューラルネットワークベースのベースコーラー2900の各空間畳み込み層は、ほぼ1億回の乗算操作を有する。次いで、5つの配列決定サイクル及び7つの空間畳み込み層のカスケード（又はシーケンス）のウィンドウに対して、空間畳み込みニューラルネットワークは、約6億2千万回の乗算操作を実行する。更に、時間畳み込みニューラルネットワークは、約1千万回の乗算操作を実行する。

40

50

【0261】

現在のスライディングウィンドウ（又はベースコールの現在の反復）におけるサイクル $N-1$ に対する画像データは、以前のスライディングウィンドウ（又はベースコールの以前の反復）におけるサイクル N として処理されるため、現在のスライディングウィンドウで行われた処理の中間結果と、後続のウィンドウで行われた処理の中間結果を記憶して、それによって、後続のスライディングウィンドウどうしの間で重なり合う配列決定サイクルに対する、入力画像データの冗長処理（又は再処理）を回避する（又は不要にする）という機会がもたらされる。

【0262】

しかしながら、中間結果は、実用的とは言えない容量のストレージを必要とする数テラバイトのデータである。この技術的問題を克服するために、開示された技術は、中間結果を、その中間結果が最初にニューラルネットワークベースのベースコーラー 2900 によって生成される際に圧縮し、その後のスライディングウィンドウでその圧縮された中間結果を再目的化して利用して冗長な計算を回避し、それによって、中間結果を再生成しない（又は、一度だけ生成する）ということを提案する。

【0263】

図 35 において、一実施態様によれば、圧縮論理 3530（又は圧縮ネットワーク又は圧縮サブネットワーク又は圧縮層又はスクイーズ層）は、クラスター特徴フィルタリング論理 3030 の出力を処理し、出力の圧縮表現を生成する。一実施態様では、圧縮ネットワーク 3530 は、クラスター特徴フィルタリング論理 3030 によって生成された特徴マップの深さ次元数を低減する圧縮畳み込み層を備える。

【0264】

例えば、フィルタリングされたクラスターごとの空間マップ 3054、3064、3074、3084、及び 3094 の深度次元数が 14（すなわち、空間出力ごとに 14 個の特徴マップ又は 14 個のチャンネル）であると考える。圧縮ネットワーク 3530 は、フィルタリングされたクラスターごとの空間マップ 3054、3064、3074、3084、及び 3094 を、圧縮されたクラスター特徴 3504 と総称されるそれぞれの配列決定サイクル $N-2$ 、 $N-1$ 、 N 、 $N+1$ 、及び $N+2$ について、それぞれの圧縮されフィルタリングされたクラスターごとの空間マップ 3554、3564、3574、3584、及び 3594 に減衰させる。一実施態様では、圧縮され、フィルタリングされたクラスターごとの空間マップ 3554、3564、3574、3584、及び 3594 の各々は、2（すなわち、圧縮された空間出力ごとに 2 つの特徴マップ又は 2 つのチャンネル）の深度次元数を有する。他の実施態様では、圧縮され、フィルタリングされたクラスターごとの空間マップ 3554、3564、3574、3584、及び 3594 は、3 又は 4（すなわち、圧縮された空間出力ごとに 3 つ又は第 4 の特徴マップ又は 3 つ又は第 4 のチャンネル）の深度次元数を有することができる。更に他の実施態様では、圧縮され、フィルタリングされたクラスターごとの空間マップ 3554、3564、3574、3584、及び 3594 は、1（すなわち、圧縮された空間出力ごとに 1 つの特徴マップ又は 1 つのチャンネル）の深度次元数を有することができる。一実施態様では、圧縮層 3530 は、ReLU のような活性化関数を含まない。他の実施態様では、圧縮層 3530 は、活性化関数を含むことができる。他の実施態様では、圧縮論理 3530 は、対応する圧縮空間マップセットを、各々が 4 つより多くの特徴マップを有するように構成することができる。

【0265】

ここで、圧縮論理 3530 が圧縮出力をどのように生成するかを論じる。

【0266】

一実施態様では、圧縮論理 3530 は、非線形性を導入しながら、特徴マップの数（すなわち、深さ次元、又はチャンネルの数）を減少させるために 1×1 の畳み込みを使用する。 1×1 の畳み込みは、1 のカーネルサイズを有する。 1×1 の畳み込みは、空間次元を変化させることなく、体積深さを別のスクイーズされた又は拡張された表現に変換することができる。 1×1 の畳み込みは、入力チャンネルにわたる、完全接続線形層のように動作

10

20

30

40

50

する。これは、多くのチャネルを有する特徴マップから、より少数の特徴マップにマッピングするのに有用である。単一の 1×1 の畳み込みが、2つの特徴マップを有する入力テンソルに適用され得る。 1×1 の畳み込みは、2チャネル入力を単一チャネルの出力に圧縮する。

【0267】

圧縮層108によって生成された圧縮出力（又は圧縮特徴マップ若しくは圧縮空間マップ若しくは圧縮時間マップ）の数は、圧縮層3530内に構成された 1×1 の畳み込みフィルタ（又は圧縮畳み込みフィルタ若しくは圧縮フィルタ）の数の関数である。一実施態様では、圧縮層3530は、2つの 1×1 畳み込みフィルタを有することができる。第1の 1×1 の畳み込みフィルタは、14個の特徴マップを有する空間特徴ボリュームを処理し、 101×101 の空間次元数を保存しながら第1の特徴マップを生成し得る。第2の 1×1 の畳み込みフィルタはまた、14個の特徴マップを有する空間特徴ボリュームを処理し、 101×101 の空間次元数を保存しながら第2の特徴マップを生成することができる。したがって、圧縮層3530は、14個の特徴マップを有する空間特徴ボリュームを、2つの空間特徴マップを有する圧縮出力（すなわち、圧縮比=7）に減縮する。

【0268】

いくつかの実施態様では、開示される技術は、ニューラルネットワークベースのベースコーラー2900の空間ネットワークにおいて約80%の畳み込みを節約する。一実施態様では、80%の節約は、その後の配列決定サイクルにおいて、圧縮論理及び圧縮特徴マップの再目的化しての利用が5回の配列決定サイクルの入力ウィンドウ（例えば、サイクルN、サイクルN+1、サイクルN-1、サイクルN+2、及びサイクルN-2）に対して使用されるときに、空間畳み込みにおいて観察される。別の実施態様では、その後の配列決定サイクルにおける圧縮論理及び圧縮特徴マップの再目的化しての利用が、10回の配列決定サイクルの入力ウィンドウ（例えば、サイクルN、サイクルN+1、サイクルN-1、サイクルN+2、サイクルN-2、サイクルN+3、及びサイクルN-3）に使用される場合、空間畳み込みにおいて90%の節約が観察される。すなわち、ウィンドウサイズが大きいほど、圧縮論理の使用及び圧縮特徴マップの再目的化からの節約が大きくなり、ウィンドウサイズが大きいほど、追加の隣接サイクルからのより大きなコンテキストの組み込みによるベースコール性能が良好である。より大きなウィンドウに対するより大きな節約は、所与の計算能力に対する全体的な性能を改善する。

【0269】

圧縮論理によってもたらされる計算効率及びコンパクトな計算フットプリントは、中央演算処理装置（Central Processing Unit、CPU）、グラフィック処理ユニット（Graphics Processing Unit、GPU）、フィールドプログラマブルゲートアレイ（FPGA）、粗粒度再構成可能アーキテクチャ（Coarse-Grained Reconfigurable Architecture、CGRA）、特定用途向け集積回路（ASIC）、特定用途向け命令セットプロセッサ（Application Specific Instruction-set Processor、ASIP）、及びデジタル信号プロセッサ（DSP）のようなリソース制約プロセッサ上のニューラルネットワークベースのベースコーラー2900のハードウェア実装を容易にする。

【0270】

圧縮論理によって節約された計算は、ニューラルネットワークベースのベースコーラー2900により多くの畳み込み演算子を組み込むことを可能にする。例としては、空間畳み込み層及び時間畳み込み層により多くの畳み込みフィルタを追加すること、畳み込みフィルタのサイズを増加させること、空間畳み込み層及び時間畳み込み層の数を増加させることが挙げられる。追加の畳み込み演算は、ニューラルネットワークベースのベースコーラー2900の強度パターン検出及び全体的なベースコール精度を改善する。

【0271】

圧縮論理及びその圧縮された出力に関する更なる詳細は、2021年2月18日に出願された「DATA COMPRESSION FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLING」と題する米国特許非仮出願

10

20

30

40

50

第 17 / 179, 395 号 (代理人整理番号 I L L M 1029-2 / I P-1964-US) に見出すことができ、その全体が参照により本明細書に組み込まれる。

【0272】

図 35 では、現在の配列決定サイクル N についてのクラスターごとの状態 3042 が、ニューラルネットワークベースのベースコーラー 2900 の中間出力、すなわち、圧縮されフィルタリングされたクラスターごとの空間マップ 3554、3564、3574、3584、及び 3594 と組み合わせられる。

【0273】

いくつかの実施態様では、クラスターごとの状態 3042 の次元数は、圧縮されフィルタ処理されたクラスターごとの空間マップ 3554、3564、3574、3584、及び 3594 の次元数と一致するように修正される (例えば、次元をトリミングすること、次元を追加すること、パディングすること (例えば、0 パディングすること)、クローニングすることなどによって)。異なる実施態様では、クラスター単位の状態 3042 並びに圧縮されフィルタリングされたクラスター単位の空間マップ 3554、3564、3574、3584、及び 3594 のそれぞれの 5 つのインスタンスは、連結、合計、要素単位の乗算、要素単位の乗算及び合計 (畳み込み) など、上記で説明した技法を使用してそれぞれ組み合わせることができる。

【0274】

次いで、図 35 に示される例によれば、それぞれの圧縮されフィルタリングされたクラスターごとの空間マップ 3554、3564、3574、3584、及び 3594 と、クラスターごとの状態 3042 のそれぞれの 5 つのインスタンスとのそれぞれの 5 つの組み合わせが、時相論理 2928 によって処理されて、現在の配列決定サイクル N についてのベースコール 3080 が生成される。ベースコール 3080 は、クラスターピクセル、したがってクラスター特徴 3004 に対応するクラスターに対して行われる。

【0275】

次に、クラスターごとの状態 3042 の異なる変形例、及びこれらの変形例が開示された技術の異なる実施態様に従ってどのように決定されるかについて説明する。

【0276】

図 36 は、クラスターピクセルに対応する圧縮空間畳み込み特徴の過去の特徴値を使用してクラスターごとの状態を生成する (3600) 一実施態様を示す。図 36 では、クラスターごとの状態 3042 は、例えば、最小値選択関数、最大値選択関数、平均化関数、指数加重平均化関数などを使用して、上記で説明したように、以前の配列決定サイクル及び現在の配列決定サイクルからのクラスターピクセルに対応する圧縮空間畳み込み特徴の過去の特徴値の要約統計値を蓄積することによって決定される。次いで、クラスターごとの状態 3042 は、クラスターピクセル及び現在のベースコール反復 / 演算の圧縮空間畳み込み特徴と組み合わせられて、対応するクラスターに対するベースコールを生成する。これらのステップは、図 36 のアクション 3602、3612、3622、3632、及び 3642 によって示される。

【0277】

図 37 において、クラスターごとの状態 3042 は、一実施態様による、現在の配列決定サイクルについて R T A によって決定されたクラスターごとの変動補正係数である。次いで、クラスターごとの状態 3042 は、クラスターピクセル及び現在のベースコール反復 / 演算の圧縮空間畳み込み特徴と組み合わせられて、対応するクラスターに対するベースコールを生成する。これらのステップは、図 37 のアクション 3702、3712、3722、3732、及び 3742 によって示される。

【0278】

図 38 は、クラスターごとの状態 3042 を圧縮特徴に組み込むために、対応するクラスターピクセルの過去の強度値を使用してクラスターごとの状態 3042 を生成する (3800) 一実施態様を示す。図 38 では、クラスターごとの状態 3042 は、例えば、最小値選択関数、最大値選択関数、平均化関数、指数加重平均化関数などを使用して、上記

で説明したように、以前の配列決定サイクル及び現在の配列決定サイクルからのクラスターピクセルの過去の強度値の要約統計値を蓄積することによって決定される。次いで、クラスターごとの状態 3042 は、クラスターピクセル及び現在のベースコール反復／演算の圧縮空間畳み込み特徴と組み合わせられて、対応するクラスターについてのベースコールを生成する。これらのステップは、図 38 のアクション 3802、3812、3822、3832、及び 3842 によって示される。

【0279】

ウェル状態データからピクセル状態データへの変換

図 39 は、疎なウェルごとの状態から密なピクセルごとの状態を生成する一実施態様を示す。図 39 では、ウェルごとの状態 3902 は疎であり、転置畳み込み又は補間（例えば線形、双線形、三次補間）3904 を使用して密なピクセルごとの状態 3906 に変換することができる。

【0280】

「疎」及び「密」は、配列（例えば、ベクトル又は行列）中の 0 要素対非 0 要素の数を指す。疎なアレイは、大部分が 0 であり、少数の非ゼロエントリを含むアレイである。密な配列は、ほとんどゼロでないものを含む。高密度アレイでは、一実施態様では、各ピクセルごとの状態は、少なくとも 2 つ、4 つ、又は 8 つの隣接するピクセルごとの状態によって囲まれ得る。疎なアレイでは、一実施態様では、各ウェルごとの状態は、2 つ、4 つ、又は 8 つの隣接するウェルごとの状態によってさえ取り囲まれない。状態の集合は、最初の行と最後の行との間の全ての行が非 0 要素で定義され、値が与えられる場合、密と呼ばれる。行にギャップ又は 0 要素がある場合、状態の集合は疎と呼ばれる。疎な性は、多くの要素（例えば、1 つおき、3 つおき、又は 4 つおき）が 0 であるか、又は 0 に非常に近いことを意味し得る。

【0281】

図 40 は、ベースコール動作を実行するためのベースコーラーへの入力として、疎なウェルごとの状態、密なピクセルごとの状態、及びピクセルごとの強度値を提供する一実施態様を示す。図 40 において、疎なウェルごとの状態 3902 は、ベースコーラー 144、例えば、ニューラルネットワークベースのベースコーラー 2900 への入力として、直接入力として、並びに、例えば、残留接続及び／又はスキップ接続を使用する冗長入力として提供され得る。いくつかの実施態様では、疎なウェルごとの状態 3902 は、密なピクセルごとの状態 3906 及び密なピクセルごとの強度値 4012（すなわち、クラスター画像）とともに補足入力として提供され、これらは、ベースコールを生成するためにベースコーラー 144 によって処理される。

【0282】

図 41 は、ベースコール動作を実行するためのニューラルネットワークベースのベースコーラー 2900 への入力として、疎なウェルごとの状態、密なピクセルごとの状態、及びピクセルごとの強度値を使用する一実施態様を示す。ニューラルネットワークベースのベースコーラー 2900 への入力は、密なピクセルごとの状態 3906 及び密なピクセルごとの強度値 4012（すなわち、クラスター画像）である。

【0283】

配列決定システム

図 42 A 及び図 42 B は、配列決定システム 4200 A の一実施態様を示す。配列決定システム 4200 A は、構成可能プロセッサ 4246 を含む。構成可能プロセッサ 4246 は、本明細書に開示されるベースコール技術を実施する。配列決定システムは、「シーケンサ」とも称される。

【0284】

配列決定システム 4200 A は、生物学的物質又は化学物質のうちの少なくとも 1 つに関連する任意の情報又はデータを得るよう動作し得る。いくつかの実施態様では、配列決定システム 4200 A は、ベンチトップデバイス又はデスクトップコンピュータと同様であり得るワークステーションである。例えば、所望の反応を実施するためのシステム及

10

20

30

40

50

び構成要素の大部分（又は全て）は、共通のハウジング 4 2 0 2 内にあってもよい。

【0 2 8 5】

特定の実施態様では、配列決定システム 4 2 0 0 A は、*de novo sequencing*、全ゲノム又は標的ゲノム領域の再配列、及びメタゲノミクスを含むがこれらに限定されない、様々な用途のために構成された核酸配列決定システムである。シーケンスはまた、DNA 又は RNA 分析に使用されてもよい。いくつかの実施態様では、配列決定システム 4 2 0 0 A はまた、バイオセンサ内に反応部位を生成するように構成されてもよい。例えば、配列決定システム 4 2 0 0 A は、サンプルを受容し、サンプル由来のクローン的に増幅された核酸の表面結合クラスターを生成するように構成され得る。各クラスターは、バイオセンサ内の反応部位を構成するか、又はその一部であってもよい。

10

【0 2 8 6】

例示的な配列決定システム 4 2 0 0 A は、バイオセンサ 4 2 1 2 と相互作用して、バイオセンサ 4 2 1 2 内で所望の反応を行うように構成されたシステム受け部又はインターフェース 4 2 1 0 を含んでもよい。図 4 2 A に関して以下の説明では、バイオセンサ 4 2 1 2 はシステム受け部 4 2 1 0 内に装填される。しかしながら、バイオセンサ 4 2 1 2 を含むカートリッジは、システム受け部 4 2 1 0 に挿入されてもよく、一部の状態では、カートリッジは一時的又は永久的に除去され得ることが理解される。上述のように、カートリッジは、とりわけ、流体制御及び流体貯蔵構成要素を含んでもよい。

【0 2 8 7】

特定の実施態様では、配列決定システム 4 2 0 0 A は、バイオセンサ 4 2 1 2 内で多数の平行反応を行うように構成されている。バイオセンサ 4 2 1 2 は、所望の反応が生じ得る 1 つ以上の反応部位を含む。反応部位は、例えば、バイオセンサの固体表面に固定化されてもよく、又はバイオセンサの対応する反応チャンバ内に位置するビーズ（又は他の可動基材）に固定化されてもよい。反応部位は、例えば、クローン的に増幅された核酸のクラスターを含むことができる。バイオセンサ 4 2 1 2 は、固体撮像デバイス（例えば、CCD 又は CMOS 撮像素子）及びそれに取り付けられたフローセルを含んでもよい。フローセルは、配列決定システム 4 2 0 0 A から溶液を受容し、溶液を反応部位に向けて方向付ける 1 つ以上の流路を含んでもよい。任意選択的に、バイオセンサ 4 2 1 2 は、熱エネルギーをフローチャネルの内外に伝達するための熱要素と係合するように構成することができる。

20

30

【0 2 8 8】

配列決定システム 4 2 0 0 A は、相互に相互作用して、生物学的又は化学的分析のための所定の方法又はアッセイプロトコルを実行する、様々な構成要素、アセンブリ、及びシステム（又はサブシステム）を含んでもよい。例えば、配列決定システム 4 2 0 0 A は、配列決定システム 4 2 0 0 A の様々な構成要素、アセンブリ、及びサブシステムと通信し得るシステムコントローラ 4 2 0 6 と、更にバイオセンサ 4 2 1 2 と、を含む。例えば、システム受け部 4 2 1 0 に加えて、配列決定システム 4 2 0 0 A はまた、配列決定システム 4 2 0 0 A 及びバイオセンサ 4 2 1 2 の流体ネットワーク全体にわたる流体の流れを制御するための流体制御システム 4 2 0 8 と、バイオアッセイシステムによって使用され得る全ての流体（例えば、流体、ガス又は液体）を保持するように構成された流体貯蔵システム 4 2 1 4、流体ネットワーク、流体貯蔵システム 4 2 1 4、及び／又はバイオセンサ 4 2 1 2 内の流体の温度を調整し得る温度制御システム 4 2 0 4 と、バイオセンサ 4 2 1 2 を照明するように構成された照明システム 4 2 1 6 と、を含み得る。上述のように、バイオセンサ 4 2 1 2 を有するカートリッジがシステム受け部 4 2 1 0 内に装填される場合、カートリッジはまた、流体制御及び流体貯蔵構成要素を含んでもよい。

40

【0 2 8 9】

また、配列決定システム 4 2 0 0 A は、ユーザと対話するユーザインターフェース 4 2 1 8 を含んでもよい。例えば、ユーザインターフェース 4 2 1 8 は、ユーザから情報を表示又は要求するディスプレイ 4 2 2 0 と、ユーザ入力を受け取るためのユーザ入力デバイス 4 2 2 2 と、を含むことができる。いくつかの実施態様では、ディスプレイ 4 2 2 0 及

50

びユーザ入力デバイス 4222 は、同じデバイスである。例えば、ユーザインターフェース 4218 は、個々のタッチの存在を検出し、またディスプレイ上のタッチの場所を識別するように構成されたタッチ検知ディスプレイを含んでもよい。しかしながら、マウス、タッチパッド、キーボード、キーパッド、ハンドヘルドスキャナ、音声認識システム、動き認識システムなどの他のユーザ入力デバイス 4222 が使用されてもよい。以下でより詳細に説明するように、配列決定システム 4200A は、所望の反応を実施するために、バイオセンサ 4212（例えば、カートリッジの形態）を含む様々な構成要素と通信してもよい。配列決定システム 4200A はまた、バイオセンサから得られたデータを分析して、ユーザに所望の情報を提供するように構成されてもよい。

【0290】

システムコントローラ 4206 は、マイクロコントローラ、低減命令セットコンピュータ（Reduced Instruction Set Computer、RISC）、特定用途向け集積回路（Application Specific Integrated Circuit、ASIC）、フィールドプログラマブルゲートアレイ（Field Programmable Gate Array、FPGA）、粗粒化再構成可能構造（Coarse-Grained Reconfigurable Architecture、CGRA）、論理回路、及び本明細書に記載される機能を実行することができる任意の他の回路又はプロセッサを含んでもよい。上記の実施例は、例示的なものに過ぎず、したがって、システムコントローラという用語の定義及び／又は意味を制限することを意図するものではない。例示的な実施態様では、システムコントローラ 4206 は、検出データを取得し分析する少なくとも 1 つのために、1 つ以上の記憶要素、メモリ、又はモジュール内に記憶された命令のセットを実行する。検出データは、ピクセル信号の複数のシーケンスを含むことができ、それにより、数百万個のセンサ（又はピクセル）のそれぞれからのピクセル信号のシーケンスを、多くのベースコールサイクルにわたって検出することができる。記憶要素は、配列決定システム 4200A 内の情報源又は物理メモリ要素の形態であってもよい。

【0291】

命令セットは、本明細書に記載される様々な実施態様の方法及びプロセスなどの特定の動作を実行するように配列決定システム 4200A 又はバイオセンサ 4212 に指示する様々なコマンドを含んでもよい。命令のセットは、有形の非一時的コンピュータ可読媒体又は媒体の一部を形成し得るソフトウェアプログラムの形態であってもよい。本明細書で使用する時、用語「ソフトウェア」及び「ファームウェア」は互換可能であり、RAM メモリ、ROM メモリ、EPROM メモリ、EEPROM メモリ、及び不揮発性 RAM（NVRAM）メモリを含むコンピュータによって実行されるメモリに記憶された任意のコンピュータプログラムを含む。上記メモリタイプは、例示的なものに過ぎず、したがって、コンピュータプログラムの記憶に使用可能なメモリのタイプに限定されない。

【0292】

ソフトウェアは、システムソフトウェア又はアプリケーションソフトウェアなどの様々な形態であってもよい。更に、ソフトウェアは、別個のプログラムの集合、又はより大きいプログラム内のプログラムモジュール若しくはプログラムモジュールの一部分の形態であってもよい。ソフトウェアはまた、オブジェクト指向プログラミングの形態のモジュール式プログラミングを含んでもよい。検出データを取得した後、検出データは、ユーザ入力に応じて処理された配列決定システム 4200A によって自動的に処理されてもよく、又は別の処理マシン（例えば、通信リンクを介したリモート要求）によって行われる要求に応じて処理されてもよい。図示の実施態様では、システムコントローラ 4206 は分析モジュール 4244 を含む。他の別の実施態様では、システムコントローラ 4206 は分析モジュール 4244 を含まず、代わりに分析モジュール 4244 へのアクセスを有する（例えば、分析モジュール 4244 は、クラウド上で別個にホスティングされ得る）。

【0293】

システムコントローラ 4206 は、通信リンクを介して、バイオセンサ 4212 及び配列決定システム 4200A の他の構成要素に接続されてもよい。システムコントローラ 4206 はまた、オフサイトシステム又はサーバに通信可能に接続されてもよい。通信リン

10

20

30

40

50

クは、配線、コード、又は無線であってもよい。システムコントローラ 4206 は、ユーザインターフェース 4218 及びユーザ入力デバイス 4222 からユーザ入力又はコマンドを受信してもよい。

【0294】

流体制御システム 4208 は、流体ネットワークを含み、流体ネットワークを通る 1 つ以上の流体の流れを方向付けるように構成されている。流体ネットワークは、バイオセンサ 4212 及び流体貯蔵システム 4214 と流体連通していてもよい。例えば、選択された流体が流体貯蔵システム 4214 から引き出され、制御された方法でバイオセンサ 4212 に指向されてもよく、又は流体は、バイオセンサ 4212 から引き出され、例えば、流体貯蔵システム 4214 内の廃棄物リザーバに指向されてもよい。図示されていないが、流体制御システム 4208 は、流体ネットワーク内の流体の流量又は圧力を検出する流量センサを含んでもよい。センサは、システムコントローラ 4206 と通信してもよい。

10

【0295】

温度制御システム 4204 は、流体ネットワーク、流体貯蔵システム 4214 及び／又はバイオセンサ 4212 の異なる領域における流体の温度を調節するように構成されている。例えば、温度制御システム 4204 は、バイオセンサ 4212 と相互作用し、バイオセンサ 4212 内の反応部位に沿って流れる流体の温度を制御する熱循環器を含んでもよい。温度制御システム 4204 はまた、配列決定システム 4200 A 又はバイオセンサ 4212 の中実要素又は構成要素の温度を調節してもよい。図示されていないが、温度制御システム 4204 は、流体又は他の構成要素の温度を検出するためのセンサを含んでもよい。センサは、システムコントローラ 4206 と通信してもよい。

20

【0296】

流体貯蔵システム 4214 は、バイオセンサ 4212 と流体連通しており、所望の反応を行うために使用される様々な反応成分又は反応物質を貯蔵してもよい。流体貯蔵システム 4214 はまた、流体ネットワーク及びバイオセンサ 4212 を洗浄又はクリーニングし、反応物質を希釈するための流体を貯蔵してもよい。例えば、流体貯蔵システム 4214 は、サンプル、試薬、酵素、他の生体分子、緩衝液、水性、及び非極性溶液などを保存するための様々なリザーバを含んでもよい。更に、流体貯蔵システム 4214 はまた、バイオセンサ 4212 から廃棄物を受容するための廃棄物リザーバを含んでもよい。カートリッジを含む実装形態では、カートリッジは、流体貯蔵システム、流体制御システム、又は温度制御システムのうちの 1 つ以上を含み得る。したがって、これらのシステムに関する本明細書に記載される構成要素のうちの 1 つ以上は、カートリッジハウジング内に収容され得る。例えば、カートリッジは、サンプル、試薬、酵素、他の生体分子、緩衝液、水性、及び非極性溶液、廃棄物などを保存するための様々なリザーバを有し得る。したがって、流体貯蔵システム、流体制御システム、又は温度制御システムのうちの 1 つ以上は、カートリッジ又は他のバイオセンサを介してバイオアッセイシステムと取り外し可能に係合され得る。

30

【0297】

照明システム 4216 は、バイオセンサを照明するための光源（例えば、1 つ以上の LED）及び複数の光学構成要素を含んでもよい。光源の例としては、レーザ、アークランプ、LED、又はレーザダイオードが挙げられる。光学構成要素は、例えば、反射器、偏光板、ビームスプリッタ、コリメータ、レンズ、フィルタ、ウェッジ、プリズム、鏡、検出器などであってもよい。照明システムを使用する実施態様では、照明システム 4216 は、励起光を反応部位に方向付けるように構成されてもよい。一例として、蛍光団は、緑色の光の波長によって励起されてもよく、そのため、励起光の波長は約 4232 nm であり得る。一実施態様では、照明システム 4216 は、バイオセンサ 4212 の表面の表面法線に平行な照明を生成するように構成されている。別の実施態様では、照明システム 4216 は、バイオセンサ 4212 の表面の表面法線に対してオフアングルである照明を生成するように構成されている。更に別の実施態様では、照明システム 4216 は、いくつかの平行照明及びある程度のオフアングル照明を含む複数の角度を有する照明を生成する

40

50

ように構成されている。

【0298】

システム受け部又はインターフェース4210は、機械的、電氣的、及び流体的な方法のうち少なくとも1つにおいてバイオセンサ4212と係合するように構成される。システム受け部4210は、バイオセンサ4212を所望の配向に保持して、バイオセンサ4212を通る流体の流れを容易にすることができる。システム受け部4210はまた、バイオセンサ4212と係合するように構成された電気接点を含んでもよく、それにより、配列決定システム4200Aは、バイオセンサ4212と通信することができ、及び／又はバイオセンサ4212に電力を供給することができる。更に、システム受け部4210は、バイオセンサ4212と係合するように構成された流体ポート（例えば、ノズル）を含んでもよい。いくつかの実施態様では、バイオセンサ4212は、電氣的に、また流体方式で、システム受け部4210に取り外し可能に結合される。

10

【0299】

加えて、配列決定システム4200Aは、他のシステム若しくはネットワークと遠隔で、又は他のバイオアッセイシステム4200Aと通信してもよい。バイオアッセイシステム4200Aによって得られた検出データは、リモートデータベースに記憶されてもよい。

【0300】

図42Bは、図42Aのシステムで使用され得るシステムコントローラ4206のブロック図である。一実施態様では、システムコントローラ4206は、互いに通信することができる1つ以上のプロセッサ又はモジュールを含む。プロセッサ又はモジュールのそれぞれは、特定のプロセスを実行するためのアルゴリズム（例えば、有形及び／又は非一時的コンピュータ可読記憶媒体上に記憶された命令）又はサブアルゴリズムを含んでもよい。システムコントローラ4206は、モジュールの集合として概念的に例示されるが、専用ハードウェアボード、DSP、プロセッサなどの任意の組み合わせを利用して実装されてもよい。あるいは、システムコントローラ4206は、単一のプロセッサ又は複数のプロセッサを備えた既製のPCを使用して実装されてもよく、機能動作はプロセッサ間に分散される。更なるオプションとして、以下に記載されるモジュールは、特定のモジュール式機能が専用ハードウェアを利用して実行されるハイブリッド構成を利用して実装されてもよく、残りのモジュール式機能は、既製のPCなどを利用して実行される。モジュールはまた、処理ユニット内のソフトウェアモジュールとして実装されてもよい。

20

30

【0301】

動作中、通信ポート4250は、バイオセンサ4212（図42A）及び／又はサブシステム4208、4214、4204（図42A）から情報（例えば、データ）に情報（例えばコマンド）を送信してもよい。実施態様では、通信ポート4250は、ピクセル信号の複数のシーケンスを出力することができる。通信リンク4234は、ユーザインターフェース4218からユーザ入力を受信し（図42A）、ユーザインターフェース4218にデータ又は情報を送信し得る。バイオセンサ4212又はサブシステム4208、4214、4204からのデータは、バイオアッセイセッション中に、システムコントローラ4206によってリアルタイムで処理されてもよい。追加的に又は代替的に、データは、バイオアッセイセッション中にシステムメモリ内に一時的に記憶され、リアルタイム又はオフライン操作よりも遅く処理されてもよい。

40

【0302】

図42Bに示すように、システムコントローラ4206は、中央処理装置（CPU）4252とともに主制御モジュール4224と通信する複数のモジュール4226～4248を含んでもよい。主制御モジュール4224は、ユーザインターフェース4218と通信してもよい（図42A）。モジュール4226～4248は、主制御モジュール4224と直接通信するものとして示されているが、モジュール4226～4248はまた、互いに、ユーザインターフェース4218と、及びバイオセンサ4212と直接通信してもよい。また、モジュール4226～4248は、他のモジュールを介して主制御モジュール

50

ル 4 2 2 4 と通信してもよい。

【0303】

複数のモジュール 4 2 2 6 ~ 4 2 4 8 は、サブシステム 4 2 0 8、4 2 1 4、4 2 0 4 及び 4 2 1 6 とそれぞれ通信するシステムモジュール 4 2 2 8 ~ 4 2 3 2、4 2 2 6 を含む。流体制御モジュール 4 2 2 8 は、流体ネットワークを通る 1 つ以上の流体の流れを制御するために、流体制御システム 4 2 0 8 と通信して、流体ネットワークの弁及び流量センサを制御してもよい。流体貯蔵モジュール 4 2 3 0 は、流体が少ないとき、又は廃棄物リザーバが満タン容量又はほぼ満タン容量にあるときにユーザに通知してもよい。流体貯蔵モジュール 4 2 3 0 はまた、流体が所望の温度で貯蔵され得るように、温度制御モジュール 4 2 3 2 と通信してもよい。照明モジュール 4 2 2 6 は、所望の反応（例えば、結合事象）が生じた後など、プロトコル中に指定された時間で反応部位を照明するために、照明システム 4 2 1 6 と通信してもよい。いくつかの実施態様では、照明モジュール 4 2 2 6 は、照明システム 4 2 1 6 と通信して、指定された角度で反応部位を照明することができる。

10

【0304】

複数のモジュール 4 2 2 6 ~ 4 2 4 8 はまた、バイオセンサ 4 2 1 2 と通信するデバイスモジュール 4 2 3 6 と、バイオセンサ 4 2 1 2 に関連する識別情報を判定する識別モジュール 4 2 3 8 とを含んでもよい。デバイスモジュール 4 2 3 6 は、例えば、システム受け部 4 2 1 0 と通信して、バイオセンサが配列決定システム 4 2 0 0 A との電氣的及び流体的接続を確立したことを確認してもよい。識別モジュール 4 2 3 8 は、バイオセンサ 4 2 1 2 を識別する信号を受信してもよい。識別モジュール 4 2 3 8 は、バイオセンサ 4 2 1 2 の識別情報を使用して、他の情報をユーザに提供してもよい。例えば、識別モジュール 4 2 3 8 は、ロット番号、製造日、又はバイオセンサ 4 2 1 2 で動作することが推奨されるプロトコルを判定し、その後表示してもよい。

20

【0305】

複数のモジュール 4 2 2 6 ~ 4 2 4 8 はまた、バイオセンサ 4 2 1 2 から信号データ（例えば、画像データ）を受信及び分析する分析モジュール 4 2 4 4（信号処理モジュール又は信号プロセッサとも呼ばれる）も含む。分析モジュール 4 2 4 4 は、検出／画像データを記憶するためのメモリ（例えば、RAM 又はフラッシュ）を含む。検出データは、ピクセル信号の複数のシーケンスを含むことができ、それにより、数百万個のセンサ（又はピクセル）のそれぞれからのピクセル信号のシーケンスを、多くのベースコールサイクルにわたって検出することができる。信号データは、その後の分析のために記憶されてもよく、又はユーザインターフェース 4 2 1 8 に送信されて、所望の情報をユーザに表示することができる。いくつかの実施態様では、信号データは、分析モジュール 4 2 4 4 が信号データを受信する前に、固体撮像素子（例えば、CMOS 画像センサ）によって処理され得る。

30

【0306】

分析モジュール 4 2 4 4 は、複数の配列決定サイクルのそれぞれにおいて、光検出器から画像データを取得するように構成されている。画像データは、光検出器によって検出される発光信号から導出され、ニューラルネットワークベースのベースコーラー 2 9 0 0 を介して、複数の配列決定サイクルのそれぞれに対して画像データを処理し、複数の配列決定サイクルのそれぞれでクラスターのうちの少なくともいくつかに対するベースコールを生成する。光検出器は、1 つ以上のオーバーヘッドカメラ（例えば、バイオセンサ 4 2 1 2 上のクラスターの画像を上から撮影する Illumina の G A I I x の C C D カメラ）の一部であってもよく、又はバイオセンサ 4 2 1 2 自体の一部（例えば、バイオセンサ 4 2 1 2 上のクラスターの下にあり、底部からのクラスターの画像を取る Illumina の i S e q の CMOS 画像センサ）であってもよい。

40

【0307】

光検出器の出力は、それぞれクラスターの強度放射及びそれらの周囲背景を示す配列決定画像である。配列決定画像は、配列決定中にシーケンスにヌクレオチドを組み込む結果

50

として生成される強度放射を示す。強度放射は、関連するクラスター及びそれらの周囲の背景に由来するものである。配列決定画像は、メモリ 4 2 4 8 に記憶される。

【0308】

プロトコルモジュール 4 2 4 0 及び 4 2 4 2 は、主制御モジュール 4 2 2 4 と通信して、所定のアッセイプロトコルを実施する際にサブシステム 4 2 0 8、4 2 1 4 及び 4 2 0 4 の動作を制御する。プロトコルモジュール 4 2 4 0 及び 4 2 4 2 は、所定のプロトコルに従って特定の動作を実行するように配列決定システム 4 2 0 0 A に指示するための命令セットを含み得る。図示のように、プロトコルモジュールは、配列決定ごとの合成プロセスを実行するための様々なコマンドを発行するように構成された、合成による配列決定 (Sequencing-By-Synthesis、SBS) モジュール 4 2 4 0 であってもよい。SBS において、核酸テンプレートに沿った核酸プライマーの伸長を監視して、テンプレート中のヌクレオチドシーケンスを判定する。下にある化学プロセスは、重合 (例えば、ポリメラーゼ酵素により触媒される) 又はライゲーション (例えば、リガーゼ酵素により触媒される) であり得る。特定のポリマー系 SBS の実装形態では、プライマーに付加されるヌクレオチドの順序及びタイプの検出を使用してテンプレートのシーケンスを判定することができるように、蛍光標識ヌクレオチドをテンプレート依存様式でプライマー (それによってプライマーを伸長させる) に添加する。例えば、第 1 の SBS サイクルを開始するために、1 つ以上の標識されたヌクレオチド、DNA ポリメラーゼなどを、核酸テンプレートのアレイを収容するフローセル内に / それを介して送達することができる。核酸テンプレートは、対応する反応部位に位置してもよい。プライマー伸長が、組み込まれる標識ヌクレオチドを、撮像イベントを通して検出することができる、これらの反応部位が検出され得る。撮像イベントの間、照明システム 4 2 1 6 は、反応部位に励起光を提供することができる。任意選択的に、ヌクレオチドは、ヌクレオチドがプライマーに付加されると、更なるプライマー伸長を終結する可逆的終結特性を更に含むことができる。例えば、脱ブロック作用因子が送達されてその部分を除去するまで、その後の伸長が起こらないように、可逆的ターミネーター部分を有するヌクレオチド類似体をプライマーに付加することができる。したがって、可逆終端を使用する別の実装形態では、フローセル (検出が生じる前又は後) にデブロック試薬を送達するために、コマンドを与えることができる。1 つ以上のコマンドは、様々な送達ステップ間の洗浄をもたらすために与えられ得る。次いで、サイクルを n 回繰り返してプライマーを n ヌクレオチドだけ伸長させ、それによって長さ n のシーケンスを検出することができる。例示的な配列決定技術は、例えば、Bentley ら、Nature 456:53-59 (2008)、国際公開第 04/018497 号、米国特許第 7,057,026 号、国際公開第 91/06678 号、同第 07/123744 号、米国特許第 7,329,492 号、同第 7,211,414 号、同第 7,315,019 号、同第 7,405,281 号、及び米国特許出願公開第 2008/0147080 82 号に記載されており、これらの各々は、参照により本明細書に組み込まれる。

