(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2015-514239
(P2015-514239A)
(43)公表日　平成27年5月18日(2015.5.18)

(51)Int.Cl.		F I			テーマコード（参考）
G10L 15/10	(2006.01)	G10L	15/10	500Z	5L096
G06T 7/00	(2006.01)	G06T	7/00	P	
G06T 7/20	(2006.01)	G06T	7/20	B	
G10L 25/51	(2013.01)	G10L	25/51	400	
G10L 15/00	(2013.01)	G10L	15/00	200G	

審査請求 未請求　予備審査請求 有　（全 104 頁）　最終頁に続く

(21)出願番号　特願2015-505720 (P2015-505720)
(86)(22)出願日　平成25年3月7日 (2013.3.7)
(85)翻訳文提出日　平成26年12月3日 (2014.12.3)
(86)国際出願番号　PCT/US2013/029558
(87)国際公開番号　WO2013/154701
(87)国際公開日　平成25年10月17日 (2013.10.17)
(31)優先権主張番号　61/623,910
(32)優先日　平成24年4月13日 (2012.4.13)
(33)優先権主張国　米国 (US)
(31)優先権主張番号　13/664,295
(32)優先日　平成24年10月30日 (2012.10.30)
(33)優先権主張国　米国 (US)

(71)出願人　595020643
クゥアルコム・インコーポレイテッド
ＱＵＡＬＣＯＭＭ　ＩＮＣＯＲＰＯＲＡＴＥＤ
アメリカ合衆国、カリフォルニア州　９２１２１－１７１４、サン・ディエゴ、モアハウス・ドライブ　５７７５
(74)代理人　100108855
弁理士　蔵田　昌俊
(74)代理人　100109830
弁理士　福原　淑弘
(74)代理人　100103034
弁理士　野河　信久
(74)代理人　100075672
弁理士　峰　隆司

最終頁に続く

(54)【発明の名称】マルチモーダル整合方式を使用するオブジェクト認識

(57)【要約】
シーン中の１つまたは複数のオブジェクトを認識し、位置を特定するための方法、システムおよび製造品が開示される。シーンの画像および／またはビデオがキャプチャされる。シーンにおいて記録されたオーディオを使用して、キャプチャされたシーンのオブジェクト探索が狭められる。たとえば、キャプチャされた画像／ビデオ中の探索エリアを限定するために、音の到来方向（ＤＯＡ）が判断され、使用され得る。別の例では、記録されたオーディオ中で識別される音のタイプに基づいてキーポイントシグネチャが選択され得る。キーポイントシグネチャは、本システムが認識するように構成された特定のオブジェクトに対応する。次いで、キャプチャされたシーン中で識別されるキーポイントを、選択されたキーポイントシグネチャと比較する、シフト不変特徴変換（ＳＩＦＴ）分析を使用して、シーン中のオブジェクトが認識され得る。
【選択図】　図２

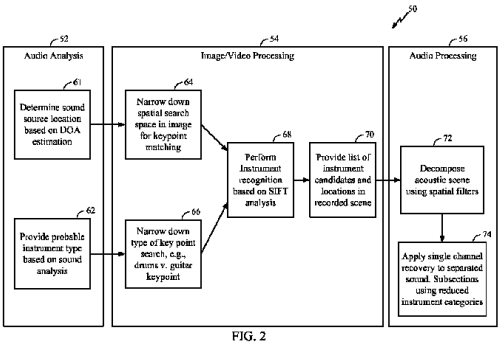


FIG. 2

【特許請求の範囲】
【請求項１】
デバイスにおいて、シーン中のオブジェクトを認識する方法であって、
前記シーンにおいて記録されたオーディオに基づいて前記オブジェクトに対応するキーポイントを選択することと、
前記選択されたキーポイントに基づいて前記オブジェクトを識別することと
を備える、方法。
【請求項２】
前記シーンにおいて記録されたオーディオに基づいて、１つまたは複数のオブジェクトに対応する１つまたは複数のキーポイントシグネチャを選択することと、
前記シーンの画像中の複数のキーポイントを識別することと、
前記オブジェクトを識別するために前記キーポイントを前記キーポイントシグネチャと比較することと
をさらに備える、請求項１に記載の方法。
【請求項３】
前記シーンにおいて記録された前記オーディオに基づいてシーン画像の一部分を選択することと、
前記画像の前記一部分内からのみ前記キーポイントを選択することと
をさらに備える、請求項１に記載の方法。
【請求項４】
前記シーンにおいて記録された前記オーディオに基づいて前記画像の一部分を選択することが、
前記オーディオからオーディオ到来方向（ＤＯＡ）を判断することと、
前記オーディオＤＯＡに基づいて前記画像の前記一部分を選択することと
を含む、請求項３に記載の方法。
【請求項５】
前記オーディオＤＯＡを判断することが、
前記シーンに位置する複数のマイクロフォンにおいて前記オーディオを受信し、それによって複数のマイクロフォン信号を生成することと、
前記マイクロフォン信号に基づいて前記オーディオＤＯＡを判断することと
を含む、請求項４に記載の方法。
【請求項６】
前記シーンのビデオ記録から複数の局所動きベクトルを計算することと、
前記局所動きベクトルを１つまたは複数のオブジェクトに対応する所定の局所動きベクトルのデータベースと比較することによって、および前記キーポイントを１つまたは複数のキーポイントシグネチャと比較することによって前記オブジェクトを識別することと
をさらに備える、請求項１に記載の方法。
【請求項７】
前記シーンにおいて記録された前記オーディオから複数の音響認識特徴を計算することと、
前記音響認識特徴を１つまたは複数のオブジェクトに対応する所定の音響認識特徴のデータベースと比較することによって、および前記キーポイントを１つまたは複数のキーポイントシグネチャと比較することによって前記オブジェクトを識別することと
をさらに備える、請求項１に記載の方法。
【請求項８】
前記音響認識特徴がメル周波数ケプストラム係数を含む、請求項７に記載の方法。
【請求項９】
前記画像中に現れる１つまたは複数のオブジェクトについての範囲情報を判断することと、
前記範囲情報に基づいて前記キーポイントを分析することと

をさらに備える、請求項１に記載の方法。
【請求項１０】
範囲情報を判断することが、オートフォーカスカメラを使用して範囲情報を判断することと、マルチカメラ画像視差推定を使用して範囲情報を判断することと、上記の任意の好適な組合せとからなるグループから選択される、請求項９に記載の方法。
【請求項１１】
シーンにおいて記録されたオーディオに基づいて前記シーン中のオブジェクトに対応するキーポイントを選択するように構成されたキーポイントセレクタと、
前記選択されたキーポイントに基づいて前記オブジェクトを識別するように構成された整合デバイスと
を備える、装置。
【請求項１２】
シーンの画像中の複数のキーポイントを識別するように構成されたキーポイント検出器をさらに備え、
前記キーポイントセレクタが、前記シーンにおいて記録されたオーディオに基づいて、１つまたは複数のオブジェクトに対応する１つまたは複数のキーポイントシグネチャを選択するように構成され、
前記整合デバイスが、前記シーン中のオブジェクトを識別するために前記キーポイントを前記キーポイントシグネチャと比較するように構成された、
請求項１１に記載の装置。
【請求項１３】
前記シーンにおいて記録された前記オーディオに基づいて前記シーンの画像の一部分を選択するように構成された第１のセレクタと、
前記画像の前記一部分内からのみ前記キーポイントを選択するように構成された第２のセレクタと
をさらに備える、請求項１１に記載の装置。
【請求項１４】
前記第１のセレクタが、
前記オーディオからオーディオ到来方向（ＤＯＡ）を判断するように構成された検出器と、
前記オーディオＤＯＡに基づいて前記画像の前記一部分を選択するように構成された第３のセレクタと
を含む、請求項１３に記載の装置。
【請求項１５】
前記検出器が、
前記オーディオを受信して、複数のマイクロフォン信号を生成するための、前記シーンに位置する複数のマイクロフォンと、
前記マイクロフォン信号に基づいて前記オーディオＤＯＡを判断するように構成されたオーディオプロセッサと
を含む、請求項１４に記載の装置。
【請求項１６】
前記シーンのビデオ記録から複数の局所動きベクトルを計算するように構成されたビデオプロセッサ
をさらに備え、
前記整合デバイスが、前記局所動きベクトルを１つまたは複数のオブジェクトに対応する所定の局所動きベクトルのデータベースと比較することによって、および前記キーポイントを１つまたは複数のキーポイントシグネチャと比較することによって前記オブジェクトを識別するように構成された、
請求項１１に記載の装置。
【請求項１７】

前記シーンにおいて記録された前記オーディオから複数の音響認識特徴を計算するように構成されたオーディオプロセッサ
をさらに備え、
前記整合デバイスが、前記音響認識特徴を１つまたは複数のオブジェクトに対応する所定の音響認識特徴のデータベースと比較することによって、および前記キーポイントを１つまたは複数のキーポイントシグネチャと比較することによって前記オブジェクトを識別するように構成された、
請求項１１に記載の装置。
【請求項１８】
前記音響認識特徴がメル周波数ケプストラム係数を含む、請求項１７に記載の装置。
【請求項１９】
前記画像中に現れる１つまたは複数のオブジェクトについての範囲情報を判断するように構成された範囲検出器と、
前記範囲情報に基づいて前記キーポイントを分析するように構成されたキーポイント検出器と
をさらに備える、請求項１１に記載の装置。
【請求項２０】
前記範囲検出器が、オートフォーカスカメラと、マルチカメラアレイと、上記の任意の好適な組合せとからなるグループから選択される検出器を含む、請求項１９に記載の装置。
【請求項２１】
シーンにおいて記録されたオーディオに基づいて前記シーン中のオブジェクトに対応するキーポイントを選択するための手段と、
前記選択されたキーポイントに基づいて前記オブジェクトを識別するための手段と
を備える、装置。
【請求項２２】
前記シーンにおいて記録されたオーディオに基づいて、１つまたは複数のオブジェクトに対応する１つまたは複数のキーポイントシグネチャを選択するための手段と、
前記シーンの画像中の複数のキーポイントを識別するための手段と、
前記シーン中の前記オブジェクトを識別するために前記キーポイントを前記キーポイントシグネチャと比較するための手段と
をさらに備える、請求項２１に記載の装置。
【請求項２３】
前記シーンにおいて記録された前記オーディオに基づいて前記シーンの画像の一部分を選択するための手段と、
前記画像の前記一部分内からのみ前記キーポイントを選択するための手段と
をさらに備える、請求項２１に記載の装置。
【請求項２４】
前記シーンにおいて記録された前記オーディオに基づいて前記画像の一部分を選択するための前記手段が、
前記オーディオからオーディオ到来方向（ＤＯＡ）を判断するための手段と、
前記オーディオＤＯＡに基づいて前記画像の前記一部分を選択するための手段と
を含む、請求項２３に記載の装置。
【請求項２５】
前記オーディオＤＯＡを判断するための手段が、
前記シーンに位置する複数のマイクロフォンにおいて前記オーディオを受信し、それによって複数のマイクロフォン信号を生成するための手段と、
前記マイクロフォン信号に基づいて前記オーディオＤＯＡを判断するための手段と
を含む、請求項２４に記載の装置。
【請求項２６】