【0309】

SBS サイクルのヌクレオチド送達ステップでは、単一のタイプのヌクレオチドのいずれかを一度に送達することができ、又は複数の異なるヌクレオチドタイプ (例えば、A、C、T、及び G を一緒に) を送達することができる。一度に単一のタイプのヌクレオチドのみが存在するヌクレオチド送達構成では、異なるヌクレオチドは、個別化された送達に固有の時間的分離に基づいて区別することができるため、異なるヌクレオチドは別個の標識を有する必要はない。したがって、配列決定方法又は装置は、単一の色検出を使用することができる。例えば、励起源は、単一の波長又は単一の波長範囲の励起のみを提供する必要がある。ある時点で、送達フローセル内に存在する複数の異なるヌクレオチドをもたらすヌクレオチド送達構成では、異なるヌクレオチドタイプを組み込む部位は、混合物中のそれぞれのヌクレオチドタイプに付着された異なる蛍光標識に基づいて区別することができる。例えば、4 つの異なる蛍光団のうちの 1 つをそれぞれ有する 4 つの異なるヌクレオチドを使用することができる。一実施形態では、4 つの異なる蛍光団は、スペクトル

の4つの異なる領域における励起を使用して区別することができる。例えば、4つの異なる励起放射線源を使用することができる。あるいは、4つ未満の異なる励起源を使用することができるが、単一源からの励起放射線の光学的濾過を使用して、フローセルにおいて異なる励起放射線の範囲を生成することができる。

【0310】

いくつかの実装形態では、4つ未満の異なる色を、4つの異なるヌクレオチドを有する混合物中で検出することができる。例えば、ヌクレオチドの対は、同じ波長で検出することができるが、対のうちの1つのメンバーに対する強度の差に基づいて、又は、対の他のメンバーについて検出された信号と比較して明らかな信号を出現又は消失させる、対の1つのメンバーへの変化（例えば、化学修飾、光化学修飾、又は物理的改質を行うことを介して）に基づいて区別され得る。4個未満の色の検出を使用して4個の異なるヌクレオチドを区別するための例示的な装置及び方法が、例えば、米国特許出願第61／538,294号及び同第61／619,878号に記載されており、それらの全体が参照により本明細書に組み込まれる。2012年9月21日に出願された米国特許出願第13／624,200号は、その全体が参照により組み込まれる。

10

【0311】

複数のプロトコルモジュールはまた、バイオセンサ4212内の製品を増幅するための流体制御システム4208及び温度制御システム4204にコマンドを発行するように構成されたサンプル調製（又は発生）モジュール4242を含んでもよい。例えば、バイオセンサ4212は、配列決定システム4200Aに係合されてもよい。増幅モジュール4242は、バイオセンサ4212内の反応チャンバに必要な増幅成分を送達するために、流体制御システム4208に命令を発行することができる。他の実装形態では、反応部位は、テンプレートDNA及び／又はプライマーなどの増幅のためのいくつかの成分を既に含有していてもよい。増幅成分を反応チャンバに送達した後、増幅モジュール4242は、既知の増幅プロトコルに従って異なる温度段階を通して温度制御システム4204にサイクルするように指示し得る。いくつかの実装形態では、増幅及び／又はヌクレオチドの取り込みは、等温的に実行される。

20

【0312】

SBSモジュール4240は、クローン性アンプリコンのクラスターがフローセルのチャンネル内の局所領域上に形成されるブリッジPCRを実行するコマンドを発行することができる。ブリッジPCRを介してアンプリコンを発生させた後、アンプリコンを「線形化」して、一本鎖テンプレートDNAを作製してもよく、sstDNA及び配列決定プライマーは、関心領域に隣接する普遍シーケンスにハイブリダイズされてもよい。例えば、合成方法による可逆的ターミネーター系配列決定を、上記のように又は以下のように使用することができる。

30

【0313】

各ベースコール又は配列決定サイクルは、例えば、修飾DNAポリメラーゼ及び4タイプのヌクレオチドの混合物を使用することによって達成することができる単一の塩基によってsstDNAを延長することができる。異なるタイプのヌクレオチドは、固有の蛍光標識を有することができるが、各ヌクレオチドは、各サイクルにおいて単一塩基の組み込みのみが生じることを可能にする可逆的ターミネーターを更に有し得る。単一の塩基をsstDNAに添加した後、励起光が反応部位に入射し、蛍光発光を検出することができる。検出後、蛍光標識及びターミネーターは、sstDNAから化学的に切断され得る。別の同様のベースコール又は配列決定サイクルは、以下の通りであってもよい。そのような配列決定プロトコルでは、SBSモジュール4240は、バイオセンサ4212を通る試薬及び酵素溶液の流れを方向付けるように流体制御システム4208に指示することができる。本明細書に記載される装置及び方法とともに利用することができる例示的な可逆性ターミネーターベースのSBS方法は、米国特許出願公開第2007／0166705（A1）号、米国特許出願公開第2006／0188901（A1）号、米国特許第7,057,026号、米国特許出願公開第2006／0240439（A1）号明細書、米国特許

40

50

出願公開第2006/02814714709(A1)号明細書、国際公開第05/065814号、米国特許出願公開第2005/014700900(A1)号明細書、国際公開第06/08B199号及び国際公開第07/01470251号(それぞれ参照によりその全体が本明細書に組み込まれる)に記載されている。可逆性ターミネーターベースのSBSの例示的な試薬が、米国特許第7,541,444号、同第7,057,026号、同第7,414,14716号、同第7,427,673号、同第7,566,537号、同第7,592,435号、及び国際公開第07/14835368号に記載されており、これらのそれぞれは、参照により本明細書に組み込まれる。

【0314】

いくつかの実装形態では、増幅及びSBSモジュールは、単一のアッセイプロトコルで動作してもよく、例えば、テンプレート核酸は増幅され、続いて同じカートリッジ内で配列される。

【0315】

配列決定システム4200Aはまた、ユーザがアッセイプロトコルを再構成することを可能にし得る。例えば、配列決定システム4200Aは、決定されたプロトコルを修正するために、ユーザインターフェース4218を通じてユーザにオプションを提供し得る。例えば、バイオセンサ4212が増幅のために使用されると判定された場合、配列決定システム4200Aは、アニーリングサイクルの温度を要求し得る。更に、配列決定システム4200Aは、選択されたアッセイプロトコルに対して一般的に許容されないユーザ入力をユーザが提供した場合に、ユーザに警告を発行し得る。

【0316】

実施態様では、バイオセンサ4212は、数百万個のセンサ(又はピクセル)を含み、それらのそれぞれは、連続するベースコールサイクルにわたって複数のピクセル信号のシーケンスを発生させる。分析モジュール4244は、センサのアレイ上のセンサの行方向及び/又は列方向の場所に従って、ピクセル信号の複数のシーケンスを検出し、それらに対応するセンサ(又はピクセル)に属させる。

【0317】

図42Cは、ベースコールセンサ出力などの配列決定システム4200Aからのセンサデータの分析のためのシステムの簡略ブロック図である。図42Cの例では、システムは、構成可能プロセッサ4246を含む。構成可能プロセッサ4246は、中央演算処理装置(CPU)4252(すなわち、ホストプロセッサ)によって実行されるランタイムプログラム/論理4280と協調して、ベースコーラー(例えば、ニューラルネットワークベースのベースコーラー2900)を実行することができる。配列決定システム4200Aは、バイオセンサ4212及びフローセルを含む。フローセルは、遺伝物質のクラスターが、遺伝物質中の塩基を識別するためにクラスター内の反応を引き起こすために使用される一連のクラスターフローに曝露される1つ以上のタイルを含み得る。センサは、タイルデータを提供するために、フローセルの各タイルにおけるシーケンスの各サイクルの反応を感知する。遺伝的配列決定はデータ集約的操作であり、このデータ集約的操作は、ベースコールセンサデータを、ベースコール動作中に感知された各遺伝物質群のベースコールのシーケンスに変換する。

【0318】

本実施例のシステムは、ベースコール動作を調整するためのランタイムプログラム/論理4280を実行するCPU4252と、タイルデータのアレイのシーケンスを記憶するメモリ4248Bと、ベースコール動作によって生成されたベースコールリードと、ベースコール動作で使用される他の情報と、を記憶する。また、この図では、システムは、構成ファイル(単数又は複数)、例えば、FPGAビットファイル、並びに構成可能プロセッサ4246を構成及び再構成し、ニューラルネットワークを実行するために使用されるニューラルネットワークのモデルパラメータを記憶するメモリ4248Aを含む。配列決定システム4200Aは、構成可能プロセッサを構成するためのプログラムを含み得、いくつかの実装形態では、ニューラルネットワークを実行する再構成可能プロセッサを含み

10

20

30

40

50

得る。

【0319】

配列決定システム4200Aは、バス4289によって構成可能プロセッサ4246に結合される。バス4289は、一例では、PCI-SIG規格（PCI Special Interest Group）によって現在維持及び開発されているPCIe規格（Peripheral Component Interconnect Express）と互換性のあるバス技術などの高スループット技術を使用して実装することができる。また、この実施例では、メモリ4248Aは、バス4293によって構成可能プロセッサ4246に結合される。メモリ4248Aは、構成可能プロセッサ4246を有する回路基板上に配設されたオンボードメモリであってもよい。メモリ4248Aは、ベースコール動作で使用される作業データの構成可能プロセッサ4246による高速アクセスに使用される。バス4293はまた、PCIe規格と互換性のあるバス技術などの高スループット技術を使用して実装することもできる。

10

【0320】

フィールドプログラマブルゲートアレイFPGA、粗いグレー構成可能な再構成可能アレイCGRAs、並びに他の構成可能かつ再構成可能なデバイスを含む構成可能プロセッサは、コンピュータプログラムを実行する汎用プロセッサを使用して達成され得るよりも、より効率的に又はより高速に様々な機能を実装するように構成することができる。構成可能プロセッサの構成は、時にはビットストリーム又はビットファイルと呼ばれる構成ファイルを生成するために機能的な説明を編集することと、構成ファイルをプロセッサ上の構成可能要素に配布することと、を含む。構成ファイルは、データフローパターンを設定するように回路を構成することにより、分散メモリ及び他のオンチップメモリリソースの使用、ルックアップテーブルコンテンツ、構成可能な論理ブロックの動作、及び構成可能な論理ブロックの動作、及び構成可能なアレイの構成可能な相互接続及び他の要素のような構成可能な実行ユニットを含む。構成ファイルがフィールド内で変更され得る場合、ロードされた構成ファイルを変更することによって構成ファイルを変更することができる場合に再構成可能である。例えば、構成ファイルは、揮発性SRAM要素内に、不揮発性読み書きメモリ素子内に記憶されてもよく、構成可能又は再構成可能プロセッサ上の構成可能要素のアレイ間に分散されたものであってもよい。様々な市販の構成可能プロセッサは、本明細書に記載されるようなベースコール動作において使用するのに好適である。例としては、Googleのテンソル処理ユニット（TPU）（商標）、GX4 Rackmount Series（商標）、GX9 Rackmount Series（商標）、NVIDIA DGX-1（商標）、Microsoft's Stratix V FPGA（商標）、GraphcoreのIntelligent Processor Unit（IPU）（商標）、QualcommのZeroth Platform（商標）（Snapdragon processors（商標）、NVIDIA Volta（商標）、NVIDIA DRIVE PX（商標）、NVIDIAのJETSON TX1/TX2 MODULE（商標）、Intel's Nirvana（商標）、Movidius VPU（商標）、Fujitsu DPI（商標）、アームDynamic IQ（商標）、IBM TrueNorth（商標）、Lambda GPU Server with Testa V100s（商標）、Xilinx Alveo（商標）U200、Xilinx Alveo（商標）U2190、Xilinx Alveo（商標）U280、Intel/Altera Stratix（商標）GX2800、Intel/Altera Stratix（商標）GX2800、及びIntel Stratix（商標）GX10M、が含まれる。いくつかの実施例では、ホストCPUは、構成可能プロセッサと同じ集積回路上に実装することができる。

20

30

40

【0321】

本明細書に記載の実施態様は、構成可能プロセッサ4246を使用して、ニューラルネットワークベースのベースコーラー2900を実装する。構成可能プロセッサ4246の構成ファイルは、高レベルの記述言語HDL又はレジスタ転送レベルRTL言語仕様を使

50

用して実行される論理機能を指定することによって実装され得る。本明細書は、選択された構成可能プロセッサが構成ファイルを発生させるように設計されたリソースを使用してコンパイルすることができる。構成可能プロセッサではない場合がある特定用途向け集積回路の設計を発生させる目的で、同じ又は類似の仕様をコンパイルすることができる。

【0322】

したがって、本明細書に記載される全ての実施態様における構成可能プロセッサ構成可能プロセッサ4246の代替例は、本明細書に記載されるようなニューラルネットワークベースのベースコール動作を実行するように構成された、特定用途向けASIC若しくは専用集積回路若しくは集積回路のセットを含む構成されたプロセッサ、又はシステムオンチップSOCデバイス、又はシステムオンチップSOCデバイス、又はグラフィック処理ユニット(GPU)プロセッサ若しくは粗粒度再構成可能アーキテクチャ(Coarse-Grained Reconfigurable Architecture、CGRA)プロセッサを含む。

10

【0323】

一般に、ニューラルネットワークの動作を実行するように構成された、本明細書に記載の構成可能プロセッサ及び構成されたプロセッサは、本明細書ではニューラルネットワークプロセッサと呼ばれる。

【0324】

構成可能プロセッサ4246は、この例では、CPU4252によって実行されるプログラムを使用して、又は構成可能要素4291のアレイを構成する他のソースによってロードされた構成ファイルによって構成される(例えば、構成論理ブロック(Configuration Logic Block、CLB)、例えばルックアップテーブル(Look Up Table、LUT)、フリップフロップ、演算処理ユニット(PMU)、及び計算メモリユニット(Compute Memory Unit、CMU)、構成可能なI/Oブロック、プログラマブル相互接続)を計算して、ベースコール機能を実行する。この実施例では、構成は、バス4289及び4293に結合され、ベースコール動作で使用される要素間でデータ及び制御パラメータを分配する機能を実行するデータフロー論理4297を含む。

20

【0325】

また、構成可能プロセッサ4246は、データフロー論理4297を用いて構成されて、ニューラルネットワークベースのベースコーラー2900を実行する。論理4297は、マルチサイクル実行クラスター(例えば、4279)を含み、この実施例では、実行クラスターXを介した実行クラスター1を含む。マルチサイクル実行クラスターの数は、動作の所望のスループットを伴うトレードオフ、及び構成可能プロセッサ4246上の利用可能なリソースに従って選択することができる。

30

【0326】

マルチサイクル実行クラスターは、構成可能プロセッサ4246上の構成可能な相互接続及びメモリリソースを使用して実装されるデータ流路4299によってデータフロー論理4297に結合される。また、マルチサイクル実行クラスターは、構成可能な相互接続及びメモリリソースを使用して、例えば構成可能プロセッサ4246上に実装される制御経路4295によってデータフロー論理4297に結合され、利用可能な実行クラスターを示す制御信号、ニューラルネットワークベースのベースコーラー2900の動作の実行のための入力ユニットを利用可能な実行クラスターに提供するための準備、ニューラルネットワークベースのベースコーラー2900の学習されたパラメータを提供するための準備、ベースコール分類データの出力パッチを提供するための準備、並びにニューラルネットワークベースのベースコーラー2900の実行に使用される他の制御データを提供するのである。

40

【0327】

構成可能プロセッサ4246は、学習されたパラメータを使用してニューラルネットワークベースのベースコーラー2900の動作を実行して、ベースコール動作の検知サイクルに関する分類データを生成するように構成されている。ニューラルネットワークベースのベースコーラー2900の動作を実行して、ベースコール動作の被験者検知サイクルの

50

分類データを生成する。ニューラルネットワークベースのベースコーラー 2900 の動作は、N 個の検知サイクルのそれぞれの検知サイクルからのタイルデータのアレイの数 N を含むシーケンスで動作し、N 回の検知サイクルは、本明細書に記載される例での時間シーケンスにおける動作ごとの 1 つの塩基位置に対する異なるベースコール動作のためのセンサデータを提供する。任意選択的に、N 個の感知サイクルのうちのいくつかは、実行される特定のニューラルネットワークモデルに従って必要に応じて、シーケンスから出ることができる。数 N は、1 を超える任意の数であり得る。本明細書に記載されるいくつかの実施例では、N 個の感知サイクルの感知サイクルは、時系列で、被験者感知サイクルに先行する少なくとも 1 つの感知サイクル、及び被験者サイクルに後続する少なくとも 1 つの感知サイクルについての感知サイクルのセットを表す。本明細書では、数 N が 5 以上の整数である実施例が記載される。

10

【0328】

データフロー論理 4297 は、N 個のアレイの空間的に位置合わせされたパッチのタイルデータを含む所与の動作のための入力ユニットを使用して、ニューラルネットワークベースのベースコーラー 2900 の動作のために、メモリ 4248A から構成可能プロセッサ 4246 に、タイルデータ及びモデルパラメータの少なくともいくつかの学習されたパラメータを移動させるように構成されている。入力ユニットは、1 回の DMA 動作におけるダイレクトメモリアクセス動作によって、又は、配備されたニューラルネットワークの実行と協調して、利用可能なタイムスロットの間に移動するより小さいユニット内で移動させることができる。

20

【0329】

本明細書に記載される感知サイクルのタイルデータは、1 つ以上の特徴を有するセンサデータのアレイを含むことができる。例えば、センサデータは、DNA、RNA、又は他の遺伝物質の遺伝的シーケンスにおける塩基位置で 4 塩基のうちの 1 つを識別するために分析される 2 つの画像を含むことができる。タイルデータはまた、画像及びセンサに関するメタデータを含むことができる。例えば、ベースコール動作の実施態様では、タイルデータは、タイル上の遺伝物質群の中心からのセンサデータのアレイ内の各ピクセルの距離を示す中心情報からの距離などの、クラスターとの画像の位置合わせに関する情報を含むことができる。

30

【0330】

以下に記載されるように、ニューラルネットワークベースのベースコーラー 2900 の実行中に、タイルデータはまた、ニューラルネットワークベースのベースコーラー 2900 の実行中に生成されるデータも含み得る。このデータは、ニューラルネットワークベースのベースコーラー 2900 の動作中に再計算されるのではなく再使用することができる中間データと称される。例えば、ニューラルネットワークベースのベースコーラー 2900 の実行中に、データフロー論理 4297 は、タイルデータのアレイの所与のパッチのセンサデータの代わりに、中間データをメモリ 4248A に書き込むことができる。このような実施態様は、以下により詳細に記載される。

【0331】

図示されているように、ベースコール動作の検知サイクルからタイルのセンサデータを含むタイルデータを記憶するランタイムプログラム／論理 4280 によってアクセス可能なメモリ（例えば、4248A）を含む、ベースコールセンサ出力の分析のためのシステムが説明される。また、システムは、メモリへのアクセスを有する構成可能プロセッサ 4246 などのニューラルネットワークプロセッサを含む。ニューラルネットワークプロセッサは、訓練されたパラメータを使用してニューラルネットワークの動作を実行して、感知サイクルのための分類データを生成するように構成される。本明細書に記載されるように、ニューラルネットワークの動作は、被験者サイクルを含む N 個の感知サイクルのそれぞれの感知サイクルからタイルデータの N 個のアレイのシーケンスで動作して、被験者サイクルの分類データを生成する。データフロー論理 4297 は、N 個の感知サイクルのそれぞれの感知サイクルからの N 個のアレイの空間的に位置合わせされたパッチのデータを

40

50

含む入力ユニットを使用して、ニューラルネットワークの実行のために、メモリからニューラルネットワークプロセッサにタイルデータ及び訓練されたパラメータを移動させるために提供される。

【0332】

また、ニューラルネットワークプロセッサがメモリへのアクセスを有し、複数の実行クラスターを含み、ニューラルネットワークを実行するように構成された複数の実行クラスター内の実行クラスターを含むシステムも説明される。データフロー論理1997は、複数の実行クラスター内の利用可能な実行クラスターにタイルデータの入力ユニットを提供することであって、入力ユニットは、被験者検知サイクルを含むそれぞれの検知サイクルからタイルデータのアレイのN個の空間的に位置合わせされたパッチを含む、ことと、実行クラスターに、N個の空間的に位置合わせされたパッチをニューラルネットワークに適用して、被験者検知サイクルの空間的に位置合わせされたパッチに対する分類データの出力パッチを生成させることであって、Nは1より大きい、ことと、を行うために、メモリへのアクセス及び複数の実行クラスターの中の実行クラスターへのアクセスを有する。

【0333】

図43Aは、ホストプロセッサによって実行されるランタイムプログラム（例えば、ランタイム論理4280）の機能を含む、ベースコール動作の態様を示す簡略図である。この図では、フローセルからの画像センサの出力は、ライン4300上で画像処理スレッド4301に提供され、画像処理スレッド4301は、個々のタイルのセンサデータのアレイ内の位置合わせ及び配置、及び画像の再サンプリングなどの画像上のプロセスを実行することができ、フローセル内の各タイルのタイルクラスターマスクを計算するプロセスによって使用することができ、フローセルの対応するタイル上の遺伝子材料のクラスターに対応するセンサデータのアレイ内のピクセルを識別するプロセスによって使用することができる。画像処理スレッド4301の出力は、CPU内のディスパッチ論理4303に、ライン4302上に提供され、それは、高速バス4304上又は高速バス4306上のデータキャッシュ4305（例えば、SSD記憶装置）に、ベースコール動作の状態に従って、図19Cの構成可能プロセッサ1946などのニューラルネットワークプロセッサハードウェア4307にタイルデータのアレイを転送する。処理され、変換された画像は、以前に使用されたサイクルを感知するために、データキャッシュ4305上に記憶され得る。ハードウェア4307は、ニューラルネットワークによって出力された分類データをディスパッチ論理4303に返し、ディスパッチ論理4303は、情報をデータキャッシュ4305に、又はライン4308上でスレッド4309に渡し、それは、分類データを使用してベースコーラー及び品質スコア計算を実行し、ベースコールリードのための標準フォーマットでデータを配置することができる。ベースコーラー及び品質スコア計算を実行するスレッド4309の出力は、ライン4310上でスレッド4311に提供され、それは、ベースコールリードを集約し、データ圧縮などの他の動作を実行し、結果として得られたベースコーラー出力を顧客による利用のために指定された宛先に書き込む。

【0334】

いくつかの実施態様では、ホストは、ニューラルネットワークを支持するハードウェア4307の出力の最終処理を実行する、スレッド（図示せず）を含むことができる。例えば、ハードウェア4307は、マルチクラスターニューラルネットワークの最終層から分類データの出力を提供することができる。ホストプロセッサは、ベースコール及び品質スコアスレッド4302によって使用されるデータを設定するために、分類データを超えて、ソフトマックス関数などの出力起動機能を実行することができる。また、ホストプロセッサは、ハードウェア4307への入力前のタイルデータのバッチ正規化などの入力動作（図示せず）を実行することができる。

【0335】

図43Bは、図19Cの構成などの構成可能プロセッサ1946の構成の簡略図である。図43Bでは、構成可能プロセッサ1946は、複数の高速PCIeインターフェースを有するFPGAを含む。FPGAは、図19Cを参照して説明されるデータフロー論理

1997を含むラッパー4390を用いて構成されている。ラッパー4390は、CPU通信リンク4377を介して、CPU内のランタイムプログラムとのインターフェース及び調整を管理し、DRAM通信リンク4397を介してオンボードDRAM4399（例えば、メモリ1448A）との通信を管理する。ラッパー4390内のデータフロー論理1997は、数Nのサイクルのために、オンボードDRAM4399上のタイルデータのアレイをクラスター4385まで横断することによって取得されたパッチデータを提供し、クラスター4385からプロセスデータ4387を取得して、オンボードDRAM4399に配信する。ラッパー4390はまた、タイルデータの入力アレイと、分類データの出力パッチの両方について、オンボードDRAM4399とホストメモリとの間のデータの転送を管理する。ラッパーは、ライン4383上でパッチデータを、割り当てられたクラスター4385に転送する。ラッパーは、クラスター4385に、ライン4381上で、オンボードDRAM4399から取得された重みやバイアスなどの訓練されたパラメータを提供する。ラッパーは、クラスター4385に、ライン4379上で、CPU通信リンク4377を介してホスト上のランタイムプログラムから提供されるか又はそれに応答して発生した構成及び制御データを提供する。クラスターはまた、タイルデータのアレイの横断を管理して空間的に位置合わせされたパッチデータを提供し、かつクラスター4385のリソースを使用してパッチデータ上でマルチサイクルニューラルネットワークを実行するために、ホストからの制御信号と協働して使用される状態信号を、ライン4389上でラッパー4390に提供することができる。

10

20

【0336】

上述のように、タイルデータの複数のパッチのうちの対応するパッチ上で実行するように構成されたラッパー4390によって管理される単一の構成可能プロセッサ上に複数のクラスターが存在し得る。各クラスターは、本明細書に記載される複数の感知サイクルのタイルデータを使用して、被験者感知サイクルにおけるベースコールの分類データを提供するように構成することができる。

【0337】

システムの例では、フィルタ重み及びバイアスのようなカーネルデータを含むモデルデータをホストCPUから構成可能プロセッサに送信することができ、その結果、モデルは、サイクル数の関数として更新され得る。ベースコール動作は、代表的な例では、数百の感知サイクルの順序で含むことができる。ベースコール動作は、いくつかの実施態様では、ペアリングされた端部読み取りを含むことができる。例えば、モデル訓練されたパラメータは、20サイクルごと（又は他の数のサイクル）ごとに、又は特定のシステム及びニューラルネットワークモデルに実装される更新パターンに従って更新されてもよい。タイル上の遺伝的クラスター内の所与のストリングのためのシーケンスが、ストリングの第1の端部から下方に（又は上方に）延在する第1の部分と、ストリングの第2の端部から上方に（又は下方に）延在する第2の部分とを含む、ペアリングされた端部リードを含むいくつかの実施態様では、学習されたパラメータは、第1の部分から第2の部分への遷移で更新され得る。

30

【0338】

いくつかの実施例では、タイルのための感知データの複数サイクルの画像データは、CPUからラッパー4390に送信され得る。ラッパー4390は、任意選択的に、感知データの一部の前処理及び変換を行い、その情報をオンボードDRAM4399に書き込むことができる。各感知サイクルの入力タイルデータは、タイル当たり感知サイクル当たり4000×3000ピクセル以上を含むセンサデータのアレイを含むことができ、2つの特徴はタイルの2つの画像の色を表し、1ピクセル当たり1つ又は2つのバイトを含むセンサデータのアレイを含むことができる。数Nが、マルチサイクルニューラルネットワークの各動作において使用される3回の検知サイクルである実施態様用では、マルチサイクルニューラルネットワークの各動作のためのタイルデータのアレイは、数当たり数百メガバイトの数で消費することができる。システムのいくつかの実施態様では、タイルデータはまた、タイルごとに1回記憶されたクラスターセンターからの距離（distance-from

40

50

-cluster center、D F C) データのアレイ、又はセンサデータ及びタイルに関する他のタイプのメタデータも含む。

【0339】

動作中、マルチサイクルクラスターが利用可能である場合、ラッパーは、パッチをクラスターに割り当てる。ラッパーは、タイルの横断面にタイルデータの次のパッチをフェッチし、適切な制御及び構成情報とともに割り当てられたクラスターに送信する。クラスターは、構成可能プロセッサ上の十分なメモリを用いて構成されて、パッチを含むデータのパッチを、定位置に処理されているいくつかのシステム内で複数サイクルから保持するのに十分なメモリを有するように構成することができ、様々な実施態様では、ピンポンバッファ技術又はラスタ走査技術を使用して処理される。

10

【0340】

割り当てられたクラスターが、現在のパッチのニューラルネットワークのその動作を完了し、出力パッチを生成すると、それはラッパーに信号を送る。ラッパーは、割り当てられたクラスターから出力パッチを読み出すか、あるいは割り当てられたクラスターは、データをラッパーにブッシュする。次いで、ラッパーは、D R A M 4 3 9 9 内の処理されたタイルのための出力パッチを組み立てることになる。タイル全体の処理が完了し、データの出力パッチが D R A M に転送されると、ラッパーは、処理された出力アレイを、特定のフォーマットでホスト／C P U に返送する。いくつかの実施態様では、オンボード D R A M 4 3 9 9 は、ラッパー 4 3 9 0 内のメモリ管理論理によって管理される。ランタイムプログラムは、リアルタイム分析を提供するために連続フローで動作する全てのサイクルについての全てのタイルデータのアレイの分析を完了するために、配列決定動作を制御することができる。

20

【0341】

図 4 4 は、C P U 上で状態データを決定し、ベースコールのために C P U から F P G A に状態データをロードする一実施態様を示す。

【0342】

用語及び追加の実施態様

ベースコールは、蛍光標識されたタグを検体とともに組み込む又は取り付けることを含む。検体は、ヌクレオチド又はオリゴヌクレオチドであってよく、タグは、特定のヌクレオチド型 (A、C、T、又は G) であってもよい。励起光は、タグを有する検体に向けられ、タグは検出可能な蛍光シグナル又は強度発光を発する。強度発光は、検体に化学的に結合された励起タグによって放出される光子を示す。

30

【0343】

特許請求の範囲を含む本出願全体を通して、「画像、画像データ、又は検体及びそれらの周囲背景の強度放射を示す画像領域が使用されるとき、それらは、検体に取り付けられたタグの強度放射を指す。当業者であれば、取り付けられたタグの強度放射は、タグが取り付けられている検体の強度放射を表すか、又はそれに相当し、したがって互換的に使用されることを理解するであろう。同様に、検体の特性は、検体に取り付けられたタグ、又は取り付けられたタグからの強度放射の特性を指す。例えば、検体の中心とは、検体に取り付けられたタグによって放出される強度放射の中心を指す。別の例では、検体の周囲の背景とは、検体に取り付けられたタグによって放出される強度放射の周囲の背景を指す。

40

【0344】

特許、特許出願、記事、書籍、論文、及びウェブページが挙げられるがこれらに限定されない、本出願において引用された全ての文献及び類似材料は、そうした文献及び類似材料の形式に関わりなく、それらの全体が参照により明示的に組み込まれる。組み込まれた文献及び類似の材料のうちの 1 つ以上が、以下に限定するわけではないが、定義された用語、用語使用、記載された技術などで、本出願とは異なる、又は矛盾する場合には、本出願が優越する。

【0345】

開示される技術は、核酸テンプレート又はその相補体、例えば、D N A 若しくは R N A

50

ポリヌクレオチド又は他の核酸試料などの核酸試料から得ることができる核酸シーケンス情報の品質及び量を改善するためにニューラルネットワークを使用する。したがって、開示される技術の特定の実施は、以前に利用可能な方法と比較して、より高いスループットのポリヌクレオチド配列決定、例えば、より高いDNA又はRNAシーケンスデータの収集速度、シーケンスデータ収集におけるより高い効率、及び／又はそのようなシーケンスデータを得る低コストを提供する。

【0346】

開示される技術は、ニューラルネットワークを使用して、固相核酸クラスターの中心を識別し、そのようなクラスターの配列決定中に生成される光信号を解析して、隣接する、隣接する、又は重複するクラスター間で曖昧さなく区別して、単一の離散したソースクラスターに配列決定シグナルを割り当てる。したがって、これら及び関連する実施態様は、高密度クラスターアレイの領域からの、シーケンスデータなどの有意義な情報の回収を可能にし、有用な情報は、重複する又は非常に近接して離間配置された隣接クラスターの影響を混乱させることに起因して、そのような領域から以前に得られなかった場合がある。重複するシグナルの効果（例えば、核酸配列決定において使用されるような）の効果を含む。

10

【0347】

以下により詳細に記載されるように、特定の実施態様では、本明細書で提供されるように、1つ又は複数の核酸クラスターに固定化された固体支持体を含む組成物が提供される。各クラスターは、同じシーケンスの複数の固定化された核酸を含み、本明細書で提供されるような検出可能な中心標識を有する識別可能な中心を有し、識別可能な中心は、クラスター内の周囲領域において固定化された核酸と区別可能である。また、識別可能な中心を有するこのようなクラスターを作製及び使用するための方法も本明細書に記載される。

20

【0348】

本開示の実施態様は、多数の状況での使用が見出され、その利点は、クラスター内の実質的に中心の位置の位置を識別、決定、アノテーション、記録、ないしは別の方法で割り当てる能力から得られる、多くの状況において使用が見出されるであろう。ハイスループット核酸配列決定、光学的又は他のシグナルを個別のソースクラスターに割り当てるための画像分析アルゴリズムの開発、及び固定化された核酸クラスターの中心の認識が望ましい及び有益である他の用途が望ましい。

30

【0349】

特定の実施態様では、本発明は、核酸配列決定（例えば、「配列決定」）などのハイスループット核酸分析に関連する方法を企図する。例示的なハイスループット核酸解析としては、非限定的に、デノボ配列決定、再配列決定、全ゲノム配列決定、遺伝子発現解析、遺伝子発現モニタリング、エピジェネティクス分析、ゲノムメチル化分析、対立遺伝子特異的プライマー伸長（Allele Specific Primer Extension、APSE）、遺伝的多様性プロファイリング、全ゲノム多型発見及び解析、単一ヌクレオチド多型解析、ハイブリダイゼーション系配列決定法などが挙げられる。当業者は、様々な異なる核酸が、本発明の方法及び組成物を使用して分析され得ることを理解するであろう。

【0350】

本発明の実施は核酸配列決定に関連して記載されているが、それらは、異なる時点で取得された画像データ、空間位置、又は他の時間的若しくは物理的観点で取得された画像データが分析される任意の分野において適用可能である。例えば、本明細書に記載される方法及びシステムは、マイクロアレイ、生物学的検体、細胞、生物などからの画像データが取得され、異なる時点又は視点で取得され、分析される、分子生物学及び細胞生物学の分野において有用である。画像は、蛍光顕微鏡法、光学顕微鏡法、共焦点顕微鏡法、光学画像化法、磁気共鳴画像化法、トモグラフィ走査などが挙げられるが、これらに限定されない、当該技術分野において既知の任意の数の技術を使用して得ることができる。別の例として、本明細書に記載される方法及びシステムは、監視、空中、又は衛星撮像技術などによって取得された画像データが、異なる時点又は視点で取得され、分析される場合に適用

40

50

することができる。本方法及びシステムは、視野内で取得された画像を分析するのに特に有用であり、この視野内で、観察される検体は、視野内の互いに対して同じ場所に留まる。しかしながら、検体は、別個の画像で異なる特性を有してもよく、例えば、検体は、視野の別々の画像において異なるように見える場合がある。例えば、検体は、異なる画像で検出された所与の検体の色とは異なるように見える場合があり、異なる画像内の所与の検体のために検出された信号の強度の変化、又は更には、1つの画像中の所与の検体の信号の外観、及び別の画像内の検体の信号の消失を示し得る。

【0351】

本明細書に記載される例は、学術分析又は商業的分析のための様々な生物学的又は化学的プロセス及びシステムにおいて使用されてもよい。より具体的には、本明細書に記載される例は、指定された反応を示すイベント、特性、品質、又は特性を検出することが望ましい様々なプロセス及びシステムにおいて使用されてもよい。例えば、本明細書に記載される例としては、光検出デバイス、バイオセンサ、及びそれらの構成要素、並びにバイオセンサとともに動作するバイオアッセイシステムが挙げられる。いくつかの実施例では、装置、バイオセンサ、及びシステムは、フローセルと、実質的に一体型構造で一緒に（取り外し可能に又は固定的に）結合された1つ以上の光センサと、を含み得る。

【0352】

装置、バイオセンサ、及びバイオアッセイシステムは、個別に又は集合的に検出され得る複数の指定された反応を実施するように構成されてもよい。装置、バイオセンサ、及びバイオアッセイシステムは、複数の指定された反応が並行して生じる多数のサイクルを実行するように構成されてもよい。例えば、装置、バイオセンサ、及びバイオアッセイシステムを使用して、酵素操作及び光又は画像検出／捕捉の反復サイクルを通して、DNA特徴の高密度シーケンスをシーケンスすることができる。したがって、デバイス、バイオセンサ、及びバイオアッセイシステム（例えば、1つ以上のカートリッジを介した）は、試薬又は他の反応成分を反応溶液中に送達する1つ以上のマイクロ流体チャネル、バイオセンサ、及びバイオアッセイシステムを含んでもよい。いくつかの実施例では、反応溶液は、約5以下、又は約4以下、又は約3以下のpHを含むなど、実質的に酸性であってもよい。いくつかの他の実施例では、反応溶液は、約8以上、又は約9以上、又は約10以上のpHを含むなど、実質的にアルカリ性／塩基性であってもよい。本明細書で使用するとき、用語「酸性」及びその文法的変異体は、約7未満のpH値を指し、用語「塩基性」、「アルカリ性」及びその文法的変異型は、約7を超えるpH値を指す。