前記シーンのビデオ記録から複数の局所動きベクトルを計算するための手段と、
前記局所動きベクトルを１つまたは複数のオブジェクトに対応する所定の局所動きベクトルのデータベースと比較することによって、および前記キーポイントを１つまたは複数のキーポイントシグネチャと比較することによって前記オブジェクトを識別するための手段と
をさらに備える、請求項２１に記載の装置。
【請求項２７】
前記シーンにおいて記録された前記オーディオから複数の音響認識特徴を計算するための手段と、
前記音響認識特徴を１つまたは複数のオブジェクトに対応する所定の音響認識特徴のデータベースと比較することによって、および前記キーポイントを１つまたは複数のキーポイントシグネチャと比較することによって前記オブジェクトを識別するための手段と
をさらに備える、請求項２１に記載の装置。
【請求項２８】
前記音響認識特徴がメル周波数ケプストラム係数を含む、請求項２７に記載の装置。
【請求項２９】
画像中に現れる１つまたは複数のオブジェクトについての範囲情報を判断するための手段と、
前記範囲情報に基づいて前記キーポイントを分析するための手段と
をさらに備える、請求項２１に記載の装置。
【請求項３０】
範囲情報を判断するための手段が、オートフォーカスカメラを使用して範囲情報を判断するための手段と、マルチカメラ画像視差推定を使用して範囲情報を判断するための手段と、上記の任意の好適な組合せとからなるグループから選択される、請求項２９に記載の装置。
【請求項３１】
シーンにおいて記録されたオーディオに基づいて前記シーン中のオブジェクトに対応するキーポイントを選択するためのコードと、
前記選択されたキーポイントに基づいて前記オブジェクトを識別するためのコードと
を備える、１つまたは複数のプロセッサによって実行可能な命令のセットを具備するコンピュータ可読媒体。
【請求項３２】
前記シーンにおいて記録されたオーディオに基づいて、１つまたは複数のオブジェクトに対応する１つまたは複数のキーポイントシグネチャを選択するためのコードと、
前記シーンの画像中の複数のキーポイントを識別するためのコードと、
前記シーン中の前記オブジェクトを識別するために前記キーポイントを前記キーポイントシグネチャと比較するためのコードと
をさらに備える、請求項３１に記載のコンピュータ可読媒体。
【請求項３３】
前記シーンにおいて記録された前記オーディオに基づいて画像の一部分を選択するためのコードと、
前記画像の前記一部分内からのみ前記キーポイントを選択するためのコードと
をさらに備える、請求項３１に記載のコンピュータ可読媒体。
【請求項３４】
前記シーンにおいて記録された前記オーディオに基づいて前記画像の一部分を選択するための前記コードが、
前記オーディオからオーディオ到来方向（ＤＯＡ）を判断するためのコードと、
前記オーディオＤＯＡに基づいて前記画像の前記一部分を選択するためのコードと
を含む、請求項３３に記載のコンピュータ可読媒体。
【請求項３５】

前記オーディオＤＯＡを判断するためのコードが、
前記シーンに位置する複数のマイクロフォンにおいて前記オーディオを受信し、それによって複数のマイクロフォン信号を生成するためのコードと、
前記マイクロフォン信号に基づいて前記オーディオＤＯＡを判断するためのコードと
を含む、請求項３４に記載のコンピュータ可読媒体。
【請求項３６】
前記シーンのビデオ記録から複数の局所動きベクトルを計算するためのコードと、
前記局所動きベクトルを１つまたは複数のオブジェクトに対応する所定の局所動きベクトルのデータベースと比較することによって、および前記キーポイントを１つまたは複数のキーポイントシグネチャと比較することによって前記オブジェクトを識別するためのコードと
をさらに備える、請求項３１に記載のコンピュータ可読媒体。
【請求項３７】
前記シーンにおいて記録された前記オーディオから複数の音響認識特徴を計算するためのコードと、
前記音響認識特徴を１つまたは複数のオブジェクトに対応する所定の音響認識特徴のデータベースと比較することによって、および前記キーポイントを１つまたは複数のキーポイントシグネチャと比較することによって前記オブジェクトを識別するためのコードと
をさらに備える、請求項３１に記載のコンピュータ可読媒体。
【請求項３８】
前記音響認識特徴がメル周波数ケプストラム係数を含む、請求項３７に記載のコンピュータ可読媒体。
【請求項３９】
画像中に現れる１つまたは複数のオブジェクトについての範囲情報を判断するためのコードと、
前記範囲情報に基づいて前記キーポイントを分析するためのコードと
をさらに備える、請求項３１に記載のコンピュータ可読媒体。
【請求項４０】
範囲情報を判断するためのコードが、オートフォーカスカメラを使用して範囲情報を判断するためのコードと、マルチカメラ画像視差推定を使用して範囲情報を判断するためのコードと、上記の任意の好適な組合せとからなるグループから選択される、請求項３９に記載のコンピュータ可読媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
米国特許法第１１９条による優先権の主張
[0001]本特許出願は、本出願の譲受人に譲渡され、参照により本明細書に明確に組み込まれる、２０１２年４月１３日に出願された米国仮出願第６１／６２３，９１０号の優先権を主張する。
【０００２】
[0002]本開示は、一般にオブジェクト認識装置および方法に関する。
【背景技術】
【０００３】
[0003]様々な適用例は、視覚シーン中のオブジェクトを識別することが可能である機械またはプロセッサを有することから恩恵を受け得る。コンピュータビジョンの分野は、シーン中のオブジェクトまたは特徴を識別することを可能にする技法および／またはアルゴリズムを提供することを試みており、オブジェクトまたは特徴は、１つまたは複数のキーポイント（keypoint）を識別する記述子によって特徴づけられ得る。ＳＩＦＴ（スケール不変特徴変換（Scale Invariant Feature Transform））など、これらの技法はまた、しばしば、適用例の中でも、オブジェクト認識、オブジェクト検出、画像整合、３次元構造

構築、ステレオ対応、および／または動き追跡、顔認識に適用される。

【０００４】

[0004]大部分のオブジェクト認識技法は、シーンからキャプチャされた視覚情報、たとえば、ビデオ、画像またはピクチャのみに依拠する。

【発明の概要】

【０００５】

[0005]この概要は、すべての企図される実施形態の包括的な概観ではなく、すべての実施形態の主要または重要な要素を識別するものでも、いずれかまたはすべての実施形態の範囲を定めるものでもない。その唯一の目的は、後で提示するより詳細な説明の導入として、１つまたは複数の実施形態のいくつかの概念を簡略化された形式で提示することである。

【０００６】

[0006]シーン中の１つまたは複数のオブジェクトを認識し、位置を特定するための改善された技法が開示される。これらの技法は、シーン中のオブジェクトを認識するのを助けるために、シーンにおいて記録されたオーディオならびに視覚情報の使用を取り入れる。これらの技法の一態様によれば、装置は、シーンにおいて記録されたオーディオに基づいてシーン中のオブジェクトに対応するキーポイントを選択するように構成されたキーポイントセレクタと、選択されたキーポイントに基づいてオブジェクトを識別するように構成されたキーポイント整合デバイスとを含む。

【０００７】

[0007]さらなる態様によれば、シーン中のオブジェクトを認識する方法は、シーンにおいて記録されたオーディオに基づいてオブジェクトに対応するキーポイントを選択することと、選択されたキーポイントに基づいてオブジェクトを識別することとを含む。

【０００８】

[0008]さらなる態様によれば、１つまたは複数のプロセッサによって実行可能な命令のセットを具備するコンピュータ可読媒体が提供される。媒体は、シーンにおいて記録されたオーディオに基づいてシーン中のオブジェクトに対応するキーポイントを選択するためのプログラミングコードと、選択されたキーポイントに基づいてオブジェクトを識別するためのプログラミングコードとを記憶する。

【０００９】

[0009]他の態様、特徴および利点は、以下の図および詳細な説明を審査すると当業者に明らかであるかまたは明らかになる。すべてのそのような追加の特徴、態様、および利点は本明細書内に含まれ、添付の特許請求の範囲によって保護されるものである。

【図面の簡単な説明】

【００１０】

[0010]図面は例示のためのものにすぎないことを理解されたい。さらに、図中の構成要素は必ずしも一定の縮尺であるとは限らず、代わりに、本明細書で説明する技法およびデバイスの原理を示すことに強調が置かれる。図において、同様の参照番号は、異なる図全体にわたって対応する部分を示す。

【図１】[0011]例示的な聴覚シーンおよび視聴覚シーン分析システムを示す概念図。

【図２】[0012]図１の視聴覚シーン分析システムを動作させる方法を示すプロセスブロック図。

【図３】[0013]図１の聴覚シーン分析システムを使用して聴覚シーンを分解し、処理する例示的な方法を示すフローチャート。

【図４】[0014]聴覚シーン中の音生成オブジェクトを識別する例示的な方法を示すフローチャート。

【図５Ａ】[0015]聴覚シーン中の音生成オブジェクトを識別する第２の例示的な方法のフローチャート。

【図５Ｂ】聴覚シーン中の音生成オブジェクトを識別する第２の例示的な方法のフローチャート。

【図６】[0016]図１の聴覚シーン分析システムとともに使用され得る例示的なオブジェクト認識サブシステムのいくつかの構成要素を示すブロック図。
【図７】[0017]記録された画像上でオブジェクト認識を実行するための機能段階を示すブロック図。
【図８】[0018]例示的な画像処理段階におけるガウススケール空間発生を示す図。
【図９】[0019]例示的な画像処理段階における特徴検出を示す図。
【図１０】[0020]例示的な画像処理段階における特徴記述子抽出を示す図。
【図１１】[0021]照度不変特徴／キーポイント検出のためのスケール空間正規化の一例を示す図。
【図１２】[0022]正規化されたスケール空間差分の１つのレベルがどのように取得され得るかの一例を示す図。
【図１３】[0023]改善された特徴検出のために正規化されたスケール空間差分を発生するための方法を示す図。
【図１４】[0024]特徴／キーポイント検出デバイスの一例を示すブロック図。
【図１５】[0025]画像整合デバイスの一例を示すブロック図。
【図１６Ａ】[0026]一般的構成による装置Ａ１００のブロック図。
【図１６Ｂ】[0027]マイクロフォンアレイＲ１００と装置Ａ１００のインスタンスとを含むデバイスＤ１０のブロック図。
【図１６Ｃ】[0028]点音源ｊから受信された信号成分の、アレイＲ１００のマイクロフォンＭＣ１０およびＭＣ２０の軸に対する到来方向θ_jを示す図。
【図１７】[0029]装置Ａ１００の実装形態Ａ１１０のブロック図。
【図１８Ａ】[0030]フィルタ更新モジュールＵＭ１０の実装形態ＵＭ２０のブロック図。
【図１８Ｂ】[0031]フィルタ更新モジュールＵＭ２０の実装形態ＵＭ２２のブロック図。
【図１９Ａ】[0032]カメラＣＭ１０をもつアレイＲ１００の４マイクロフォン実装形態Ｒ１０４の構成の一例の上面図。
【図１９Ｂ】[0033]到来方向の推定のための遠距離場モデルを示す図。
【図２０】[0034]装置Ａ１００の実装形態Ａ１２０のブロック図。
【図２１】[0035]装置Ａ１２０およびＡ２００の実装形態Ａ２２０のブロック図。
【図２２】[0036]ＤＯＡ推定にＳＲＰ－ＰＨＡＴを使用した結果によるヒストグラムの例を示す図。
【図２３】[0037]ＩＶＡ適応ルール（４０～６０度の音源分離）を使用して適応された逆混合行列の異なる出力チャネルに関する４つのヒストグラムのセットの一例を示す図。
【図２４】[0038]画像またはビデオキャプチャ中にシーン中のオブジェクトの視差を検出するように構成された例示的な画像キャプチャデバイスの図。
【図２５】[0039]図２４のデバイス中に含まれ得る例示的な画像処理システムのブロック図。
【図２６Ａ】[0040]知覚されたオブジェクト深さと相関させられたオブジェクト視差の例示的な例の図。
【図２６Ｂ】知覚されたオブジェクト深さと相関させられたオブジェクト視差の例示的な例の図。
【図２７Ａ】[0041]一般的構成による方法Ｍ１００のフローチャート。
【図２７Ｂ】[0042]方法Ｍ１００の実装形態Ｍ２００のフローチャート。
【図２７Ｃ】[0043]一般的構成による、オーディオ信号を分解するための装置ＭＦ１００のブロック図。
【図２７Ｄ】[0044]別の一般的構成による、オーディオ信号を分解するための装置Ａ１００のブロック図。
【図２８Ａ】[0045]方法Ｍ１００の実装形態Ｍ３００のフローチャート。
【図２８Ｂ】[0046]装置Ａ１００の実装形態Ａ３００のブロック図。
【図２８Ｃ】[0047]装置Ａ１００の別の実装形態Ａ３１０のブロック図。
【図２９Ａ】[0048]方法Ｍ２００の実装形態Ｍ４００のフローチャート。

【図２９Ｂ】[0049]方法Ｍ２００の実装形態Ｍ５００のフローチャート。
【図３０Ａ】[0050]方法Ｍ１００の実装形態Ｍ６００のフローチャート。
【図３０Ｂ】[0051]装置Ａ１００の実装形態Ａ７００のブロック図。
【図３１】[0052]装置Ａ１００の実装形態Ａ８００のブロック図。
【図３２】[0053]モデルＢｆ＝ｙを示す図。
【図３３】[0054]図３２のモデルの変形Ｂ’ｆ＝ｙを示す図。
【図３４】[0055]複数の音源がアクティブであるシナリオを示す図。
【発明を実施するための形態】
【００１１】
[0056]図面を参照し、組み込む以下の詳細な説明は、１つまたは複数の特定の実施形態について説明し、例示する。限定するためではなく、例示し、教示するためだけに提供されるこれらの実施形態について、当業者が特許請求の範囲を実践することを可能にするのに十分詳細に図示し、説明する。したがって、簡潔のために、説明は、当業者に知られているある情報を省略し得る。
【００１２】
[0057]「例示的」という単語は、本開示全体にわたって、「例、事例、または例示の働きをすること」を意味するために使用する。本明細書で「例示的」と記載されたものはどんなものも、必ずしも他の手法または特徴よりも好ましいまたは有利であると解釈されるべきではない。それの文脈によって明確に限定されない限り、「信号」という用語は、本明細書では、ワイヤ、バス、または他の伝送媒体上に表されたメモリロケーション（またはメモリロケーションのセット）の状態を含む、それの通常の意味のいずれかを示すために使用される。
【００１３】
[0058]本明細書で説明するオブジェクト認識技法は多くの異なるシーンに適用され得るが、本明細書で説明する例は、多くの音源、たとえば、ミュージシャン、演奏者、楽器などが単一のシーン中に存在する、音楽シーンに関係する。いくつかのビデオゲーム（たとえば、Ｇｕｉｔａｒ　Ｈｅｒｏ（登録商標）、Ｒｏｃｋ　Ｂａｎｄ（登録商標））およびコンサート音楽シーンは、複数の楽器およびボーカリストが同時にプレイすることを伴い得る。現在の商用ゲームおよび音楽生成システムでは、これらのシナリオから記録されたオーディオを別々に分析し、後処理し、アップミックスすることが可能であるように、それらのシナリオが、連続的にプレイされるか、または近接して配置されたマイクロフォンを用いてプレイされる必要がある。これらの制約は、音楽生成の場合、干渉を制御する能力および／または空間効果を記録する能力を制限し得、ビデオゲームの場合、制限されたユーザエクスペリエンスをもたらし得る。
【００１４】
[0059]どんな楽器がプレイされているのか、およびどれくらいのミュージシャン／音源がシーン中に存在するのかを狭めるのを助ける、何らかのアプリオリ（a priori）な知識または他の情報が利用可能にされた場合、音楽聴覚シーン分解は大幅に簡略化され得る。
【００１５】
[0060]本明細書で開示するオブジェクト認識技法は、複数の音源を有するシーンにおいて記録されたオーディオを分解するための従来の試みの制限の多くを克服する。概して、オブジェクトまたは特徴認識は、特徴識別および／またはオブジェクト認識のために画像中の関心ポイント（キーポイントとも呼ばれる）を識別することおよび／またはそれらのキーポイントの周りの局所的特徴（localized features）を識別することを伴い得る。本明細書で開示するシステムおよび方法では、いくつかの画像ベースの楽器およびオーディオベースのノート／楽器認識技法が組み合わせられる。いくつかの異なるデバイスは、コンピュータビジョンを使用して特徴識別および／またはオブジェクト認識を実行することが可能であり得る。そのようなデバイスの例は、電話ハンドセット（たとえば、セルラーハンドセット）、ビデオ記録が可能なハンドヘルドモバイルデバイス、オーディオおよびビデオコンテンツを記録する個人メディアプレーヤ、携帯情報端末（ＰＤＡ）または他の

ハンドヘルドコンピューティングデバイス、ならびにノートブック、ノートブックコンピュータ、ラップトップコンピュータ、タブレットコンピュータ、または他のポータブルコンピューティングデバイス中に実装され得る。さらに、複数の楽器およびボーカリストが同時にプレイすることを伴い得る、ビデオゲーム（たとえば、Ｇｕｉｔａｒ　Ｈｅｒｏ（登録商標）、Ｒｏｃｋ　Ｂａｎｄ（登録商標））およびコンサート音楽シーンを実行することが可能なデバイス。ポータブルコンピューティングデバイスの種類は現在、ラップトップコンピュータ、ノートブックコンピュータ、ウルトラポータブルコンピュータ、タブレットコンピュータ、モバイルインターネットデバイス、スマートブックおよびスマートフォンなどの名称を有するデバイスを含む。

【００１６】

[0061]第１の方法では、オーディオ知識のみに基づく情報が、シーン中で探索される音源のタイプを定義するのを助け、画像ベースのスケール不変特徴変換（ＳＩＦＴ）探索において考慮されるべきオブジェクト形状のキーポイントシグネチャの探索ユニバースを低減する。さらに、そのようなキーポイント探索は、必ずしも静止画像に制限されるとは限らないが、深さ（範囲）レイヤ探索の必要に応じて単一または複数のカメラを使用する、典型的なミュージシャンの動きパターンのための周囲ビデオフレームの分析をも伴い得る。キーポイント探索は、認識された楽器を関連する尤度で与えるために、マルチモーダルベイズ推定（multi-modal Bayesian estimation）を介して組み込まれる。

【００１７】

[0062]第２の方法では、画像のいくつかの部分中の安定した楽器キーポイントを計算するために、マルチマイクロフォン音定位情報と楽器形状認識の両方から音源（たとえば、楽器）ロケーションが推定される。この方法は、オブジェクト認識を改善するために第１の方法と組み合わせられ得る。

【００１８】

[0063]第３の方法では、第２の方法を使用して判断される情報など、関連するマルチマイクロフォン音源定位情報とともに、音声／オーディオ認識において使用されるメル周波数ケプストラム係数（ＭＦＣＣ：mel-frequency cepstral coefficient）などの音響特徴が、音源認識を行うためにマルチモーダルベイズ推定において直接使用される。第３の方法は、オブジェクト認識を改善するために第１の方法と組み合わせられ得る。

【００１９】

[0064]上記の方法は、たとえば、シーンから記録されたオーディオを分解することに対するスパース復元分解手法（sparse recovery decomposition approach）の場合、基底関数インベントリ（basis function inventory）のサイズのより改良された定義を可能にし得る。

【００２０】

[0065]図１は、例示的な聴覚シーン１０と例示的な聴覚シーン分析システム１２とを示す概念図である。聴覚シーン分析システム１２は、シーン分析システム１４と、マイクロフォンのアレイ１８と、１つまたは複数のカメラ１６とを含む。カメラ１６は、シーン１０に対して様々なロケーションおよび角度に配置された１つまたは複数の静止画像カメラおよび／または１つまたは複数のビデオカメラを含み得る。

【００２１】

[0066]シーン分析システム１４は、オブジェクト認識サブシステム２０と、音響分解サブシステム２２とを含む。オブジェクト認識サブシステム２０は、本明細書で説明する方法に従って、シーンにおいて記録されたオーディオ、（１つまたは複数の）画像および／またはビデオに基づいてシーン１０中の音源を認識し、位置を特定するように構成される。音響分解サブシステム２２は、分離されたオーディオが個々に処理され得るように、オブジェクト認識サブシステム２０からの情報に基づいて、シーンを別個の音源に分解するように構成される。

【００２２】

[0067]図２は、図１の視聴覚シーン分析システム１０を動作させる例示的な方法を示す

プロセスブロック図５０である。本プロセスは、シーン中の１つまたは複数の楽器の位置を特定し、識別するために、シーンにおいて記録された視聴覚情報を分析するためのステップを示している。本方法は、オーディオ分析ブロック５２と、画像および／またはビデオ処理ブロック５４と、オーディオ処理ブロック５６とを含む。

【００２３】

[0068]開示する方法は、キャプチャされたオーディオおよび／またはビデオ信号を一連のセグメントとして処理し得る。典型的なセグメント長は約１～１０秒にわたる。１つの特定の例では、信号は、約１秒の長さをそれぞれ有する一連の重複しないセグメントまたは「フレーム」に分割される。また、そのような方法によって処理されるセグメントは、異なる演算によって処理されるより大きいセグメントのセグメント（すなわち、「サブフレーム」）であり得、またはその逆も同様である。

【００２４】

[0069]オーディオ分析ブロック５２は、シーンにおいて記録されたオーディオ情報に基づいてシーン中の（１つまたは複数の）音源ロケーションを判断するステップを含む（ボックス６１）。オーディオ情報はマイクロフォンアレイ１８によってキャプチャされ得る。音ロケーションは、シーン中の音源および／または音源について判断された範囲情報から音の推定される到来方向（ＤＯＡ：direction of arrival）に基づいて判断され得る。音源のＤＯＡは、本明細書において以下で説明するオーディオＤＯＡ推定技法を使用して推定され得、音源の範囲は、図１８～図２９を参照しながら本明細書において以下で説明する範囲発見技法を使用して推定され得る。

【００２５】

[0070]オーディオ分析ブロック５２はまた、シーン中の各音源に音源の推定タイプを与えるステップを含む（ボックス６２）。たとえば、楽器について、シーンにおいて記録された音は、その音を生成している楽器の可能性のあるタイプと音源を整合させるために、楽器ノートライブラリを使用して分析され得る。

【００２６】

[0071]ボックス６１、６２からの音源ロケーションおよびタイプ推定は、画像／ビデオ処理ブロック５４に受け渡され、音源の視覚的識別のために探索を制限するために使用される。ボックス６４において、推定されたロケーション情報を使用して、キーポイント整合のためにシーンの記録画像中の空間探索空間を狭める。ボックス６６において、画像キーポイント探索が推定楽器タイプに基づいて狭められる。これらのステップの両方は、シーン中の（１つまたは複数の）楽器を識別する信頼性を著しく改善し得、また、（１つまたは複数の）楽器の視覚的認識を行うために必要とされる処理の量を低減し得る。

【００２７】

[0072]ボックス６８において、シーン中の（１つまたは複数の）楽器を識別するために、シーンにおいて記録された画像および／またはビデオデータ上で視覚的オブジェクト認識分析が実行される。この分析は、視覚特徴分析方式、たとえば、シーンのスケール不変特徴変換（ＳＩＦＴ）分析を伴うことができ、分析されるべき画像のキーポイントおよびエリアは、ボックス６１、６２からのオーディオ導出情報に基づいて狭められる。例示的なＳＩＦＴ分析方法の詳細については、本明細書において以下で図７～図１７に関して開示する。

【００２８】

[0073]視覚特徴分析の結果（ボックス７０）は、シーン中の音源（たとえば、楽器）候補とそれらの対応するロケーションとのリストであり、そのリストはオーディオ処理ブロック５６に与えられる。

【００２９】

[0074]オーディオ処理ブロック５６は、記録されたオーディオの品質を向上させるために、別個の音源がより良く分離され、識別され、処理され得るように、シーンから記録されたオーディオをさらに分析してオーディオを分解する。ボックス７２において、画像／ビデオ処理ブロック５２からのロケーション情報を使用して、識別された音源ロケーショ

ンサブセクタの各々のほうへそれぞれ向けられたマルチマイクロフォンアレイのための空間フィルタを発生する。これは、記録されたオーディオデータ中の音源を分離するのを支援する。ボックス７４において、楽器音源の識別を改善するために、シングルチャネル基底関数インベントリベースのスパース復元技法が、分離された音サブセクタの各々に適用される。信号チャネル復元技法は、基底関数インベントリを低減するために楽器カテゴリーノートの低減されたセットを使用することができ、この低減は、画像／ビデオ処理ブロック５４によって与えられた楽器候補のリスト（ボックス７０）によって誘導される。ボックス７０において使用され得る例示的なスパース復元技法については、本明細書において以下で図３０～図３７に関して説明する。

【００３０】

[0075]図３は、図１の聴覚シーン分析システム１２を使用して聴覚シーンを分解する例示的な方法を示すフローチャート２００である。ステップ２０２において、システム１２がオーディオおよび視覚情報（静止画像および／またはビデオ）を記録する。ステップ２０４において、オブジェクト認識サブシステム２０がシーン１０中の音生成オブジェクトのうちの１つまたは複数を識別し、その位置を特定する。ステップ２０６において、音響分解サブシステム２２は音響シーンを別個の音源に分解する。ステップ２０８において、音響分解サブシステム２２は、分離された音に信号チャネル基底関数インベントリベースのスパース復元を適用する。