【0353】

いくつかの実施例では、反応部位は、均一又は反復パターンなどの所定の方法で提供又は離間される。いくつかの他の実施例では、反応部位はランダムに分布している。反応部位のそれぞれは、関連する反応部位からの光を検出する1つ以上の光ガイド及び1つ以上の光センサと関連付けることができる。いくつかの実施例では、反応部位は、指定された反応を少なくとも部分的に区画化し得る反応凹部又はチャンバ内に位置する。

【0354】

本明細書で使用するとき、「指定された反応」は、対象となる検体などの対象となる化学物質又は生物学的物質の化学的、電気的、物理的、又は光学的特性（又は品質）のうちの少なくとも1つの変化を含む。特定の実施例では、指定された反応は、例えば、蛍光標識生体分子を対象とする蛍光標識生体分子の組み込みなどの正の結合事象である。より一般的には、指定された反応は、化学変換、化学変化、又は化学的相互作用であってもよい。指定された反応はまた、電気特性の変化であってもよい。特定の実施例では、指定された反応は、検体と蛍光標識された分子を組み込むことを含む。検体は、オリゴヌクレオチドであってもよく、蛍光標識分子は、ヌクレオチドであってもよい。励起光が標識ヌクレオチドを有するオリゴヌクレオチドに向けられ、蛍光団が検出可能な蛍光シグナルを発するときに、指定された反応が検出され得る。代替例では、検出された蛍光は、化学発光又は生物発光の結果である。指定された反応はまた、例えば、ドナー蛍光団をアクセプタ蛍光団に近接させることによって蛍光（又はF o r s t e r）共鳴エネルギー移動（fluor

escence resonance energy transfer、FRET)を増加させるか、ドナー蛍光団とアクセプタ蛍光団とを離すことによってFRETを減少させるか、消光剤を蛍光団から離すことによって蛍光を増加させるか、又は消光剤及び蛍光団を近接配置することによって蛍光を減少させることができる。

【0355】

本明細書で使用する時、「反応溶液」、「反応成分」又は「反応物質」は、少なくとも1つの指定された反応を得るために使用され得る任意の物質を含む。例えば、潜在的な反応成分としては、例えば、試薬、酵素、サンプル、他の生体分子、及び緩衝液が挙げられる。反応成分は、溶液中の反応部位に送達されてもよく、及び／又は反応部位で固定されてもよい。反応成分は、反応部位で固定化された対象検体などの別の物質と直接又は間接的に相互作用し得る。上記のように、反応溶液は、実質的に酸性であってもよい（すなわち、比較的高い酸性度を含む）（例えば、約5以下のpH、約4以下のpHを含む）、又は約3以下のpH、又は実質的にアルカリ性／塩基性（すなわち、比較的高いアルカリ性／塩基性を含む）（例えば、約8以上のpH、約9以上のpH、又は約10以上のpHを含む）。

10

【0356】

本明細書で使用する時、用語「反応部位」は、少なくとも1つの指定された反応が生じ得る局所的領域である。反応部位は、物質がその上に固定され得る反応構造又は基材の支持表面を含んでもよい。例えば、反応部位は、その上に反応成分、例えば、その上に核酸のコロニーを有する反応構造（フローセルのチャンネル内に位置付けられ得る）の表面を含んでもよい。いくつかのこのような実施例では、コロニー中の核酸は同じ配列を有し、例えば、一本鎖又は二本鎖テンプレートのクローンコピーである。しかしながら、いくつかの実施例では、反応部位は、例えば、一本鎖又は二本鎖形態で、単一の核酸分子のみを含有してもよい。

20

【0357】

複数の反応部位は、反応構造に沿ってランダムに分布してもよく、又は所定の様式で配置されてもよい（例えば、マイクロアレイなどのマトリックス内の並列）。反応部位はまた、指定された反応を区画化するように構成された空間領域又は容積を少なくとも部分的に画定する反応室又は凹部を含むことができる。本明細書で使用する時、用語「反応チャンバ」又は「反応凹部」は、支持構造体の画定された空間領域（多くの場合、流路と流体連通している）を含む。反応凹部は、周囲環境又は空間領域から少なくとも部分的に分離されてもよい。例えば、複数の反応凹部は、検出表面などの共有された壁によって互いに分離されてもよい。より具体的な例として、反応凹部は、検出表面の内面によって画定された窪み、ウェル、溝、空洞、又は窪みを含むナノセルであってもよく、ナノセルが流路と流体連通することができるように、開口部又は開口部（すなわち、開側面である）を有することができる。

30

【0358】

いくつかの実施例では、反応構造の反応凹部は、固体がその中に完全に又は部分的に挿入され得るように、固体（半固体を含む）に対してサイズ及び形状を定められる。例えば、反応凹部は、捕捉ビーズを収容するような大きさ及び形状であってもよい。捕捉ビーズは、クローン的に増幅されたDNA又はその上の他の物質を有してもよい。あるいは、反応凹部は、およそその数のビーズ又は固体基材を受容するような大きさ及び形状であってもよい。別の例として、反応凹部は、拡散又はフィルタ流体又は反応凹部に流入し得る溶液を制御するように構成された多孔質ゲル又は物質で充填されてもよい。

40

【0359】

いくつかの実施例では、光センサ（例えば、フォトダイオード）は、対応する反応部位と関連付けられる。反応部位に関連する光センサは、関連する反応部位において指定された反応が生じたときに、少なくとも1つの光ガイドを介して、関連する反応部位からの光放射を検出するように構成されている。いくつかの場合では、複数の光センサ（例えば、光検出又はカメラデバイスのいくつかのピクセル）は、単一の反応部位に関連付けられて

50

もよい。他の場合では、単一の光センサ（例えば、単一のピクセル）は、単一の反応部位に、又は反応部位の群と関連付けられてもよい。バイオセンサの光センサ、反応部位、及び他の特徴は、光の少なくとも一部が反射されることなく光センサによって直接検出されるように構成されてもよい。

【0360】

本明細書で使用する時、「生物学的又は化学物質」は、生体分子、対象試料、対象検体、及び他の化学化合物を含む。生物学的物質又は化学物質を使用して、他の化学化合物を検出、識別、若しくは分析するか、又は他の化学化合物を研究又は分析するための仲介として機能してもよい。特定の実施例では、生物学的物質又は化学物質は、生体分子を含む。本明細書で使用する時、「生体分子」は、バイオポリマー、ヌクレオチド、核酸、ポリヌクレオチド、オリゴヌクレオチド、タンパク質、酵素、ポリペプチド、抗体、抗原、リガンド、受容体、多糖類、炭水化物、ポリリン酸、細胞、組織、生物、若しくはそれらの断片、又は前述の種の類似体若しくは模倣体などの任意の他の生物学的に活性な化学化合物のうちの少なくとも1つを含む。更なる例では、生物学的若しくは化学物質又は生体分子は、酵素又は試薬などの別の反応の生成物、例えば、ピロ配列決定反応においてピロリン酸を検出するために使用される酵素又は試薬などの酵素又は試薬の生成物を検出する。ピロホスフェート検出に有用な酵素及び試薬は、例えば、参照によりその全体が組み込まれる米国特許公開第2005/0244870（A1）号に記載されている。

10

【0361】

生体分子、試料、及び生物学的物質又は化学物質は、天然に存在しても合成であってもよく、反応凹部又は領域内の溶液又は混合物中に懸濁されてもよい。生体分子、試料、及び生物学的物質又は化学物質もまた、固体相又はゲル材料に結合されてもよい。生体分子、試料、及び生物学的物質又は化学物質はまた、医薬組成物を含んでもよい。場合によっては、対象とする生体分子、試料、及び生物学的物質又は化学物質は、標的、プローブ、又は検体と呼ばれる場合がある。

20

【0362】

本明細書で使用する時、「バイオセンサ」は、反応部位又は反応部位に近接して生じる指定された反応を検出するように構成された複数の反応部位を有する反応構造を含む装置を含む。バイオセンサは、固体光検出装置又は「撮像」デバイス（例えば、CCD又はCMOS光検出デバイス）、及び任意選択的に、それに取り付けられたフローセルを含んでもよい。フローセルは、反応部位と流体連通する少なくとも1つの流路を含み得る。1つの特定の例として、バイオセンサは、バイオアッセイシステムに流体的かつ電氣的に結合するように構成される。バイオアッセイシステムは、所定のプロトコル（例えば、配列番号合成）に従って反応部位に反応溶液を送達し、複数の撮像事象を実施してもよい。例えば、バイオアッセイシステムは、反応溶液を反応部位に沿って流すことができる。反応溶液のうちの少なくとも1つは、同じ又は異なる蛍光標識を有する4種類のヌクレオチドを含んでもよい。ヌクレオチドは、反応部位の対応するオリゴヌクレオチドなどに結合してもよい。次いで、バイオアッセイシステムは、励起光源（例えば、発光ダイオード（LED）などの固体光源）を使用して反応部位を照明することができる。励起光は、波長の範囲を含む所定の波長又は複数の波長を有してもよい。入射励起光によって励起された蛍光標識は、光センサによって検出され得る発光信号（例えば、励起光とは異なる波長又は波長の光、及び潜在的に互いに異なる）を提供することができる。

30

40

【0363】

本明細書で使用する時、用語「固定化された」は、生体分子又は生物学的物質又は化学物質に関して使用されるとき、生体分子又は生物学的物質又は化学物質を、光検出デバイス又は反応構造の検出表面などの表面に実質的に付着させることを含む。例えば、生体分子又は生物学的物質又は化学物質は、非共有結合（例えば、静電力、ファンデルワールス、及び疎水性界面の脱水）を含む吸着技術、並びに官能基又はリンカーが生体分子の表面への結合を促進する共有結合技術を含む吸着技術を用いて、反応構造の表面に固定化されてもよい。生体分子又は生物学的物質若しくは化学物質を表面に固定化することは、表

50

面の特性、生体分子又は生物学的物質若しくは化学物質を担持する液体媒体、並びに生体分子又は生物学的物質若しくは化学物質自体の特性に基づいてもよい。いくつかの場合において、表面は、生体分子（又は生物学的物質又は化学物質）を表面に固定化するのを容易にするために、表面を官能化（例えば、化学的又は物理的に修飾）してもよい。

【0364】

いくつかの実施例では、核酸は、その反応凹部の表面などの反応構造に固定化することができる。特定の実施例では、本明細書に記載される装置、バイオセンサ、バイオアッセイシステム及び方法は、天然ヌクレオチド、及び天然ヌクレオチドと相互作用するように構成された酵素の使用を含んでもよい。天然ヌクレオチドとしては、例えば、リボヌクレオチド又はデオキシリボヌクレオチドが挙げられる。天然ヌクレオチドは、一リン酸、二リン酸、又は三リン酸形態であってよく、アデニン（A）、チミン（T）、ウラシル（U）、グアニン（G）、又はシトシン（C）から選択される塩基を有することができる。しかしながら、上記ヌクレオチドの非天然ヌクレオチド、修飾ヌクレオチド、又は類似体を使用することができることが理解されるであろう。

10

【0365】

上記のように、生体分子又は生物学的物質又は化学物質は、反応構造の反応凹部内の反応部位に固定されてもよい。このような生体分子又は生物学的物質は、干渉嵌め、接着、共有結合、又は捕捉によって、反応凹部に物理的に保持又は固定化されてもよい。反応凹部に配置され得る物品又は固体の例としては、ポリマービーズ、ペレット、アガロースゲル、粉末、量子ドット、又は反応チャンバ内で圧縮及び／又は保持され得る他の固体が挙げられる。特定の実施態様では、反応凹部は、DNAオリゴヌクレオチドに共有結合することができるヒドロゲル層でコーティング又は充填されてもよい。特定の実施例では、DNAボールなどの核酸超構造は、例えば、反応凹部の内面に取り付けることによって、又は反応凹部に液体中に滞留することによって、反応凹部内又は反応凹部に配置することができる。DNAボール又は他の核酸超構造を実施することができ、次いで、反応凹部内又は反応凹部に配置することができる。あるいは、DNAボールは、反応凹部においてその場で合成することができる。反応凹部内に固定された物質は、固体、液体、又は気体状態であり得る。

20

【0366】

本明細書で使用するとき、用語「検体」は、相対位置に従って他の点又は領域と区別することができるパターンの点又は領域を意味することを意図する。個々の検体は、特定のタイプの1つ以上の分子を含むことができる。例えば、検体は、特定のシーケンスを有する単一の標的核酸分子を含むことができ、又は検体は、同じシーケンス（及び／又はその相補的シーケンス）を有するいくつかの核酸分子を含むことができる。パターンの異なる検体である異なる分子は、パターン内の検体の場所に従って互いに分化させることができる。例示的な検体としては、基材中のウェル、基材中又は基材上のビーズ（又は他の粒子）、基材からの突出部、基材上の隆起部、基材上のゲル材料のパッド、又は基材内のチャンネルが挙げられる。

30

【0367】

検出、特徴付け、又は識別される様々な標的検体のいずれも、本明細書に記載される装置、システム、又は方法で 사용할 ことができる。例示的な検体としては、限定するものではないが、核酸（例えば、DNA、RNA又はそれらの類似体）、タンパク質、多糖類、細胞、抗体、エピトープ、受容体、リガンド、酵素（例えば、キナーゼ、ホスファターゼ又はポリメラーゼ）、小分子薬物候補、細胞、ウイルス、生物などが挙げられるが、これらに限定されない。

40

【0368】

用語「検体」、「核酸」、「核酸分子」、及び「ポリヌクレオチド」という用語は、本明細書において互換的に使用される。様々な実施態様では、核酸は、特定のタイプの核酸分析のために、本明細書で提供されるようなテンプレート（例えば、核酸テンプレート、又は核酸テンプレートに相補的な核酸相補体）として使用されてもよく、核酸増幅、核酸

50

発現解析、及び／又は核酸配列決定、又はこれらの好適な組み合わせが挙げられるが、これらに限定されない。特定の実施態様における核酸としては、例えば、3' - 5' ホスホジエステル中のデオキシリボヌクレオチドの直鎖ポリマー、又はデオキシリボ核酸 (DNA)、例えば、一本鎖及び二本鎖 DNA、ゲノム DNA、コピー DNA 若しくは相補的 DNA (cDNA)、組換え DNA、又は任意の形態の合成 DNA 若しくは修飾 DNA が挙げられる。他の実施態様では、核酸としては、例えば、3' - 5' ホスホジエステル中のリボヌクレオチドの直鎖ポリマー、又はリボ核酸 (RNA) などの他の結合、例えば、一本鎖及び二本鎖 RNA、メッセンジャー (mRNA)、コピー RNA 又は相補的 RNA (cRNA)、選択的にスプライシングされた mRNA、リボソーム RNA、小核 RNA (snRNA)、マイクロ RNA (miRNA)、低干渉 RNA (siRNA)、piwi RNA (piRNA)、又は任意の形態の合成 RNA 若しくは修飾 RNA が挙げられる。本発明の組成物及び方法において使用される核酸は、長さが変化してもよく、無傷又は完全長の分子若しくはフラグメント、又はより大きい核酸分子のより小さい部分であってもよい。特定の実施態様では、核酸は、本明細書の他の箇所に記載されるように、1 つ以上の検出可能な標識を有してもよい。

10

20

30

40

50

【0369】

用語「検体」、「クラスター」、「核酸クラスター」、「核酸コロニー」、及び「DNA クラスター」は互換的に使用され、固体支持体に結合された核酸テンプレート及び／又はその相補体の複数のコピーを指す。典型的には、特定の好ましい実施態様では、核酸クラスターは、それらの 5' 末端を介して固体支持体に結合されたテンプレート核酸及び／又はその相補体の複数のコピーを含む。核酸クラスターを構成する核酸鎖のコピーは、一本鎖又は二本鎖形態であってもよい。クラスター内に存在する核酸テンプレートのコピーは、例えば、標識部分の存在に起因して、互いに異なる対応する位置にヌクレオチドを有することができる。対応する位置はまた、異なる化学構造を有するが、ウラシル及びチミンの場合など、類似の Watson-Crick 塩基対形成特性を有するアナログ構造を含むことができる。

【0370】

核酸のコロニーはまた、「核酸クラスター」と称されることもある。核酸コロニーは、本明細書の他の箇所で更に詳細に記載されるように、クラスター増幅又はブリッジ増幅技術によって任意に作成することができる。標的シーケンスの複数の反復は、ローリングサークル増幅手順を使用して作成された混乱剤などの単一の核酸分子中に存在し得る。

【0371】

本発明の核酸クラスターは、使用される条件に応じて、異なる形状、サイズ、及び密度を有することができる。例えば、クラスターは、実質的に円形、多面、ドーナツ形、又はリング形状の形状を有することができる。核酸クラスターの直径は、約 0.2 μm ~ 約 6 μm 、約 0.3 μm ~ 約 4 μm 、約 0.4 μm ~ 約 3 μm 、約 0.5 μm ~ 約 2 μm 、約 0.75 μm ~ 約 1.5 μm 、又は任意の介在直径であるように設計することができる。特定の実施態様において、核酸クラスターの直径は、約 0.5 μm 、約 1 μm 、約 1.5 μm 、約 2 μm 、約 2.5 μm 、約 3 μm 、約 4 μm 、約 5 μm 、又は約 6 μm である。核酸クラスターの直径は、クラスターの産生において実施される増幅サイクルの数、核酸テンプレートの長さ、又はクラスターが形成される表面に付着したプライマーの密度を含むが、これらに限定されない多数のパラメータによって影響され得る。核酸クラスターの密度は、典型的には、0.1 / mm^2 、1 / mm^2 、10 / mm^2 、100 / mm^2 、1,000 / mm^2 、10,000 / mm^2 ~ 100,000 / mm^2 の範囲であるように設計することができる。本発明は、一部では、より高密度の核酸クラスター、例えば、100,000 / mm^2 ~ 1,000,000 / mm^2 、及び 1,000,000 / mm^2 ~ 10,000,000 / mm^2 を更に企図する。

【0372】

本明細書で使用するとき、「検体」は、検体又は視野内の対象領域である。マイクロアレイデバイス又は他の分子分析デバイスに関連して使用される場合、検体は、類似又は同

一の分子によって占有される領域を指す。例えば、検体は、増幅オリゴヌクレオチド、又は同じ又は類似の配列を有するポリヌクレオチド又はポリペプチドの任意の他の群であり得る。他の実施態様では、検体は、試料上の物理的領域を占有する任意の要素又は要素群であり得る。例えば、検体は、ランドのパセル、水の本体などであってもよい。検体が撮像されると、各検体は、一部の領域を有する。したがって、多くの実施態様では、検体は、単に1つのピクセルではない。

【0373】

検体間の距離は、任意の数の方式で説明することができる。いくつかの実施態様では、検体間の距離は、1つの検体の中心から別の検体の中心まで説明することができる。他の実施態様では、距離は、1つの検体の縁部から別の検体の縁部まで、又は各検体の最も外側の識別可能な点間に記載することができる。検体の縁部は、チップ上の理論的若しくは実際の物理的境界、又は検体の境界内のいくつかの点として説明することができる。他の実施態様では、距離は、試料上の固定点、又は試料の画像に関して説明することができる。

10

【0374】

一般に、分析方法に関して、いくつかの実施態様が本明細書に記載される。自動又は半自動化方式で方法を実行するためのシステムも提供されることが理解されるであろう。したがって、本開示は、ニューラルネットワークベースのテンプレート生成及びベースコールシステムを提供し、システムは、プロセッサと、記憶デバイスと、画像分析のためのプログラムと、を含み、プログラムは、本明細書に記載の方法のうちの1つ以上を実行するための命令を含む。したがって、本明細書に記載される方法は、例えば、本明細書に記載されるか又は当該技術分野において既知の構成要素を有するコンピュータ上で実行することができる。

20

【0375】

本明細書に記載される方法及びシステムは、様々な物体のうちのいずれかを分析するのに有用である。特に有用な物体は、固体担体又は付着した検体を有する固相表面である。本明細書に記載される方法及びシステムは、 x y 平面における検体の繰り返しパターンを有する物体とともに使用される場合、利点を提供する。一例は、細胞、ウイルス、核酸、タンパク質、抗体、炭水化物、小分子（薬物候補など）、生物学的活性分子、又は他の対象検体の集合を有するマイクロアレイである。

30

【0376】

核酸及びポリペプチドなどの生物学的分子を有する検体を有するアレイの用途の数が増えてきた。このようなマイクロアレイは、典型的には、デオキシリボ核酸（DNA）又はリボ核酸（RNA）プローブが挙げられる。これらは、ヒト及び他の生物に存在するヌクレオチドシーケンスに特異的である。特定の用途では、例えば、個々のDNA又はRNAプローブをアレイの個々の検体に取り付けることができる。既知のヒト又は生物からのものなどの試験試料は、標的核酸（例えば、遺伝子フラグメント、mRNA、又はアンプリコン）が配列中のそれぞれの検体で相補的プローブにハイブリダイズするように、アレイに曝露することができる。プローブは、標的的特異的プロセス（例えば、標的核酸上に存在する標識に起因して、又は検体においてハイブリダイズした形態で存在するプローブ又は標的の酵素標識に起因して）標識することができる。次いで、検体の上の特定の光の周波数を走査して、どの標的核酸が試料中に存在するかを識別することによって検査することができる。

40

【0377】

生物学的マイクロアレイは、遺伝子配列決定及び類似の用途に使用され得る。一般に、遺伝子配列決定は、DNA又はRNAのフラグメントなどの標的核酸の長さのヌクレオチドの順序を決定することを含む。比較的短いシーケンスは、典型的には、各検体において配列決定され、得られたシーケンス情報は、様々な生物情報科学法において使用されて、フラグメントが由来する多くの広範な長さの遺伝物質のシーケンスを確実に決定するために、様々な生物情報科学法において使用されてもよい。特徴的フラグメントの自動化され

50

たコンピュータベースのアルゴリズムが開発されており、ゲノムマッピング、遺伝子の識別、及びそれらの機能などにおいて、より最近使用されてきた。マイクロアレイは、多数の変異体が存在するため、ゲノム含有量を特徴付けるのに特に有用であり、このことは、個々のプローブ及び標的に対して多くの実験を実施する代替物であるため、ゲノム含有量を特徴付けるのに特に有用である。マイクロアレイは、実用的な様式でこのような調査を実施するための理想的なフォーマットである。

【0378】

当該技術分野において既知の様々な検体アレイ（「マイクロアレイ」とも称される）のいずれも、本明細書に記載される方法又はシステムにおいて使用することができる。典型的なアレイは、各々が個々のプローブ又はプローブの集団を有する検体を含む。後者の場合、各検体におけるプローブの集団は、典型的には、単一種のプローブを有する均質である。例えば、核酸配列の場合、各検体は、各々、共通の配列を有する複数の核酸分子を有することができる。しかしながら、いくつかの実施態様では、アレイの各検体における集団は、不均質であり得る。同様に、タンパク質配列は、単一のタンパク質又はタンパク質の集団を有する検体を有することができ、典型的には、同じアミノ酸配列を有するが、必ずしもそうではない。プローブは、例えば、プローブを表面に共有結合することによって、又はプローブと表面との非共有相互作用を介して、アレイの表面に取り付けることができる。いくつかの実施態様では、核酸分子などのプローブは、ゲル層を介して表面に付着させることができ、例えば、米国特許出願第13／784,368号及び米国特許出願公開第2011／0059865（A1）号に記載されており、これらの各々は、参照により本明細書に組み込まれる。

【0379】

例示的なアレイとしては、限定するものではないが、Illumina, Inc (San Diego, Calif.) から入手可能なBeadChipアレイ又は他のもの、例えば、プローブが、表面上に存在するビーズ（例えば、表面上のウェル内のビーズ）に取り付けられる、以下、米国特許第6,266,459号、同第6,355,431号、同第6,770,441号、同第6,859,570号、若しくは米国特許第7,622,294号、又は国際公開第00／63437号に、記載されたものなどの他のものが挙げられ、これらの各々は、参照により本明細書に組み込まれる。使用することができる市販のマイクロアレイの更なる例としては、例えば、VLSIPS（商標）（Very Large Scale Immobilized Polymer Synthesis）技術と称されることがある技術に従って合成されたAffymetrix（登録商標）GeneChip（登録商標）マイクロアレイ又は他のマイクロアレイが挙げられる。スポットティングされたマイクロアレイはまた、本開示のいくつかの実施態様による方法又はシステムにおいて使用することができる。例示的なスポットティングされたマイクロアレイは、Amersham Biosciencesから入手可能なCodeLink（商標）Arrayである。有用な別のマイクロアレイは、Agilent Technologiesから入手可能なSurePrint（商標）Technologyなどのインクジェット印刷法を使用して製造されるものである。

【0380】

他の有用な配列としては、核酸配列決定用途で使用されるものが挙げられる。例えば、ゲノムフラグメント（多くの場合、クラスターと称される）のアンプリコンを有するアレイは、Bentley et al., Nature 456:53-59 (2008)、国際公開第04／018497号、国際公開第91／06678号、同第07／123744号、米国特許第7,329,492号、同第7,211,414号、同第7,315,019号、米国特許7,405,281号若しくは米国特許7,057,026号、又は米国特許出願公開第2008／0108082（A1）号に記載されており、これらの各々は、参照により本明細書に組み込まれる。核酸配列決定に有用な別のタイプの配列は、エマルションPCR技術から生成される粒子の配列である。実施例は、Dressman et al., Proc. Natl. Acad. Sci. USA 100:881

7-8822(2003)、国際公開第05/010145号、米国特許出願公開第2005/0130173号又は米国特許出願公開第2005/0064460号に記載されており、これらの各々は、その全体が参照により本明細書に組み込まれる。

【0381】

核酸配列に使用される配列は、多くの場合、核酸検体のランダムな空間パターンを有する。例えば、Illumina Incから入手可能なHiSeq又はMiSeq配列決定プラットフォーム(San Diego, Calif.)は、核酸配列がランダム播種、続いてブリッジ増幅によって形成されるフローセルを利用する。しかしながら、パターン化されたアレイは、核酸配列決定又は他の分析用途にも使用することができる。パターン化されたアレイの例、それらの製造方法及びその使用法は、米国特許出願第13/787,396号、米国特許第13/783,043号、米国特許第13/784,368号、米国特許出願公開第2013/0116153(A1)号、及び米国特許出願公開第2012/0316086(A1)号に記載されており、これらの各々は、参照により本明細書に組み込まれる。このようなパターン化された配列の検体を使用して、単一の核酸テンプレート分子を捕捉して、例えば、ブリッジ増幅を介して、均質なコロニーの後続の形成を行うことができる。このようなパターン化されたアレイは、核酸配列決定用途に特に有用である。

【0382】

アレイ(又は本明細書の方法又はシステムで使用される他の物体)上の検体のサイズは、特定の用途に適するように選択することができる。例えば、いくつかの実施態様では、アレイの検体は、単一の核酸分子のみを収容するサイズを有することができる。このサイズ範囲の複数の検体を有する表面は、単一分子分解能で検出するための分子の配列を構築するのに有用である。このサイズ範囲の検体もまた、核酸分子のコロニーをそれぞれ含む検体を有するアレイでの使用にも有用である。したがって、アレイの検体は各々、約 1 m^2 以下、約 $500\text{ }\mu\text{m}^2$ 以下、約 $100\text{ }\mu\text{m}^2$ 以下、約 $10\text{ }\mu\text{m}^2$ 以下、約 $1\text{ }\mu\text{m}^2$ 以下、約 500 nm^2 以下、又は約 100 nm^2 以下、約 10 nm^2 以下、約 5 nm^2 以下、又は約 1 nm^2 以下の面積を有することができる。代替的に又は追加的に、アレイの検体は、約 1 mm^2 以上、約 $500\text{ }\mu\text{m}^2$ 以上、約 $100\text{ }\mu\text{m}^2$ 以上、約 $10\text{ }\mu\text{m}^2$ 以上、約 $1\text{ }\mu\text{m}^2$ 以上、約 500 nm^2 以上、約 100 nm^2 以上、約 10 nm^2 以上、約 5 nm^2 以上、又は約 1 nm^2 以上である。実際に、検体は、上記に例示したものから選択される上限と下限との間の範囲内の大きさを有することができる。表面の検体のいくつかのサイズ範囲が核酸及び核酸のスケールに関して例示されてきたが、これらのサイズ範囲の検体は、核酸を含まない用途に使用され得ることが理解されるであろう。検体のサイズは、核酸用途に使用されるスケールに必ずしも限定される必要はないことが更に理解されるであろう。

【0383】

検体のアレイなどの複数の検体を有する物体を含む実施例では、検体は、互いの間で分離されている、別個のものとするすることができる。本発明において有用なアレイは、最大で $100\text{ }\mu\text{m}$ 、 $50\text{ }\mu\text{m}$ 、 $10\text{ }\mu\text{m}$ 、 $5\text{ }\mu\text{m}$ 、 $1\text{ }\mu\text{m}$ 、 $0.5\text{ }\mu\text{m}$ 以下の縁部から縁部までの距離によって分離される検体を有することができる。代替的に又は追加的に、アレイは、少なくとも $0.5\text{ }\mu\text{m}$ 、 $1\text{ }\mu\text{m}$ 、 $5\text{ }\mu\text{m}$ 、 $10\text{ }\mu\text{m}$ 、 $50\text{ }\mu\text{m}$ 、 $100\text{ }\mu\text{m}$ 以上の縁部から縁部までの距離によって分離される検体を有することができる。これらの範囲は、検体の平均縁部間隔及び縁部間隔、並びに最小又は最大間隔に適用することができる。

【0384】

いくつかの実施態様では、アレイの検体は、別個である必要はなく、代わりに、隣接する検体は互いに当接することができる。検体が別個であるか否かに関わらず、検体及び/又は検体のピッチの大きさは、アレイが所望の密度を有することができるように変化し得る。例えば、規則的なパターンにおける平均検体ピッチは、最大で $100\text{ }\mu\text{m}$ 、 $50\text{ }\mu\text{m}$ 、 $10\text{ }\mu\text{m}$ 、 $5\text{ }\mu\text{m}$ 、 $1\text{ }\mu\text{m}$ 、 $0.5\text{ }\mu\text{m}$ 以下であり得る。代替的に又は追加的に、規則

的なパターンにおける平均検体ピッチは、少なくとも $0.5\ \mu\text{m}$ 、 $1\ \mu\text{m}$ 、 $5\ \mu\text{m}$ 、 $10\ \mu\text{m}$ 、 $50\ \mu\text{m}$ 、 $100\ \mu\text{m}$ 以上であり得る。これらの範囲は、規則的なパターンの最大ピッチ又は最小ピッチにも適用することができる。例えば、規則的なパターンの最大検体ピッチは、 $100\ \mu\text{m}$ 以下、 $50\ \mu\text{m}$ 以下、 $10\ \mu\text{m}$ 以下、 $5\ \mu\text{m}$ 以下、 $1\ \mu\text{m}$ 以下、 $0.5\ \mu\text{m}$ 以下とすることができ、かつ／又は規則的なパターンにおける最小検体ピッチは、少なくとも $0.5\ \mu\text{m}$ 、 $1\ \mu\text{m}$ 、 $5\ \mu\text{m}$ 、 $10\ \mu\text{m}$ 、 $50\ \mu\text{m}$ 、 $100\ \mu\text{m}$ 以上であり得る。

【0385】

アレイ内の検体の密度は、単位面積あたりに存在する検体の数に関しても理解され得る。例えば、アレイに関する検体の平均密度は、少なくとも約 1×10^3 検体/ mm^2 、 1×10^4 検体/ mm^2 、 1×10^5 検体/ mm^2 、 1×10^6 検体/ mm^2 、 1×10^7 検体/ mm^2 、 1×10^8 検体/ mm^2 、又は 1×10^9 検体/ mm^2 以上であり得る。代替的に又は追加的に、アレイに関する検体の平均密度は、最大で約 1×10^9 検体/ mm^2 、 1×10^8 検体/ mm^2 、 1×10^7 検体/ mm^2 、 1×10^6 検体/ mm^2 、 1×10^5 検体/ mm^2 、 1×10^4 検体/ mm^2 、又は 1×10^3 検体/ mm^2 以下であり得る。

【0386】

上記の範囲は、例えば、検体のアレイの全て又は一部を含む規則的なパターンの全て又は一部に適用することができる。

【0387】

パターン内の検体は、様々な形状のいずれかを有することができる。例えば、アレイの表面上などの二次元平面で観察される場合、検体は、丸みを帯びた、円形、楕円形、矩形、正方形、対称、非対称、三角形、多角形などに見える場合がある。検体は、例えば六角形又は直線パターンを含む規則的な繰り返しパターンで配置することができる。パターンは、所望のレベルのパッキングを達成するように選択され得る。例えば、円形検体は、六角形の配置で最適に充填される。当然のことながら、他のパッキング構成もまた、円形検体のために使用することができ、逆もまた同様である。

【0388】

パターンは、パターンの最小幾何学的単位を形成するサブセット内に存在する検体の数に関して特徴付けることができる。サブセットは、例えば、少なくとも約 2、3、4、5、6、10 個以上の検体を含み得る。検体のサイズ及び密度に応じて、幾何学的単位は、 $1\ \text{mm}^2$ 、 $500\ \mu\text{m}^2$ 、 $100\ \mu\text{m}^2$ 、 $50\ \mu\text{m}^2$ 、 $10\ \mu\text{m}^2$ 、 $1\ \mu\text{m}^2$ 、 $500\ \text{nm}^2$ 、 $100\ \text{nm}^2$ 、 $50\ \text{nm}^2$ 、 $10\ \text{nm}^2$ 以下の面積を占めることができる。代替的に又は追加的に、幾何学的単位は、 $10\ \text{nm}^2$ 、 $50\ \text{nm}^2$ 、 $100\ \text{nm}^2$ 、 $500\ \text{nm}^2$ 、 $1\ \mu\text{m}^2$ 、 $10\ \mu\text{m}^2$ 、 $50\ \mu\text{m}^2$ 、 $100\ \mu\text{m}^2$ 、 $500\ \mu\text{m}^2$ 、 $1\ \text{mm}^2$ 以上の面積を占めることができる。形状、サイズ、ピッチなどの幾何学的単位における検体の特性は、アレイ又はパターンの検体に関して、より一般的に本明細書に記載されるものから選択することができる。

【0389】

検体の規則的なパターンを有するアレイは、検体の相対的な場所に対して順序付けられるが、各検体の 1 つ以上の他の特性に対してランダムであってもよい。例えば、核酸シーケンスの場合、核酸検体は、それらの相対的な位置に関して規則的であるが、任意の特定の検体に存在する核酸種に関するシーケンスの知識に関してランダムであってもよい。より具体的な例として、テンプレート核酸を有する検体の反復パターンを播種し、各検体でテンプレートを増幅して、検体においてテンプレートのコピーを形成することによって形成される核酸シーケンス（例えば、クラスター増幅又はブリッジ増幅を介して、核酸検体の規則的なパターンを有するが、シーケンスにわたる核酸のシーケンスの分布に関してランダムであろう。したがって、アレイ上の核酸材料の存在の検出は、検体の繰り返しパターンをもたらすことができるのに対し、シーケンス特異的検出は、アレイにわたる信号の非反復分布をもたらすことができる。

【0390】

本明細書におけるパターン、順序、ランダム性などの説明は、アレイ上の検体などの物体上の検体にも関するだけでなく、画像中の検体にも関連することが理解されるであろう。したがって、パターン、順序、ランダム性などは、限定するものではないが、グラフィカルユーザーインターフェース又は他の出力デバイスなどのコンピュータ可読媒体又はコンピュータ構成要素を含むがこれらに限定されない、画像データを記憶、操作、又は通信するために使用される様々なフォーマットのうちのいずれかに存在することができる。

【0391】

本明細書で使用する時、用語「画像」は、物体の全て又は一部の表現を意味することを意図する。表現は、光学的に検出された再現であり得る。例えば、蛍光、発光、散乱、又は吸収信号から画像を得ることができる。画像内に存在する物体の部分は、物体の表面又は他のx y面であり得る。典型的には、画像は二次元表現であるが、場合によっては、画像内の情報は、3つ以上の次元から導出することができる。画像は、光学的に検出された信号を含む必要はない。非光信号を代わりに存在させることができる。画像は、本明細書の他の箇所に記載されるもののうちの1つ以上などの、コンピュータ可読フォーマット又は媒体に提供することができる。

【0392】

本明細書で使用する時、「画像」は、試料又は他の物体の少なくとも一部分の再現又は表現を指す。いくつかの実施態様では、再現は、例えばカメラ又は他の光学検出器によって生成される光再現である。再現は、非光学的再現、例えば、ナノ細孔検体のアレイから得られる電気信号の表現、又はイオン感应性CMOS検出器から得られた電気信号の表現であり得る。特定の実施態様では、非光学的再現性は、本明細書に記載される方法又は装置から除外され得る。画像は、例えば、100 μm 、50 μm 、10 μm 、5 μm 、1 μm 、又は0.5 μm 未満離れたものを含む、様々な間隔のいずれかで存在する検体の検体を区別することができる解像度を有することができる。

【0393】

本明細書で使用する時、「取得」、「取得」、及び同様の用語は、画像ファイルを取得するプロセスの任意の部分の部分を指す。いくつかの実施態様では、データ取得は、標本の画像を生成することと、標本内の信号を探すことと、信号の画像を探すか又は生成するように検出デバイスに指示することと、画像ファイルの更なる分析又は変換のための命令、及び画像ファイルの任意の数の変換又は操作のための命令を与えることと、を含むことができる。

【0394】

本明細書で使用する時、用語「テンプレート」は、信号又は検体間の場所又は関係の表現を指す。したがって、いくつかの実施態様では、テンプレートは、検体中の検体に対応する信号の表現を有する物理的グリッドである。いくつかの実施態様では、テンプレートは、チャート、テーブル、テキストファイル、又は検体に対応する場所を示す他のコンピュータファイルであり得る。本明細書に提示される実施態様では、異なる基準点で捕捉された試料の画像のセットにわたって検体の場所を追跡するためにテンプレートが生成される。例えば、テンプレートは、x、y座標のセット、又は別の検体に対する1つの検体の方向及び／若しくは距離を記述する値のセットであり得る。

【0395】

本明細書で使用する時、用語「標本」は、画像が取り込まれる物体又は物体の領域を指すことができる。例えば、画像が土壌の表面から撮影される実施例では、ランドのセルは、標本であり得る。生体分子の分析がフローセル内で行われる他の実施態様では、フローセルは、任意の数のサブディビジョンに分割されてもよく、これらの各々は検体であってもよい。例えば、フローセルは、様々な流路又はレーンに分割されてもよく、各レーンは、画像化される2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、140、160、180、200、400、600、800、1000個以上の別個の領域に更に分割され得る。フローセル

の一例は8つのレーンを有し、各レーンは120個の標本又はタイルに分割されている。別の実施態様では、試料は、複数のタイル、又は更にはフローセル全体で作製されてもよい。したがって、各検体の画像は、撮像されるより大きい表面の領域を表すことができる。

【0396】

本明細書に記載される範囲及び連続数リストへの言及は、列挙された数だけではなく、列挙された数字間の全ての実数を含むことが理解されるであろう。

【0397】

本明細書で使用する時、「参照点」は、画像間の任意の時間的又は物理的区別を指す。好ましい別の実施態様では、基準点は時間点である。より好ましい実施態様では、参照点は、配列決定反応中の時点又はサイクルである。しかしながら、用語「基準点」は、画像を区別又は分離することができる、角度、回転、時間、又は他の態様などの画像を区別又は分離する他の態様を含むことができる。

【0398】

本明細書で使用する時、「画像のサブセット」は、セット内の画像のグループを指す。例えば、サブセットは、画像のセットから選択される1、2、3、4、6、8、10、12、14、16、18、20、30、40、50、60又は任意の数の画像を含んでもよい。特定の別の実施態様では、サブセットは、1、2、3、4、6、8、10、12、14、16、18、20、30、40、50、60以下、又は画像のセットから選択される任意の数の画像を含んでもよい。好ましい別の実施態様では、画像は、各サイクルに相関する4つの画像を有する1つ以上の配列決定サイクルから得られる。したがって、例えば、サブセットは、4サイクルにわたって取得された16画像のグループであり得る。

【0399】

塩基は、ヌクレオチド塩基又はヌクレオチド、(アデニン)、C(シトシン)、T(チミン)、又はG(グアニン)を指す。本出願は、「塩基」及び「ヌクレオチド」を互換的に使用する。

【0400】

用語「染色体」は、DNA及びタンパク質成分(特にヒストン)を含むクロマチンストランドに由来する、生きている細胞の本発明の有効性を有する遺伝子キャリアを指す。従来の国際的に認識されている個々のヒトゲノム染色体番号付けシステムが本明細書で使用される。

【0401】

「部位」という用語は、参照ゲノム上の固有の位置(例えば、染色体ID、染色体位置及び配向)を指す。いくつかの実施態様では、部位は、残基、シーケンスタグ、又はシーケンス上のセグメントの位置であってもよい。用語「遺伝子座」は、参照染色体上の核酸シーケンス又は多型の特定の位置を指すために使用されてもよい。

【0402】

本明細書における用語「試料」は、典型的には、配列決定及び／又はフェーズドされる核酸を含有する生物液、細胞、組織、器官、又は生物に由来する試料、又は配列決定及び／又はフェーズドされる核酸シーケンスを少なくとも1つ含有する核酸の混合物に由来する試料を指す。このような試料としては、痰／口腔流体、羊水、血液、血液画分、細針生検試料(例えば、外科生検、針生検など)、尿、腹膜流体、胸膜流体、組織外植片、臓器培養物、及びこれらの任意の他の組織若しくは細胞調製物、又はこれらの画分若しくは誘導体が挙げられるが、これらに限定されない。試料は、多くの場合、ヒト被験者(例えば、患者)から採取されるが、試料は、イヌ、ネコ、ウマ、ヤギ、ヒツジ、ウシ、ブタなどを含むがこれらに限定されない、染色体を有する任意の生物から採取することができる。試料は、生物学的源から得られるように、又は試料の特性を修正する前処理後に、直接使用することができる。例えば、このような前処理は、血漿を血液から調製すること、粘性流体を希釈することなどを含んでもよい。前処理の方法には、濾過、沈殿、希釈、蒸留、混合、遠心分離、凍結、凍結乾燥、濃縮、増幅、核酸フラグメント化、干渉成分の不活性

化、試薬の添加、溶解などを含んでもよいが、これらに限定されない。

【0403】

用語「配列」は、互いに結合されたヌクレオチドの鎖を含むか、又はそれを表す。ヌクレオチドは、DNA又はRNAに基づくことができる。1つのシーケンスは、複数のサブシーケンスを含んでもよいことを理解されたい。例えば、単一のシーケンス（例えば、PCRアンプリコン）は、350ヌクレオチドを有してもよい。試料読み取りは、これらの350ヌクレオチド内に複数のサブシーケンスを含んでもよい。例えば、試料読み取りは、例えば、20～50ヌクレオチドを有する第1及び第2のフランキングサブシーケンスを含んでもよい。第1及び第2の隣接するサブシーケンスは、対応するサブシーケンス（例えば、40～100ヌクレオチド）を有する反復セグメントの両側に位置してもよい。隣接するサブシーケンスの各々は、プライマーサブシーケンス（例えば、10～30ヌクレオチド）を含んでもよい（又はその一部分を含み得る）。読み取りを容易にするために、用語「サブシーケンス」は「シーケンス」と称されるが、2つのシーケンスは、共通のストランド上で互いに別個である必要はないことが理解される。本明細書に記載の様々なシーケンスを区別するために、シーケンスは、異なる標識（例えば、標的シーケンス、プライマーシーケンス、隣接シーケンス、参照シーケンスなど）を与えられてもよい。「対立遺伝子」などの他の用語は、同様の物体を区別するために異なるラベルを与えられてもよい。アプリケーションは、「読み取り」及び「シーケンス読み取り」を互換的に使用する。

10

【0404】

用語「ペアエンド配列決定（paired end sequencing）」は、標的フラグメントの両端をシーケンスする配列決定方法を指す。ペアエンド配列決定は、ゲノム再構成及び反復セグメントの検出、並びに遺伝子融合及び新規転写物の検出を容易にし得る。ペアエンド配列決定の方法は、国際公開第07010252号、国際出願第GB2007/003798号、及び米国特許出願公開第2009/0088327号に記載されており、これらは各々参照により本明細書に組み込まれる。一例では、一連の動作は、以下のように実行され得る。（a）核酸のクラスターを生成し、（b）核酸を線形化し、（c）上記のように、第1の配列決定プライマーをハイブリダイズし、伸長、走査及びデブロッキングのサイクルを繰り返し実行し、（d）相補的なコピーを合成することによって、フローセル表面上の標的核酸を「反転」させ、（e）再合成された鎖を線形化し、（f）上記のよう

20

30

【0405】

用語「参照ゲノム」又は「参照シーケンス」は、対象からの識別されたシーケンスを参照するために使用され得る任意の生物の部分的又は完全ないずれかの特定の既知のゲノムシーケンスを指す。例えば、ヒト被験者に使用される参照ゲノム、並びに多くのその他の生物が、National Center for Biotechnology Information (ncbi.nlm.nih.gov)で見出される。「ゲノム」とは、核酸シーケンスで発現される、生物又はウイルスの完全な遺伝子情報を意味する。ゲノムは、DNAの遺伝子及び非コードシーケンスの両方を含む。参照シーケンスは、それに位置合わせされたリードよりも大きくてもよい。例えば、それは、少なくとも約100倍大きい、又は少なくとも約1000倍大きい、又は少なくとも約10,000倍大きい、又は少なくとも約105倍大きい、又は少なくとも約106倍大きい、又は少なくとも約107倍大きい場合がある。一例では、参照ゲノムシーケンスは、完全長ヒトゲノムのものである。別の例では、参照ゲノムシーケンスは、染色体13などの特定のヒト染色体に限定される。いくつかの実施態様では、参照染色体は、ヒトゲノムバージョンhg19からの染色体シーケンスである。このようなシーケンスは、染色体参照シーケンスと称される場合があるが、用語参照ゲノムは、そのようなシーケンスを網羅することを意図している。参照シーケンスの他の例としては、その他の種のゲノム、並びに任意の種の染色体

40

50

、部分染色体領域（ストランドなど）などが挙げられる。様々な実施態様では、参照ゲノムは、複数の個体に由来するコンセンサスシーケンス又は他の組み合わせである。しかしながら、特定の用途では、参照シーケンスは、特定の個体から採取されてもよい。他の実施態様では、「ゲノム」はまた、ゲノム配列の特定の記憶形式及び表現を使用するいわゆる「グラフゲノム」も被覆する。一実施態様では、グラフゲノムは、線形ファイルにデータを記憶する。別の実施態様では、グラフゲノムは、代替的な配列決定（例えば、小さい差異を有する染色体の異なるコピー）がグラフ内の異なる経路として記憶されている表現を指す。グラフゲノムの実装に関する追加情報は、<https://www.biorxiv.org/content/biorxiv/early/2018/03/20/194530.full.pdf> にて参照することができ、その内容は、その全体が参照により本明細書に組み込まれる。

10

【0406】

用語「読み取り」は、ヌクレオチド試料又は参照のフラグメントを記述するシーケンスデータの集合を指す。用語「読み取り」は、試料読み取り及び／又は参照読み取りを指し得る。典型的には、必ずしもそうではないが、読み取りは、試料又は参照における連続的な塩基対の短いシーケンスを表す。読み取りは、試料又は参照フラグメントのベース対シーケンス（A C T G）によって記号的に表されてもよい。読み取りが参照シーケンスと一致するか、又は他の基準を満たすかを判定するために、メモリデバイスに記憶され、適切に処理されてもよい。リードは、シーケンシング装置から直接、又は試料に関する記憶されたシーケンス情報から間接的に得られてもよい。場合によっては、例えば、染色体又はゲノム領域又は遺伝子に位置合わせされ、特異的に割り当てられ得る、より大きいシーケンス又は領域を識別するために使用することができる十分な長さ（例えば、少なくとも約25bp）のDNAシーケンスである。

20

【0407】

次世代配列決定法としては、例えば、合成技術（Illumina）、ピロ配列決定（454）、イオン半導体技術（Ion Torrent sequencing）、単分子リアルタイム配列決定及びライゲーションによる配列決定（SOLiD配列決定）が挙げられる。配列決定法に応じて、各読み取りの長さは、約30bp～10,000bpを超えて変化し得る。例えば、SOLiDシーケンサを用いたDNA配列決定法は、約50bpの核酸リードを生成する。別の例では、Ion Torrent Sequencingは、最大400bpの核酸リードを生成し、454のピロ配列は、約700bpの核酸リードを生成する。更に別の例では、単分子リアルタイム配列決定法は、10,000bp～15,000bpのリードを生成し得る。したがって、特定の実施態様では、核酸シーケンスのリードは、30～100bp、50～200bp、又は50～400bpの長さを有する。

30

【0408】

用語「試料読み取り」、「試料シーケンス」又は「試料フラグメント」は、試料から対象とするゲノムシーケンスに関するシーケンスデータを指す。例えば、試料読み取りは、順方向及び逆方向プライマーシーケンスを有するPCRアンプリコンからのシーケンスデータを含む。シーケンスデータは、任意の選択シーケンス方法論から得ることができる。試料の読み取りは、例えば、配列単位合成（sequencing-by-synthesis、SBS）反応、配列決定・ライゲーション反応、又は反復要素の長さ及び／若しくは同一性を決定することが望ましい任意の他の好適な配列決定方法論であり得る。試料読み取りは、複数の試料読み取りに由来するコンセンサス（例えば、平均又は加重）シーケンスであり得る。特定の実施態様では、参照シーケンスを提供することは、PCRアンプリコンのプライマーシーケンスに基づいて目的の遺伝子座を識別することを含む。

40

【0409】

用語「生フラグメント」は、試料読み取り又は試料フラグメント内の対象とする指定位置又は二次位置に少なくとも部分的に重なり合う、対象とするゲノムシーケンスの一部のシーケンスデータを指す。生フラグメントの非限定的な例としては、二重ステッチされた

50

フラグメント、単純なステッチされたフラグメント、及び単純な非ステッチのフラグメントが挙げられる。用語「生」は、生フラグメントが試料読み取りにおけるシーケンスデータとのいくつかの関係を有するシーケンスデータを含むことを示すために使用され、生フラグメントが、試料読み取りにおける潜在の変異体に対応し、かつそれを認証又は確認する支持変異体を示すかどうかに関わらず、使用される。用語「生フラグメント」は、フラグメントが、試料読み取りにおける変異型コールを検証する支持変異体を必ずしも含むことを示すものではない。例えば、試料読み取りが、第1の変異体を呈するために、変異型コールアプリケーションによって判定されるとき、この変異型コールアプリケーションは、1つ以上の生フラグメントが、試料読み取りにおける変異体を考慮して、そうでなければ発生することが予想され得る、対応するタイプの「支持」変異体を欠くと判定することができる。

10

【0410】

用語「マッピング」、「整列された」、「整列している」、又は「整列する」という用語は、読み取り又はタグを参照シーケンスと比較し、それによって、参照シーケンスが読み取りシーケンスを含むかどうかを判定するプロセスを指す。参照シーケンスが読み取られた場合、読み取りは参照シーケンスにマップされてもよく、又は特定の別の実施態様では、参照シーケンス内の特定の位置にマッピングされてもよい。いくつかの場合において、整列は、リードが特定の参照シーケンスのメンバーであるか否か（すなわち、リードが参照シーケンス中に存在するか又は存在していないか）を単に伝える。例えば、ヒト染色体13についての参照シーケンスに対するリードの位置合わせは、染色体13の参照シーケンス中にリードが存在するかどうかを伝える。本情報を提供するツールは、セットメンバーシップテスタ (set membership tester) と呼ばれる場合がある。場合によっては、位置合わせは、読み取り又はタグマップがある参照シーケンス内の位置を更に示す。例えば、参照シーケンスがヒトゲノムシーケンス全体である場合、位置合わせは、染色体13上にリードが存在することを示してもよく、更に、リードが染色体13の特定の鎖及び／又は部位にあることを更に示してもよい。

20

【0411】

用語「インデル」は、生物のDNA中の塩基の挿入及び／又は欠失を指す。マイクロインデルは、1～50ヌクレオチドの正味変化をもたらすインデルを表す。インデルの長さが3の倍数でない限り、ゲノムの領域をコードする際に、フレームシフト変異が生じる。インデルは、点突然変異と対比することができる。インデル挿入物は、配列からヌクレオチドを欠失させるが、点変異は、DNA中の全体的な数を変えことなくヌクレオチドのうちの1つを置き換える置換の形態である。インデルはまた、隣接するヌクレオチドにおける置換として定義され得る Tandem Base 変異 (Tandem Base Mutation、TBM) と対比することもできる（主に2つの隣接するヌクレオチドで置換されるが、隣接する3つのヌクレオチドでの置換が観察された）。

30

【0412】

用語「変異体」は、核酸参照とは異なる核酸シーケンスを指す。典型的な核酸シーケンス変異体としては、限定するものではないが、単一ヌクレオチド多型 (Single Nucleotide Polymorphism、SNP)、短い欠失及び挿入多型 (Indel)、コピー数変動 (Copy Number Variation、CNV)、マイクロ衛星マーカー、又は短いタンDEM反復及び構造変異が挙げられる。体細胞変異体コールは、DNA試料中に低頻度で存在する変異体を識別するための努力である。体細胞変異体のコーリングは、癌治療の文脈において対象となる。癌は、DNAにおける変異の蓄積によって引き起こされる。腫瘍由来のDNA試料は、一般的に不均質であり、いくつかの正常細胞、癌進行の早期段階（より少ない突然変異を伴う）、及び一部の後期細胞（より多くの変異を有する）を含む。この不均一性のため、腫瘍を配列決定するとき（例えば、FFPE試料から）、体細胞変異は、多くの場合、低頻度で現れる。例えば、SNVは、所与の塩基を被覆するリードの10%のみに見られ得る。変異体分類子によって体細胞又は生殖細胞系として分類される変異体は、本明細書では「試験中の変異体」とも称される。

40

50

【0413】

用語「ノイズ」は、配列決定プロセス及び／又は変異型コールアプリケーションにおける1つ以上の誤差から生じる、誤りのある変異型コールを指す。

【0414】

用語「変異体頻度」は、集団内の特定の遺伝子座における対立遺伝子（遺伝子の変異体）の相対頻度を表し、分画又は割合として表される。例えば、分画又は割合は、その対立遺伝子を保有する集団中の全ての染色体の割合であってもよい。一例として、試料変異体頻度は、個体から対象となるゲノムシーケンスについて得られたリード及び／又は試料の数に対応する「集団」にわたって、対象とするゲノムシーケンスに沿った特定の遺伝子座／位置における対立遺伝子／変異体の相対頻度を表す。別の例として、ベースライン変異体頻度は、1つ以上のベースラインゲノムシーケンスに沿った特定の遺伝子座／位置における対立遺伝子／変異体の相対頻度を表し、ここで、「集団」は、正常な個人の集団から1つ以上のベースラインゲノムシーケンスについて得られたリード及び／又は試料の数に対応する。

10

【0415】

用語「変異型対立遺伝子頻度（Variant Allele Frequency、VAF）」は、変異体を標的位置での全体的な被覆率で割った、配列決定されたリードの割合を指す。VAFは、変異体を保有する配列決定されたリードの割合の尺度である。

【0416】

用語「位置」、「指定位置」、及び「遺伝子座」は、ヌクレオチドシーケンス内の1つ以上のヌクレオチドの位置又は座標を指す。用語「位置」、「指定位置」、及び「遺伝子座」はまた、ヌクレオチドのシーケンスにおける1つ以上の塩基対の位置又は座標を指す。

20

【0417】

用語「ハプロタイプ」は、互いに遺伝する染色体上の隣接する部位における対立遺伝子の組み合わせを指す。ハプロタイプは、存在する場合、座の所与のセットの間に生じた組換えイベントの数に応じて、1つの遺伝子座、いくつかの座、又は染色体全体であってもよい。

【0418】

本明細書における「閾値」という用語は、試料、核酸、又はその一部（例えば、読み取り）を特徴付けるためのカットオフとして使用される数値又は数値を指す。閾値は、経験的分析に基づいて変化してもよい。閾値は、そのような値を生じさせる源が特定の様式で分類されるべきかどうかを判定するために、測定値又は計算値と比較することができる。閾値は、経験的又は分析的に識別することができる。閾値の選択は、ユーザが分類を行う必要があることを望む信頼度に依存する。閾値は、特定の目的（例えば、感度及び選択性のバランスのために）選択されてもよい。本明細書で使用する時、用語「閾値」は、分析の過程が変化し得る点、及び／又はアクションがトリガされ得る点を示す。閾値は、所定の数である必要はない。その代わりに、閾値は、例えば、複数の因子に基づく関数であってもよい。閾値は、状況に適応し得る。更に、閾値は、上限、下限、又は限界間の範囲を示し得る。

30

40

【0419】

いくつかの実施態様では、配列決定データに基づく指標又はスコアは、閾値と比較され得る。本明細書で使用する時、用語「メトリック」又は「スコア」は、配列決定データから決定された値又は結果を含んでもよく、又は配列決定データから決定された値又は結果に基づく関数を含んでもよい。閾値と同様に、指標又はスコアは、状況に適応し得る。例えば、指標又はスコアは、正規化された値であってもよい。スコア又はメトリックの一例として、1つ以上の実施態様は、データを分析する際にカウントスコアを使用することができる。計数スコアは、試料読み取りの数に基づいてもよい。試料読み取りは、試料読み取りが少なくとも1つの共通の特性又は品質を有するように、1つ以上のフィルタリング段階を経ていてもよい。例えば、計数スコアを決定するために使用される試料読み取り

50

の各々は、参照シーケンスと整列されていてもよく、又は潜在的な対立遺伝子として割り当てられてもよい。共通の特性を有する試料読み取りの数をカウントして、リードカウントを決定することができる。カウントスコアは、読み取りカウントに基づいてもよい。いくつかの実施態様では、計数スコアは、リードカウントと等しい値であってもよい。他の実施例では、計数スコアは、リードカウント及び他の情報に基づいてもよい。例えば、計数スコアは、遺伝子座の特定の対立遺伝子のリードカウント、及び遺伝子座の合計リード数に基づいてもよい。いくつかの実施態様では、計数スコアは、遺伝子座のリードカウント及び以前に得られたデータに基づいてもよい。いくつかの実施態様では、計数スコアは、所定の値間の正規化スコアであってもよい。計数スコアはまた、試料の他の遺伝子座からのリードカウントの関数、又は対象となる試料と同時に動作された他の試料からのリードカウントの関数であってもよい。例えば、計数スコアは、特定の対立遺伝子の読み取りカウント及び試料中の他の遺伝子座の読み取りカウント、及び／又は他の試料からのリードカウントの関数であってもよい。一例として、他の遺伝子座からのリードカウント及び／又は他の試料からのリードカウントを使用して、特定の対立遺伝子についての計数スコアを正規化してもよい。

10

【0420】

用語「被覆率」又は「フラグメント被覆率」は、配列の同じフラグメントに対する多数の試料読み取りの計数又は他の尺度を指す。リードカウントは、対応するフラグメントをカバーするリードの数のカウントを表し得る。代替的に、被覆率は、履歴知識、試料の知識、遺伝子座の知識などに基づく指定された因子を、リードカウントに掛けることによって決定されてもよい。

20

【0421】

用語「読み取り深さ」（従来、「×」が続く数）は、標的位置における重複アライメントを伴う配列決定されたリードの数を指す。これは、多くの場合、間隔のセット（エキソン、遺伝子、又はパネルなど）のカットオフを超える平均又は百分率として表現される。例えば、臨床報告は、パネル平均被覆率が、標的化されたベースカバー>100×の98%を有する1,105×であると言うことができる。

【0422】

用語「ベースコール品質スコア」又は「Qスコア」は、単一の配列決定された塩基が正しい確率に反比例する0-50からの範囲のPHREDスケールされた確率を指す。例えば、20のQを有するTベースコールは、99.99%の確率で正しいと考えられる。Q<20での任意のベースコールは、低品質であると思なされるべきであり、変異体を支持する配列決定されたリードのかなりの割合が低い場合に識別される任意の変異体は、潜在的に偽陽性であると思なされるべきである。

30

【0423】

用語「変異体リード」又は「変異体リード番号」は、変異体の存在を支持する配列決定されたリードの数を指す。

【0424】

「撚れ性 (strandedness)」（又はDNA撚れ性）に関して、DNA中の遺伝的メッセージは、文字A、G、C、及びTの文字、例えば、5'-AGGACAA-3'として表すことができる。多くの場合、シーケンスは、本明細書に示される方向、すなわち、5'端を左に、3'端を右に書き込む。DNAは、(特定のウイルスのように)一本鎖分子として生じる場合があるが、通常、二本鎖単位としてDNAを見つける。これは、2つの抗平行ストランドを有する二重螺旋構造を有する。この場合、「逆平行」という語は、2つのストランドが平行に動作するが、反対の極性を有することを意味する。二本鎖DNAは、塩基とペアリングによって一緒に保持され、ペアリングは、アデニン(A)対がチミン(T)及びシトシン(C)対とグアニン(G)との対となるように、常に保持される。このペアリングは相補性と称され、一本のDNA鎖は、他方の相補体であると言われる。したがって、二本鎖DNAは、5'-AGGACAA-3'及び3'-TCCGTGT-5'と同様に、2つのストリングとして表され得る。2つのストランドは、反対の極性を有すること

40

50

に留意されたい。したがって、2つのDNA鎖のストランド性は、基準ストランド及びその補体、順方向及び逆方向ストランド、トップ及びボトムストランド、センス及びアンチセンスストランド、又はワトソン及びクリックストランドと称され得る。

【0425】

リードアライメント（リードマッピングとも呼ばれる）は、ゲノム中のシーケンスが由来する場合に、参照するプロセスである。アライメントが行われると、所与の読み取りの「マッピング品質」又は「マッピング品質スコア（mapping quality score、MAPQ）」は、ゲノム上のその位置が正しい確率を定量化する。マッピング品質は、位相スケールで符号化され、Pはアライメントが正しくない確率である。確率は $P = 10^{(-MAPQ/10)}$ のように計算され、式中MAPQはマッピング品質である。例えば、-4の電力に対する $40 = 10$ のマッピング品質は、読み取りが不正確に位置合わせされた0.01%の機会が存在することを意味する。したがって、マッピング品質は、読み取りの基本品質、参照ゲノムの複雑性、及びペアエンド情報などのいくつかの位置合わせ因子と関連付けられる。最初に、読み取りの基本品質が低い場合、観察されたシーケンスが誤っている可能性があり、したがってそのアライメントが誤っていることを意味する。第2に、マッピング能力はゲノムの複雑さを指す。反復領域は、これらの領域に含まれるマップ及びリードをマッピングすることがより困難であり、通常、マッピング品質が低くなる。この文脈では、MAPQは、リードが一意的に整列されておらず、それらの実際の原点を決定することができないという事実を反映する。第3に、パリッドエンド配列決定データの場合、コンコダント対は、より良好に整列される可能性が高い。マッピング品質が高いほど、アライメントがより良好である。良好なマッピング品質と整合された読み取りは、通常、読み出しシーケンスが良好であり、高いマッピング可能領域内ではわずかな不一致と位置合わせされたことを意味する。MAPQ値は、アライメント結果の品質管理として使用することができる。20よりも高いMAPQと位置合わせされたリードの割合は、通常、下流分析のためである。

10

20

【0426】

本明細書で使用するとき、「信号」は、例えば画像内の発光、好ましくは発光などの検出可能なイベントを指す。したがって、好ましい別の実施態様では、信号は、画像内に捕捉された任意の検出可能な発光（すなわち、「スポット」）を表すことができる。したがって、本明細書で使用するとき、「信号」は、検体の検体からの実際の放出の両方を指すことができ、実際の検体と相関しない擬似発光を指すことができる。したがって、信号はノイズから生じ得、試験片の実際の検体を代表しないように後に廃棄することができる。

30

【0427】

本明細書で使用するとき、用語「クランプ」は、一群の信号を指す。特定の実施態様では、信号は、異なる検体に由来する。好ましい別の実施態様では、信号クランプは、一緒にクラスター化する信号群である。より好ましい実施態様では、信号クランプは、1つの増幅オリゴヌクレオチドによって覆われた物理的領域を表す。各信号クランプは、理想的には、いくつかの信号（テンプレートサイクル当たり1つ、恐らくはクロストークによってより多く）として観察されるべきである。したがって、2つ（以上）の信号が同じ信号のクランプからテンプレートに含まれる、重複する信号が検出される。

40

【0428】

本明細書で使用するとき、「最小」、「最大」、「最小化」、「最大化」、及びその文法的変異形などの用語は、絶対最大値又は最小値ではない値を含むことができる。いくつかの実施態様では、値は、最大値及び最小値付近を含む。他の実施例では、値は、局所的な最大値及び／又は局所最小値を含むことができる。いくつかの実施態様では、値は、絶対最大値又は最小値のみを含む。

【0429】

本明細書で使用するとき、「クロストーク」は、別個の画像においても検出される1つの画像内の信号の検出を指す。好ましい別の実施態様では、クロストークは、放射された信号が2つの別個の検出チャンネルで検出されるときに発生し得る。例えば、放射された信

50

号が1つの色で発生する場合、その信号の放射スペクトルは、別の色で別の放射された信号と重なってもよい。好ましい実施態様では、ヌクレオチド塩基 A、C、G、及び T の存在を示すために使用される蛍光分子は、別個のチャンネルで検出される。しかし、A 及び C の発光スペクトルは重複するため、色チャンネルを使用した検出中に、C 色信号の一部が検出され得る。したがって、A 信号と C 信号との間のクロストークにより、1つのカラー画像からの信号が他のカラー画像に現れることを可能にする。いくつかの実施態様では、G 及び T クロストークがある。いくつかの実施態様では、チャンネル間のクロストークの量は非対称である。チャンネル間のクロストークの量は、とりわけ、適切な放射スペクトルを有する信号分子の選択、並びに検出チャンネルのサイズ及び波長範囲の選択によって制御され得ることが理解されるであろう。

10

【0430】

本明細書で使用する時、「位置合わせする (register)」、「位置合わせすること (registering)」、「位置合わせ (registration)」、及び同様の用語は、画像又はデータセット内の信号を、最初の時点又は視点からの画像又はデータセット内の信号と関連させるための任意のプロセスを指す。例えば、位置合わせを使用して、画像のセットからの信号を整列させてテンプレートを形成することができる。別の例では、位置合わせを使用して、他の画像からの信号をテンプレートに位置合わせすることができる。1つの信号は、別の信号に直接又は間接的に位置合わせされてもよい。例えば、画像「S」からの信号は、画像「G」に直接位置合わせされてもよい。別の例として、画像「N」からの信号は、画像「G」に直接位置合わせされてもよく、代替的に、画像「N」からの信号は、以前に画像「G」に位置合わせされた画像「S」に位置合わせされてもよい。したがって、画像「N」からの信号は、画像「G」に間接的に位置合わせされる。

20

【0431】

本明細書で使用する時、用語「基準」は、物体内又は物体上の区別可能な基準点を意味することを意図する。基準点は、例えば、マーク、第2の物体、形状、縁部、領域、不規則性、チャンネル、ピット、ポストなどであり得る。基準点は、物体の画像内に、又は物体を検出することに由来する別のデータセット内に存在することができる。基準点は、物体の平面内の x 及び y 又は x y 座標によって指定することができる。代替的に又は追加的に、基準点は、例えば、物体と検出器との相対位置によって定義される、x y 平面に直交する z 座標によって指定することができる。基準点に対する1つ以上の座標は、物体の1つ以上の他の検体、又は物体に由来する画像又は他のデータセットに対して指定することができる。

30

【0432】

本明細書で使用する時、用語「光信号」は、例えば、蛍光、発光、散乱、又は吸収信号を含むことを意図する。光信号は、紫外線 (Ultraviolet、UV) 範囲 (約 200 ~ 390 nm)、可視 (Visible、VIS) 範囲 (約 391 ~ 770 nm)、赤外線 (Infrared、IR) 範囲 (約 0.771 ~ 25 マイクロメートル)、又は電磁スペクトルの他の範囲で検出することができる。これらの範囲のうちの1つ以上の全て又は一部を除外する方式で、光信号を検出することができる。

【0433】

本明細書で使用する時、用語「信号レベル」は、所望又は所定の特性を有する検出されたエネルギー又は符号化された情報の量又は量を意味することを意図する。例えば、光信号は、強度、波長、エネルギー、周波数、電力、輝度などのうちの1つ以上によって定量化することができる。他の信号は、電圧、電流、電界強度、磁場強度、周波数、電力、温度などの特性に従って定量化することができる。信号の不在は、ゼロの信号レベル、又はノイズとは有意に区別されない信号レベルであると理解される。

40

【0434】

本明細書で使用する時、用語「シミュレートする」は、物又はアクションの特性を予測する物又はアクションの表現又はモデルを作成することを意味することを意図する。表現又はモデルは、多くの場合、物又はアクションと区別可能であり得る。例えば、表現又

50

はモデルは、色、被加工、サイズ、又は形状の全て又は一部から検出される信号の強度などの1つ以上の特性に対するものと区別することができる。特定の実施態様では、表現又はモデルは、物又はアクションと比較して、理想化、誇張、ミュート、又は不完全であり得る。したがって、いくつかの実施態様では、モデルの表現は、例えば、上記の特性のうちの少なくとも1つに関して表すものであることを表すものであることができる。表現又はモデルは、本明細書の他の箇所に記載されるもののうちの1つ以上などのコンピュータ可読フォーマット又は媒体に提供され得る。

【0435】

本明細書で使用する時、用語「特定の信号」は、背景エネルギー又は情報などの他のエネルギー又は情報にわたって選択的に観察される、検出されたエネルギー又は符号化情報を意味することを意図する。例えば、特定の信号は、特定の強度、波長、又は色で検出された光信号；特定の周波数、電力、又は電界強度で検出された電気信号；又は分光法及び分析検出に関連する当該技術分野で既知の他の信号であり得る。

【0436】

本明細書で使用する時、用語「スワス (swath)」は、物体の矩形部分を意味することを意図する。スワスは、ストリップの最長寸法に平行な方向に、物体と検出器との間の相対移動によって走査される細長いストリップであり得る。一般に、矩形部分又はストリップの幅は、その全長に沿って一定である。物体の複数のスワスは、互いに平行であってもよい。物体の複数のスワスは、互いに重なり合い、互いに隣接するか、又は間質領域によって互いに分離され得る。

【0437】

本明細書で使用する時、用語「分散」は、予想される差、及び観察される差、又は2つ以上の観測結果間の差を意味することを意図する。例えば、分散は、期待値と測定値との間の不一致であり得る。標準偏差、標準偏差の二乗、変動係数などの統計関数を使用して、分散を表すことができる。

【0438】

本明細書で使用する時、用語「x y 座標」は、x y 平面内の位置、サイズ、形状、及び／又は向きを指定する情報を意味することを意図する。情報は、例えば、デカルト座標系における数値座標であり得る。座標は、x 軸及び y 軸の一方又は両方に対して提供することができる、又は x y 平面内の別の場所に対して提供することができる。例えば、物体の検体の座標は、対象物の基準又は他の検体の場所に対する検体の場所を指定することができる。

【0439】

本明細書で使用する時、用語「x y 平面」は、直線軸 x 及び y によって画定される二次元領域を意味することを意図する。検出器及び検出器によって観察される物体を参照して使用される場合、領域は検出器と検出されている物体との間の観測方向に直交するように更に指定することができる。

【0440】

本明細書で使用する時、用語「z 座標」は、x y 平面に直交する軸に沿った点、線、又は領域の位置を指定する情報を意味することを意図する。特定の実施態様では、z 軸は、検出器によって観察される物体の領域に直交する。例えば、光学系の焦点の方向は、z 軸に沿って指定されてもよい。

【0441】

いくつかの実施態様では、獲得された信号データは、アフィン変換を用いて変換される。いくつかのそのような実施態様では、テンプレートの生成は、色チャンネル間のアフィン変換が動作間に一貫しているという事実を使用する。この一貫性のため、標本中の検体の座標を決定する際に、デフォルトオフセットのセットを使用することができる。例えば、デフォルトオフセットファイルは、A チャンネルなどの1つのチャンネルに対する異なるチャンネルに対する相対変換（シフト、スケール、スキュー）を含むことができる。しかしながら、他の実施態様では、動作中及び／又は動作間の色チャンネルドリフト間のオフセットは

、オフセット駆動型テンプレート生成を困難にする。このような実施例では、本明細書で提供される方法及びシステムは、オフセットしたテンプレート生成を利用することができ、これについては以下で更に説明する。

【0442】

上記の実施態様のいくつかの実施態様では、システムはフローセルを含み得る。いくつかの実施態様では、フローセルは、レーン、又は他の構成のタイルを含み、タイルの少なくとも一部は、1つ以上の検体群を含む。いくつかの実施態様では、検体は、核酸などの複数の分子を含む。特定の態様では、フローセルは、標識されたヌクレオチド塩基を核酸の配列に送達し、それによって、核酸を含む検体に対応する信号を生成するように、検体内の核酸にハイブリダイズするプライマーを伸長させるように構成される。好ましい実施態様では、検体内の核酸は、互いに同一又は実質的に同一である。

10

【0443】

本明細書に記載される画像分析システムのいくつかにおいて、画像のセット内の各画像は、色信号を含み、異なる色は、異なるヌクレオチド塩基に対応する。いくつかの実施態様では、画像のセットの各画像は、少なくとも4つの異なる色から選択される単一の色を有する信号を含む。いくつかの実施態様では、画像のセット内の各画像は、4つの異なる色から選択される単一の色を有する信号を含む。本明細書に記載されるシステムのいくつかにおいて、核酸は、4つの異なる画像を生成するように、4つの異なる標識ヌクレオチド塩基を分子の配列に提供することにより、核酸を配列決定することができ、各画像は単一の色を有するシグナルを含み、信号色が、4つの異なる画像の各々に対して異なることにより、核酸内の特定の位置に存在する4つの可能なヌクレオチドに対応する4つのカラー画像のサイクルを生成する。特定の態様では、システムは、追加の標識ヌクレオチド塩基を分子の配列に送達するように構成されているフローセルを含み、それによって複数のカラー画像のサイクルを生成する。

20

【0444】

好ましい実施態様形態では、本明細書で提供される方法は、プロセッサがアクティブにデータを収集しているか、又はプロセッサが低活動状態にあるかどうかを判定することを含み得る。多数の高品質画像を収集及び記憶することは、典型的には、大量の記憶容量を必要とする。更に、収集され記憶されると、画像データの分析はリソース集約的になり得、追加の画像データの収集及び記憶などの他の機能の処理能力を妨げる可能性がある。したがって、本明細書で使用する時、用語「低活動状態」は、所与の時間におけるプロセッサの処理能力を指す。いくつかの実施態様では、低活動状態は、プロセッサがデータを収集及び／又は記憶していないときに生じる。いくつかの実施態様では、一部のデータ収集及び／又は記憶が行われる場合には、低いアクティビティ状態が生じるが、他の機能に干渉することなく画像分析が同時に生じ得るように、追加の処理能力が残る。