【００３１】

[0076]図４は、聴覚シーン中の音生成オブジェクトを識別する第１の例示的な方法を示すフローチャート３００である。この方法はオブジェクト認識サブシステム２０によって実行され得る。ステップ３０２において、キャプチャされた画像中のキーポイントを識別する。ステップ３０４において、シーンにおいて記録されたオーディオに基づいて、楽器などの音生成オブジェクトに対応する１つまたは複数のキーポイントシグネチャを選択する。ステップ３０６において、画像中のキーポイントを、選択されたキーポイントシグネチャと比較することによって、シーン中の少なくとも１つのオブジェクトを識別する。

【００３２】

[0077]図５Ａ～図５Ｂに、聴覚シーン中の音生成オブジェクトを識別する第２の例示的な方法のフローチャート４００を示す。この方法はオブジェクト認識サブシステム２０によって実行され得る。ステップ４０２において、キャプチャされた画像中のキーポイントを識別する。ステップ４０４において、識別されたキーポイントから安定したキーポイントを選択する。ステップ４０６において、シーンから記録されたオーディオに基づいて、シーンの画像中の関心領域（ＲＯＩ：region of interest）を選択する。ステップ４０８において、ＲＯＩ中の安定したキーポイントを選択する。

【００３３】

[0078]ステップ４１０において、シーンのビデオから局所動きベクトル（ＬＭＶ：local motion vector）を計算する。ステップ４１２において、ＲＯＩ中のＬＭＶを選択する。

【００３４】

[0079]ステップ４１４において、シーンにおいて記録されたオーディオに基づいて、楽器などの音生成オブジェクトに対応する１つまたは複数のキーポイントシグネチャを選択する。

【００３５】

[0080]ステップ４１６において、シーンからの記録されたオーディオに基づいてオーディオ信頼性値（ＣＶ：confidence value）を計算する。オーディオＣＶは、オーディオ特徴整合デバイス、たとえば、ＭＦＣＣ分類器の出力に基づき得る。オーディオＣＶはベクトルであり得、ベクトルの各要素は、オブジェクトが特定のタイプのオブジェクト、たとえば、トランペット、ピアノなどである尤度を示す。

【００３６】

[0081]ステップ４１８において、シーンのキャプチャされたデジタル画像に基づいて画

像信頼性値（ＣＶ）を計算する。画像ＣＶは、整合デバイス、たとえば、ＳＩＦＴ整合デバイスの出力に基づき得る。ＳＩＦＴ整合デバイスは、画像ＣＶを生成するために、ＲＯＩ中の安定したキーポイントを、選択されたキーポイントシグネチャと比較する。画像ＣＶはベクトルであり得、ベクトルの各要素は、オブジェクトが特定のタイプのオブジェクト、たとえば、トランペット、ピアノなどである尤度を示す。

【００３７】

[0082]ステップ４２０において、シーンからの記録されたビデオに基づいてビデオ信頼性値（ＣＶ）を計算する。ビデオＣＶは、ＲＯＩ中で選択されたＬＭＶを比較するヒストグラム整合プロセスの出力に基づき得る。ビデオＣＶはベクトルであり得、ベクトルの各要素は、オブジェクトが特定のタイプのオブジェクト、たとえば、トランペット、ピアノなどである尤度を示す。

【００３８】

[0083]オーディオＣＶ、画像ＣＶおよびビデオＣＶはそれぞれ正規化され得る。

【００３９】

[0084]ステップ４２２において、オーディオＣＶと画像ＣＶとビデオＣＶとに基づいてシーン中のオブジェクトを識別する。たとえば、最終ＣＶは、オーディオＣＶと画像ＣＶとビデオＣＶとの重み付き和として計算され得る。各ＣＶの重み付け係数は、それぞれの録音モダリティの信号対雑音比（ＳＮＲ）に基づくことができ、特に現在の録音フレームのＳＮＲの関数であり得る。

【００４０】

[0085]モダリティＣＶがベクトルである場合、最終ＣＶもベクトルであり、ベクトルの各要素は、オブジェクトが特定のタイプのオブジェクト、たとえば、トランペット、ピアノなどである尤度を示す。最大尤度を示す要素がオブジェクトを識別する。

【００４１】

[0086]図６は、図１の聴覚シーン分析システム１２とともに使用され得る例示的なオブジェクト認識サブシステム５００のいくつかの構成要素を示すブロック図である。サブシステム５００は、オーディオプロセッサ５０２と、画像プロセッサ５０４と、ビデオプロセッサ５０６と、ＳＩＦＴ整合デバイス５３２と、キーポイントシグネチャデータベース（ＤＢ）５３４と、音響特徴データベース５３６と、音響特徴整合デバイス５３８と、ヒストグラム整合デバイス５４０と、オブジェクト局所動きベクトル（ＬＭＶ）ヒストグラムデータベース５４２と、マルチモーダル分類器５４４とを含む。

【００４２】

[0087]オーディオプロセッサ５０２は、シーンにおいてマイクロフォンアレイ１８からオーディオ信号を受信し、記録する。画像プロセッサ５０４は、シーンのピクチャを撮っている１つまたは複数のカメラ５０８から、シーンの１つまたは複数の画像を受信し、記録する。ビデオプロセッサ５０６は、シーンを記録している１つまたは複数のビデオカメラ５１０から、ビデオ信号を受信し、記録する。

【００４３】

[0088]オーディオプロセッサ５０２は、到来方向（ＤＯＡ）検出器５１２と、関心領域（ＲＯＩ）セレクタ５１４と、音分類器５１６と、音響特徴抽出器５１８とを含む。マイクロフォンアレイ１８から受信されたマイクロフォン信号から、ＤＯＡ検出器５１２は、シーン内に位置する音源から出ている音の到来方向を判断する。ＤＯＡ検出器５１２の例示的な構成要素および機能については、本明細書において図１８～図２５に関して説明する。ＤＯＡとアレイの位置とから、シーン中の音源のロケーションの推定が判断され得る。このＤＯＡ情報はＲＯＩセレクタ５１４に受け渡される。ＲＯＩセレクタ５１４は、ＤＯＡ情報とマイクロフォンアレイ１８の既知の位置とに基づいて音源のロケーションを推定する。ＲＯＩセレクタ５１４は、次いで、ロケーション情報に基づいてシーンの画像の特定の部分を選択する。選択された部分またはＲＯＩは、音源を含んでおり、したがって、キーポイント探索とＬＭＶ計算をシーンの一部分のみに制限するために使用され得る。

【００４４】

[0089]音分類器５１６は、記録されたオーディオの特性に基づいて音源のタイプを分類する。たとえば、音源として楽器のタイプを識別するために、分類器５１６によって楽器ノートライブラリが使用され得る。

【００４５】

[0090]音分類器５１６の出力はオーディオ信頼性値であり、それはキーポイントシグネチャデータベース５３４への入力として与えられる。オーディオ信頼性値に基づいてキーポイントシグネチャデータベース５３４から１つまたは複数のキーポイントシグネチャが選択される。これらの選択されたキーポイントシグネチャはＳＩＦＴ整合デバイス５３２に与えられる。

【００４６】

[0091]音響特徴抽出器５１８は、ＭＦＣＣなど、マイクロフォン信号から導出された音響特性を計算する。これらの抽出された特徴は音響特徴整合デバイス５３８に与えられ、音響特徴整合デバイス５３８は、抽出された特徴を様々なタイプの音源の音響特徴のデータベース５３６と比較することによって音源を識別する。音響特徴整合デバイスの出力は音響特徴信頼性値であり得、この音響特徴信頼性値は、他のＣＶについて上記で説明したのと同様の要素を有するベクトルであり得る。

【００４７】

[0092]画像プロセッサ５０４は、キーポイント検出器５２０と、安定キーポイント検出器５２２と、ＲＯＩキーポイントセレクタ５２４とを含む。キーポイント検出器５２０は、本明細書で説明する方法を使用して、シーンのキャプチャされたデジタル画像中のキーポイントを判断する。安定キーポイント検出器５２２は、キーポイント探索を改善し、安定しているそれらの検出されたキーポイントのみを選択する。ＲＯＩキーポイントセレクタ５２４は、ＲＯＩセレクタ５１４から、キャプチャされた画像中のＲＯＩを識別する座標情報を受信する。この座標情報に基づいて、ＲＯＩキーポイントセレクタは、画像キーポイント選択を、ＲＯＩ内に位置するそれらの安定したキーポイントに狭める。

【００４８】

[0093]ＲＯＩ内で検出された安定したキーポイントは、次いで、ＳＩＦＴ整合デバイス５３２に与えられる。

【００４９】

[0094]本質的に、ＳＩＦＴ整合デバイス５３２は、画像ＣＶを発生するために、安定したＲＯＩキーポイントを、キーポイントシグネチャデータベース５３４から取り出されたキーポイントシグネチャと比較する。

【００５０】

[0095]ビデオプロセッサ５０６は、ＬＭＶ計算器５２６と、ＲＯＩ　ＬＭＶセレクタ５２８と、ＲＯＩ　ＬＭＶヒストグラム計算器５３０とを含む。ＬＭＶ計算器５２６は、（１つまたは複数の）カメラ５１０からデジタルビデオ信号を受信し、シーンの所定の録音持続時間についてＬＭＶを計算する。ＬＭＶは、次いで、ＲＯＩ　ＬＭＶセレクタ５２８に受け渡される。ＲＯＩ　ＬＭＶセレクタ５２８は、ＲＯＩセレクタ５１４からＲＯＩの座標情報を受信し、その座標情報に基づいてＲＯＩ内のそれらのＬＭＶを選択する。

【００５１】

[0096]ＲＯＩ内のＬＭＶは、次いで、ＲＯＩ　ＬＭＶヒストグラム計算器５３０に受け渡され、ＲＯＩ　ＬＭＶヒストグラム計算器５３０は、ＲＯＩからＬＭＶヒストグラムを計算する。シーンのＬＭＶヒストグラムは、次いで、ヒストグラム整合デバイス５４０に受け渡される。ヒストグラム整合デバイス５４０は、最も近接した整合を見つけるために、シーンＬＭＶヒストグラムを、オブジェクトＬＭＶヒストグラムデータベース５４２に記憶されたオブジェクトＬＭＶヒストグラムと比較する。ヒストグラム整合デバイス５４０は、この比較に基づいてビデオＣＶを出力する。

【００５２】

[0097]マルチモーダル分類器５４４は、ＳＩＦＴ整合デバイス５３２と、音分類器５１６と、音響特徴整合デバイス５３８と、ヒストグラム整合デバイス５４０との出力に基づ

いてシーン中のオブジェクトを識別する。マルチモーダル分類器５４４は、オーディオＣＶと画像ＣＶとビデオＣＶと音響特徴ＣＶとの重み付き和であり得る、最終信頼性値ベクトルを計算することによってこれを達成することができる。分類器５４４は、認識された楽器を関連する尤度で与えるためにベイズ推定を実行し得る。ＣＶの重み付け係数は、図４Ａ～図４Ｂに関して説明したものと同様であり得る。

【００５３】

[0098]さらに、サブシステム５００はまた、シーン中の認識された各オブジェクトについて改善されたオブジェクトロケーションを出力し得る。改善された（１つまたは複数の）オブジェクトロケーションは、マルチモーダル分類器５４４、カメラ５０８から出力と、オーディオプロセッサ５０２のＲＯＩセレクタ５１４からの推定オブジェクトロケーションとに基づくことができる。改善された（１つまたは複数の）オブジェクトロケーションは、関心領域またはオブジェクトロケーションを推定する際のそれらの精度および速度を改善するためにＤＯＡ検出器５１２および／またはＲＯＩセレクタ５１４にフィードバックされ得、たとえば、前のビデオ／画像フレームにおいて判断された推定ＤＯＡまたはオブジェクトロケーションは、オーディオプロセッサ５０２がそれのＲＯＩ選択プロセスにおいて使用する初期座標として次のフレームに手渡され得る。