30

【0445】

本明細書で使用する時、「競合を識別する」とは、複数のプロセスがリソースに対して競合する状況を識別することを指す。いくつかのそのような実施態様では、1つのプロセスは、別のプロセスに対して優先度を与えられる。いくつかの実施態様では、競合は、時間、処理能力、記憶能力、又は優先度が与えられる任意の他のリソースの割り当てに対する優先度を与える必要性に関連し得る。したがって、いくつかの実施態様では、処理時間又は容量が、データセットを分析し、データセットを取得及び／又は記憶するかのいずれかなどの2つのプロセス間に分散される場合、2つのプロセス間の不一致が存在し、プロセスのうちの1つに優先度を与えることによって解決することができる。

40

【0446】

本明細書では、画像分析を実行するためのシステムも提供される。システムは、プロセッサと、記憶容量と、画像分析用のプログラムと、を含むことができ、プログラムは、記憶のための第1のデータセット及び分析のための第2のデータセットを処理するための命令を含み、処理は、記憶装置上の第1のデータセットを取得及び／又は記憶することと、プロセッサが第1のデータセットを取得していないときに第2のデータセットを分析する

50

ことと、を含む。特定の態様では、プログラムは、第1のデータセットを収集及び／又は記憶することと、第2のデータセットを解析することとの間の競合の少なくとも1つのインスタンスを識別するための命令を含み、第1のデータセットを収集及び／又は記憶することが優先度を与えられるように、画像データを取得及び／又は記憶することが優先される。特定の態様では、第1のデータセットは、光学撮像デバイスから収集された画像ファイルを含む。特定の態様では、システムは、光学撮像デバイスを更に備える。いくつかの実施態様では、光学撮像装置は、光源と検出デバイスとを備える。

【0447】

本明細書で使用する時、用語「プログラム」は、タスク又はプロセスを実行するための命令又はコマンドを指す。用語「プログラム」は、用語「モジュール」と互換的に使用され得る。特定の態様では、プログラムは、同じコマンドセットの下で実行される様々な命令のコンパイルであり得る。他の実施態様では、プログラムは、別個のバッチ又はファイルを参照することができる。

【0448】

以下に記載されるのは、本明細書に記載される画像分析を実行するための方法及びシステムを利用する驚くべき効果の一部である。いくつかの配列決定の実現例では、配列決定システムの有用性の重要な尺度は、その全体的な効率である。例えば、1日当たりに生成されるマッピング可能なデータの量、並びに機器の設置及び動作の総コストは、経済的な配列決定ソリューションの重要な態様である。マッピング可能なデータを生成し、システムの効率を高めるための時間を短縮するために、リアルタイムのベースコールを機器コンピュータ上で有効にすることができ、配列決定ケミストリ及び画像化と並行して動作することができる。これにより、配列決定ケミストリ仕上げの前に、データ処理及び分析が完了することを可能にする。更に、中間データに必要な記憶を低減し、ネットワークを横切って移動する必要があるデータの量を制限することができる。

【0449】

シーケンス出力が増加している間、本明細書で提供されるシステムからネットワークに転送された動作ごとのデータ、及び二次分析処理ハードウェアは、実質的に減少している。機器コンピュータ（取得コンピュータ）上でデータを変換することにより、ネットワークロードが劇的に低減される。これらのオン機器、オフネットワークデータ低減技術に伴わずに、DNA配列決定機器のフレッツの画像出力は、ほとんどのネットワークをクリップするであろう。

【0450】

ハイスループットDNA配列決定機器の広範な採用は、使用の容易さ、用途の範囲に対する支持、及び実質的に任意のラボ環境に対する適合性によって、部分的に駆動されてきた。本明細書に提示される高度に効率的なアルゴリズムは、配列決定機器を制御することができる単純なワークステーションに、有意な分析機能を加えることを可能にする。計算ハードウェアの必要条件のこの低減は、配列決定出力レベルが増加し続けるにつれて、更に重要となる、いくつかの実用的な利点を有する。例えば、単純なタワー、熱生成、実験室設置面積、及び電力消費を最小限に抑えるために、画像分析及びベースコールを行うことによって、最小に保たれる。対照的に、他の商業的な配列決定技術は、1次分析のために、最大5回の処理電力で、そのコンピューティングインフラストラクチャを最近ランブアップして、熱出力及び電力消費の増加を開始する。したがって、いくつかの実施態様では、本明細書で提供される方法及びシステムの計算効率は、サーバハードウェアを最小限に抑えながら、それらの配列決定スループットを増加させることを可能にする。

【0451】

したがって、いくつかの実施態様では、本明細書に提示される方法及び／又はシステムは、状態マシンとして機能し、各試料の個々の状態の追跡を保ち、試料が次の状態に進む準備ができていることを検出すると、適切な処理を行い、試料をその状態に前進させる。状態マシンがファイルシステムを監視して、好ましい実施態様に従って試料が次の状態に進む準備ができているかを判定する方法のより詳細な例が、以下の実施例1に記載されて

10

20

30

40

50

いる。

【0452】

好ましい実施態様では、本明細書で提供される方法及びシステムは、マルチスレッドであり、構成可能な数のスレッドと協働することができる。したがって、例えば、核酸配列決定の文脈において、本明細書で提供される方法及びシステムは、リアルタイム分析のためのライブ配列決定動作中に背景において作用することができ、又はオフライン分析のために既存の画像データセットを使用して動作することができる。特定の好ましい実施態様では、方法及びシステムは、各スレッドを、それが関与する検体のそれ自体のサブセットを与えることによって、マルチスレッドを取り扱う。これにより、スレッド保持の可能性が最小限に抑えられる。

10

【0453】

本開示の方法は、検出装置を使用して物体の標的画像を取得する工程を含むことができ、この画像は、物体上の検体の繰り返しパターンを含む。表面の高解像度撮像が可能な検出装置が特に有用である。特定の実施態様では、検出装置は、本明細書に記載される密度、ピッチ、及び／又は検体サイズにおける検体を区別するのに十分な分解能を有するであろう。表面から画像又は画像データを得ることができる検出装置が特に有用である。例示的な検出器は、物体と検出器とを静的関係に維持しつつ、面積画像を取得するように構成されたものである。走査装置も使用することができる。例えば、連続領域画像を取得する装置（例えば、「工程及びショット」検出器と呼ばれる）を使用することができる。また、物体の表面上の点又は線を連続的に走査して、表面の画像を構築するためにデータを蓄積するデバイスも有用である。点走査検出器は、表面の $x-y$ 平面内のラスタ運動を介して物体の表面上の点（すなわち、小さい検出領域）を走査するように構成することができる。線走査検出器は、物体の表面の y 次元に沿った線を走査するように構成することができる。この線の最長寸法は、 x 次元に沿って生じる。検出デバイス、物体、又はその両方を移動させて、走査検出を達成することができることが理解されるであろう。例えば核酸配列決定用途において特に有用な検出装置は、米国特許出願公開第 2012/0270305 (A1) 号、第 2013/0023422 (A1) 号及び同第 2013/0260372 (A1) 号、並びに米国特許第 5,528,050 号、同第 5,719,391 号、米国特許第 8,158,926 号、及び米国特許第 8,241,573 号に記載されており、これらの各々は、参照により本明細書に組み込まれる。

20

30

【0454】

本明細書に開示される実施態様は、ソフトウェア、ファームウェア、ハードウェア、又はそれらの任意の組み合わせを生成するためのプログラミング技術又は工学技術を使用して、製造方法、装置、システム、又は物品として実装されてもよい。本明細書で使用するとき、用語「製造物品」は、光学記憶デバイスなどのハードウェア又はコンピュータ可読媒体、並びに揮発性又は不揮発性メモリデバイス内に実装されるコード又は論理を指す。そのようなハードウェアとしては、フィールドプログラマブルゲートアレイ (field programmable gate array、FPGA)、粗粒度再構成可能構造 (coarse grained reconfigurable architecture、CGRA)、特定用途向け集積回路 (application-specific integrated circuit、ASIC)、複合プログラマブル論理デバイス (Complex Programmable Logic Device、CPLD)、プログラマブルロジックアレイ (Programmable Logic Array、PLA)、マイクロプロセッサ、又は他の同様の処理装置が挙げられるが、これらに限定されない。特定の実施態様では、本明細書に記載される情報又はアルゴリズムは、非一過性記憶媒体中に存在する。

40

【0455】

特定の実施態様形態では、本明細書に記載されるコンピュータ実装の方法は、物体の複数の画像が取得されている間、リアルタイムで発生することができる。このようなリアルタイム分析は、核酸配列が流体及び検出工程の繰り返しサイクルに供される核酸配列決定用途に特に有用である。配列決定データの分析は、多くの場合、本明細書に記載される方法をリアルタイム又は背景で実行するのに有益であり得る一方で、他のデータ収集又は分

50

析アルゴリズムがプロセス中である間に、本明細書に記載される方法を実行することが有益であり得る。本方法で 사용할 ことができるリアルタイム分析法の例は、Illumina, Inc (San Diego, Calif) から市販されており、及び／又は、参照により本明細書に組み込まれる米国特許出願公開第 2012/0020537 (A1) 号に記載されている MiSeq 及び HiSeq 配列決定機器に使用されるものである。

【0456】

1つ以上のプログラムされたコンピュータによって形成され、本明細書に記載される方法の1つ以上の工程を実行するために実行されるコードを有するプログラミングが、1つ以上の機械可読媒体上に記憶されている、例示的データ分析システム。一実施態様では、例えば、システムは、標的物体からデータを取得するように構成されている1つ以上の検出システム（例えば、光学撮像システム）へのシステムのネットワーキングを可能にするように設計されたインターフェースを含む。インターフェースは、適切な場合には、データを受信及び条件することができる。特定の実施態様では、検出システムは、例えば、アレイ又は他の物体の画像と一緒に形成する個々の画像要素又はピクセルを表す画像データを出力する。プロセッサは、処理コードによって定義された1つ以上のルーチンに従って、受信した検出データを処理する。処理コードは、様々なタイプのメモリ回路に記憶されてもよい。

【0457】

現時点で企図される実施態様によれば、検出データ上で実行される処理コードは、検出データを分析して、データ内で可視又は符号化された個々の検体の場所、及び検体が検出されない場所（すなわち、検体が存在しないか、又は、既存の検体から有意な信号が検出されない場所）及びメタデータを判定するように設計されたデータ分析ルーチンを含む。特定の実施態様では、アレイ内の検体位置は、典型的には、撮像された検体に付着した蛍光染料の存在に起因して、非検体位置よりも明るく見える。検体は、例えば、検体におけるプローブの標的が検出されているアレイ内に存在しない場合、検体は、それらの周囲領域よりも明るく見える必要はないことが理解されるであろう。個々の検体が現れる色は、使用される染料、並びに撮像目的のために撮像システムによって使用される光の波長の関数であり得る。標的が結合されていない、又は特定のラベルを有さない検体は、マイクロアレイ内の予想される場所などの他の特性に従って識別することができる。

【0458】

データ分析ルーチンがデータ中に個々の検体を配置すると、値割り当てが実行され得る。一般に、値割り当ては、対応する場所における検出器構成要素（例えば、ピクセル）によって表されるデータの特性に基づいて、各検体にデジタル値を割り当てる。すなわち、例えば、画像化データが処理されるとき、値割り当てルーチンは、特定の場所で特定の色又は波長の光が検出されたことを認識するように、その場所でピクセルのグループ又はクラスターによって示されるように、設計されてもよい。典型的なDNA画像化用途では、例えば、4つの共通ヌクレオチドは、4つの別個の区別可能な色によって表される。次いで、各色は、そのヌクレオチドに対応する値を割り当てられてもよい。

【0459】

本明細書で使用するとき、用語「モジュール」、「システム」、又は「システムコントローラ」は、1つ以上の機能を実行するように動作するハードウェア及び／又はソフトウェアシステム及び回路を含み得る。例えば、モジュール、システム、又はシステムコントローラは、コンピュータメモリなどの有形及び非一時的コンピュータ可読記憶媒体上に記憶された命令に基づいて動作を実行する、コンピュータプロセッサ、コントローラ、又は他のログベースのデバイスを含んでもよい。代替的に、モジュール、システム、又はシステムコントローラは、有線論理及び回路に基づいて動作を実行する、有線デバイスを含んでもよい。添付の図面に示されるモジュール、システム、又はシステムコントローラは、ソフトウェア又は結線命令に基づいて動作するハードウェア及び回路、ハードウェアを動作させるように命令するソフトウェア、又はそれらの組み合わせを表し得る。モジュール、システム、又はシステムコントローラは、1つ又はコンピュータマイクロプロセッサな

どの1つ以上のプロセッサを含む、及び／又はそれと接続されるハードウェア回路又は回路を含むか、又は表すことができる。

【0460】

本明細書で使用する時、用語「ソフトウェア」及び「ファームウェア」は互換可能であり、RAMメモリ、ROMメモリ、EPROMメモリ、EEPROMメモリ、及び不揮発性RAM (non-volatile RAM、NVRAM) メモリを含むコンピュータによって実行されるメモリに記憶された任意のコンピュータプログラムを含む。上記メモリタイプは単なる例であり、コンピュータプログラムの記憶に使用可能なメモリの種類に限定されるものではない。

【0461】

分子生物学分野では、使用中の核酸配列決定のためのプロセスの1つは、配列番号合成である。この技術は、非常に平行な配列決定プロジェクトに適用することができる。例えば、自動プラットフォームを使用することにより、数百万の配列決定反応を同時に行うことが可能である。したがって、本発明の実施態様の1つは、核酸配列決定中に生成された画像データを収集、記憶、及び分析するための機器及び方法に関する。

【0462】

収集及び記憶することができるデータ量の莫大なゲインは、合理化された画像分析方法を更により有益にする。例えば、本明細書に記載される画像分析方法は、設計者及びエンドユーザの両方が、既存のコンピュータハードウェアの効率的な使用を行うことを可能にする。したがって、本明細書では、迅速に増加するデータ出力の面における処理データの計算量を低減する方法及びシステムが本明細書に提示される。例えば、DNA配列決定の分野では、収率は最近の過程で15倍に拡大され、DNA配列決定デバイスの単一の動作において何百ものギガベースに達する可能性がある。計算インフラストラクチャの要件が比例的に増加した場合、大規模なゲノム規模の実験は、ほとんどの研究者に到達していない。したがって、より多くの生シーケンスデータの生成は、二次分析及びデータ記憶の必要性を増加させ、データ輸送及び記憶の最適化を非常に有益にする。本明細書に提示される方法及びシステムのいくつかの実施態様は、使用可能なシーケンスデータを生成するために必要な時間、ハードウェア、ネットワーキング、及び実験室インフラストラクチャ要件を低減することができる。

【0463】

本開示は、方法を実行するための様々な方法及びシステムを説明する。方法のいくつかの例は、一連の工程として記載される。しかしながら、実施態様は、本明細書に記載される特定の工程及び／又は工程の順序に限定されないことを理解されたい。工程は省略されてもよく、工程は修正されてもよく、及び／又は他の工程が追加されてもよい。更に、本明細書に記載される工程を組み合わせることができ、工程は同時に実行されてもよく、工程は同時に実行されてもよく、工程は複数のサブ工程に分割されてもよく、工程は、異なる順序で実行されてもよく、又は工程（又は一連の工程）は、反復的に再実行されてもよい。加えて、本明細書には異なる方法が記載されているが、他の実施態様では、異なる方法（又は異なる方法の工程）を組み合わせてもよいことを理解されたい。

【0464】

いくつかの実施態様では、タスク又は動作を実行するように「構成された」処理ユニット、プロセッサ、モジュール、又はコンピューティングシステムは、タスク又は動作を実行するように特に構造化されていると理解され得る（例えば、タスク又は動作を実行するように調整又は意図された、及び／又はタスク若しくは動作を実行するように調整若しくは意図された1つ以上のプログラム又は命令を有すること、及び／又はタスク若しくは動作を実行するように調整又は意図された処理回路の配置を有する）。明確さ及び疑義の回避のために、汎用コンピュータ（適切にプログラムされた場合にタスク又は動作を実行するように構成された）となり得る）汎用コンピュータは、タスク又は動作を実行するために具体的にプログラム又は構造的に変更されない限り、タスク又は動作を実行するように「構成されている」ように構成されていない）。

【0465】

更に、本明細書に記載される方法の操作は、操作が、商業的に妥当な時間期間内に、当業者には、平均的なヒト又は当業者によって実施されることができないように、十分に複雑であり得る。例えば、本方法は、そのような人が商業的に妥当な時間内で方法を完了することができないように、比較的複雑な計算に依存し得る。

【0466】

本出願全体を通して、様々な刊行物、特許、又は特許出願が参照されている。これらの出版物の全体の開示は、本発明が属する技術分野の状態をより完全に説明するために、本出願において参照により本明細書に組み込まれる。

【0467】

用語「含む (comprising)」は、本明細書では、列挙された要素のみならず、任意の追加の要素を更に包含する、オープンエンドであることが意図される。

【0468】

本明細書で使用する時、用語「各々」は、項目の集合を参照して使用されるとき、集合内の個々の項目を識別することを意図しているが、必ずしも集合内の全ての項目を指すものではない。明示的な開示又は文脈がそうでないことを明確に指示する場合、例外が生じ得る。

【0469】

上記の実施例を参照して本発明を説明してきたが、本発明から逸脱することなく様々な修正を行うことができることを理解されたい。

【0470】

本出願のモジュールは、ハードウェア又はソフトウェアで実装することができ、図に示されるように、正確に同じブロックで分割される必要はない。いくつかは、異なるプロセッサ若しくはコンピュータ上に実装されてもよく、又は多数の異なるプロセッサ若しくはコンピュータの中で広がることもできる。加えて、モジュールの一部は、達成される機能に影響を及ぼすことなく、図に示されるものとは並行して、又は異なる順序で操作されることが理解されるであろう。また、本明細書で使用する時、用語「モジュール」は、モジュールを構成するために本明細書で考慮することができる、「サブモジュール」を含むことができる。モジュールとして指定された図のブロックはまた、方法におけるフローチャート工程と考えることができる。

【0471】

本明細書で使用する時、情報項目の「識別」は、その情報の項目の直接仕様を必ずしも必要としない。情報は、単に、一方向の1つ以上の層を通じた実際の情報を単に参照することによって、又は情報の実際の項目を決定するのに十分である異なる情報の1つ以上のアイテムを識別することによって、フィールド内で「識別され得る」ことができる。加えて、用語「指定する」は、本明細書では、「識別する」と同じであることを意味する。

【0472】

本明細書で使用する時、所与の信号、イベント又は値は、「先行する信号、イベント又は先行する信号の値、所与の信号、イベント、又は値によって影響されるイベント又は値に依存する。介在処理要素、工程又は期間が存在する場合、所与の信号、イベント、又は値は、「先行する信号、イベント又は値」に依存して「存在する」ことができる。介在処理要素又は工程が2つ又はそれ以上の信号、イベント、又は値を組み合わせる場合、処理要素又は工程の信号出力は、「信号、イベント、又は値入力のそれぞれ」に依存していると思なされる。所与の信号、イベント又は値が先行する信号、イベント又は値と同じである場合、これは単に、所与の信号、イベント、又は値が「先行する信号、イベント又は値」に「依存して (in dependence upon) 若しくは (dependent on)」又は先行する信号、イベント又は値に「基づいて」いると思なされる。別の信号、イベント又は値に対する所与の信号、イベント、又は値の「応答性」は、同様に定義される。

【0473】

本明細書で使用する時、「並行して」又は「並行して」は、正確な同時性を必要とし

10

20

30

40

50

ない。個人の1人の評価が、個人の別の評価が完了する前に開始する場合に十分である。

【0474】

本出願は、「配列決定画像」、「クラスター画像」及び「クラスター強度画像」を互換的に指す。

【0475】

他の実施態様

一実施態様では、現在の配列決定サイクルについて、状態データは、例えば、ニューラルネットワークベースのベースコーラー144の中間層への残留接続及び／又はスキップ接続を使用して、ベースコーラー2900への入力として冗長的に提供され得る。いくつかの実施態様では、残留接続及び／又はスキップ接続は、状態データがニューラルネットワークベースのベースコーラー2900の第1の層への入力として提供されることに加えて、状態データを入力として提供する。

10

【0476】

一実施態様では、現在の配列決定サイクルについて、状態データは、全ての以前の配列決定サイクルとは対照的に、以前の配列決定サイクルのサブセットから決定することができる。

【0477】

一実施態様では、現在の配列決定サイクルについて、各インスタンスが以前の配列決定サイクルの異なるサブセットから決定されるように、状態データの複数のインスタンスを決定することができる。

20

【0478】

一実施態様では、現在の配列決定サイクルについて、状態データの複数のコピーが、ベースコーラー144への入力として、例えば複数の入力チャンネルとして提供することができる。

【0479】

いくつかの実施態様では、状態情報は、時間レイヤへの入力として与えられ得る。

【0480】

発明性及び非自明性の客観的な指標としての性能結果

図45は、非ニューラルネットワークベースのベースコーラーRTA、状態情報を伴わないニューラルネットワークベースのベースコーラーDeepRTA(5ci_k14)、並びに追加／補足入力としてピクセルごとのMIN状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State(5ci_k14_DC)のベースコールパフォーマンスを比較する。図45において、x軸は配列決定ランの配列決定サイクルの進行を示し、y軸はベースコール誤差率を示す。また、図45は、第1のリード(リード1)の配列決定及び第2のリード(リード2)の配列決定のための2つのチャートを有する。

30

【0481】

実証されるように、状態のベースコール精度を有するDeepRTAは、リード1において状態情報を有しないDeepRTAに対して10%の改善を示し、リード2において8%の改善を示す。状態を有するDeepRTAはまた、RTAよりも著しく性能が優れている。

40

【0482】

図46異なる状態チャンネルにわたって、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State(5ci_k14_DC)のベースコールパフォーマンスを比較する。図46では、異なる状態チャンネルは、DC(MIN)_blue、DC(MIN)_green、DC(MIN)_avg、DC(MIN)_min、及びDC(MIN)である。DC(MIN)_blue状態チャンネルは、青色チャンネルについてのみピクセルごとのMIN状態値を含む単一チャンネルである。DC(MIN)_green状態チャンネルは、緑色チャンネルについてのみピクセルごとのMIN状態値を含む単一チャンネルである。DC(MIN)_avg状態チャンネルは、

50

青色チャネル及び緑色チャネルのMIN状態値の平均として決定されるピクセルごとのAVG状態値及びパンチャネルAVG状態値を含む単一のチャネルである。DC(MIN)___min状態チャネルは、青色チャネル及び緑色チャネルのMIN状態値から選択されたピクセルごとのMIN状態値及びパンチャネルMIN状態値を含む単一チャネルである。DC(MIN)状態チャネルは2つのチャネルであり、1つは青色チャネルのためのピクセルごと及びチャネルごとのMIN状態値を含み、もう1つは緑色チャネルのためのピクセルごと及びチャネルごとのMIN状態値を含む。

【0483】

補助入力としてDC(MIN)状態チャネルを用いると、DeepRTA with Stateのベースコール精度は、リード1におけるDeepRTAに対して10%の改善を示す。補助入力としてDC(MIN)___blue状態チャネルを用いると、DeepRTA with Stateのベースコール精度は、リード1におけるDeepRTAに対して7%の改善を示す。補助入力としてDC(MIN)___green状態チャネルを用いると、DeepRTA with Stateのベースコール精度は、リード1におけるDeepRTAに対して4%の改善を示す。補助入力としてDC(MIN)___avg状態チャネルを用いると、DeepRTA with Stateのベースコール精度は、リード1におけるDeepRTAに対して8%の改善を示す。補助入力としてDC(MIN)___min状態チャネルを用いると、DeepRTA with Stateのベースコール精度は、リード1におけるDeepRTAに対して7%の改善を示す。

【0484】

図47は、k-mer、特に5merに対する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State(DC)のベースコールパフォーマンスを示す。5merは、5つの塩基位置(例えば、ACGCGG、GGGGG、TCGCG)についての反復塩基パターンを指す。図47において、横軸は、5merの1024通りの組み合わせ全てを示し、縦軸は、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State(DC)のベースコールパフォーマンスを示す。

【0485】

特に、縦軸は、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State(DC)のベースコール誤差を、状態情報のないDeepRTAのベースコール誤差と比較し、その結果、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State(DC)のベースコール誤差が小さいほど、縦軸に沿った負の値が大きくなる。示されるように、Gの高度に誤りがちなk-mer(アクティブ状態の欠如(すなわち、Gの暗状態のみ)に起因する)は、状態を伴う開示される状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA(DC)を用いて、小さいベースコール誤差率4702を経験する。

【0486】

図48は、非ニューラルネットワークベースのベースコーラーRTA、状態情報を伴わないニューラルネットワークベースのベースコーラーDeepRTA(5ci_k14)、追加/補足入力としてピクセルごとのMIN状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State(5ci_k14___DC___MIN)、並びに追加/補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State(5ci_k14___DC___MIN___MAX)のベースコールパフォーマンスを比較する。

【0487】

図48は、状態情報を有しないDeepRTA(5ci_k14)がRTAに対して32%の改善を実証し、追加/補足入力としてピクセルごとのMIN状態情報を使用する開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA

with State (5 c i _ k 1 4 _ D C _ M I N) が R T A に対して 3 9 % の改善を実証し、追加／補足入力としてピクセルごとの M I N 及び M A X 状態情報を使用する開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep R T A with State (5 c i _ k 1 4 _ D C _ M I N _ M A X) が R T A に対して 5 3 % の改善を実証することを示す。

【0488】

図49は、非ニューラルネットワークベースのベースコーラー R T A、状態情報及び5サイクル隣接入力ウィンドウを伴わないニューラルネットワークベースのベースコーラー Deep R T A (5 c i _ k 1 4)、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとの M I N 状態情報を使用する開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep R T A with State (5 c i _ k 1 4 _ D C _ M I N)、3サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとの M I N 状態情報及び M A X 状態情報を使用する開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep R T A with State (3 c i _ k 1 4 _ D C _ M I N _ M A X)、並びに5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとの M I N 状態情報及び M A X 状態情報を使用する開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep R T A with State (5 c i _ k 1 4 _ D C _ M I N _ M A X) のベースコールパフォーマンスを比較する。

【0489】

図49は、3サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとの M I N 及び M A X 状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep R T A with State (3 c i _ k 1 4 _ D C _ M I N _ M A X) が、R T A に対して 4 5 % の改善を実証することを示す。これは、状態情報を有しない Deep R T A (5 c i _ k 1 4) (及び Deep R T A with State (5 c i _ k 1 4 _ D C _ M I N) (両方とも5サイクル隣接入力ウィンドウを有する))と比較して、より小さい入力ウィンドウを用いても、Deep R T A with State (3 c i _ k 1 4 _ D C _ M I N _ M A X) (3サイクル隣接入力ウィンドウ)がより良好に機能することを示す。すなわち、状態情報を用いて、より少ない画像データを処理し、それによって計算を節約しながら、より良好なベースコール精度を得る。

【0490】

図50は、k-mer (特に5mer) 上の5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとの M I N 及び M A X 状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep R T A with State (5 c i _ k 1 4 _ D C _ M I N _ M A X) のベースコールパフォーマンスを、状態情報及び5サイクル隣接入力ウィンドウなしのニューラルネットワークベースのベースコーラー Deep R T A (5 c i _ k 1 4) のベースコールパフォーマンスと比較する。5merは、5つの塩基位置 (例えば、A C G C G、G G G G G、T C G C G) についての反復塩基パターンを指す。図50において、横軸は、5merの1024通りの組み合わせ全てを示し、縦軸は、開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep R T A with State (D C) のベースコールパフォーマンスを示す。

【0491】

特に、図50において、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとの M I N 及び M A X 状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラー Deep R T A with State (5 c i _ k 1 4 _ D C _ M I N _ M A X) は、状態情報及び5サイクル隣接入力ウィンドウを伴わないニューラルネットワークベースのベースコーラー Deep R T A (5 c i _ k 1 4) よりも長いネガティブラインを有し、これは、5サイクル隣接入力ウィンドウとともに

に追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA (5 c i _ k 1 4 _ D C _ M I N _ M A X) が、x 軸方向に5merについてより低いベースコール誤差率を有することを示す。

【0492】

図51は、等化器実装を伴う非ニューラルネットワークベースのベースコーラーRTA、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5 c i _ k 1 4 _ D C _ M I N _ M A X)、並びに5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのAVG状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5 c i _ k 1 4 _ D C _ A V G) のベースコールパフォーマンスを比較する。

10

【0493】

図51は、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5 c i _ k 1 4 _ D C _ M I N _ M A X) が、等化器実装を伴うRTAに対して36.7%の改善を実証することを示す。図51はまた、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのAVG状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5 c i _ k 1 4 _ D C _ A V G) が、等化器実装を伴うRTAに対して34.6%の改善を実証することを示す。

20

【0494】

図52は、等化器実装を伴う非ニューラルネットワークベースのベースコーラーRTAと、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN、MAX、及びAVG状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5 c i _ k 1 4 _ D C _ M I N _ M A X _ A V G) とのベースコールパフォーマンスを比較する。

【0495】

30

図52は、5サイクル隣接入力ウィンドウとともに追加／補足入力としてピクセルごとのMIN、MAX、及びAVG状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5 c i _ k 1 4 _ D C _ M I N _ M A X _ A V G) が、等化器実装を伴うRTAに対して36.9%の改善を実証することを示す。

【0496】

図53は、等化器実装を伴う非ニューラルネットワークベースのベースコーラーRTA、5サイクル隣接入力ウィンドウ及びサイズ14 (K=14) のフィルタバンクとともに追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5 c i _ k 1 4 _ D C _ M I N _ M A X)、並びに5サイクル隣接入力ウィンドウ及びサイズ32 (K=32) のフィルタバンクとともに追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5 c i _ k 1 4 _ D C _ M I N _ M A X) のベースコールパフォーマンスを比較する。

40

【0497】

図53は、5サイクル隣接入力ウィンドウ及びサイズ14 (K=14) のフィルタバンクとともに追加／補足入力としてピクセルごとのMIN及びMAX状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラーDeepRTA with State (5 c i _ k 1 4 _ D C _ M I N _ M A X) が、リード1につ

50

いて等化器実装を用いた R T A に対して 3 6 . 7 % の改善及びリード 2 について 2 7 . 2 % の改善を実証することを示す。図 5 3 はまた、5 サイクル隣接入力ウィンドウ及びサイズ 3 2 (K = 3 2) のフィルタバンクとともに、追加／補足入力としてピクセルごとの M I N 及び M A X 状態情報を使用する、開示された状態ベース及びニューラルネットワークベースのベースコーラー D e e p R T A w i t h S t a t e (5 c i _ k 1 4 _ D C _ M I N _ M A X) が、リード 1 について等化器実装を伴う R T A に対して 4 4 . 8 % の改善及びリード 2 について 3 6 . 9 % の改善を実証することを示す。

【 0 4 9 8 】

図 5 4 は、上記で議論されるように、指数加重平均実施の異なるアルファパラメータ構成例えば、0 . 0 5 、 0 . 0 7 、 0 . 1 0 、 0 . 1 2 について、開示される状態ベース及びニューラルネットワークベースのベースコーラー D e e p R T A と S t a t e とのベースコールパフォーマンスを比較する。示されるように、0 . 1 2 のアルファパラメータは、リード 1 及びリード 2 の両方について最良のベースコール精度を有する。

【 0 4 9 9 】

図 5 5 及び 5 6 は、開示された技術の一実施態様による、指数加重平均によって決定される状態値を追跡するグラフである。図 5 5 及び 5 6 において、青線は生強度であり、指数加重平均推定値は、サイクル 1 ~ 2 0 からの平均強度及びアルファ = 0 . 0 2 を使用してサイクル 2 1 で初期化される。x 軸はサイクル指数であり、y 軸は測定された振幅（青色）+ 推定された平均強度（橙色）である。オレンジ色の推定された振幅は、所望の強度／平均強度の包絡線に従う。平均強度を回復するために O N / O F F 強度にわたって平均化するため、ノイズが中程度である。次の例では、振幅はサイクル 7 5 で突然ステップダウンする。指数移動平均は、この振幅変化を非常に迅速に推定することができる。ピクセル強度平均は、大きな O N / O F F ステップ並びにより小さな振幅変動にわたって平均しなければならないので、非常に雑音の多い統計値である。

【 0 5 0 0 】

図 5 7 は、コールされたリードの数、リード 1 及びリード 1 についてのリードミスマッチ誤差率及びそれらの平均、単一ヌクレオチド多型 (S N P) 再現性、S N O 精度、S N P コール精度 (F 1 スコア) 、挿入／欠失 (I n d e l) 再現性、I n d e l 精度、及び I n d e l コール精度 (F 1 スコア) のような二次解析タスク及びメトリックスの観点から、等化器実装を有する R T A (R T A + E q) 、状態情報を有しない D e e p R T A (D e e p R T A (k 1 4 _ 1 m)) 、及び開示された D e e p R T A w i t h S t a t e (D e e p R T A _ e x t r a c h a n n e l s (k 1 4 _ 1 , _ o n o f f _ e x p a v g)) のベースコールパフォーマンスを比較する。開示された D e e p R T A _ e x t r a c h a n n e l s (k 1 4 _ 1 , _ o n o f f _ e x p a v g) モデルは、上述の指数加重平均実装から決定された状態情報を使用する。実証されるように、開示された D e e p R T A _ e x t r a c h a n n e l s (k 1 4 _ 1 , _ o n o f f _ e x p a v g) モデルは、様々な二次解析タスク及びメトリックに関して他のものよりも性能が優れている。

【 0 5 0 1 】

条項

開示された技術は、システム、方法、又は製品として実施することができる。実施態様の 1 つ以上の特徴を、塩基実施態様と組み合わせることができる。相互に排他的でない実施態様は、組み合わせ可能であると教示されている。実施態様の 1 つ以上の特徴を他の実施態様と組み合わせることができる。本開示は、これらのオプションのユーザを定期的に通知する。これらの選択肢を繰り返す列挙のいくつかの実施態様からの省略は、前述のセクションで教示されている組み合わせを制限するものとして解釈されるべきではない。これらの記載は、以下の実施のそれぞれに参照することにより本明細書に組み込まれる。

【 0 5 0 2 】

開示された技術、又はその要素の 1 つ以上の実施態様及び条項は、示された方法ステップを実行するためのコンピュータ使用可能なプログラムコードを備えた非一時的コンピュ

10

20

30

40

50

ータ可読記憶媒体を含むコンピュータ製品の形態で実装することができる。更に、開示された技術、又はその要素の1つ以上の実施態様及び条項は、メモリと、メモリに結合され、例示的な方法ステップを実行するように動作する少なくとも1つのプロセッサと、を含む装置の形態で実装することができる。更に、別の態様では、開示された技術又はその要素の1つ以上の実施態様及び条項は、本明細書に記載の方法ステップのうちの1つ以上を実行するための手段の形態で実装ことができ、この手段は、(i) ハードウェアモジュール、(ii) 1つ以上のハードウェアプロセッサ上で実行されるソフトウェアモジュール、又は(iii) ハードウェア及びソフトウェアモジュールの組み合わせ、を含むことができ、(i) ~ (iii) のいずれかが、本明細書に記載の特定の技術を実施し、ソフトウェアモジュールは、コンピュータ可読記憶媒体（又は複数のそのような媒体）に記憶される。

10

【0503】

このセクションで説明される条項は、特徴として組み合わせることができる。簡潔性の目的で、特徴の組み合わせは、個別に列挙されず、特徴の各ベースセットで繰り返されない。読者は、このセクションで説明される条項において識別された特徴が、本出願の他のセクションにおける実施態様として識別された基本特徴のセットと容易に組み合わせることができる方法を理解するであろう。これらの条項は、相互排他的、網羅的、又は制限的であることを意味せず、開示された技術は、これらの条項に限定されず、むしろ、特許請求される技術及びその均等物の範囲内の全ての可能な組み合わせ、修正、及び変形を包含する。

20

【0504】

このセクションで説明される条項の他の実施態様は、このセクションで説明される条項のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される条項の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行してこのセクションで説明される条項のいずれかを実行するように動作可能な1つ以上のプロセッサとを含むシステムを含むことができる。

【0505】

本発明者らは、以下の項目を開示する。

条項セット1ーピクセルごとの状態チャンネル

30

1. システムであって、

配列決定ランの現在の配列決定サイクルについて、複数のピクセル内のそれぞれのピクセルについてのそれぞれの現在の強度値を記憶するメモリであって、

当該メモリが、現在の配列決定サイクルに先行する配列決定ランの1つ以上の以前の配列決定サイクルについて、それぞれのピクセルについてのそれぞれの以前の強度値のシーケンスを記憶する、メモリと、

メモリへのアクセスを有し、それぞれのピクセルについてのそれぞれの現在の状態値を

(i) それぞれの現在の強度値、及び

(ii) それぞれの以前の強度値のシーケンス内のそれぞれの以前の強度値、に依存して生成し、

40

それぞれの現在の状態値をメモリに記憶するように構成された、状態生成器と、

メモリへのアクセスを有し、

(i) それぞれの現在の強度値、及び

(ii) それぞれの現在の状態値を処理することに応答して現在の配列決定サイクルについてのベースコールを生成するように構成されたベースコーラーと、を備える、システム。

2. それぞれのピクセルについてのそれぞれの強度値が、複数のチャンネルについてのチャンネル固有強度値によってそれぞれ特徴付けられる、条項1に記載のシステム。

3. それぞれのピクセルについてのそれぞれの状態値が、複数のチャンネルのチャンネル固

50

有状態値によってそれぞれ特徴付けられる、条項 2 に記載のシステム。

4. それぞれの現在の状態値が、それぞれの現在の強度値を用いてピクセル単位で符号化される、条項 3 に記載のシステム。

5. 複数のチャンネル内のチャンネルのサブセットについてのチャンネル固有状態値が、それぞれの現在の強度値を用いてピクセル単位で符号化される、条項 3 に記載のシステム。

6. チャンネル固有状態値が、それぞれのパンチャネル現在状態値を生成するために、複数のチャンネル内のチャンネルにわたって平均化され、それぞれのパンチャネル現在状態値が、それぞれの現在の強度値を用いてピクセル単位で符号化される、条項 3 に記載のシステム。

7. ピクセル単位符号化が、ピクセル単位の連結を含む、条項 4 に記載のシステム。

8. ピクセル単位符号化が、ピクセル単位の合計を含む、条項 4 に記載のシステム。

9. それぞれの現在の状態値が、

(i) それぞれの以前の強度値、及び

(ii) それぞれの現在の強度値から現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されるそれぞれの現在の平均強度である、条項 1 に記載のシステム。

10. それぞれの現在の平均強度が各々、チャンネル固有の現在の平均強度によって特徴付けられる、条項 9 に記載のシステム。

11. それぞれの現在の状態値が、

(i) それぞれの以前の強度値、及び

(ii) それぞれの現在の強度値から現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されるそれぞれの現在の最大強度である、条項 10 に記載のシステム。

12. それぞれの現在の最大強度が各々、チャンネル固有の現在の最大強度によって特徴付けられる、条項 11 に記載のシステム。

13. それぞれの現在の状態値が、

(i) それぞれの以前の強度値、及び

(ii) それぞれの現在の強度値から現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されるそれぞれの現在の最小強度である、条項 12 に記載のシステム。

14. それぞれの現在の最小強度が各々、チャンネル固有の現在の最小強度によってそれぞれ特徴付けられる、条項 13 に記載のシステム。

15. それぞれの現在の状態値が、

(i) それぞれの以前の強度値、及び

(ii) それぞれの現在の強度値から現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されるそれぞれの現在の指数加重平均強度である、条項 14 に記載のシステム。

16. それぞれの現在の指数加重平均強度が、前の配列決定サイクルよりも最近の配列決定サイクルを加重することに基づいて決定される、条項 15 に記載のシステム。

17. それぞれの現在の指数加重平均強度が各々、チャンネル固有の現在の指数加重平均強度によって特徴付けられる、条項 16 に記載のシステム。

18. それぞれの現在の状態値が、

(i) それぞれの現在の強度値、及び

(ii) それぞれの以前の強度値のローリングサブセットから現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されるそれぞれの現在の移動平均強度である、条項 17 に記載のシステム。

19. それぞれの現在の移動平均強度が、チャンネル固有の現在の移動平均強度によってそれぞれ特徴付けられる、条項 18 に記載のシステム。

20. それぞれの現在の強度値の各々が、それぞれの現在の強度値とそれぞれの現在のアクティブ状態強度及びそれぞれの現在の非アクティブ状態強度との比較に基づいて、アクティブ状態バケット又は非アクティブ状態バケットに帰属される、条項 19 に記載のシステム。

21. それぞれの現在のアクティブ状態強度が、

(i) それぞれの以前の強度値から現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されるそれぞれの現在の大域的 maximum 強度である、条項 20 に記載のシステム。

22. それぞれの現在の大域的 maximum 強度が各々、チャンネル固有の現在の大域的 maximum 強度によってそれぞれ特徴付けられる、条項 21 に記載のシステム。

23. それぞれの現在の非アクティブ状態強度が、

(i) それぞれの以前の強度値から現在の配列決定サイクルにおけるそれぞれのピクセルについて決定されるそれぞれの現在の大域的 maximum 強度である、条項 22 に記載のシステム。

24. それぞれの現在の大域的 minimum 強度が、チャンネル固有の現在の大域的 minimum 強度によってそれぞれ特徴付けられる、条項 23 に記載のシステム。 10

25. それぞれの現在の状態値が、現在の配列決定サイクルにおいてそれぞれのピクセルについて生成されたそれぞれの現在のアクティブ状態値及びそれぞれの現在の非アクティブ状態値を更に含む、条項 24 に記載のシステム。

26. それぞれの現在のアクティブ状態値が各々、チャンネル固有の現在のアクティブ状態値によって特徴付けられる、条項 25 に記載のシステム。

27. それぞれの現在の非アクティブ状態値が各々、チャンネル固有の現在の非アクティブ状態値によって特徴付けられる、条項 26 に記載のシステム。

28. 対象チャンネル内のピクセルについての現在のアクティブ状態値が、

(i) 現在の配列決定サイクルにおいて検出され、アクティブ状態バケットに帰属する対象チャンネル内のピクセルについての現在の強度値、及び 20

(ii) 以前の配列決定サイクルにおいて検出された対象チャンネル内のピクセルについての以前の強度値から現在の配列決定サイクルにおいて決定される、条項 27 に記載のシステム。

29. 現在の強度値が、現在の強度値と、対象チャンネル内の以前の強度値から決定された大域的 maximum 値及び大域的 minimum 値との比較に基づいて、アクティブ状態バケットに帰属する、条項 28 に記載のシステム。

30. 現在の強度値が、現在の強度値と、対象チャンネル内の先行する配列決定サイクルで決定された以前のアクティブ状態値及び以前の非アクティブ状態値との比較に基づいて、アクティブ状態バケットに帰属する、条項 29 に記載のシステム。 30

31. 現在のアクティブ状態値が、現在の強度値及び以前の強度値から決定された指数加重平均である、条項 30 に記載のシステム。

32. 現在のアクティブ状態値が、現在の強度値及び以前の強度値から決定される平均である、条項 31 に記載のシステム。

33. 現在のアクティブ状態値は、現在の強度値及び以前の強度値のローリングサブセットから決定される移動平均である、条項 32 に記載のシステム。

34. 現在のアクティブ状態値が、現在の強度値及び以前の強度値から決定される最小値である、条項 33 に記載のシステム。

35. 現在のアクティブ状態値が、現在の強度値及び以前の強度値から決定される最大値である、条項 34 に記載のシステム。 40

36. 現在のアクティブ状態値が、先行する配列決定サイクルから搬送され、現在の強度値が現在の配列決定サイクルにおける非アクティブ状態バケットに帰属する場合、現在の配列決定サイクルにおいて再決定されない、条項 35 に記載のシステム。

37. 対象チャンネル内のピクセルについての現在の非アクティブ状態値が、

(i) 現在の配列決定サイクルにおいて検出され、非アクティブ状態バケットに帰属する対象チャンネル内のピクセルについての現在の強度値、及び

(ii) 以前の配列決定サイクルにおいて検出された対象チャンネル内のピクセルについての以前の強度値から現在の配列決定サイクルにおいて決定される、条項 36 に記載のシステム。

38. 現在の強度値が、現在の強度値と、対象チャンネル内の以前の強度値から決定され 50

た大域的 maximum 値及び大域的 minimum 値との比較に基づいて、非アクティブ状態バケットに帰属する、条項 37 に記載のシステム。

39. 現在の強度値が、現在の強度値と、対象チャネル内の先行する配列決定サイクルで決定された以前の非アクティブ状態値及び以前の非アクティブ状態値との比較に基づいて、非アクティブ状態バケットに帰属する、条項 38 に記載のシステム。

40. 現在の非アクティブ状態値が、現在の強度値及び以前の強度値から決定された指数加重平均である、条項 39 に記載のシステム。

41. 現在の非アクティブ状態値が、現在の強度値及び以前の強度値から決定される平均である、条項 40 に記載のシステム。

42. 現在の非アクティブ状態値は、現在の強度値及び以前の強度値のローリングサブセットから決定される移動平均である、条項 41 に記載のシステム。

43. 現在の非アクティブ状態値が、現在の強度値及び以前の強度値から決定される最小値である、条項 42 に記載のシステム。

44. 現在の非アクティブ状態値が、現在の強度値及び以前の強度値から決定される最大値である、条項 43 に記載のシステム。

45. 現在の非アクティブ状態値が、先行する配列決定サイクルから搬送され、現在の強度値が現在の配列決定サイクルにおけるアクティブ状態バケットに帰属する場合、現在の配列決定サイクルにおいて再決定されない、条項 44 に記載のシステム。

46. それぞれの現在状態値が、配列決定ランの各配列決定サイクルにおいて反復的に生成される、条項 1 に記載のシステム。

47. メモリが、配列決定ランの次の配列決定サイクルについて、それぞれのピクセルについてのそれぞれの次の強度値のシーケンスを記憶するように更に構成される、条項 1 に記載のシステム。

48. ベースコーラーが、

(i) それぞれの現在の強度値、

(ii) 以前の配列決定サイクルのうちの 1 つ以上についてのそれぞれの以前の強度値のシーケンスにおけるそれぞれの以前の強度値、

(iii) 次の配列決定サイクルのうちの 1 つ以上についてのそれぞれの次の強度値のシーケンスにおけるそれぞれの次の強度値、及び

(iv) それぞれの状態値を処理することに応答して、現在の配列決定サイクルについてのベースコールを生成するように更に構成される、条項 47 に記載のシステム。

49. それぞれの状態値が、それぞれの以前の強度値及びそれぞれの次の強度値を用いてピクセル単位で符号化される、条項 48 に記載のシステム。

50. チャネルが、特定のレーザによる照射と特定の光学フィルタを通した撮像との組み合わせに対応する、条項 3 に記載のシステム。

51. チャネルが、フィルタ波長帯域に対応する、条項 49 に記載のシステム。

52. チャネルが、配列決定サイクルにおける撮像イベントに対応する、条項 51 に記載のシステム。

53. ベースコーラーが、ニューラルネットワークである、条項 1 に記載のシステム。

54. ニューラルネットワークが、畳み込みニューラルネットワークである、条項 53 に記載のシステム。

55. 畳み込みニューラルネットワークは、複数の空間畳み込み層と、複数の時間畳み込み層と、を含む、条項 54 に記載のシステム。

56. それぞれの状態値が、それぞれのピクセルの過去の強度パターンを特徴付けるように構成される、条項 3 に記載のシステム。

57. 過去の強度パターンが、ベースコーラーのベースコール精度の損失を補償するように構成される、条項 56 に記載のシステム。

58. 過去の強度パターンが、 $k-mer$ の塩基をベースコールするときのベースコーラーのベースコール精度の損失を補償するように構成される、条項 57 に記載のシステム。

10

20

30

40

50

59. 現在の配列決定サイクルについてのベースコールが、それぞれの現在の信号値及びそれぞれの以前の信号値が検出される1つ以上のクラスターについてのベースコールを含む、条項1に記載のシステム。

60. それぞれの状態値が、クラスターのそれぞれの信号プロファイルを識別するように構成される、条項59に記載のシステム。

61. ベースコーラーが、既に行われた配列決定ランについてオフラインで生成された配列決定画像を使用して訓練される、条項1に記載のシステム。

62. それぞれのピクセルについてのそれぞれの状態値が、ベースコーラーによるベースコールに先立って、既に行われた配列決定ランの各配列決定サイクルについてオフラインで計算される、条項61に記載のシステム。

63. ベースコーラーが、それぞれの状態値を使用してベースコール精度の損失を補償するように訓練される、条項62に記載のシステム。

64. システムであって、

配列決定ランの現在の配列決定サイクルについて、複数の信号単位内のそれぞれの信号単位についてのそれぞれの現在の信号値を記憶するメモリであって、

当該メモリが、現在の配列決定サイクルに先行する配列決定ランの1つ以上の以前の配列決定サイクルについて、それぞれの信号単位についてのそれぞれの以前の信号値のシーケンスを記憶する、メモリと、

メモリへのアクセスを有し、それぞれの信号単位についてのそれぞれの現在の状態値を

(i) それぞれの現在の信号値、及び

(ii) それぞれの以前の強度値のシーケンス内のそれぞれの以前の強度値、に依存して生成し、

それぞれの現在の状態値をメモリに記憶するように構成された、状態生成器と、

メモリへのアクセスを有し、

(i) それぞれの現在の信号値、及び

(ii) それぞれの現在の状態値を処理することに応じて現在の配列決定サイクルについてのベースコールを生成するように構成されたベースコーラーと、を備える、システム。

65. それぞれの信号ユニットが、画像のそれぞれのピクセルである、条項64に記載のシステム。

66. それぞれの現在の信号値が、現在の配列決定サイクルにおいて検出されたそれぞれのピクセルのそれぞれの現在のピクセル強度である、条項65に記載のシステム。

67. それぞれの以前の信号値が、以前の配列決定サイクルにおいて検出されたそれぞれのピクセルについてのそれぞれの以前のピクセル強度である、条項66に記載のシステム。

68. 現在の配列決定サイクルについてのベースコールが、それぞれの現在の信号値及びそれぞれの以前の信号値が検出される1つ以上のクラスターについてのベースコールを含む、条項67に記載のシステム。

69. それぞれの現在の状態値が、クラスターのそれぞれの信号プロファイルを識別するように構成される、条項68に記載のシステム。

70. メモリが、配列決定ランの次の配列決定サイクルについて、それぞれの信号単位のそれぞれの次の信号値のシーケンスを記憶するように更に構成される、条項69に記載のシステム。

71. ベースコーラーが、

(i) それぞれの現在の信号値、

(ii) 以前の配列決定サイクルのうちの1つ以上についてのそれぞれの以前の信号値のシーケンス内のそれぞれの以前の信号値、

(iii) 次の配列決定サイクルのうちの1つ以上についてのそれぞれの次の信号値のシーケンスにおけるそれぞれの次の信号値、及び

10

20

30

40

50

(i v) それぞれの状態値を処理することに対応して、現在の配列決定サイクルについてのベースコールを生成するように更に構成される、条項 7 0 に記載のシステム。

7 2 . それぞれの現在の信号値が、現在の配列決定サイクルにおいて検出されたそれぞれの現在の電圧値である、条項 6 4 に記載のシステム。

7 3 . それぞれの以前の信号値が、以前の配列決定サイクルにおいて検出されたそれぞれの以前の電圧値である、条項 7 2 に記載のシステム。

7 4 . それぞれの現在の信号値が、現在の配列決定サイクルにおいて検出されたそれぞれの現在の現在の値である、条項 6 4 に記載のシステム。

7 5 . それぞれの以前の信号値が、以前の配列決定サイクルにおいて検出されたそれぞれの以前の現在の値である、条項 7 4 に記載のシステム。

7 6 . システムであって、

中央配列決定サイクル、1 つ以上の右隣接配列決定サイクル、及び 1 つ以上の左隣接配列決定サイクルについての配列決定画像を含む配列決定サイクルの第 1 のスライディングウィンドウを処理するように構成されたニューラルネットワークベースのベースコーラーと、

(i) 配列決定サイクルの第 1 のスライディングウィンドウ内の中央配列決定サイクルに関する配列決定画像内のピクセル強度、及び (i i) 配列決定サイクルの第 1 のスライディングウィンドウ内の中央配列決定サイクルに先行する配列決定サイクルに関する配列決定画像内のピクセル強度に基づいて、配列決定サイクルの第 1 のスライディングウィンドウ内の配列決定画像のピクセルごとの状態を決定するように構成された状態生成器と、

ニューラルネットワークベースのベースコーラーが、配列決定サイクルの第 1 のスライディングウィンドウ内の中央配列決定サイクルをベースコールするために、ピクセルごとの状態とともに、配列決定サイクルの第 1 のスライディングウィンドウ内の配列決定画像を処理するように更に構成される、システム。

7 7 . 配列決定サイクルの第 2 のスライディングウィンドウが、中央配列決定サイクル、1 つ以上の右隣接配列決定サイクル、及び 1 つ以上の左隣接配列決定サイクルについての配列決定画像を含み、

状態生成器が、(i) 配列決定サイクルの第 2 のスライディングウィンドウ内の中央配列決定サイクルについての配列決定画像内のピクセル強度、及び (i i) 配列決定サイクルの第 2 のスライディングウィンドウ内の中央配列決定サイクルに先行する配列決定サイクルについての配列決定画像内のピクセル強度に基づいて、配列決定サイクルの第 2 のスライディングウィンドウ内の配列決定画像のピクセルごとの状態を決定するように構成され、

ニューラルネットワークベースのベースコーラーが、配列決定サイクルの第 2 のスライディングウィンドウ内の配列決定画像をピクセルごとの状態とともに処理して、配列決定サイクルの第 2 のスライディングウィンドウ内の中央配列決定サイクルをベースコールするように更に構成される、条項 7 6 に記載のシステム。

7 8 . 配列決定サイクルの第 3 のスライディングウィンドウが、中央配列決定サイクル、1 つ以上の右隣接配列決定サイクル、及び 1 つ以上の左隣接配列決定サイクルについての配列決定画像を含み、

状態生成器が、(i) 配列決定サイクルの第 3 のスライディングウィンドウ内の中央配列決定サイクルに関する配列決定画像内のピクセル強度、及び (i i) 配列決定サイクルの第 3 のスライディングウィンドウ内の中央配列決定サイクルに先行する配列決定サイクルに関する配列決定画像内のピクセル強度に基づいて、配列決定サイクルの第 3 のスライディングウィンドウ内の配列決定画像のピクセルごとの状態を決定するように構成され、

ニューラルネットワークベースのベースコーラーが、配列決定サイクルの第 3 のスライディングウィンドウ内の配列決定画像をピクセルごとの状態とともに処理して、配列決定サイクルの第 3 のスライディングウィンドウ内の中央配列決定サイクルをベースコールするように更に構成される、条項 7 7 に記載のシステム。

10

20

30

40

50

79. システムであって、

複数の強度パターン検出器を有するニューラルネットワークベースのベースコーラーであって、複数の強度パターン検出器内の強度パターン検出器が、配列決定画像の現在のスライディングウィンドウ内のピクセルの強度値からの強度パターン検出に基づいて対象ベースコール動作を実行するように構成され、強度パターン検出器の配列決定画像の現在のウィンドウへの閉じ込めによるベースコール精度の損失を伴う、ニューラルネットワークベースのベースコーラーと、

(i) 配列決定画像の現在のスライディングウィンドウ内の配列決定画像に先行する配列決定画像内のピクセルの過去の強度値と、(ii) 配列決定画像の現在のスライディングウィンドウ内のターゲット配列決定画像内のピクセルのターゲット強度値とに基づいて強度コンテキストデータを決定するように構成された強度コンテキスト化論理と、

10

強度コンテキストデータをピクセルの強度値に付加して、強度コンテキスト化ピクセルを生成するように構成されたデータフロー論理と、を含み、

ニューラルネットワークベースのベースコーラーが、強度パターン検出器を強度コンテキスト化ピクセルに適用し、対象ベースコール動作のためのベースコールを生成するように更に構成され、強度コンテキスト化ピクセル内の強度コンテキストデータが、ベースコール精度の損失を補償する、システム。

80.

配列決定サイクルの現在のスライディングウィンドウ内の第1の複数の配列決定サイクルに関する現在のウィンドウ配列決定信号データにアクセスすることと、

20

第1の複数の配列決定サイクルにおける配列決定サイクルに先行する第2の複数の配列決定サイクルについての以前の配列決定信号データにアクセスすることと、

現在のウィンドウ配列決定信号データ及び以前の配列決定信号データに基づいて状態信号データを生成することと、

現在のウィンドウ配列決定信号データ及び状態信号データを処理することに応答して、第1の複数の配列決定サイクルにおける少なくとも1つの配列決定サイクルをベースコールすることと、を含む、コンピュータ実装方法。

81.

配列決定ランの現在の配列決定サイクルを、現在の配列決定サイクルに先行する配列決定ランの1つ以上の以前の配列決定サイクルから決定された状態情報に基づいて、ベースコールすることを含む、コンピュータ実装方法。

30

82. システムであって、

配列決定ランの現在の配列決定サイクルについてのクラスターについての現在の配列決定画像、及び現在の配列決定サイクルに先行する配列決定ランの1つ以上の以前の配列決定サイクルについてのクラスターについての以前の配列決定画像を記憶するメモリと、

メモリへのアクセスを有し、現在の配列決定サイクルについて、現在の配列決定画像及び以前の配列決定画像に基づいて配列決定画像内のピクセルについての現在のピクセルごとの状態を生成するように構成されたホストプロセッサと、

構成可能プロセッサであって、メモリへのアクセスを有し、ニューラルネットワークを実行してベースコール分類スコアを生成するように構成された、構成可能プロセッサと、

40

データフロー論理であって、メモリ、ホストプロセッサ、及び構成可能プロセッサへのアクセスを有し、

現在のピクセルごとの状態及び現在の配列決定画像をニューラルネットワークに提供し、ニューラルネットワークに、現在のピクセルごとの状態及び現在の配列決定画像を処理することに応答して、現在の配列決定サイクルについてのクラスターについての現在のベースコール分類スコアを生成させ、

現在のベースコール分類スコアをホストに提供する、データフロー論理と、を備える、システム。

83. メモリが、以前の配列決定サイクルにおいてホストプロセッサによって生成されたピクセルに対する前のピクセルごとの状態を更に記憶する、条項82に記載のシステム

50

。

84. ホストプロセッサが、現在の配列決定画像及び以前のピクセルごとの状態に基づいて、現在のピクセルごとの状態を生成するように更に構成される、条項83に記載のシステム。

【0506】

条項セット2-ウェルごとの状態チャンネル

1. システムであって、

検体の配列決定画像を処理し、空間畳み込み特徴を生成するように構成された空間畳み込みニューラルネットワークと、

空間畳み込み特徴から、検体の中心を含む空間畳み込み特徴のサブセットを選択するように構成されたフィルタリング論理と、

空間畳み込み特徴のサブセットに対する検体ごとの状態を生成するように構成された状態生成論理と、

空間畳み込み特徴のサブセット及び検体ごとの状態を処理し、時間的にコンパクトな出力を生成するように構成された時間的処理論理と、

時間的にコンパクトな出力に基づいてベースコールを生成するように構成されたベースコール論理と、を備える、システム。

2. 検体ごとの状態が、配列決定画像を独立して処理することに対応して線形モデルによって検体ごとに生成される強度変動補正係数である、条項1に記載のシステム。

3. 強度変動補正係数が、検体の強度プロファイル間のスケール変動を補償する増幅係数を含む、条項2に記載のシステム。

4. 強度変動補正係数が、複数のチャンネルに沿った検体の強度プロファイル間のシフト変動を補償するチャンネル固有オフセット係数を含む、条項3に記載のシステム。

5. 強度変動補正係数が、シフト変動を補償する共通オフセット係数を含む、条項4に記載のシステム。

6. 強度変動補正係数が、配列決定ランの現在の配列決定サイクルにおいて決定された現在の強度統計値の分析と、現在の配列決定サイクルに先行する配列決定ランの1つ以上の以前の配列決定サイクルにおいて決定された過去の強度統計値の分析とを組み合わせることに基づいて、検体ごとに生成される、条項5に記載のシステム。

7. 検体の強度プロファイルが、複数のチャンネル内のチャンネルに対するチャンネル固有強度値によって特徴付けられる、条項1に記載のシステム。

8. 検体ごとの状態が、チャンネルについてのチャンネル固有状態値によって特徴付けられる、条項7に記載のシステム。

9. 検体ごとの状態が、中心的特徴ごとに空間畳み込み特徴のサブセットを用いて符号化される、条項8に記載のシステム。

10. チャンネルのサブセットについてのチャンネル固有状態値が、中心的特徴ごとに空間畳み込み特徴のサブセットを用いて符号化される、条項8に記載のシステム。

11. チャンネル固有状態値が、パンチャンネル状態値を生成するためにチャンネルにわたって平均化され、パンチャンネル状態値が、中心的特徴ごとに空間畳み込み特徴のサブセットを用いて符号化される、条項8に記載のシステム。

12. 検体ごとの状態が、中心的特徴ごとに空間畳み込み特徴のサブセットと連結される、条項9に記載のシステム。

13. 検体ごとの状態が、中心的特徴ごとに空間畳み込み特徴のサブセットと合計される、条項9に記載のシステム。

14. 検体ごとの状態が、現在の配列決定サイクル及び以前の配列決定サイクルにおいて空間処理論理によって生成された空間畳み込み特徴のサブセットの特徴値に基づく、条項1に記載のシステム。

15. 検体ごとの状態が、特徴値の平均である、条項14に記載のシステム。

16. 検体ごとの状態が、特徴値のうちの最大のものである、条項14に記載のシステム。

10

20

30

40

50

17. 検体ごとの状態が、特徴値のうちの最小のものである、条項14に記載のシステム。

18. 検体ごとの状態が、特徴値の指数加重平均である、条項14に記載のシステム。

19. 指数加重平均が、前の配列決定サイクルよりも最近の配列決定サイクルを加重することに基づいて決定される、条項18に記載のシステム。

20. 検体ごとの状態が、特徴値の移動平均である、条項14に記載のシステム。

21. 移動平均が、以前の配列決定サイクルからの特徴値のローリングサブセットを使用する、条項20に記載のシステム。

22. 検体ごとの状態が、アクティブ状態値及び非アクティブ状態値を含む、条項14に記載のシステム。

23. アクティブ状態値及び非アクティブ状態値が、チャンネル固有である、条項22に記載のシステム。

24. アクティブ状態値及び非アクティブ状態値が、先行するベースコールから決定される、条項23に記載のシステム。

25. アクティブ状態値及び非アクティブ状態値が、特徴値の大域的最大値及び大域的最小値に基づいて決定される、条項23に記載のシステム。

26. アクティブ状態値及び非アクティブ状態値が、特徴値の指数加重平均に基づいて決定される、条項23に記載のシステム。

27. 入力が、現在の配列決定サイクル、1つ以上の以前の配列決定サイクル、及び1つ以上の次の配列決定サイクルについての配列決定画像のスライディングウィンドウを含む、条項1に記載のシステム。

28. 特徴セットが、現在の配列決定サイクル、以前の配列決定サイクル、及び次の隣接配列決定サイクルについての配列決定画像のスライディングウィンドウ内の配列決定画像のピクセルから抽出される、条項27に記載のシステム。

29. 中心の特徴セットが、現在の配列決定サイクル、以前の配列決定サイクル、及び次の配列決定サイクルについての特徴セットから選別される、条項28に記載のシステム。

30. 検体ごとの状態セットが、現在の配列決定サイクル、以前の配列決定サイクル、及び次の隣接配列決定サイクルの中心の特徴セットについて生成され、それを用いて符号化される、条項29に記載のシステム。

31. 空間処理論理及び時間的処理論理が、ニューラルネットワークである、条項1に記載のシステム。

32. ニューラルネットワークが、畳み込みニューラルネットワークである、条項31に記載のシステム。

33. 検体ごとの状態が、ピクセルの過去の強度パターンを特徴付けるように構成される、条項1に記載のシステム。

34. 過去の強度パターンが、ベースコール精度の損失を補償するように構成される、条項33に記載のシステム。

35. 過去の強度パターンが、k-merの塩基をベースコールするときのベースコール精度の損失を補償するように構成される、条項34に記載のシステム。

36. 検体ごとの状態が、検体の強度プロファイルを識別するように構成される、条項1に記載のシステム。

37. 空間処理論理及び時間的処理論理が、既に行われた配列決定ランについてオフラインで生成された配列決定画像を使用して訓練される、条項1に記載のシステム。

38. 検体ごとの状態が、ベースコールに先立って、既に行われた配列決定ランの各配列決定サイクルについてオフラインで計算される、条項37に記載のシステム。

39. 空間処理論理及び時間的処理論理が、検体ごとの状態を使用して、ベースコール精度の損失を補償するように訓練される、条項38に記載のシステム。

40. システムであって、

検体の強度プロファイルが配列決定画像のピクセルにわたって分散される入力进行处理し

10

20

30

40

50

、検体の強度プロファイルがピクセルから抽出された特徴に集約される空間的にコンパクトな出力を生成するように構成された空間処理論理と、

検体の中心で検出されたピーク強度を特徴付ける特徴から、検体ごとの中心的特徴を選別するように構成された検体集束論理と、

検体ごとの中心的特徴について検体ごとの状態を生成するように構成された状態生成論理と、

検体ごとの中心的特徴及び検体ごとの状態を処理し、時間的にコンパクトな出力を生成するように構成された時間的処理論理と、

時間的にコンパクトな出力に基づいてベースコールを生成するように構成されたベースコール論理と、を備える、システム。

10

【0507】

条項セット3－圧縮特徴状態チャンネルごと

1. システムであって、

クラスターの配列決定画像を処理し、空間畳み込み特徴を生成するように構成された空間畳み込みニューラルネットワークと、

空間畳み込み特徴から、クラスターの中心を含む空間畳み込み特徴のサブセットを選択するように構成されたフィルタリング論理と、

空間畳み込み特徴のサブセットを圧縮特徴のセットに圧縮するように構成された圧縮論理であって、空間畳み込み特徴のサブセットがM個のチャンネルを有し、圧縮特徴のセットがN個のチャンネルを有し、 $M > N$ である、圧縮論理と、

20

圧縮特徴のセット内の圧縮特徴についての状態情報にアクセスし、状態情報を圧縮特徴に付加してステートフルな圧縮特徴のセットを生成するように構成されたコンテキスト化論理と、

ステートフルな圧縮特徴のセットを処理し、時間畳み込みステートフル特徴を生成するように構成された時間畳み込みニューラルネットワークと、

時間畳み込みステートフル特徴に基づいてクラスターについてのベースコールを生成するように構成されたベースコール論理と、を備える、システム。

2. チャンネルごとの状態が、配列決定画像を独立して処理することに応答して線形モデルによって検体ごとに生成される強度変動補正係数である、条項1に記載のシステム。

3. 強度変動補正係数が、検体の強度プロファイル間のスケール変動を補償する増幅係数を含む、条項2に記載のシステム。

30

4. 強度変動補正係数が、複数のチャンネルに沿った検体の強度プロファイル間のシフト変動を補償するチャンネル固有オフセット係数を含む、条項3に記載のシステム。

5. 強度変動補正係数が、シフト変動を補償する共通オフセット係数を含む、条項4に記載のシステム。

6. 強度変動補正係数が、配列決定ランの現在の配列決定サイクルにおいて決定された現在の強度統計値の分析と、現在の配列決定サイクルに先行する配列決定ランの1つ以上の以前の配列決定サイクルにおいて決定された過去の強度統計値の分析とを組み合わせることに基づいて、検体ごとに生成される、条項5に記載のシステム。

7. 検体の強度プロファイルが、複数のチャンネル内のチャンネルに対するチャンネル固有強度値によって特徴付けられる、条項1に記載のシステム。

40

8. チャンネルごとの状態が、チャンネルについてのチャンネル固有状態値によって特徴付けられる、条項7に記載のシステム。

9. チャンネルごとの状態が、検体ごとのチャンネル固有ビットを用いてビットごとに符号化される、条項8に記載のシステム。

10. チャンネルのサブセットに関するチャンネル固有状態値が、検体ごとのチャンネル固有ビットを用いてビットごとに符号化される、条項8に記載のシステム。

11. チャンネル固有状態値が、チャンネルにわたって平均化されて、パンチャンネル状態値を生成し、パンチャンネル状態値が、検体ごとのチャンネル固有ビットを用いてビットごとに符号化される、条項8に記載のシステム。

50

12. チャンネルごとの状態が、検体ごとのチャンネル固有ビットとビットごとに連結される、条項9に記載のシステム。

13. チャンネルごとの状態が、検体ごとのチャンネル固有ビットとビットごとに合計される、条項9に記載のシステム。

14. チャンネルごとの状態が、現在の配列決定サイクル及び以前の配列決定サイクルにおいて空間処理論理によって生成された検体ごとのチャンネル固有ビットの圧縮値に基づく、条項1に記載のシステム。

15. チャンネルごとの状態が、圧縮値の平均である、条項14に記載のシステム。

16. チャンネルごとの状態が、圧縮値のうちの最大のものである、条項14に記載のシステム。

17. チャンネルごとの状態が、圧縮値のうちの最小のものである、条項14に記載のシステム。

18. チャンネルごとの状態が、圧縮値の指数加重平均である、条項14に記載のシステム。

19. 指数加重平均が、前の配列決定サイクルよりも最近の配列決定サイクルを加重することに基づいて決定される、条項18に記載のシステム。

20. チャンネルごとの状態が、圧縮値の移動平均である、条項14に記載のシステム。

21. 移動平均が、以前の配列決定サイクルからの圧縮値のローリングサブセットを使用する、条項20に記載のシステム。

22. チャンネルごとの状態が、アクティブ状態値及び非アクティブ状態値を含む、条項14に記載のシステム。

23. アクティブ状態値及び非アクティブ状態値が、チャンネル固有である、条項22に記載のシステム。

24. アクティブ状態値及び非アクティブ状態値が、先行するベースコールから決定される、条項23に記載のシステム。

25. アクティブ状態値及び非アクティブ状態値が、圧縮値の大域的最大値及び大域的最小値に基づいて決定される、条項23に記載のシステム。

26. アクティブ状態値及び非アクティブ状態値が、圧縮値の指数加重平均に基づいて決定される、条項23に記載のシステム。

27. 入力が、現在の配列決定サイクル、1つ以上の以前の配列決定サイクル、及び1つ以上の次の配列決定サイクルについての配列決定画像のスライディングウィンドウを含む、条項1に記載のシステム。

28. 特徴セットが、現在の配列決定サイクル、以前の配列決定サイクル、及び次の隣接配列決定サイクルについての配列決定画像のスライディングウィンドウ内の配列決定画像のチャンネル固有ピクセルから抽出され、検体ごとの中心的特徴セットが、現在の配列決定サイクル、以前の配列決定サイクル、及び次の配列決定サイクルについての特徴セットから選別される、条項27に記載のシステム。

29. チャンネル固有ビットセットが、現在の配列決定サイクル、以前の配列決定サイクル、及び次の配列決定サイクルについての検体ごとの中心的特徴セットから圧縮される、条項28に記載のシステム。

30. チャンネルごとの状態セットが、現在の配列決定サイクル、以前の配列決定サイクル、及び次の隣接配列決定サイクルのチャンネル固有ビットセットについて生成され、それを用いて符号化される、条項29に記載のシステム。

31. 空間処理論理及び時間的処理論理が、ニューラルネットワークである、条項1に記載のシステム。

32. ニューラルネットワークが、畳み込みニューラルネットワークである、条項31に記載のシステム。

33. チャンネルごとの状態が、チャンネル固有ピクセルの過去の強度パターンを特徴付けるように構成される、条項1に記載のシステム。

34. 過去の強度パターンが、ベースコール精度の損失を補償するように構成される、

10

20

30

40

50

条項 3 3 に記載のシステム。

3 5 . 過去の強度パターンが、k - m e r の塩基をベースコールするときのベースコール精度の損失を補償するように構成される、条項 3 4 に記載のシステム。

3 6 . 検体ごとの状態が、検体の強度プロファイルを識別するように構成される、条項 1 に記載のシステム。

3 7 . 空間処理論理及び時間的処理論理が、既に行われた配列決定ランについてオフラインで生成された配列決定画像を使用して訓練される、条項 1 に記載のシステム。

3 8 . チャンネルごとの状態が、ベースコールに先立って、既に行われた配列決定ランの各配列決定サイクルについてオフラインで計算される、条項 3 7 に記載のシステム。

3 9 . 空間処理論理及び時間的処理論理が、チャンネルごとの状態を使用してベースコール精度の損失を補償するように訓練される、条項 3 8 に記載のシステム。

4 0 . システムであって、

検体の強度プロファイルが配列決定画像のチャンネル固有ピクセルにわたって分散される入力を処理し、検体の強度プロファイルがチャンネル固有ピクセルから抽出された特徴に集約される空間的にコンパクトな出力を生成するように構成された空間処理論理と、

検体の中心で検出されたピーク強度を特徴付ける特徴から、検体ごとの中心的特徴を識別するように構成された検体集束論理と、

検体ごとの中心的特徴を検体ごとのチャンネル固有ビットに蒸留するように構成された圧縮論理であって、検体ごとの中心的特徴が M 個の特徴チャンネルを有し、検体ごとのチャンネル固有ビットが N 個のビットチャンネルを有し、 $M > N$ である、圧縮論理と、

検体ごとのチャンネル固有ビットについてチャンネルごとの状態を生成するように構成された状態生成論理と、

検体ごとのチャンネル固有ビット及びチャンネルごとの状態を処理し、時間的にコンパクトな出力を生成するように構成された時間的処理論理と、

時間的にコンパクトな出力に基づいてベースコールを生成するように構成されたベースコール論理と、を備える、システム。

【 0 5 0 8 】

条項セット 4 - ウェルごとの状態からピクセルごとの状態への変換

1 . システムであって、

検体について配列決定画像を処理することに対応して検体ごとの状態を生成し、検体ごとの状態を疎な表現で符号化するように構成された、検体ごとの状態生成論理と、

疎な表現を処理することに対応して、検体ごとの状態を配列決定画像内のピクセルのピクセルごとの状態に変換し、ピクセルごとの状態を高密度表現に符号化するように構成された、ピクセルごとの状態生成論理と、

密表現において符号化されたピクセルごとの状態で補足された配列決定画像内のピクセルを処理することに対応して、ベースコールを生成するように構成された、ベースコーラーと、を備える、システム。

2 . ピクセルごとの状態生成論理が、転置畳み込みを使用する、条項 1 に記載のシステム。

3 . 転置畳み込みが、疎表現で符号化された検体ごとの状態を密表現で符号化されたピクセルごとの状態に変換するように訓練される、条項 2 に記載のシステム。

4 . ベースコーラーが、空間畳み込み層及び時間畳み込み層を含む、条項 2 に記載のシステム。

5 . 密表現で符号化されたピクセル単位の状態が、ベースコーラーの第 1 の空間畳み込み層による処理のために、配列決定画像内のピクセルとともにピクセル単位で符号化される、条項 4 に記載のシステム。