【００５４】

キーポイント選択およびＳＩＦＴ整合デバイス

[0099]例示的なキーポイント検出器５２０、キーポイントセレクタ５２２およびＳＩＦＴ整合デバイス５３２の動作について以下のように説明する。

【００５５】

[00100]概して、オブジェクトまたは特徴認識は、オブジェクト認識のために画像中の関心ポイント（キーポイントとも呼ばれる）を識別することおよび／またはそれらのキーポイントの周りの局所的特徴を識別することを伴い得る。画像データ中のそのような特徴的な要素を本明細書では「キーポイント」と呼ぶが、本明細書で使用するキーポイントという用語は、個々のピクセル、ピクセルのグループ、分数ピクセル部分、１つまたは複数の記述子、他の画像成分、あるいはそれらの任意の組合せを指し得ることを理解されたい。特徴の高い安定性および再現性を有することは、これらの認識アルゴリズムでは非常に重要である。したがって、キーポイントは、それらが画像スケール変化および／または回転に対して不変であり、ひずみ、視点の変化、および／または雑音および照度の変化の実質的な範囲にわたってロバスト（robust）な整合を与えるように、選択および／または処理され得る。さらに、オブジェクト認識などのタスクに好適であるように、特徴記述子は、好ましくは、単一の特徴が、複数のターゲット画像からの特徴の大規模データベースに対して高い確率で正しく整合され得るという意味において特徴的であり得る。

【００５６】

[00101]画像中のキーポイントが検出され、位置を特定された後、それらのキーポイントは、様々な記述子を使用することによって識別または記述され得る。たとえば、記述子は、画像特性の中でも、形状、色、テクスチャ、回転、および／または動きなど、画像中のコンテンツの視覚特徴を表し得る。キーポイントに対応し、記述子によって表される個々の特徴は、次いで、既知のオブジェクトからの特徴のデータベースに整合される。

【００５７】

[00102]画像のためのキーポイントを識別し、選択することの一部として、選択されたいくつかのポイントは、精度または信頼性の欠如により廃棄される必要があり得る。たとえば、いくつかの初期に検出されたキーポイントは、エッジ沿いの不十分なコントラストおよび／または不十分な定位を理由に拒否され得る。そのような拒否は、照度、雑音および配向の変動に対してキーポイント安定性を増加させる際に重要である。また、特徴整合の再現性を減少させ得る、誤ったキーポイント拒否を最小限に抑えることが重要である。

【００５８】

[00103]概して、画像中の照度は、空間的に変動する関数によって表され得ることを認識されたい。したがって、照度の影響（たとえば、シェーディング、明るい画像、暗い画

像など）は、照度関数を排除する正規化プロセスによって特徴／キーポイント検出のために無効にされ得る。たとえば、画像は、画像の平滑化スケール空間（smoothened scale space）Ｌを発生するために平滑化ファクタの範囲をもつ関数Ｇ（すなわち、カーネルまたはフィルタ）を使用して、画像を漸進的（progressively）にぼかすことによって処理され得る。次いで、平滑化スケール空間レベルの隣接するペア間の差分（$L_i - L_{i-1}$）を取ることによって、画像のためのスケール空間差分Ｄが取得され得る。次いで、スケール空間レベルの特定の差分D_iを取得するために使用されるスケール空間レベルL_iのうち最も平滑なスケール空間レベルと同程度に平滑であるかそれよりも平滑である平滑化スケール空間レベルL_kでスケール空間レベルの各差分D_iを除算することによって、スケール空間Ｌの差分の正規化が達成される。

【００５９】

[00104]図７は、記録された画像上でオブジェクト認識を実行するための機能段階を示すブロック図である。画像キャプチャ段階７０２において、関心画像７０８（すなわち、記録画像）がキャプチャされ得る。画像７０８は、デジタルキャプチャ画像を取得するために、１つまたは複数の画像センサーおよび／またはアナログデジタル変換器を含み得る、画像キャプチャデバイスによってキャプチャされ得る。画像センサー（たとえば、電荷結合デバイス（ＣＣＤ）、相補型金属半導体（ＣＭＯＳ））は光を電子に変換し得る。電子はアナログ信号を形成し得、次いで、そのアナログ信号は、アナログデジタル変換器によってデジタル値に変換される。このようにして、画像Ｉ（ｘ，ｙ）を、たとえば、対応する色、照度、および／または他の特性をもつ複数のピクセルとして定義し得るデジタルフォーマットで画像７０８はキャプチャされ得る。

【００６０】

[00105]画像処理段階７０４において、キャプチャされた画像７０８は、次いで、対応するスケール空間７１０（たとえば、ガウススケール空間）を発生し、特徴検出７１２を実行し、特徴記述子抽出７１６を実行することによって処理される。特徴検出７１２は、キャプチャされた画像７０８について高度に特徴的なキーポイントおよび／または幾何学的関心キーポイントを識別し得、それらのキーポイントは、その後、特徴記述子抽出７１６において複数の記述子を取得するために使用され得る。画像比較段階７０６において、これらの記述子は、既知の記述子のデータベースとの（たとえば、キーポイントおよび／またはキーポイントの他の特性あるいはキーポイントを囲むパッチを比較することによる）特徴整合７２２を実行するために使用される。次いで、特徴整合が正しいことを確認するために、キーポイント整合に対する幾何学的検証または一致性検査７２４が実行されて、整合結果７２６が与えられる。このようにして、記録画像が、ターゲット画像のデータベースと比較されおよび／またはそれから識別され得る。

【００６１】

[00106]画像中の照度の変化は、画像のための特徴／キーポイント認識の安定性および／または再現性に有害な影響を及ぼし得ることが観測されている。すなわち、画像中の局所および／または大域（global）照度変化は、画像のための特徴／キーポイントの検出に影響を及ぼすことがある。たとえば、特徴／キーポイントの数および／またはロケーションが、画像中の照度（たとえば、シェーディング、コントラストなど）に応じて変化し得る。したがって、画像中の特徴／キーポイント検出より前に、局所および／または大域照度変化の影響を実質的になくすかまたは最小限に抑えることが有益であろう。

【００６２】

[00107]これを行うための１つの方法は、特徴／キーポイント検出を開始するより前に、局所および／または大域照度変化を除去または補償するように画像自体を処理することであり得る。しかしながら、そのようなプロセスは計算集約的であり得る。さらに、画像中に局所および／または大域照度変化が存在するかどうかを判断することがしばしば困難である。そのようなプロセスは、データベース中の画像にも適用されなければならないであろう。照度変化を補正するためにターゲット画像とデータベース画像の両方を最初に処理することなしには、特徴／キーポイント整合は成功しないことがある。しかし、照度が

特定の画像にどのように影響を及ぼし得るかの事前知識なしには、このプロセスは自動的に実装することがかなり困難である。

【００６３】

[00108]したがって、実質的な処理オーバーヘッドなしに実行され得る代替形態が必要とされる。一例によれば、特徴検出の目的での画像上の（一様あるいは非一様な）照度の影響は、スケール空間差分に特徴／キーポイント検出が実行されるより前にスケール空間差分を正規化することによって、なくされるかまたは低減され得る。この正規化プロセスは、すでに利用可能である平滑化スケール空間を使用して実行され、したがって、追加の計算が最小限に抑えられ得る。

【００６４】

[00109]一例によれば、スケール空間正規化器７１４は、照度変化が画像中のキーポイント／特徴認識に及ぼす影響を低減するかまたはなくすために、スケール空間発生７１０の一部として実装され得る。

【００６５】

[00110]図８に、例示的な画像処理段階７０４におけるガウススケール空間発生を示す。画像中の特徴検出を実行するために、スケール不変特徴変換（ＳＩＦＴ）など、いくつかのアルゴリズムが開発されている。画像中の特定のオブジェクトの検出への第１のステップは、それの局所特徴に基づいて記録されたオブジェクトを分類することである。その目的は、たとえば、照度、画像雑音、回転、スケーリング、および／または視点の小さい変化に対して不変および／またはロバストである特徴を識別し、選択することである。すなわち、クエリ画像と比較ターゲット画像との間に照度、画像雑音、回転、スケール、および／または視点の差があるにもかかわらず、これらの２つの画像間の整合が発見されなければならない。これを行うための１つの方法は、高度に特徴的な特徴（たとえば、画像中の特徴的なポイント、ピクセル、および／または領域）を識別するために画像のパッチ上の極値検出（たとえば、極大値または極小値）を実行することである。

【００６６】

[00111]ＳＩＦＴは、照度、画像雑音、回転、スケーリングの変化、および／または視点の小さい変化に対して適度に不変である局所特徴を検出し、抽出するための１つの手法である。ＳＩＦＴの画像処理段階７０４は、（ａ）スケール空間極値検出、（ｂ）キーポイント定位、（ｃ）配向割当て、および／または（ｄ）キーポイント記述子の発生を含み得る。特に、高速ロバスト特徴（ＳＵＲＦ：Speed Up Robust Features）、勾配位置および配向ヒストグラム（ＧＬＯＨ：Gradient Location and Orientation Histogram）、局所エネルギーベース形状ヒストグラム（ＬＥＳＨ：Local Energy based Shape Histogram）、勾配の圧縮ヒストグラム（ＣＨｏＧ：Compressed Histogram of Gradients）を含む、特徴検出と、後続の特徴記述子発生とのための代替アルゴリズムは、本明細書で説明する特徴からも恩恵を受け得ることが明らかなはずである。

【００６７】

[00112]ガウススケール空間発生７１０において、デジタル画像Ｉ（ｘ，ｙ）７０８は漸進的にガウスぼかし（すなわち、平滑化）されて、ガウスピラミッド７５２が構成される。ガウスぼかし（平滑化）は、概して、元の画像Ｉ（ｘ，ｙ）をスケールｃｓにおけるガウスぼかし／平滑化関数Ｇ（ｘ，ｙ，ｃｓ）で畳み込み、したがって、ガウスぼかし／平滑化関数Ｌ（ｘ，ｙ，ｃｓ）がＬ（ｘ，ｙ，ｃｓ）＝Ｇ（ｘ，ｙ，ｃｓ）＊Ｉ（ｘ，ｙ）として定義されることを伴う。ここで、Ｇはガウスカーネルであり、ｃｓは、画像Ｉ（ｘ，ｙ）をぼかすために使用されるガウス関数の標準偏差を示す。乗数ｃが変化するにつれて（$c_0 < c_1 < c_2 < c_3 < c_4$）、標準偏差ｃｓは変化し、漸進的なぼかし／平滑化が得られる。シグマｓは、ベーススケール変数（たとえば、ガウスカーネルの幅）である。高いスケール（すなわち、低い解像度）ほど、より低いスケール（すなわち、より高い解像度）よりもぼかされる／平滑化される。したがって、広いスケールレベル（すなわち、低い解像度）ほど、画像はより平滑になる（よりぼかされる）。

【００６８】

[00113]ぼけた画像Ｌを生成するために初期画像Ｉ（ｘ，ｙ）がガウシアンＧで増分的に畳み込まれるとき、ぼけた画像Ｌは、スケール空間において定数ファクタｃだけ分離される。ガウスぼかしされた（平滑化された）画像Ｌの数が増加し、ガウスピラミッド７５２のために与えられる近似が連続空間に近づくにつれて、これらの２つのスケールも１つのスケールに近づく。一例では、畳み込まれた画像Ｌはオクターブによってグループ化され得、１オクターブは、標準偏差ｓの値の倍化に対応し得る。その上、乗数ｃ（たとえば、$c_0 < c_1 < c_2 < c_3 < c_4 \ldots$）の値は、固定数の畳み込まれた画像Ｌがオクターブごとに取得されるように選択される。スケーリングの各オクターブは、明示的な画像サイズ変更に対応する。したがって、元の画像Ｉ（ｘ，ｙ）が漸進的ぼかし／平滑化関数によってぼかされる／平滑化されるにつれて、ピクセルの数は漸進的に低減される。本明細書では説明のためにガウス平滑化関数を使用したが、他のタイプの平滑化カーネル／関数が採用され得ることに留意されたい。

【００６９】

[00114]ガウスピラミッド７５２中の任意の２つの連続するガウスぼかし画像の差分を計算することによって、ガウス差分（ＤｏＧ：difference of Gaussian）ピラミッド７５４が構成される。ＤｏＧ空間７５４において、Ｄ（ｘ，ｙ，ａ）＝Ｌ（ｘ，ｙ，$c_n s$）－Ｌ（ｘ，ｙ，$c_{n-1} s$）である。ＤｏＧ画像Ｄ（ｘ，ｙ，ｓ）は、スケール$c_n s$および$c_{n-1} s$における２つの隣接するガウスぼかし画像Ｌ間の差分である。Ｄ（ｘ，ｙ，ｓ）のスケールは、$c_n s$と$c_{n-1} s$との間のどこかにある。ＤｏＧ画像Ｄが、オクターブごとに隣接するガウスぼかし画像Ｌから取得され得る。各オクターブ後に、ガウス画像が２分の１にダウンサンプリングされ、次いでこのプロセスが繰り返される。このようにして、画像は、並進、回転、スケール、および／または他の画像パラメータおよび／またはひずみに対してロバストまたは不変である局所特徴に変換され得る。