6 . 密表現で符号化されたピクセルごとの状態が、ベースコーラーのその後の空間畳み込み層への入力として供給される、条項 5 に記載のシステム。

7 . 検体ごとの状態が、配列決定画像を処理することに対応して線形モデルによって検体ごとに生成される強度変動補正係数である、条項 1 に記載のシステム。

８．強度変動補正係数が、検体の強度プロファイル間のスケール変動を補償する増幅係数を含む、条項７に記載のシステム。

９．強度変動補正係数が、複数のチャンネルに沿った検体の強度プロファイル間のシフト変動を補償するチャンネル固有オフセット係数を含む、条項８に記載のシステム。

１０．強度変動補正係数が、シフト変動を補償する共通オフセット係数を含む、条項９に記載のシステム。

１１．強度変動補正係数が、配列決定ランの現在の配列決定サイクルにおいて決定された現在の強度統計値の分析と、現在の配列決定サイクルに先行する配列決定ランの１つ以上の以前の配列決定サイクルにおいて決定された過去の強度統計値の分析とを組み合わせることに基づいて、検体ごとに生成される、条項１０に記載のシステム。

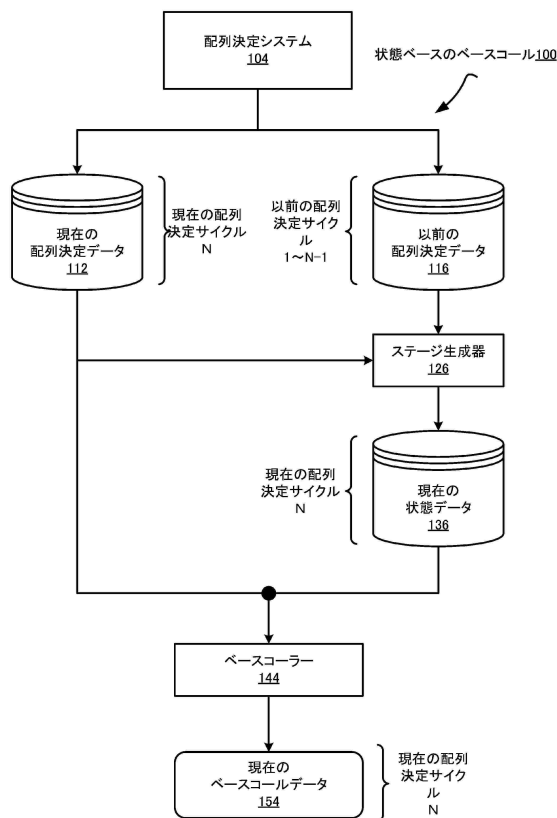
１２．検体の強度プロファイルが、複数のチャンネル内のチャンネルに対するチャンネル固有強度値によって特徴付けられる、条項１に記載のシステム。

１３．検体ごとの状態が、チャンネルについてのチャンネル固有状態値によって特徴付けられる、条項１２に記載のシステム。

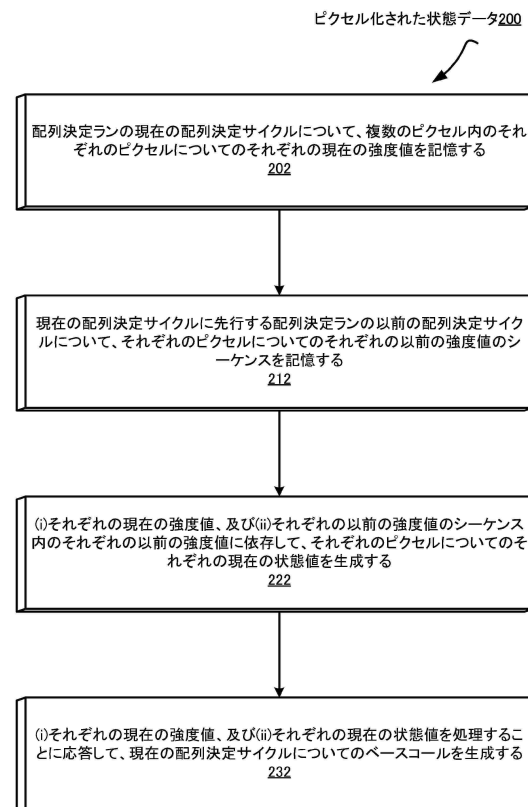
１４．ピクセルごとの状態生成論理が、補間を使用する、条項１に記載のシステム。

【図面】

【図１】



【図２】



10

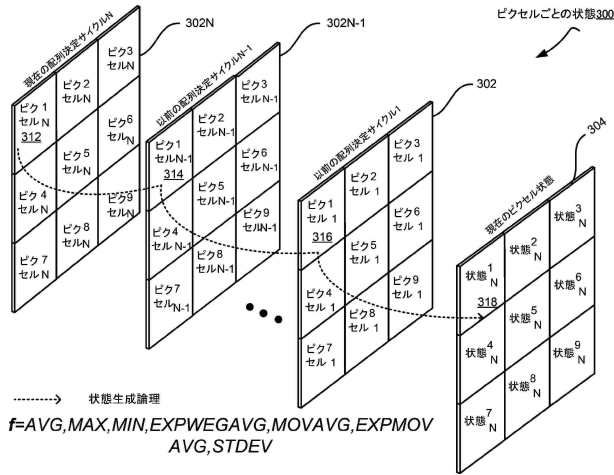
20

30

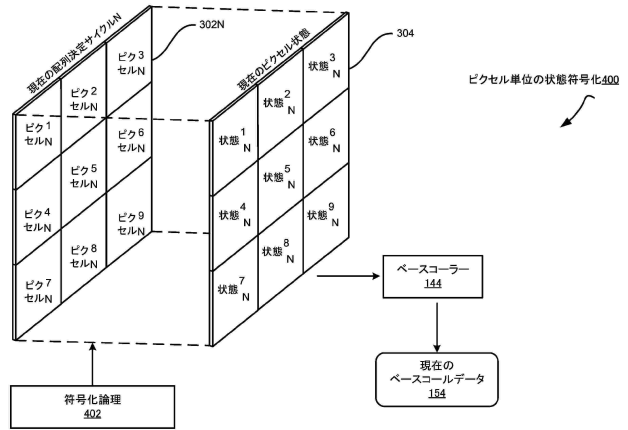
40

50

【図 3】



【図 4】



【図 5】

チャンネルごとの状態500

ピクセルP	チャンネル1	チャンネル2
サイクル1	$\begin{matrix} 1 \\ v_1 \end{matrix} \quad 512$	$\begin{matrix} 2 \\ v_1 \end{matrix} \quad 522$
サイクル2	$\begin{matrix} 1 \\ v_2 \end{matrix} \quad 513$	$\begin{matrix} 2 \\ v_2 \end{matrix} \quad 523$
サイクル3	$\begin{matrix} 1 \\ v_3 \end{matrix} \quad 514$	$\begin{matrix} 2 \\ v_3 \end{matrix} \quad 524$
サイクル4	$\begin{matrix} 1 \\ v_4 \end{matrix} \quad 515$	$\begin{matrix} 2 \\ v_4 \end{matrix} \quad 525$
サイクル5	$\begin{matrix} 1 \\ v_5 \end{matrix} \quad 516$	$\begin{matrix} 2 \\ v_5 \end{matrix} \quad 526$
サイクル5 状態	$\{ \begin{matrix} 1 \\ v_1 \end{matrix} \quad \begin{matrix} 1 \\ v_2 \end{matrix} \quad \begin{matrix} 1 \\ v_3 \end{matrix} \quad \begin{matrix} 1 \\ v_4 \end{matrix} \quad \begin{matrix} 1 \\ v_5 \end{matrix} \}$	$\{ \begin{matrix} 2 \\ v_1 \end{matrix} \quad \begin{matrix} 2 \\ v_2 \end{matrix} \quad \begin{matrix} 2 \\ v_3 \end{matrix} \quad \begin{matrix} 2 \\ v_4 \end{matrix} \quad \begin{matrix} 2 \\ v_5 \end{matrix} \}$

532 536

$$f = \text{AVG, MAX, MIN, EXPWEGAVG, MOVAVG, EXPMOVAVG, STDEV}$$

【図 6】

チャンネルごとのMIN状態600

現在の配列決定サイクルN	チャンネル1 強度値602	チャンネル2 強度値603	チャンネル1 MIN 状態値604	チャンネル2 MIN 状態値605
ピクセル801				
(1, 1)	x_N^1	y_N^1	$\text{MIN}\{x_N^1 \dots x_N^1\}$	$\text{MIN}\{y_N^1 \dots y_N^1\}$
(1, 2)	x_N^2	y_N^2	$\text{MIN}\{x_N^2 \dots x_N^2\}$	$\text{MIN}\{y_N^2 \dots y_N^2\}$
(1, 3)	x_N^3	y_N^3	$\text{MIN}\{x_N^3 \dots x_N^3\}$	$\text{MIN}\{y_N^3 \dots y_N^3\}$
(1, 4)	x_N^4	y_N^4	$\text{MIN}\{x_N^4 \dots x_N^4\}$	$\text{MIN}\{y_N^4 \dots y_N^4\}$
(1, 5)	x_N^5	y_N^5	$\text{MIN}\{x_N^5 \dots x_N^5\}$	$\text{MIN}\{y_N^5 \dots y_N^5\}$
...	...	...	...	...
(i, j)	x_N^k	y_N^k	$\text{MIN}\{x_N^k \dots x_N^k\}$	$\text{MIN}\{y_N^k \dots y_N^k\}$

10

20

30

40

50

【図 7】

現在の配列決定サイクルN

パンチャネルMIN状態700

ピクセル601	チャネル1 強度値602	チャネル2 強度値603	パンチャネルAVG MIN状態値 708	
(1, 1)	x_N^1	y_N^1		
(1, 2)	x_N^2	y_N^2		
(1, 3)	x_N^3	y_N^3	図6の列604及び605 の ピクセル単位の平均	
(1, 4)	x_N^4	y_N^4		
(1, 5)	x_N^5	y_N^5		
...	...	...	...	...
(i,j)	x_N^k	y_N^k		

【図 8】

現在の配列決定サイクルN

チャネルごとのMAX状態800

ピクセル601	チャネル1 強度値602	チャネル2 強度値603	チャネル1 MAX状態値 804	チャネル2 MAX状態値 805
(1, 1)	x_N^1	y_N^1	$\text{MAX}\{x_N^1 \dots x_1^1\}$	$\text{MAX}\{y_N^1 \dots y_1^1\}$
(1, 2)	x_N^2	y_N^2	$\text{MAX}\{x_N^2 \dots x_1^2\}$	$\text{MAX}\{y_N^2 \dots y_1^2\}$
(1, 3)	x_N^3	y_N^3	$\text{MAX}\{x_N^3 \dots x_1^3\}$	$\text{MAX}\{y_N^3 \dots y_1^3\}$
(1, 4)	x_N^4	y_N^4	$\text{MAX}\{x_N^4 \dots x_1^4\}$	$\text{MAX}\{y_N^4 \dots y_1^4\}$
(1, 5)	x_N^5	y_N^5	$\text{MAX}\{x_N^5 \dots x_1^5\}$	$\text{MAX}\{y_N^5 \dots y_1^5\}$
...	...	...	...	...
(i,j)	x_N^k	y_N^k	$\text{MAX}\{x_N^k \dots x_1^k\}$	$\text{MAX}\{y_N^k \dots y_1^k\}$

10

【図 9】

現在の配列決定サイクルN

パンチャネルAVG MAX状態900

ピクセル601	チャネル1 強度値602	チャネル2 強度値603	パンチャネルAVG MAX状態値 908	
(1, 1)	x_N^1	y_N^1		
(1, 2)	x_N^2	y_N^2		
(1, 3)	x_N^3	y_N^3	図6の列604及び605 の ピクセル単位の平均	
(1, 4)	x_N^4	y_N^4		
(1, 5)	x_N^5	y_N^5		
...	...	...	...	...
(i,j)	x_N^k	y_N^k		

【図 10】

現在の配列決定サイクルN

チャネルごとのAVG状態1000

ピクセル601	チャネル1 強度値602	チャネル2 強度値603	チャネル1 AVG状態値 1004	チャネル2 AVG状態値 1005
(1, 1)	x_N^1	y_N^1	$\text{AVG}\{x_N^1 \dots x_1^1\}$	$\text{AVG}\{y_N^1 \dots y_1^1\}$
(1, 2)	x_N^2	y_N^2	$\text{AVG}\{x_N^2 \dots x_1^2\}$	$\text{AVG}\{y_N^2 \dots y_1^2\}$
(1, 3)	x_N^3	y_N^3	$\text{AVG}\{x_N^3 \dots x_1^3\}$	$\text{AVG}\{y_N^3 \dots y_1^3\}$
(1, 4)	x_N^4	y_N^4	$\text{AVG}\{x_N^4 \dots x_1^4\}$	$\text{AVG}\{y_N^4 \dots y_1^4\}$
(1, 5)	x_N^5	y_N^5	$\text{AVG}\{x_N^5 \dots x_1^5\}$	$\text{AVG}\{y_N^5 \dots y_1^5\}$
...	...	...	...	...
(i,j)	x_N^k	y_N^k	$\text{AVG}\{x_N^k \dots x_1^k\}$	$\text{AVG}\{y_N^k \dots y_1^k\}$

20

【図 11】

現在の配列決定サイクルN

パンチャネルAVG AVG状態1100

ピクセル601	チャネル1 強度値602	チャネル2 強度値603	パンチャネルAVG AVG状態値 1108	
(1, 1)	x_N^1	y_N^1		
(1, 2)	x_N^2	y_N^2		
(1, 3)	x_N^3	y_N^3	図10の列1004及び1005 の ピクセル単位の平均	
(1, 4)	x_N^4	y_N^4		
(1, 5)	x_N^5	y_N^5		
...	...	...	...	...
(i,j)	x_N^k	y_N^k		

【図 12】

現在の配列決定サイクルN

チャネルごとのMIN MAX状態1200

ピクセル 601	チャネル1 強度値602	チャネル2 強度値603	チャネル1 MIN 状態値604	チャネル2 MIN 状態値605	チャネル1 MAX 状態値604	チャネル2 MAX 状態値605
(1, 1)	x_N^1	y_N^1				
(1, 2)	x_N^2	y_N^2				
(1, 3)	x_N^3	y_N^3	600 (図6)		800 (図8)	
(1, 4)	x_N^4	y_N^4				
(1, 5)	x_N^5	y_N^5				
...	...	...	...	...	...	...
(i,j)	x_N^k	y_N^k				

30

40

50

【図 13】

現在の配列決定サイクルN

チャンネルごとのMIN AVG状態1300

ピクセル 601	チャンネル1 強度値602	チャンネル2 強度値603	チャンネル1 MIN 状態値604	チャンネル2 MIN 状態値605	チャンネル1 AVG 状態値1004	チャンネル2 AVG 状態値1005
(1, 1)	x_N^1	y_N^1				
(1, 2)	x_N^2	y_N^2				
(1, 3)	x_N^3	y_N^3				
(1, 4)	x_N^4	y_N^4	600 図6		1000 図10	
(1, 5)	x_N^5	y_N^5				
...	...	...	...	...	...	...
(i,j)	x_N^k	y_N^k				

【図 14】

現在の配列決定サイクルN

チャンネルごとのMAX AVG状態1400

ピクセル 601	チャンネル1 強度値602	チャンネル2 強度値603	チャンネル1 MAX 状態値804	チャンネル2 MAX 状態値805	チャンネル1 AVG 状態値1004	チャンネル2 AVG 状態値1005
(1, 1)	x_N^1	y_N^1				
(1, 2)	x_N^2	y_N^2				
(1, 3)	x_N^3	y_N^3				
(1, 4)	x_N^4	y_N^4	800 図8		1000 図10	
(1, 5)	x_N^5	y_N^5				
...	...	...	...	...	...	...
(i,j)	x_N^k	y_N^k				

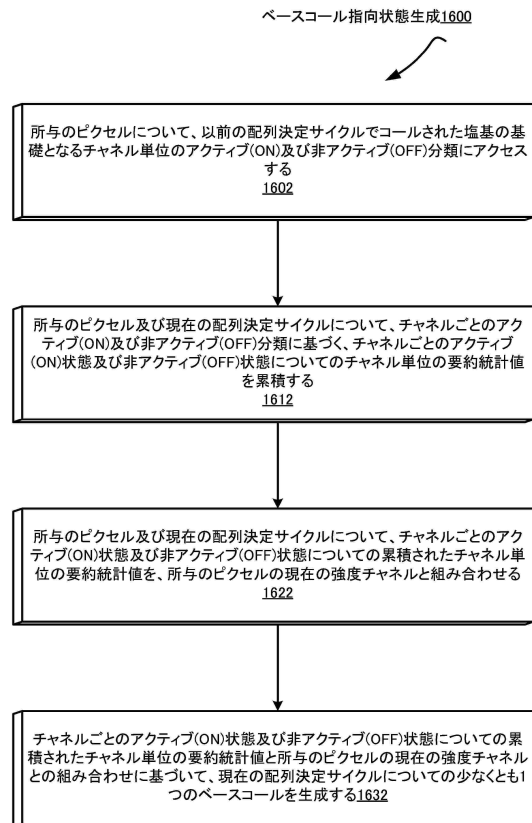
【図 15】

現在の配列決定サイクルN

チャンネルごとのMIN MAX AVG状態1500

ピクセル 601	チャンネル1 強度値602	チャンネル2 強度値603	チャンネル1 MIN 状態値604	チャンネル2 MIN 状態値605	チャンネル1 MAX 状態値804	チャンネル2 MAX 状態値805	チャンネル1 AVG 状態値1004	チャンネル2 AVG 状態値1005
(1, 1)	x_N^1	y_N^1						
(1, 2)	x_N^2	y_N^2						
(1, 3)	x_N^3	y_N^3						
(1, 4)	x_N^4	y_N^4	600 図6		800 図8		1000 図10	
(1, 5)	x_N^5	y_N^5						
...	...	...	...	...	...	...	...	...
(i,j)	x_N^k	y_N^k						

【図 16】



10

20

30

40

50

【図 17 A】

ベースコール指向ON及びOFF状態生成1700		1754		1756	
ピクセルP		チャンネル1		チャンネル2	
サイクル1	ベースコールG	OFF x_1^P		OFF y_1^P	
サイクル2	ベースコールC	OFF x_2^P		ON y_2^P	
サイクル3	ベースコールT	ON x_3^P		ON y_3^P	
サイクル4	ベースコールA	ON x_4^P		OFF y_4^P	
サイクル5	ベースコールC	OFF x_5^P		ON y_5^P	
サイクル6	ベースコールG	OFF x_6^P		OFF y_6^P	
サイクル7	ベースコールT	ON x_7^P		ON y_7^P	
サイクル8	ベースコールA	ON x_8^P		OFF y_8^P	
サイクル9	ベースコールC	OFF x_9^P		ON y_9^P	
サイクル10	ベースコールT	ON x_{10}^P		ON y_{10}^P	

【図 17 B】

サイクル11についてのチャンネルごとの状態計算		$f=AVG,MAX,MIN,EXPWEGAVG,MOVAVG,EXPMOVAVG,STDEV$	
チャンネル1 非アクティブ (OFF)状態	1702	$f\{x_1^P, x_2^P, x_5^P, x_6^P, x_9^P\}$	
チャンネル1 アクティブ (ON)状態	1712	$f\{x_3^P, x_4^P, x_7^P, x_8^P, x_{10}^P\}$	
チャンネル2 非アクティブ (OFF)状態	1722	$f\{y_1^P, y_4^P, y_6^P, y_8^P\}$	
チャンネル2 アクティブ (ON)状態	1732	$f\{y_2^P, y_3^P, y_5^P, y_7^P, y_9^P, y_{10}^P\}$	

【図 17 C】

サイクル 11	チャンネル1	チャンネル2	1702	1712	1722	1732
	x_{11}^P	y_{11}^P				

【図 18 A】

OFF状態更新1800		1854		1856	
ピクセルP		チャンネル1		チャンネル2	
サイクル11	ベースコールG	OFF x_{11}^P		OFF y_{11}^P	

【図 18 B】

サイクル12についてのチャンネルごとの状態計算		$f=AVG,MAX,MIN,EXPWEGAVG,MOVAVG,EXPMOVAVG,STDEV$	
チャンネル1 非アクティブ(OFF)状態	1802	$f\{1702, x_{11}^P\}$	
チャンネル1 アクティブ(ON)状態	1812	1712 (変化なし)	
チャンネル2 非アクティブ(OFF)状態	1822	$f\{1722, y_{11}^P\}$	
チャンネル2 アクティブ(ON)状態	1832	1732 (変化なし)	

【図 18 C】

サイクル 12	チャンネル1	チャンネル2	1802	1812	1822	1832
	x_{12}^P	y_{12}^P				

10

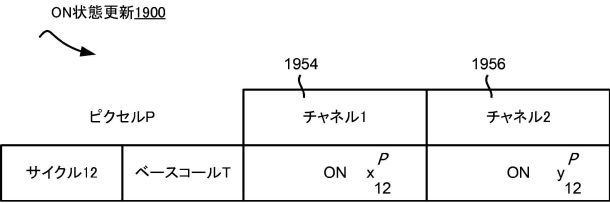
20

30

40

50

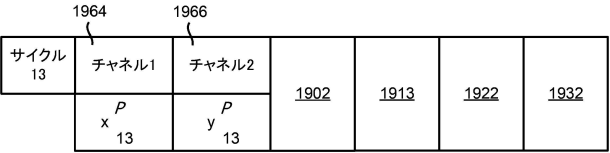
【図 19 A】



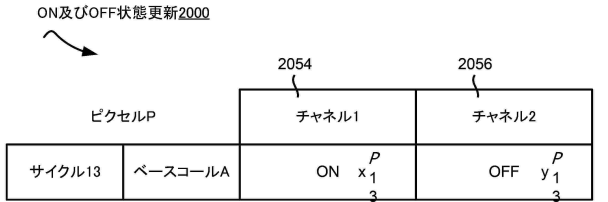
【図 19 B】

サイクル13についてのチャンネルごとの状態計算		$f=AVG,MAX,MIN,EXPWEGAVG,MOVAVG,EXPMOVAVG,STDEV$
チャンネル1 非アクティブ(OFF)状態	1902	1802 (変化なし)
チャンネル1 アクティブ(ON)状態	1912	$f\{1812, x_{12}^P\}$
チャンネル2 非アクティブ(OFF)状態	1922	1822 (変化なし)
チャンネル2 アクティブ(ON)状態	1932	$f\{1832, y_{12}^P\}$

【図 19 C】



【図 20 A】



【図 20 B】

サイクル14についてのチャンネルごとの状態計算		$f=AVG,MAX,MIN,EXPWEGAVG,MOVAVG,EXPMOVAVG,STDEV$
チャンネル1 非アクティブ(OFF)状態	2002	1902 (変化なし)
チャンネル1 アクティブ(ON)状態	2012	$f\{1912, x_{13}^P\}$
チャンネル2 非アクティブ(OFF)状態	2022	$f\{1922, y_{13}^P\}$
チャンネル2 アクティブ(ON)状態	2032	1932 (変化なし)

【図 20 C】



10

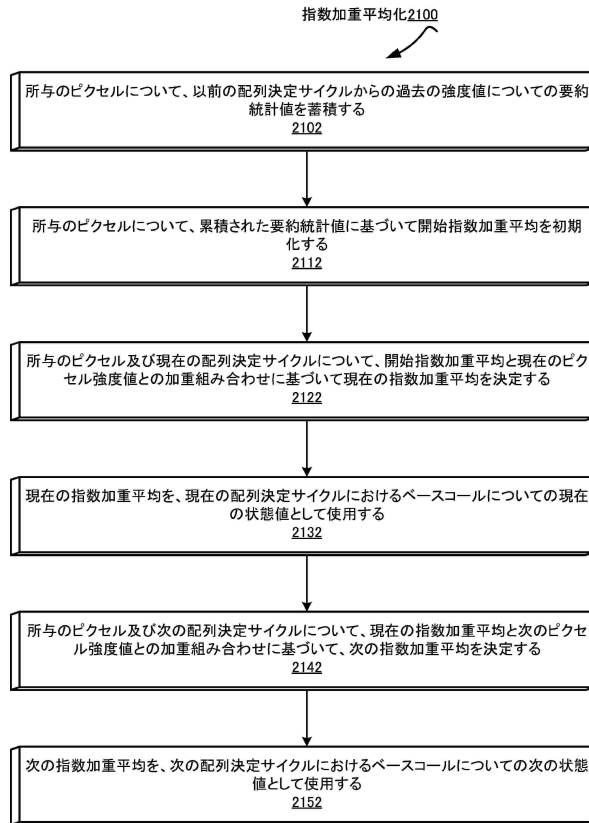
20

30

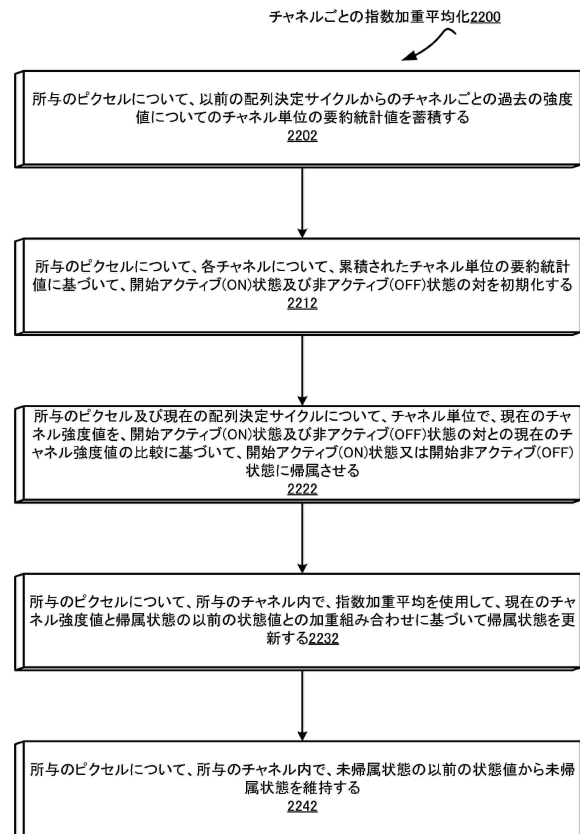
40

50

【図 2 1】



【図 2 2】



10

20

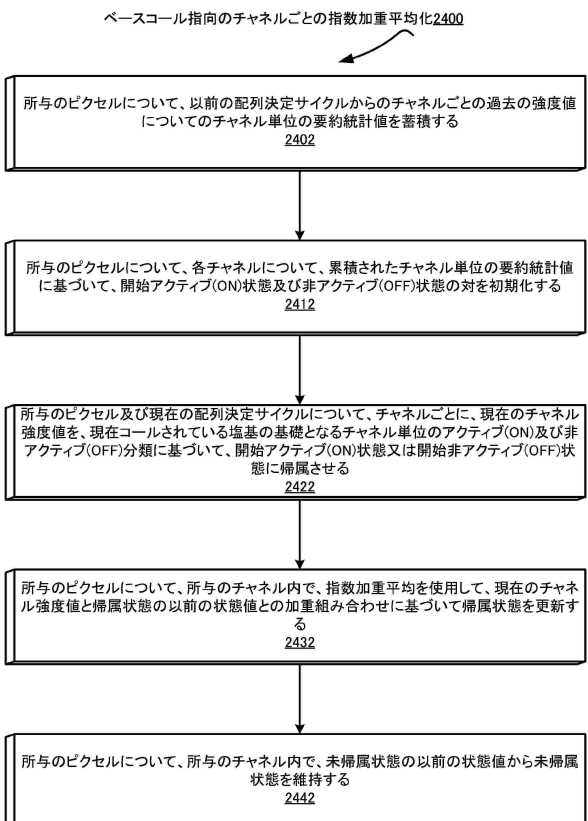
【図 2 3】

ピクセルP

大域的MAX及びMIN並びに指数加重平均を使用したチャンネルごとのON及びOFF状態決定2300

	チャンネル1	チャンネル2
サイクル1	171.4	619.2
サイクル2	-150.2	5.9
サイクル3	1345.5	821.3
サイクル4	933.3	540.1
サイクル5	-63.8	-16.0
サイクル6	大域的Max=1345.5 大域的Min=-150.2 2322 1788(アクティブ状態への帰属)	大域的Max=821.3 大域的Min=-16.0 2326 -90.1(アクティブ状態への帰属)
2332	アクティブ状態 指数加重平均 [1345.5*0.12]+[1768*0.88]	非アクティブ状態 2314 大域的Min=-150.2 (変化なし)
2342	EWA6=1717.3	EWA6=-81.208
サイクル7	2352 5.9(アクティブ状態への帰属)	2356 151.8(アクティブ状態への帰属)
2342	アクティブ状態 EWA6=1717.3 (変化なし)	非アクティブ状態 2316 大域的Max=821.3 (変化なし)
2372	EWA7=1717.3	EWA7=-81.208
2364	指数加重平均 [-150.2*0.12]+[5.9*0.88]	指数加重平均 [821.3*0.12]+[151.8*0.88]
2374	EWA7=-12832	2376

【図 2 4】



30

40

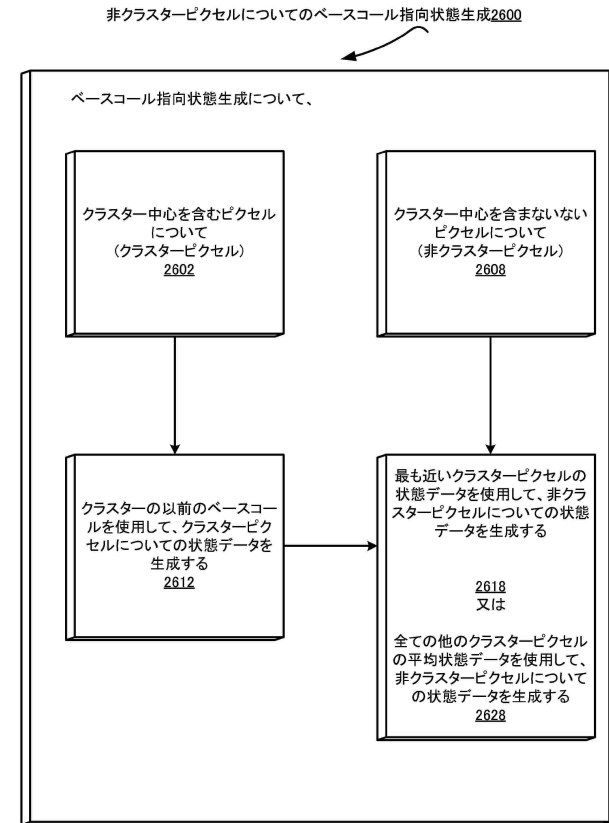
50

【図 25】

ピクセルP ← 指数加重平均を使用するベースコール指向のチャネルごとのON及びOFF状態決定2300

	チャネル1		チャネル2
サイクル1	171.4		619.2
サイクル2	-150.2		5.9
サイクル3	1345.5	2374	821.3
サイクル4	933.3		540.1
サイクル5	-63.8		-16.0
サイクル6	2322 (1768(アクティブ状態への移行))	2326 (90.1(アクティブ状態への移行))	
ベースコールA	2504 ON		2506 OFF
2502	アクティブ状態	非アクティブ状態	アクティブ状態
2332	指数加重平均 [1345.5*0.12]+[11768*0.88]	2344 大域的min=-150.2 (変化なし)	2346 大域的Max=821.3 (変化なし)
2342	EWA6=1717.3	EWA6=-150.2	EWA6=821.3
2348	EWA6=81.208		
サイクル7	2352 (5.5(アクティブ状態への移行))	2356 (151.5(アクティブ状態への移行))	
ベースコールA	2514 OFF		2516 ON
2512	アクティブ状態	非アクティブ状態	アクティブ状態
2342	EWA6=1717.3 (変化なし)	2344 指数加重平均 [-150.2*0.12]+[5.9*0.88]	2346 指数加重平均 [821.3*0.12]+[151.5*0.88]
2372	EWA7=1717.3	EWA7=-12832	EWA7=231.876
2378	EWA7=81.208		

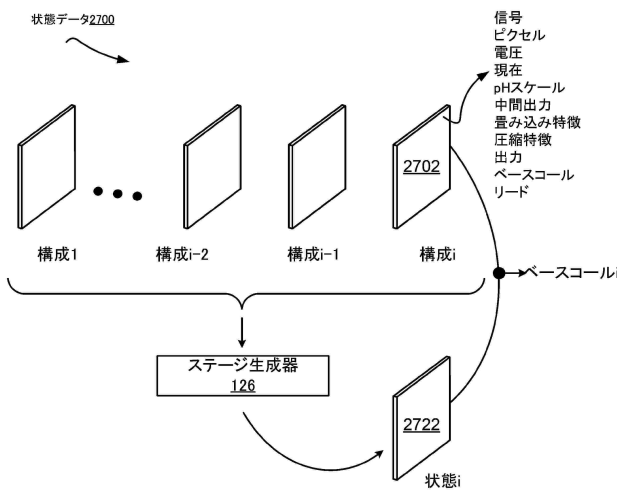
【図 26】



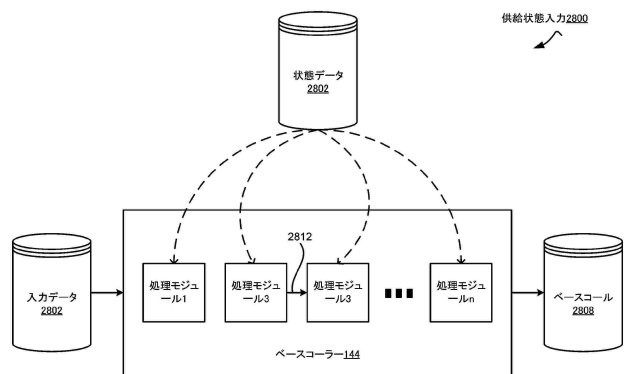
10

20

【図 27】



【図 28】

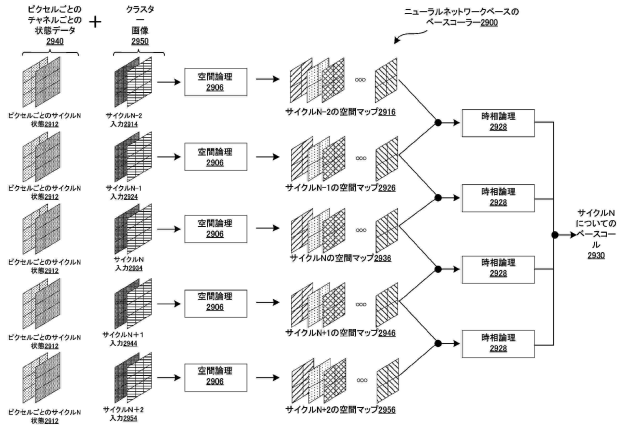


30

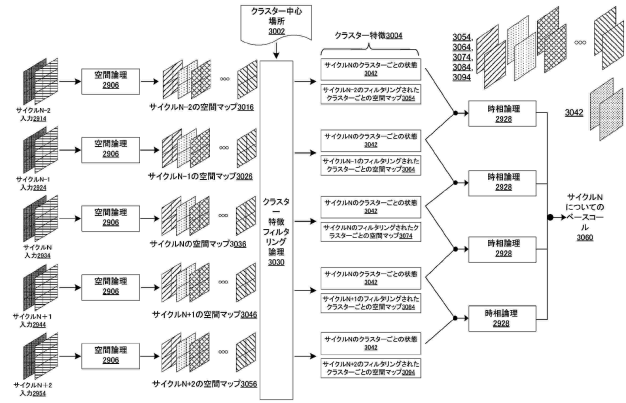
40

50

【図 29】

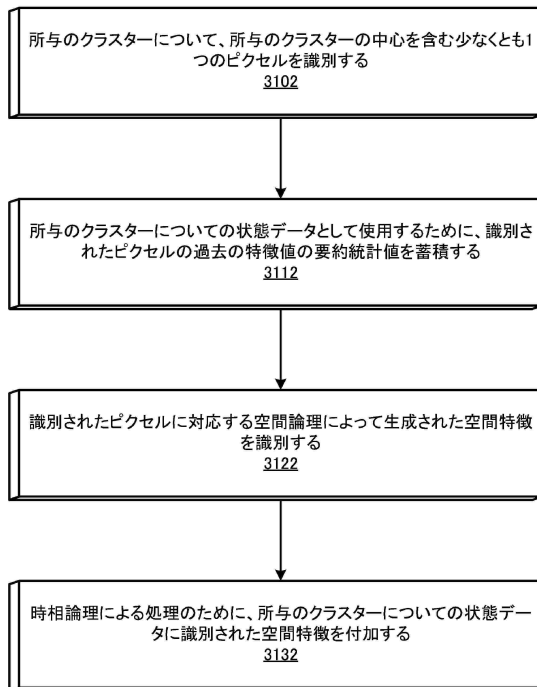


【図 30】



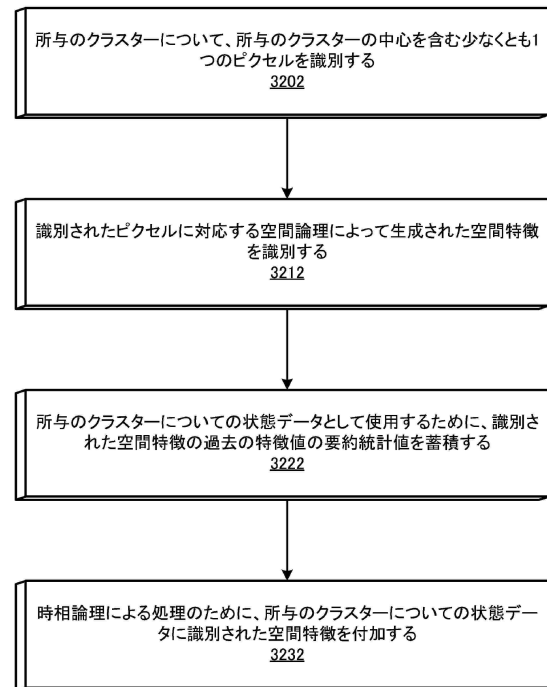
【図 31】

クラスターごとのピクセルベースの状態生成3100



【図 32】

クラスターごとの空間特徴ベースの状態生成3200



10

20

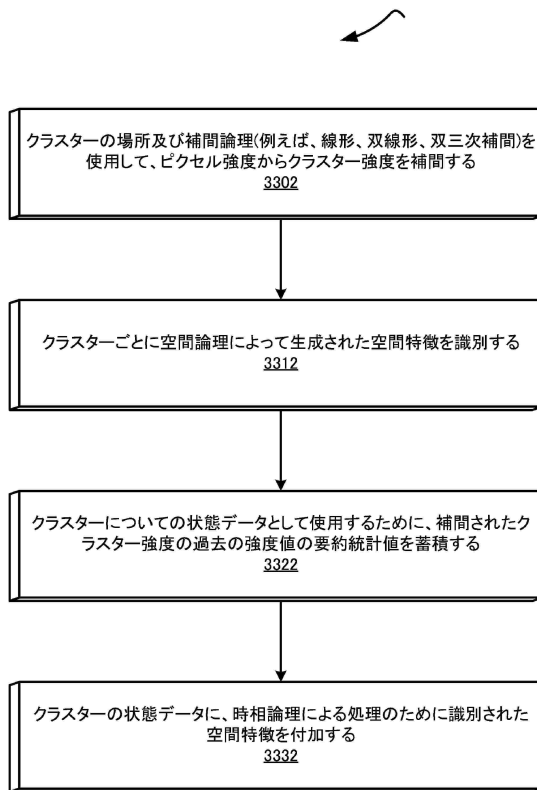
30

40

50

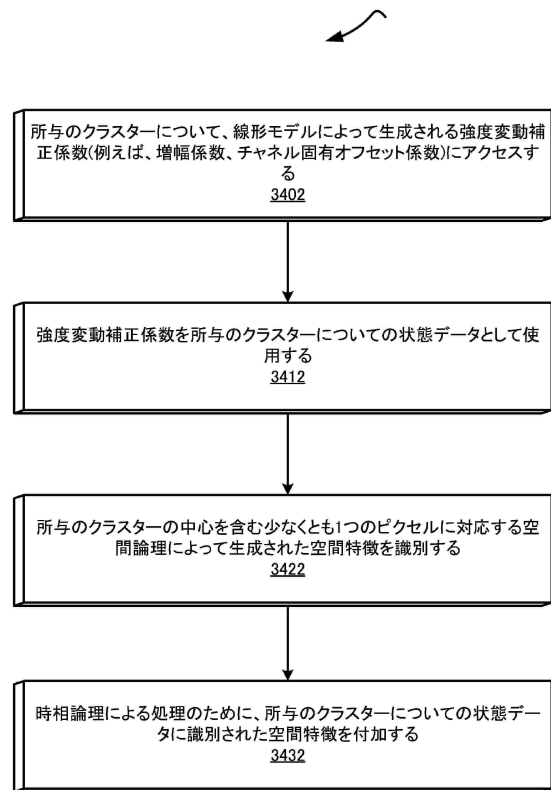
【図 3 3】

クラスターごとに補間された特徴ベースの状態生成3300



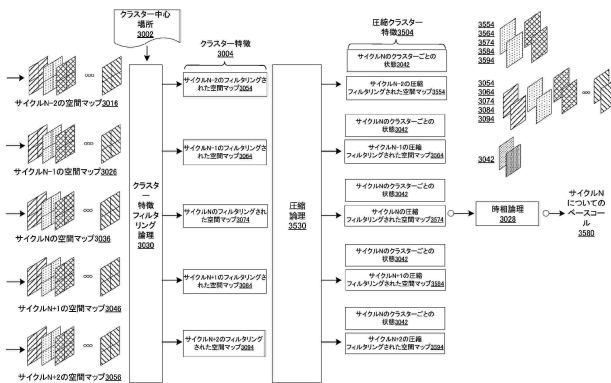
【圖 3 4】

クラスターごとのRTA係数ベースの状態生成3400



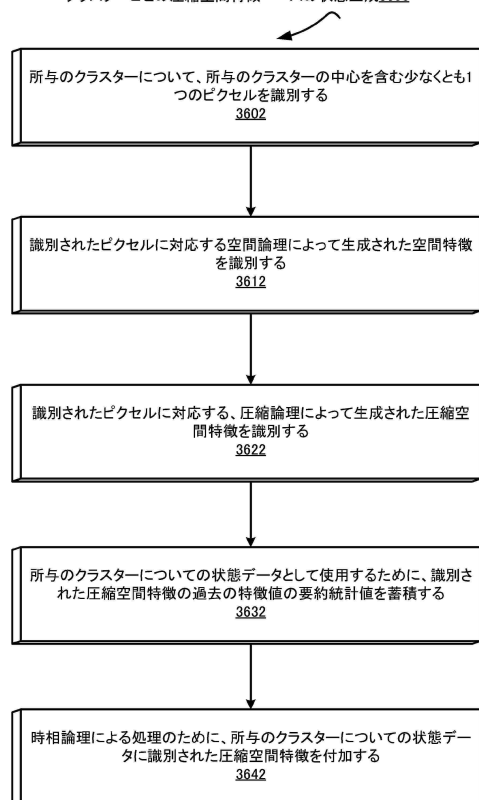
【図 3 5】

クラスターごとの圧縮状態データベースのベースコール1500



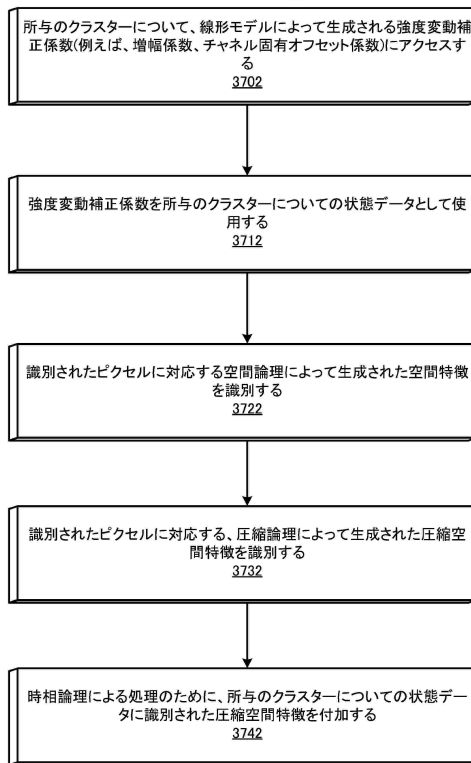
【圖 3 6】

クラスターごとの圧縮空間特徴ベースの状態生成3600



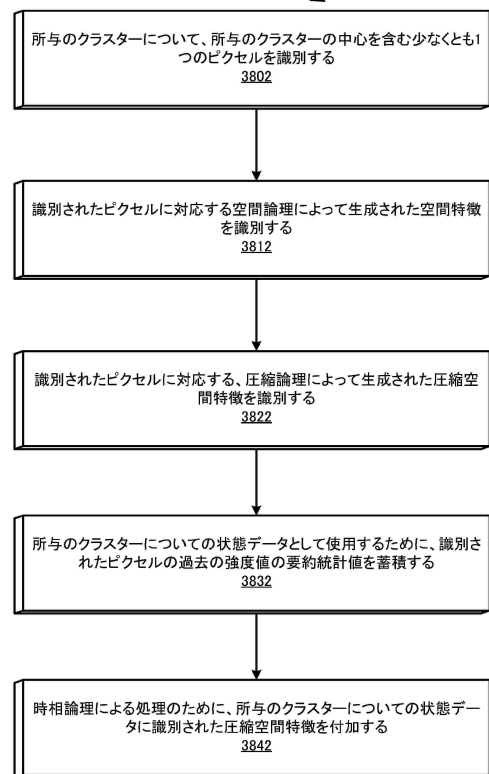
【図 37】

クラスターごとのRTA圧縮特徴ベースの状態生成3700



【図 38】

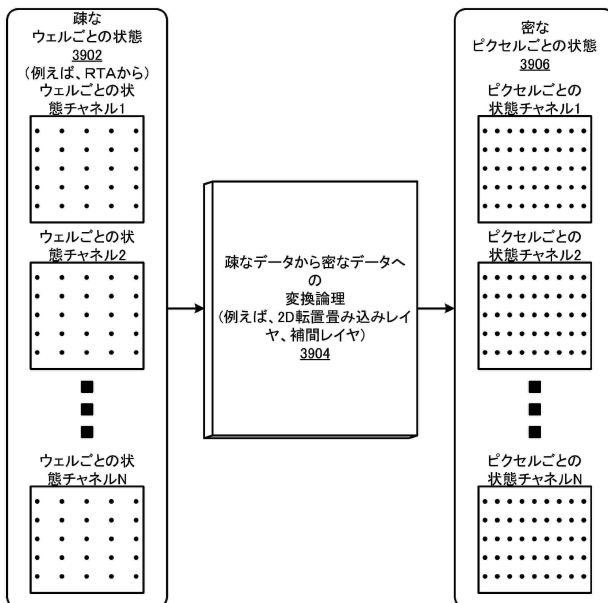
クラスターごとの圧縮ピクセルベースの状態生成3800



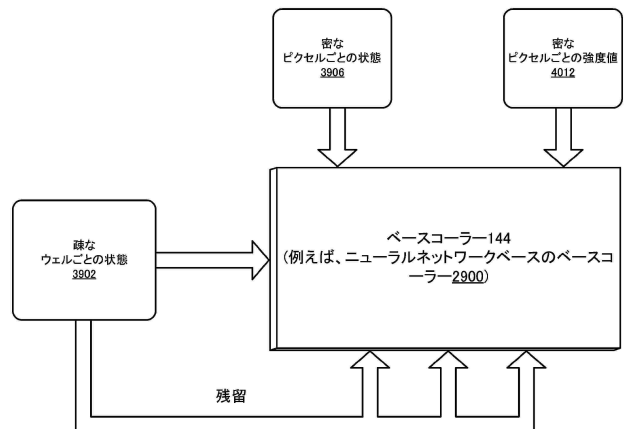
10

20

【図 39】



【図 40】

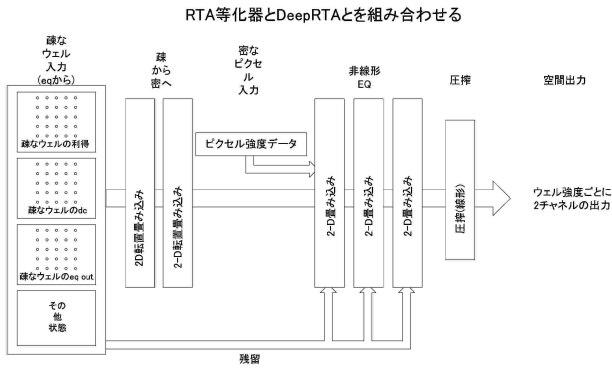


30

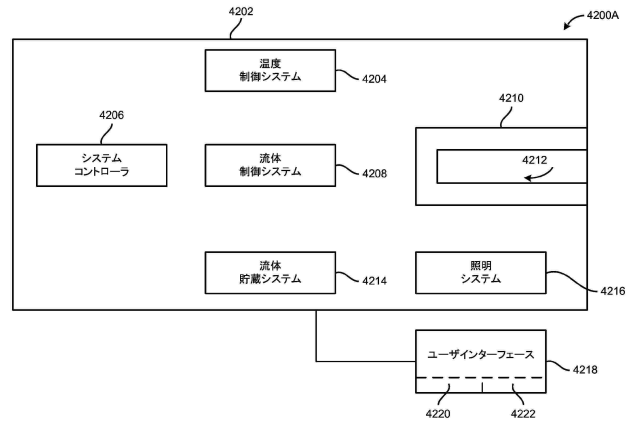
40

50

【図 4 1】

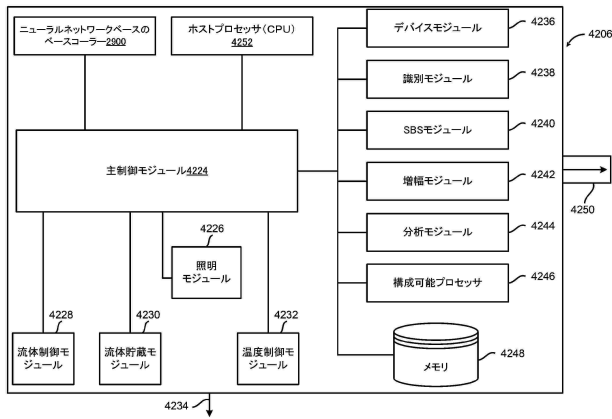


【図 4 2 A】

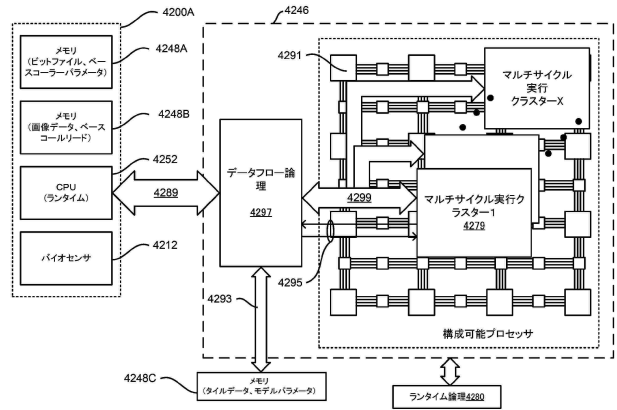


10

【図 4 2 B】

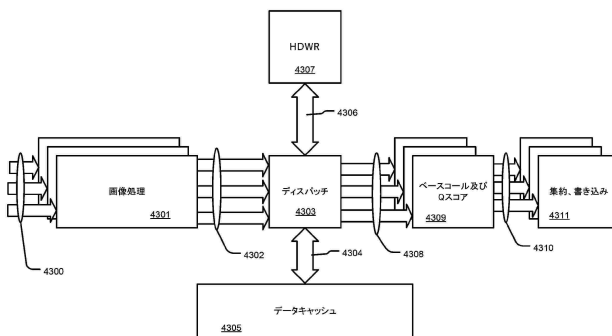


【図 4 2 C】

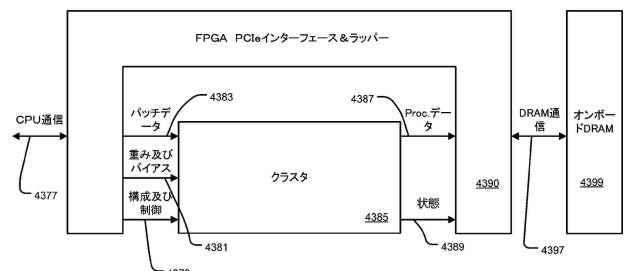


20

【図 4 3 A】



【図 4 3 B】

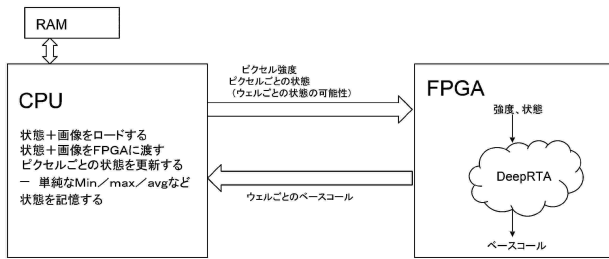


30

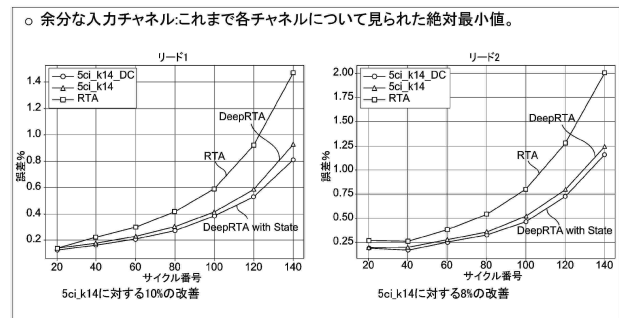
40

50

【図 4 4】

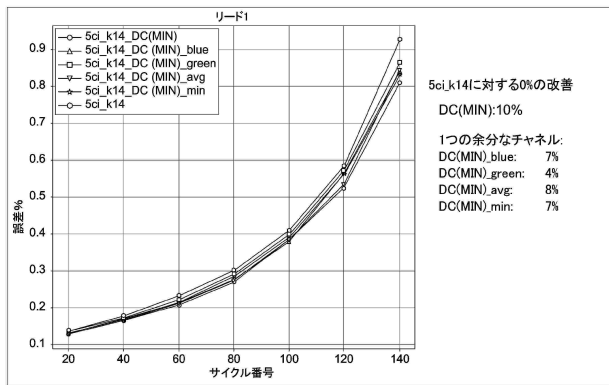


【図 4 5】

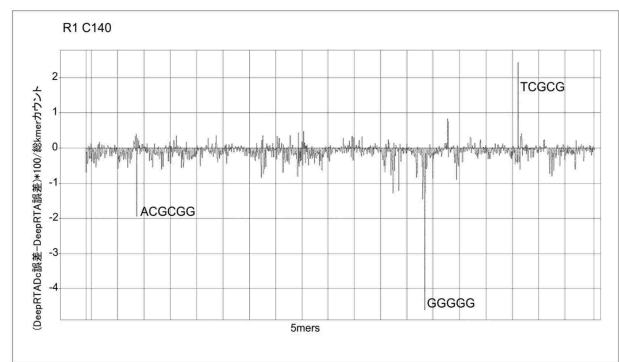


10

【図 4 6】

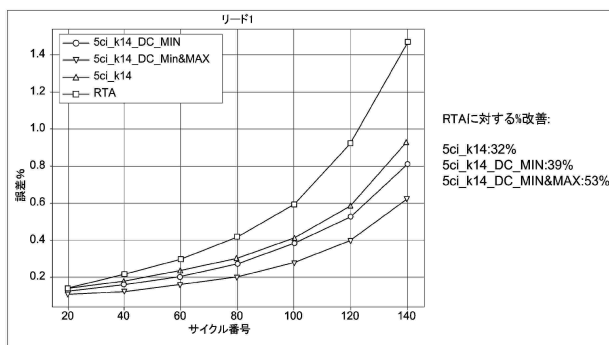


【図 4 7】



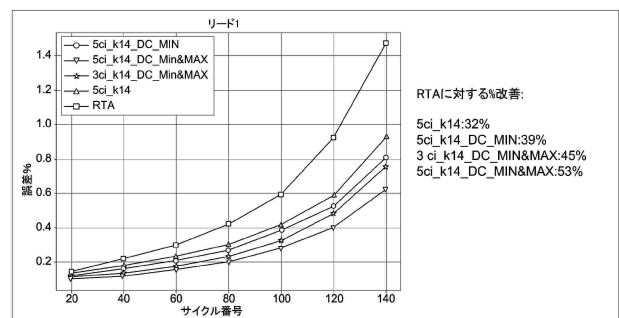
20

【図 4 8】



状態を伴わないDeepRTAシステムと比較して、
30%の誤差率の低減が実証されている
-「計算」コストは無視できる
-ハードウェア互換性

【図 4 9】

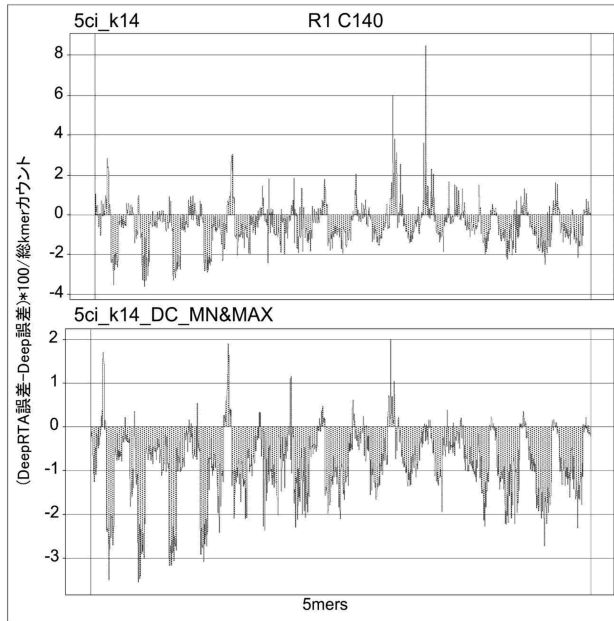


30

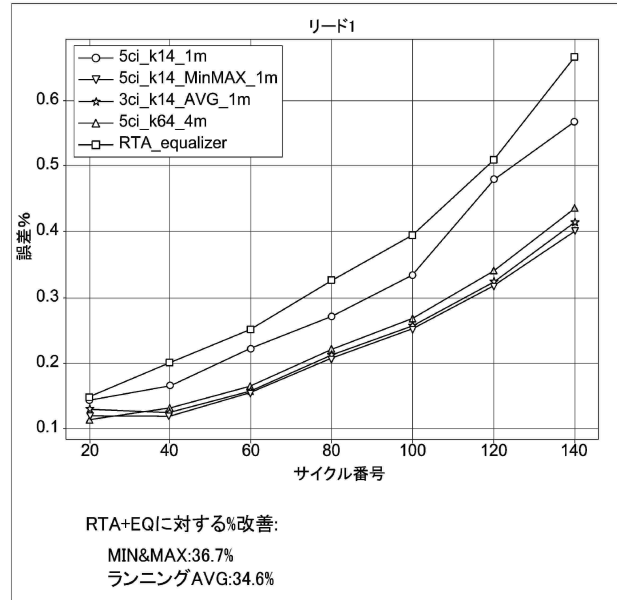
40

50

【図 50】

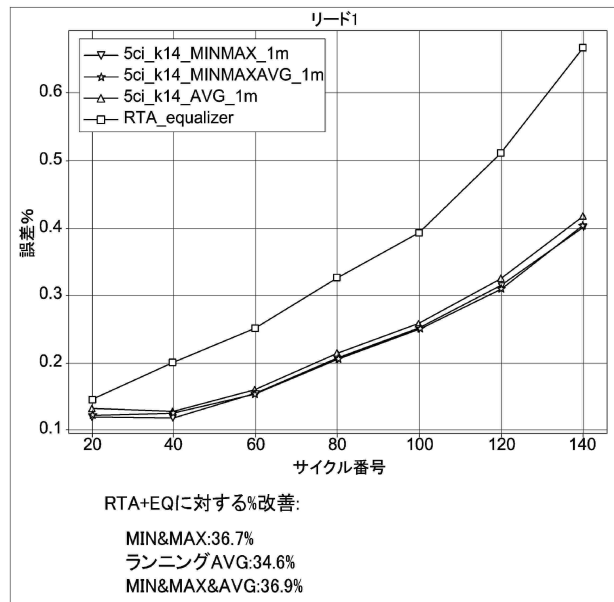


【図 51】

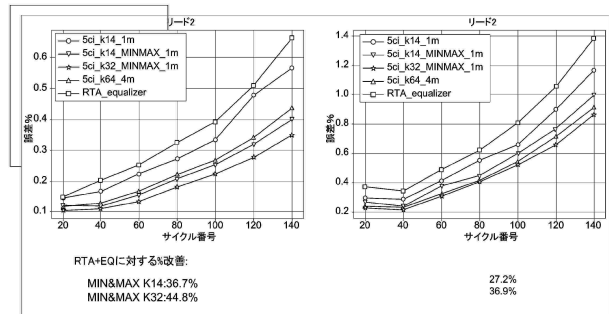


10

【図 52】



【図 53】



20

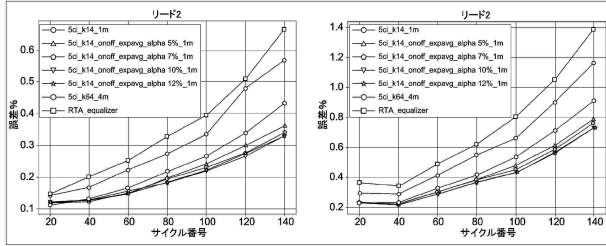
30

40

50

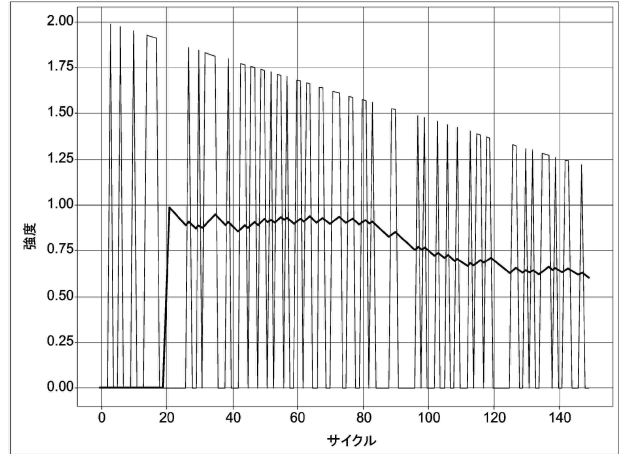
【図 5 4】

クラスターごとの情報の使用



【図 5 5】

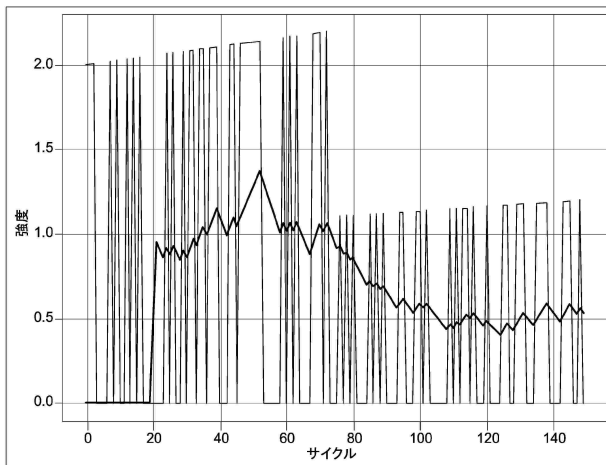
生の強度がリードにわたって減衰するときの指数移動平均



10

【図 5 6】

ステップ変化を伴う指数移動平均



【図 5 7】

25X Human Buildについての推論結果

	リードの数	読取率 (R1,R2,avg)	SNP 再現性	SNP精度	SNP F1	Indel 再現性	Indel精度	Indel F1
RTA+Eq	605,717,554**	0.41,0.72,0.56	0.990142	0.998558	0.994332	0.971573	0.978162	0.974856
DeepRTA (k14_1m)	602,022,776	0.36,0.63,0.50	0.990882	0.998508	0.994680	0.976982	0.977608	0.977295
DeepRTA_extrachannels (k14_1m_onoff_expavg)	609,209,246	0.31,0.54,0.42	0.990985	0.99849	0.994724	0.977727	0.977748	0.977737

20

30

40

50

【手続補正書】

【提出日】令和6年3月25日(2024.3.25)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

システムであって、

クラスターの配列決定画像を処理し、空間畳み込み特徴を生成するように構成された空間畳み込みニューラルネットワークと、

前記空間畳み込み特徴から、前記クラスターの中心を含む空間畳み込み特徴のサブセットを選択するように構成されたフィルタリング論理と、

空間畳み込み特徴の前記サブセットを圧縮特徴のセットに圧縮するように構成された圧縮論理であって、空間畳み込み特徴の前記サブセットがM個のチャンネルを有し、圧縮特徴の前記セットがN個のチャンネルを有し、 $M > N$ である、圧縮論理と、

圧縮特徴の前記セット内の圧縮特徴についての状態情報にアクセスし、前記状態情報を前記圧縮特徴に付加してステートフルな圧縮特徴のセットを生成するように構成されたコンテキスト化論理と、

ステートフルな圧縮特徴の前記セットを処理し、時間畳み込みステートフル特徴を生成するように構成された時間畳み込みニューラルネットワークと、

前記時間畳み込みステートフル特徴に基づいて前記クラスターについてのベースコールを生成するように構成されたベースコール論理と、を備える、システム。

【請求項2】

前記状態情報が、クラスターの前記配列決定画像を独立して処理することに応答して線形モデルによってクラスターごとに生成される強度変動補正係数を含む、請求項1に記載のシステム。

【請求項3】

前記強度変動補正係数が、

前記クラスターの強度プロファイル間のスケール変動を補償する増幅係数、

複数のチャンネルに沿った前記クラスターの前記強度プロファイル間のシフト変動を補償するチャンネル固有オフセット係数、又は

前記シフト変動を補償する共通オフセット係数のうちの1つ以上を含む、請求項2に記載のシステム。

【請求項4】

前記強度変動補正係数が、配列決定ランの現在の配列決定サイクルにおいて決定された現在の強度統計値の分析と、前記現在の配列決定サイクルに先行する前記配列決定ランの1つ以上の以前の配列決定サイクルにおいて決定された過去の強度統計値の分析とを組み合わせることに基づいて、クラスターごとに生成される、請求項3に記載のシステム。

【請求項5】

前記クラスターの前記強度プロファイルが、前記複数のチャンネル内のチャンネルについてのチャンネル固有強度値によって特徴付けられる、請求項4に記載のシステム。

【請求項6】

前記状態情報が、各チャンネルについてのチャンネル固有状態値によって特徴付けられるチャンネルごとの状態を含む、請求項5に記載のシステム。

【請求項7】

前記状態情報が、クラスターごとのチャンネル固有ビットを用いてビットごとに符号化されるチャンネルごとの状態を含む、請求項6に記載のシステム。

【請求項8】

10

20

30

40

50

前記チャンネルごとの状態が、前記クラスターごとのチャンネル固有ビットとビットごとに連結されるか、又は

前記チャンネルごとの状態が、前記クラスターごとのチャンネル固有ビットとビットごとに合計される、請求項 7 に記載のシステム。

【請求項 9】

前記チャンネルごとの状態が、前記現在の配列決定サイクル及び前記 1 つ以上の以前の配列決定サイクルにおいて空間処理論理によって生成されたクラスターごとのチャンネル固有ビットの圧縮値に基づく、請求項 8 に記載のシステム。

【請求項 10】

前記チャンネルごとの状態が、前記圧縮値の平均、前記圧縮値の移動平均、前記圧縮値の最大値、前記圧縮値の最小値、又は前記圧縮値の指数加重平均のうちの少なくとも 1 つを含む、請求項 9 に記載のシステム。

10

【請求項 11】

前記チャンネルごとの状態が、前記指数加重平均を含み、前記指数加重平均が、前の配列決定サイクルよりも最近の配列決定サイクルを加重することに基づいて決定される、請求項 10 に記載のシステム。

【請求項 12】

前記チャンネルごとの状態が、前記移動平均を含み、前記移動平均が、前記 1 つ以上の以前の配列決定サイクルからの圧縮値のローリングサブセットを使用する、請求項 10 に記載のシステム。

20

【請求項 13】

命令を含む非一時的コンピュータ可読記憶媒体であって、少なくとも 1 つのプロセッサによって実行されると、コンピューティングデバイスに、

空間畳み込みニューラルネットワークを使用して、空間畳み込み特徴を生成するようにクラスターの配列決定画像を処理させ、

前記空間畳み込み特徴から、前記クラスターの中心を含む空間畳み込み特徴のサブセットを選択させ、

M 個のチャンネルを有する空間畳み込み特徴の前記サブセットを、N 個のチャンネルを有する圧縮特徴のセットに圧縮させ（ここで、 $M > N$ である）、

圧縮特徴の前記セット内の圧縮特徴についての状態情報にアクセスさせ、

前記状態情報を前記圧縮特徴に付加させて、ステートフルな圧縮特徴のセットを生成させ、

30

時間畳み込みニューラルネットワークを使用して、ステートフルな圧縮特徴の前記セットを処理させて、時間畳み込みステートフル特徴を生成させ、

前記時間畳み込みステートフル特徴に基づいて、前記クラスターについてのベースコールを生成させる、非一時的コンピュータ可読記憶媒体。

【請求項 14】

前記状態情報が、各チャンネルについてのチャンネル固有状態値によって特徴付けられるチャンネルごとの状態を含み、

前記チャンネルごとの状態が、現在の配列決定サイクル及び 1 つ以上の以前の配列決定サイクルにおいて生成されたクラスターごとのチャンネル固有ビットの圧縮値に基づく、請求項 13 に記載の非一時的コンピュータ可読記憶媒体。

40

【請求項 15】

前記チャンネルごとの状態が、アクティブ状態値及び非アクティブ状態値を含む、請求項 14 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 16】

前記アクティブ状態値及び前記非アクティブ状態値が、先行するベースコール、

前記圧縮値の大域的 maximum 値及び大域的 minimum 値、又は

前記圧縮値の指数加重平均、のうちの少なくとも 1 つに基づいて決定される、請求項 1

50

5に記載の非一時的コンピュータ可読記憶媒体。

【請求項17】

コンピュータ実装方法であって、

空間畳み込みニューラルネットワークを使用して、空間畳み込み特徴を生成するようにクラスタの配列決定画像を処理することと、

前記空間畳み込み特徴から、前記クラスタの中心を含む空間畳み込み特徴のサブセットを選択することと、

空間畳み込み特徴の前記サブセットを圧縮特徴のセットに圧縮することであって、空間畳み込み特徴の前記サブセットがM個のチャンネルを有し、圧縮特徴の前記セットがN個のチャンネルを有し、 $M > N$ である、圧縮することと、

圧縮特徴の前記セット中の圧縮特徴についての状態情報にアクセスすることと、

前記状態情報を前記圧縮特徴に付加させて、ステートフルな圧縮特徴のセットを生成することと、

時間畳み込みニューラルネットワークを使用して、ステートフルな圧縮特徴の前記セットを処理して、時間畳み込みステートフル特徴を生成することと、

前記時間畳み込みステートフル特徴に基づいて、前記クラスタについてのベースコードを生成することと、を含む、コンピュータ実装方法。

【請求項18】

前記空間畳み込みニューラルネットワークへの入力が、現在の配列決定サイクル、1つ以上の以前の配列決定サイクル、及び1つ以上の次の配列決定サイクルについての配列決定画像のスライディングウィンドウを含む、請求項17に記載のコンピュータ実装方法。

【請求項19】

特徴セットが、前記現在の配列決定サイクル、前記1つ以上の以前の配列決定サイクル、及び前記1つ以上の次の配列決定サイクルについての配列決定画像の前記スライディングウィンドウ内の配列決定画像のチャンネル固有ピクセルから抽出され、クラスタごとの中心的特徴セットが、前記現在の配列決定サイクル、前記1つ以上の以前の配列決定サイクル、及び前記1つ以上の次の配列決定サイクルについての前記特徴セットから選別される、請求項18に記載のコンピュータ実装方法。

【請求項20】

チャンネル固有ビットセットが、前記現在の配列決定サイクル、前記1つ以上の以前の配列決定サイクル、及び前記1つ以上の次の配列決定サイクルについての前記クラスタごとの中心的特徴セットから圧縮される、請求項19に記載のコンピュータ実装方法。

10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2022/044293

A. CLASSIFICATION OF SUBJECT MATTER
INV. G16B40/20 C12Q1/6869 G16B30/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G16B G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2021/265016 A1 (VESSERE GERY [US] ET AL) 26 August 2021 (2021-08-26) par. 124, 129, 137, 138, 144; figure 1B; 1C; 19A; 19 B; -----	1-30
Y	WO 2020/191387 A1 (ILLUMINA INC [US]) 24 September 2020 (2020-09-24) paragraphs [0186], [0369], [0760]; claim 1 -----	1-30
Y	US 2020/302297 A1 (JAGANATHAN KISHORE [US] ET AL) 24 September 2020 (2020-09-24) paragraph [0179] -----	1-29

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

20 December 2022

Date of mailing of the international search report

04/01/2023

Name and mailing address of the ISA/
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Eberhardt, Anja

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2022/044293

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2021265016 A1	26-08-2021	AU 2021222050 A1	08-09-2022
		AU 2021224872 A1	08-09-2022
		AU 2021224873 A1	08-09-2022
		BR 112022016416 A2	11-10-2022
		BR 112022016495 A2	11-10-2022
		CA 3168731 A1	26-08-2021
		CN 115280418 A	01-11-2022
		CN 115398549 A	25-11-2022
		CN 115443508 A	06-12-2022
		EP 4107738 A1	28-12-2022
		EP 4107739 A1	28-12-2022
		EP 4107740 A1	28-12-2022
		IL 295584 A	01-10-2022
		IL 295585 A	01-10-2022
		IL 295587 A	01-10-2022
		KR 20220144373 A	26-10-2022
		KR 20220144374 A	26-10-2022
		KR 20220144375 A	26-10-2022
		US 2021264266 A1	26-08-2021
		US 2021264267 A1	26-08-2021
WO 2020191387 A1	24-09-2020	US 2021265016 A1	26-08-2021
		WO 2021168356 A1	26-08-2021
		WO 2021168358 A1	26-08-2021
		WO 2021168360 A1	26-08-2021
WO 2020191387 A1	24-09-2020	WO 2020191387 A1	24-09-2020
		WO 2020191389 A1	24-09-2020
		WO 2020191390 A2	24-09-2020
		WO 2020191391 A2	24-09-2020
		WO 2020205296 A1	08-10-2020
US 2020302297 A1	24-09-2020	US 2020302224 A1	24-09-2020
		US 2020302225 A1	24-09-2020
		US 2020302297 A1	24-09-2020
		US 2020327377 A1	15-10-2020

フロントページの続き

(32)優先日 令和4年9月14日(2022.9.14)

(33)優先権主張国・地域又は機関
米国(US)

(31)優先権主張番号 17/944,948

(32)優先日 令和4年9月14日(2022.9.14)

(33)優先権主張国・地域又は機関
米国(US)

(81)指定国・地域 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CV,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IQ,IR,IS,IT,JM,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

(72)発明者 エリック・ジョン・オジャーダ

アメリカ合衆国・カリフォルニア・9 2 1 2 2・サン・ディエゴ・イルミナ・ウェイ・5 2 0 0

(72)発明者 ドルナ・カシェファギギ

アメリカ合衆国・カリフォルニア・9 4 0 2 5・メンロー・パーク・イーストリッジ・アベニュー
・2 2 3 0