【００７０】

[00115]記録画像のＤｏＧ空間７５４は、発生されると、関心特徴を識別する（たとえば、画像中の高度に特徴的なポイントを識別する）ための極値検出のために利用され得る。これらの高度に特徴的なポイントは、本明細書ではキーポイントと呼ばれる。これらのキーポイントは、各キーポイントを囲むパッチまたは局所領域の特性によって識別され得る。記述子が、キーポイントおよびそれの対応するパッチごとに生成され得、それは、クエリ画像と記憶されたターゲット画像との間のキーポイントの比較のために使用され得る。「特徴」は、記述子（すなわち、キーポイントおよびそれの対応するパッチ）を指し得る。特徴（すなわち、キーポイントおよび対応するパッチ）のグループはクラスタと呼ばれることがある。

【００７１】

[00116]図９に、例示的な画像処理段階７０４における特徴検出を示す。特徴検出７１２において、ＤｏＧ空間７５４を使用して画像Ｉ（ｘ，ｙ）のキーポイントを識別し得る。特徴検出７１２は、画像中の特定のサンプルポイントかピクセルの周りの局所領域またはパッチが、（幾何学的に言って）潜在的に関心のあるパッチであるかどうかを判断しようとする。

【００７２】

[00117]概して、ＤｏＧ空間７５４中に極大値および／または極小値が識別され、これらの極大値および極小値のロケーションがＤｏＧ空間７５４中のキーポイントロケーションとして使用される。図９に示す例では、キーポイント７６０はパッチ７５８で識別されている。極大値および極小値を発見すること（局所的極値検出としても知られる）は、ＤｏＧ空間７５４中の各ピクセル（たとえば、キーポイント７６０に対するピクセル）を、それの８つの隣接するピクセルと、同じスケールで比較し、ならびに（隣接するパッチ７５６および７６２中の）９つの隣接するピクセルと、キーポイント８０８の両側に隣接するスケールの各々で比較し、合計２６個のピクセル（９×２＋８＝２６）に対して比較することによって達成され得る。ここで、パッチは３×３ピクセル領域として定義される。概して、キーポイント７５８に対するピクセル値が、パッチ７５８、７５６、および７６

０中のすべての２６個の比較されたピクセルの間で最大値または最小値である場合、それがキーポイントとして選択される。キーポイントは、それらのロケーションがより正確に識別されるようにさらに処理され得、低コントラストキーポイントおよびエッジキーポイントなど、キーポイントのうちのいくつかが廃棄され得る。

【００７３】

[00118]図１０に、例示的な画像処理段階７０４における特徴記述子抽出を示す。概して、特徴（たとえば、キーポイントおよびそれの対応するパッチ）は記述子によって表され得、記述子は、（クエリ画像からの）特徴と、ターゲット画像のデータベースに記憶された特徴との効率的な比較を可能にする。特徴記述子抽出７１６の一例では、各キーポイントは、局所画像勾配の方向に基づいて、１つまたは複数の配向、または方向を割り当てられ得る。局所画像特性に基づいて各キーポイントに一貫した配向を割り当てることによって、キーポイント記述子は、この配向に対して表され、したがって、画像回転に対する不変性を達成することができる。ガウスぼかし画像Ｌ中でおよび／またはキーポイントスケールにおいて、キーポイント７６０の周りの隣接する領域中のピクセルごとに大きさおよび方向の計算が実行され得る。（ｘ，ｙ）に位置するキーポイント７６０に対する勾配の大きさはｍ（ｘ，ｙ）として表され得、（ｘ，ｙ）におけるキーポイントに対する勾配の配向または方向はΓ（ｘ，ｙ）として表され得る。キーポイントのスケールを使用して、すべての計算がスケール不変方式で実行されるように、キーポイント７６０のスケールに最も近いスケールで、ガウス平滑化された画像Ｌを選択する。各画像サンプルＬ（ｘ、ｙ）について、このスケールで、勾配の大きさｍ（ｘ，ｙ）と配向Γ（ｘ，ｙ）とが、ピクセル差分を使用して計算される。たとえば、大きさｍ（ｘ，ｙ）は次のように計算され得る。

【数１】

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \qquad (式1)$$

【００７４】

方向または配向Γ（ｘ，ｙ）は次のように計算され得る。

【数２】

$$\Gamma(x,y)=\arctan\left[\frac{(L(x,y+1)L(x,y-1)}{(L(x+1,y)-L(x_{\cdot}-1,y)}\right] \qquad (式2)$$

【００７５】

ここで、Ｌ（ｘ，ｙ）は、キーポイントのスケールでもあるスケールｓにおける、ガウスぼかし画像Ｌ（ｘ，ｙ，ｓ）のサンプルである。

【００７６】

[00119]キーポイント７６０に対する勾配は、ＤｏＧ空間中のキーポイントの平面より上に、より高いスケールで存在するガウスピラミッド中の平面に対して、またはキーポイントより下に、より低いスケールで存在するガウスピラミッドの平面中のいずれかで、一貫して計算され得る。どちらにしても、各キーポイントについて、勾配は、キーポイントを囲む矩形エリア（たとえば、パッチ）中ですべて１つの同じスケールで計算される。その上、画像信号の周波数は、ガウスぼかし画像のスケールに反映される。しかし、ＳＩＦＴは、単にパッチ（たとえば、矩形エリア）中のすべてのピクセルにおいて勾配値を使用する。パッチがキーポイントの周りで定義され、サブブロックがブロック内で定義され、

サンプルがサブブロック内で定義され、この構成は、キーポイントのスケールが異なるときでさえ、すべてのキーポイントに対して同じままである。したがって、画像信号の周波数は、同じオクターブ中のガウス平滑化フィルタの連続適用とともに変化する一方で、異なるスケールにおいて識別されたキーポイントは、スケールで表される画像信号の周波数の変化にかかわらず、同じサンプル数でサンプリングされ得る。

【００７７】

[00120]キーポイント配向を特徴づけるために、（ＳＩＦＴでは）キーポイント７６０の近傍において（キーポイントのスケールに最も近接したスケールにおけるガウス画像を使用して）勾配配向のベクトルが生成され得る。しかしながら、キーポイント配向は、たとえば、勾配の圧縮ヒストグラム（ＣＨｏＧ）を使用することによって、勾配配向ヒストグラム（図１０参照）によっても表され得る。隣接する各ピクセルの寄与は、勾配の大きさとガウス窓とによって重み付けされ得る。ヒストグラムのピークは支配的な配向に対応する。キーポイントのすべての特性はキーポイント配向に対して測定され得、これにより、回転に対する不変性が与えられる。

【００７８】

[00121]一例では、各ブロックについてガウス重み付け勾配の分布が計算され得、各ブロックは、２サブブロック×２サブブロックで合計４サブブロックである。ガウス重み付け勾配の分布を計算するために、いくつかのビンをもつ配向ヒストグラムが形成され、各ビンはキーポイントの周りのエリアの部分をカバーする。たとえば、配向ヒストグラムは３６個のビンを有し得、各ビンは配向の３６０度範囲のうちの１０度をカバーする。代替的に、ヒストグラムは８つのビンを有し得、各ビンは３６０度範囲のうちの４５度をカバーする。本明細書で説明するヒストグラムコーディング技法は、任意の数のビンのヒストグラムに適用可能であることが明らかなはずである。ヒストグラムを最終的に生成する他の技法も使用され得ることに留意されたい。

【００７９】

[00122]勾配分布および配向ヒストグラムは様々な方法で取得され得る。たとえば、２次元勾配分布（ｄｘ，ｄｙ）（たとえば、ブロック８０６）が１次元分布（たとえば、ヒストグラム８１４）に変換される。キーポイント７６０は、キーポイント７６０を囲むパッチ８０６（セルまたは領域とも呼ばれる）の中心に位置する。各レベルのピラミッドについて事前計算された勾配が、各サンプルロケーション８０８において小さい矢として示されている。図示のように、サンプル８０８の４×４領域はサブブロック８１０を形成し、サブブロックの２×２領域はブロック８０６を形成する。ブロック８０６は記述子窓と呼ばれることもある。ガウス重み付け関数は、円８０２で示され、各サンプルポイント８０８の大きさに重みを割り当てるために使用される。円形窓８０２中の重みは平滑に低下する。ガウス窓８０２の目的は、窓の位置の小さな変化によって記述子が突然変化することを回避し、記述子の中心から遠い勾配にあまり重点を与えないことである。２×２サブブロックから配向ヒストグラム８１２の２×２＝４アレイが取得され、ヒストグラムの各ビン中に８つの配向があり、それにより（２×２）×８＝３２次元の特徴記述子ベクトルが得られる。たとえば、配向ヒストグラム８１３および８１５は、サブブロック８１０に対する勾配分布に対応し得る。しかしながら、各ヒストグラム中に８つの配向をもつヒストグラム（８ビンヒストグラム）の４×４アレイを使用して、それにより各キーポイントについて（４×４）×８＝１２８次元の特徴記述子ベクトルが得られると、より良好な結果が与えられ得る。勾配分布を取得するために、他のタイプの（たとえば、異なるボロノイセル構造を用いた）量子化ビンコンスタレーションも使用され得ることに留意されたい。

【００８０】

[00123]本明細書で使用するヒストグラムは、ビンとして知られている様々な独立したカテゴリーに分類される観測、サンプル、または出現（たとえば、勾配）の数を計数するマッピングk_iである。ヒストグラムのグラフは、ヒストグラムを表すための１つの方法にすぎない。したがって、ｋが観測、サンプル、または出現の総数であり、ｍがビンの総

数である場合、ヒストグラムk_iにおける周波数は下記の条件を満たす。
【数３】

$$n = \sum_{i=1}^{m} k_i \qquad (式3)$$

【００８１】
ただし、Σは総和演算子である。
【００８２】
[00124]キーポイントに対する特徴記述子ベクトルを取得するために、サブブロックからのヒストグラムは連結され得る。１６個のサブブロックからの８ビンヒストグラム中の勾配が使用される場合、１２８次元の特徴記述子ベクトルが得られ得る。
【００８３】
[00125]このようにして、記述子は、識別されたキーポイントごとに取得され得、そのような記述子は、ロケーション（ｘ，ｙ）と、配向と、ガウス重み付け勾配の分布の記述子とによって特徴づけられ得る。画像は、１つまたは複数のキーポイント記述子（画像記述子とも呼ばれる）によって特徴づけられ得ることに留意されたい。さらに、記述子はまた、ロケーション情報（たとえば、キーポイントの座標）、スケール（たとえば、キーポイントが検出されたガウススケール）、およびクラスタ識別子などの他の情報などを含み得る。
【００８４】
[00126]ガウス差分空間７５４中で演算することによって、画像のルミナンスのいかなるレベルシフト（ルミナンスへの空間的に一様な加法的バイアス）も完全に無視される。しかし、ルミナンスのスケールシフトは、キーポイントが判定され、最終的に選択または拒否される方法に影響を及ぼす。これは、一様な乗法的ルミナンスファクタ、ならびに空間的に変動する乗法的ルミナンスファクタの両方に当てはまる。キーポイント検出とまさに同程度に重要であるのが、画像内のそれの定位である。オブジェクトは、それの特徴の幾何学的コンテンツと、それらの空間的相互関係とによってカテゴリー分類される。したがって、キーポイントが検出された場合でも、それの定位はルミナンススケール変化に関して不変の方法で計算されるべきであるように、キーポイントの計算されたロケーションは重要な要素である。
【００８５】
[00127]したがって、キーポイントを識別し、記述子を生成するより前に、キーポイントが検出されたスケール空間から照度の影響を低減、除去、および／またはフィルタ処理するために、１つの特徴がガウス差分空間７５４を正規化することを行う。
【００８６】
スケール空間正規化の例示的な差分
[00128]図１１に、照度不変特徴（illumination invariant feature）／キーポイント検出のためのスケール空間正規化の一例を示す。画像Ｉ（ｘ，ｙ）８２２は、平滑化されたスケール空間ピラミッド８２６を発生するために、異なるスケール$c_i\sigma$において平滑化カーネルＧ（ｘ，ｙ，$c_i\sigma$）８２４で畳み込まれ得、ただし、ｉは０とｎとの間の整数である。平滑化カーネルはガウスカーネルおよび／または他のタイプの平滑化関数であり得ることに留意されたい。スケール空間差分８２８を取得するために、平滑化されたスケール空間ピラミッド８２６の２つの隣接するスケール空間の間の差分が取られ得る。
【００８７】
[00129]最初に、スケール空間差分８２８の各レベルは、画像Ｉ（ｘ，ｙ）８２２で畳み込まれた異なるスケールにおける平滑化カーネル８２４の差分（たとえば、Ｇ（ｘ，ｙ，$c_{j+1}\sigma$）－Ｇ（ｘ，ｙ，$c_j\sigma$））として定義され得ることがわかる。これは、２つの対応する平滑化スケール空間差分（たとえば、Ｌ（ｘ，ｙ，$c_{j+1}\sigma$）－Ｌ（ｘ，ｙ，$c_j\sigma$））に等しい。したがって、２つの平滑化スケール空間差分は次のように表され得る。

【数４】

$$D(x,y,\sigma)=(G(x,y,c_{j+1}\sigma)-G(x,y,c_j\sigma))*I(x,y)=L(x,y,c_{j+1}\sigma)-L(x,y,c_j\sigma)$$

(式4)

【００８８】

[00130]また、照度がスケーリング関数Ｓ（ｘ，ｙ）として表される場合、２つの平滑化スケール空間差分に対する照度変化は次のように表され得ることがわかる。

【数５】

$$D(x,y,\sigma)=(G(x,y,c_{j+1}\sigma)-G(x,y,c_j\sigma))*(I(x,y)S(x,y))$$

(式5)

【００８９】

ここで、一般的な場合、照度スケーリング関数Ｓ（ｘ，ｙ）は、空間的に変動するか、または空間的に一定であり得る。

【００９０】

[00131]しかしながら、照度スケーリング関数Ｓ（ｘ，ｙ）を取得するために照度をランタイムでモデル化することは実際的および／または実現可能でない。したがって、本明細書では、特徴選択および／またはプルーニングが一般に実行される特徴空間（たとえば、ＤｏＧ空間８２８）からの照度によってバイアスされない、基礎をなす特徴（たとえば、キーポイント）を引き出す代替手法が開示される。この代替手法によれば、画像Ｉ（ｘ，ｙ）８２２のルミナンス分布は、画像Ｉ（ｘ，ｙ）８２２から抽出されたスケール空間情報を利用することによって正規化される。照度に関する事前情報は必要とされない。この方法は、何らかの大きい計算および処理を導入することなしに、異なる照度変化にわたって一貫したレベルで安定した特徴を選定することと再現性を高めることとを可能にする。

【００９１】

[00132]これを行うために、特徴検出が行われるスケール空間差分８２８は、より広いスケール空間によって正規化され得る。この手法は次式によって定義され得る。

【数６】

$$D'(x,y,\sigma)=\left[\frac{[G(x,y,c_{j+1}\sigma)-G(x,y,c_j\sigma)]*[I(x,y)S(x,y)]}{G(x,y,c_{j+1+h}\sigma)*[I(x,y)S(x,y)]}\right]$$

(式6)

【００９２】

ただし、

第１のガウス平滑化カーネルＧ（ｘ，ｙ，$c_{j+1}\sigma$）は第２のガウス平滑化カーネルＧ（ｘ，ｙ，$c_j\sigma$）よりも広く（すなわち、スケール$c_{j+1}s$はスケール$c_j s$よりも広く、ただし、ｊは０とｎとの間の正の整数であり）、

Ｉ（ｘ，ｙ）は、処理されている画像またはそれの派生物（たとえば、画像の反射特性）であり、

Ｓ（ｘ，ｙ）は照度スケーリング関数であり、

Ｇ（ｘ，ｙ，$c_{j+1+h}\sigma$）は、第２の平滑化カーネルＧ（ｘ，ｙ，$c_{j+1}\sigma$）と同程度に広いかまたはそれよりも広いスケール空間を有する第３の平滑化カーネルであり、ただし、ｈは０とｎとの間の正の整数である。スケール空間差分８２８のあらゆるレベル上でこのプロセスを繰り返すことによって、正規化されたスケール空間８３０の差分が発生され得る。たとえば、Ｇ（ｘ，ｙ，$c_{j+1}\sigma$）とＧ（ｘ，ｙ，$c_{j}\sigma$）とによって定義される差分スケール空間では、正規化関数はＧ（ｘ，ｙ，$c_{j+1}\sigma$）またはそれより高い任意のもの（すなわち、Ｇ（ｘ，ｙ，$c_{j+2}\sigma$）、Ｇ（ｘ，ｙ，$c_{j+3}\sigma$）、．．．）であり得る。正規化関数は、差分スケール空間中で使用される両方の平滑化カーネルよりも大きい必要はなく、それは平滑器である必要のみがある。別の例では、正規化関数は、使用される第１の平滑化カーネルと第２の平滑化カーネルとの和（すなわち、Ｇ（ｘ，ｙ，$c_{j+1}\sigma$）＋Ｇ（ｘ，ｙ，$c_{j}\sigma$））であり得、したがって、

【数７】

$$D'(x,y,\sigma)=\left[\frac{[G(x,y,c_{j+1}\sigma)-G(x,y,c_j\sigma)]*[I(x,y)S(x,y)]}{[G(x,y,c_{j+1}\sigma)+G(x,y,c_j\sigma)]*[I(x,y)S(x,y)]}\right] \quad (式7)$$

【００９３】

[00133]式６は次のようにも表され得ることに留意されたい。

【数８】

$$D'(x,y,\sigma)=\left[\frac{[L(x,y,c_{j+1}\sigma)-L(x,y,c_j\sigma)]*S(x,y)}{L(x,y,c_{j+1+h}\sigma)*S(x,y)}\right] \quad (式8)$$

【００９４】

照度スケーリング関数Ｓ（ｘ，ｙ）は（式６、式７および式８の）分子と分母の両方に現れるので、それのスケーリングの影響は相殺される。すなわち、照度スケーリング関数Ｓ（ｘ，ｙ）は正規化のために使用される平滑化画像Ｌ（ｘ，ｙ，$c_{j+1+h}\sigma$）＊Ｓ（ｘ，ｙ）中に存在するので、それは、スケール空間差分［Ｌ（ｘ，ｙ，$c_{j+1}\sigma$）－Ｌ（ｘ，ｙ，$c_{j}\sigma$）］＊Ｓ（ｘ，ｙ）における照度スケーリング関数Ｓ（ｘ，ｙ）の影響を完全にまたは実質的に相殺する。前述のように、Ｌ（ｘ，ｙ，$c_{j+1+h}\sigma$）は、Ｌ（ｘ，ｙ，$c_{j+1}\sigma$）またはより高いスケール画像（すなわち、Ｌ（ｘ，ｙ，$c_{j+2}\sigma$）、Ｌ（ｘ，ｙ，$c_{j+3}\sigma$）、．．．）に等しくなり得る。このようにして、分母中の画像コンテンツは、それがごくわずかな空間アーティファクトしか導入しない程度まで平滑化される。

【００９５】

[00134]スケール空間差分を正規化する際に、正規化する平滑化画像Ｌ（ｘ，ｙ，$c_{j+1+h}\sigma$）は、（キーポイント／特徴を識別する）局所的極値位置をシフトしないように、特徴空間（すなわち、スケール空間差分）をあまりに多く変化させないように選択されなければならない。すなわち、スケール不変特徴を達成するためにはスケール空間差分が最良であることが知られているので、スケール空間差分の密接な近似が正規化後に保持されなければならない。この目的で、平滑化画像Ｌ（ｘ，ｙ，$c_{j+1+h}\sigma$）は、高周波数成分が平均されるようにそれのスケールレベルが十分に平滑であるように選択される。すなわち、平滑化画像Ｌ（ｘ，ｙ，$c_{j+1+h}\sigma$）が十分に平坦である場合、スケール空間Ｌの差分（ｘ，ｙ，$c_{j+1}\sigma$）－Ｌ（ｘ，ｙ，$c_{j}\sigma$）の形状はほとんど変化しない（すなわち、特徴／キーポイントの位置は変化しない）。一実施形態では、正規化されている差分スケールレベルを取得するために使用されるスケールレベルに近接した（それと同じであるかまたはそれの次に最も高い）スケールレベルにおける正規化関数を選択することは、多すぎる雑音を導入することを回避するので、好適であり得ることに留意されたい。たとえば、

Ｇ（ｘ，ｙ，ｃ$_{j+1}$σ）とＧ（ｘ，ｙ，ｃ$_{j}$σ）とによって定義される差分スケール空間のためにＧ（ｘ，ｙ，ｃ$_{j+1}$σ）のような平滑スケールを選ぶことによって、スケール空間中のその特定のレベルについて典型的な局所不規則性が維持され得る。

【００９６】

[00135]前記のように、画像中で検出される特徴の数は、画像の乗法的ルミナンススケール変化によって大幅に影響を受け得る（たとえば、低減され得る）。ルミナンスによって生じるスケーリングは、幾何学的変換がなくても最終の特徴空間中のコンテンツを大幅に低減する、画像上のマスクのように働く傾向がある。したがって、式６および式７の適用によって達成される正規化により、照度変化にかかわらず幾何学的有意性が「等しい」特徴が検出され、それによって再現性が増加することが保証される。

【００９７】

[00136]図１２に、正規化されたスケール空間差分の１つのレベルがどのように取得され得るかの一例を示す。ここで、画像Ｉ（ｘ，ｙ）８５２は、第１の平滑化スケール空間画像Ｌ（ｘ，ｙ，ｃ$_{j}$σ）８５８を取得するために、第１の平滑化カーネルＧ（ｘ，ｙ，ｃ$_{j}$σ）８５４で畳み込まれ得る。画像Ｉ（ｘ，ｙ）８５２はまた、第２の平滑化スケール空間画像Ｌ（ｘ，ｙ，ｃ$_{j+1}$σ）８６０を取得するために、第２の平滑化カーネルＧ（ｘ，ｙ，ｃ$_{j+1}$σ）８５６で畳み込まれ得る。第２の平滑化画像８６０と第１の平滑化画像８５８との間の差分が取られて、スケール空間レベルの差分Ｄ$_{j}$（ｘ，ｙ，σ）８６２が取得され得る。このスケール空間レベルの差分Ｄ$_{j}$（ｘ，ｙ，σ）８６２は、より高いスケール平滑化カーネルＧ（ｘ，ｙ，ｃ$_{j+1+h}$σ）８６６または平滑化スケール空間画像Ｌ（ｘ，ｙ，ｃ$_{j+1+h}$σ）８６８に基づいて（すなわち、式６および／または式７に従って）正規化されて、正規化スケール空間レベルＤ’$_{j}$（ｘ，ｙ，σ）８６４が取得され得る。このプロセスは、（スケーリングファクタｃ$_{j}$σによって設定される）異なる幅の異なる平滑化カーネルを画像Ｉ（ｘ，ｙ）に適用することによって繰り返され、それにより平滑化されたスケール空間ピラミッドが構築され得る。スケール空間差分（たとえば、図１１中の８２８）は、平滑化されたスケール空間ピラミッド（たとえば、図１１中の８２６）の隣接するレベル間の差分を取ることによって構築され得る。正規化されたスケール空間差分（たとえば、図１１中の８３０）は、式６および／または式７に従って発生され得る。

【００９８】

[00137]図１３に、照度の変化に対して耐性がある改善された特徴検出のために正規化されたスケール空間差分を発生するための方法を示す。９０２において、ｉ＝０～ｎについて、平滑化されたスケール空間ピラミッドを構成する複数の平滑化画像Ｌ（ｘ，ｙ，ｃ$_{i}$σ）を取得するために、（ｉ＝０～ｎについて、異なるｃ$_{i}$σによって設定される）異なるスケーリング幅の平滑化カーネルＧ（ｘ，ｙ，ｃ$_{i}$σ）で画像Ｉ（ｘ，ｙ）を畳み込む。画像Ｉ（ｘ，ｙ）は、照度関数Ｓ（ｘ，ｙ）によって完全にまたはピクセルごとに変更されているベース画像Ｉ$_{0}$（ｘ，ｙ）によって特徴づけられ得る。一例では、平滑化カーネルＧ（ｘ，ｙ，ｃ$_{i}$σ）は、平滑化されたスケール空間ピラミッドがガウススケール空間ピラミッドであるように、ガウスカーネルであり得る。

【００９９】

[00138]次に、９０４において、ｊ＝０～ｎ－１について、平滑化されたスケール空間ピラミッドにわたって平滑化画像の隣接するペアの差分Ｌ（ｘ，ｙ，ｃ$_{j+1}$σ）－Ｌ（ｘ，ｙ，ｃ$_{j}$σ）を取ることによって、スケール空間差分Ｄ$_{j}$（ｘ，ｙ，σ）を発生する。このプロセスは、複数のレベルを有するスケール空間差分を取得するために、隣接する平滑化画像の複数のセットについて繰り返される。第２の平滑化画像Ｌ（ｘ，ｙ，ｃ$_{j+1}$σ）を取得するために使用される第２の平滑化カーネルＧ（ｘ，ｙ，ｃ$_{j+1}$σ）は、第１の平滑化画像Ｌ（ｘ，ｙ，ｃ$_{j+1}$σ）を取得するために使用される第１の平滑化カーネルＧ（ｘ，ｙ，ｃ$_{j}$σ）よりも広くなり得ることに留意されたい。

【０１００】

[00139]次いで９０６において、ｊ＝０～ｎ－１について、スケール空間の各差分Ｄ$_{j}$（

ｘ，ｙ，σ）レベルを対応する平滑化画像Ｌ（ｘ，ｙ，c_{j+1+h}σ）で除算することによって正規化されたスケール空間差分Ｄ’$_j$（ｘ，ｙ，σ）を発生し、各平滑化画像Ｌ（ｘ，ｙ，c_{j+1+h}σ）は、画像Ｌ（ｘ，ｙ，c_{j+1}σ）およびＬ（ｘ，ｙ，c_jσ）の２つの異なる平滑化バージョンのうちの平滑なほうと同程度に平滑であるかまたはそれよりも平滑である。すなわち、正規化する平滑化画像Ｌ（ｘ，ｙ，c_{j+1+h}σ）は、画像Ｌ（ｘ，ｙ，c_{j+1}σ）およびＬ（ｘ、ｙ、c_jσ）の２つの異なる平滑化バージョンのためのスケール（たとえば、平滑化カーネル）のうちの大きいほうに等しいかまたはそれよりも広いスケール（たとえば、平滑化カーネル）を有し得る。

【０１０１】

[00140]次いで９０８において、ｊ＝０～ｎ－１について、正規化されたスケール空間差分Ｄ’$_j$（ｘ，ｙ，σ）を使用して画像Ｉ（ｘ，ｙ）の特徴を識別する。たとえば、特徴がその周りで定義され得るキーポイントとして局所的極値（すなわち、極小値または極大値）が識別され得る。次いで９１０において、識別された特徴に基づいて画像Ｉ（ｘ，ｙ）のための記述子を発生する。

【０１０２】

[00141]図１１、図１２、および図１３に示す方法は、画像の照度に関する事前情報を必要としない。この方法は、何らかの大きい（有意な）計算および処理を導入することなしに、異なる照度変化にわたって一貫したレベルで画像中の安定した特徴を選定することと再現性を高めることとを可能にする。すなわち、平滑化スケール空間は、スケール空間差分Ｄ$_j$（ｘ，ｙ，σ）を正規化するために使用される平滑化画像Ｌ（ｘ，ｙ，c_{j+1+h}σ）をすでに含むので、正規化のために除算演算の他に追加の処理は必要とされない。

【０１０３】

[00142]さらに、特徴が選択される信頼性を適応させることによって、特徴が検出されるスケール（たとえば、平滑化レベル）に従ってより安定した特徴が取得され得る。すなわち、より高いスケールは、概して、より平滑な（すなわち、よりぼかされた）バージョンの画像を含み、そのようなスケールにおいて検出されたキーポイント／特徴は、より高い程度の信頼性を有する。

【０１０４】

[00143]図１４は、照度不変特徴検出デバイスの一例を示すブロック図である。特徴検出デバイス１２００は、デジタルクエリ画像１２０２を受信または取得し得る。次いで、スケール空間発生器１２０４（たとえば、ガウススケール空間発生器）が、クエリ画像１２０２を異なるスケール幅の複数の異なる平滑化カーネル１２０３で畳み込んで、スケール空間を発生し得る。スケール空間は、異なるスケーリング幅に平滑化された画像の複数の平滑化バージョンを備え得る。次いで、スケール空間差分発生器１２０６が、スケール空間からスケール空間差分を発生する。次いで、スケール空間差分正規化器１２０８が、たとえば、スケール空間レベルの各差分を対応する平滑化画像で除算することによって、スケール空間差分を正規化し、そのような平滑化画像は、除算されるスケール空間差分を発生するために使用される平滑化画像のうちの大きいほうと同程度に広いかまたはそれよりも広いスケールを有する。次いで、キーポイント発生器１２１０が、正規化されたスケール空間差分中のキーポイントを識別または検出する。これは、たとえば、正規化されたスケール空間差分のピクセルの間で局所的極値（すなわち、極大値または極小値）を見つけることによって行われ得る。特徴発生器１２１２が、次いで、たとえば、識別されたキーポイントの周りの局所ピクセルを特徴づけることによって、特徴を発生し得る。キーポイント発生器１２１０と特徴発生器１２１２との機能は特徴検出器によって実行され得ることに留意されたい。次いで、特徴記述子発生器１２１４が、各特徴について記述子を発生して、クエリ画像を識別するように働くことができる複数の画像記述子１２１６を与える。図１４に示す機能は、別個の回路によってあるいは１つまたは複数のプロセッサによって実行され得る。

【０１０５】

[00144]図１５は、特徴検出のために正規化されたスケール空間差分を使用する画像整

合デバイスの一例を示すブロック図である。画像整合デバイス１３００は、通信インターフェース１３０４、画像キャプチャデバイス１３０６、および／または記憶デバイス１３０８に結合された、処理回路１３０２を含み得る。通信インターフェース１３０４は、ワイヤード／ワイヤレスネットワーク上で通信し、画像および／または１つまたは複数の画像のための特徴記述子を受信するように適合され得る。画像キャプチャデバイス１３０６は、たとえば、クエリ画像をキャプチャすることができるデジタルカメラであり得る。処理回路１３０２は、画像から特徴を抽出する画像処理回路１３１４と、クエリ画像をターゲット画像のデータベース１３１０におよび／またはクエリ画像記述子を記述子データベース１３１２に整合させるために、抽出された特徴を使用する画像整合回路１３１６とを含み得る。例示的な一実装形態によれば、画像整合アプリケーションが、クエリ画像を画像データベース中の１つまたは複数の画像に整合させることを試みる。画像データベースは、データベース１３１０に記憶された１つまたは複数の画像に関連する何百万もの特徴記述子を含み得る。

【０１０６】

[00145]画像処理回路１３１４は、ガウススケール空間発生器１３２２、スケール空間差分発生器１３２４、スケール空間差分正規化器１３２６、キーポイント検出器１３２８、特徴発生器１３３０、および／または特徴記述子発生器１３３２を含む、特徴識別回路１３２０を含み得る。ガウススケール空間発生器１３２２は、たとえば、図８および図１１に示すように、複数の異なるスケール空間を発生するために画像をぼかし関数（たとえば、平滑化カーネル）で畳み込むように働き得る。次いで、スケール空間差分発生器１３２４がスケール空間からスケール空間差分を発生する。次いで、スケール空間差分正規化器１３２６が、たとえば、スケール空間レベルの各差分を対応する平滑化画像で除算することによって、スケール空間差分を正規化し、そのような平滑化画像は、（図１２に示した）除算されるスケール空間差分を発生するために使用される平滑化画像のいずれよりも広い。次いで、キーポイント発生器１３２８が、正規化されたスケール空間差分中のキーポイントを識別または検出する。これは、たとえば、正規化されたスケール空間差分のピクセルの間で局所的極値（すなわち、極大値または極小値）を見つけることによって行われ得る。特徴発生器１３３０が、次いで、たとえば、（図９に示した）識別されたキーポイントの周りの局所ピクセルを特徴づけることによって、特徴を発生し得る。次いで、特徴記述子発生器１３３２が、各特徴について記述子を発生して、（図１０に示した）クエリ画像を識別するように働くことができる複数の画像記述子を与える。

【０１０７】

[00146]次いで、画像整合回路１３１６が、特徴記述子に基づいてクエリ画像を画像データベース１３１０中の画像に整合させることを試み得る。整合結果は、（たとえば、画像または特徴記述子を送るモバイルデバイスに）通信インターフェースを介して与えられ得る。

【０１０８】

[00147]いくつかの実装形態では、クエリ画像のためのキーポイントに関連する特徴記述子のセットは画像整合デバイスによって受信され得ることに留意されたい。この状況では、クエリ画像は、（記述子を取得するために）すでに処理されている。したがって、画像処理回路１３１４は、画像整合デバイス１３００からバイパスされるかまたは除去され得る。

【０１０９】

ＤＯＡ検出器およびオーディオシーン分解

[00148]本明細書で開示するシステムおよび方法のいくつかの構成では、例示的なＤＯＡ検出器５１２の機能と、空間フィルタ７２を使用してオーディオシーンを分解するプロセスとは、このセクションにおいて説明する技法を使用して達成され得る。

【０１１０】

[00149]遠距離場オーディオ処理（たとえば、オーディオ音源強調）の適用は、１つまたは複数の音源が録音デバイスから比較的大きい距離（たとえば２メートル以上の距離）

に位置するときに生じ得る。

【０１１１】

[00150]遠距離場使用事例の第１の例では、いくつかの異なる音源を含む音響シーンの記録を分解して、個別の音源のうちの１つまたは複数からそれぞれの音響成分を取得する。たとえば、異なる音源（たとえば、異なる音声および／または楽器）からの音が分離されるように、生の音楽演奏を記録することが望ましいことがある。別のそのような例では、「ロックバンド」タイプのビデオゲームなどのビデオゲームの２人以上の異なるプレーヤからの音声入力（たとえば、命令および／または歌唱）を区別することが望ましいことがある。

【０１１２】

[00151]遠距離場使用事例の第２の例では、マルチマイクロフォンデバイスを使用して、（「ズームインマイクロフォン（zoom-in microphone）」とも呼ばれる）ビューの音場を狭めることによって遠距離場オーディオ音源強調を実行する。カメラを通じてシーンを見ているユーザは、カメラのレンズのズーム機能を使用して、たとえば、個々の話者または他の音源に対するビューの視界を選択的にズームし得る。相補的音響「ズームイン」効果をもたらすために、視覚的ズーム動作と同期して、記録される音響領域も被選択音源に狭められるように、カメラを実装することが望ましいことがある。

【０１１３】

[00152]特定の遠くの音源から到来する音響成分を区別することは、単にビームパターンを特定の方向に狭めることではない。ビームパターンの空間幅は、フィルタのサイズを増加させることによって（たとえば、初期係数値のより長いセットを使用してビームパターンを定義することによって）狭められ得るが、音源の単一の到来方向にのみ依存すると、実際にはフィルタが音源エネルギーの大部分を逃すことになり得る。残響などの影響により、たとえば、音源信号は通常、異なる周波数においてやや異なる方向から到来し、結果的に、遠くの音源の到来方向は一般にはっきりしない。したがって、信号のエネルギーは、特定の方向に集中するのではなく、角度範囲にわたって拡散することがあり、特定の音源の到来角を、単一の方向におけるピークとしてではなく周波数範囲にわたる重心として特徴づけることがより有用であり得る。

【０１１４】

[00153]フィルタのビームパターンが、単一の方向（たとえば、任意の１つの周波数における最大エネルギーによって示される方向）だけでなく、異なる周波数における方向の集中の幅をカバーすることが望ましいことがある。たとえば、ビームが、様々な対応する周波数において、そのような集中の幅内で、わずかに異なる方法に向くことを可能にすることが望ましいことがある。

【０１１５】

[00154]適応ビームフォーミングアルゴリズムを使用して、１つの周波数における特定の方向での最大応答と、別の周波数における異なる方向での最大応答とを有するフィルタを取得し得る。適応ビームフォーマーは一般に、正確な音声アクティビティ検出に依存するが、これは遠距離場の話者にとって達成するのが難しい。そのようなアルゴリズムはまた、所望の音源からの信号と干渉音源からの信号とが類似のスペクトルを有するとき（たとえば、２つの音源の両方が人々の話声であるとき）パフォーマンスが芳しくないことがある。また、適応ビームフォーマーの代替として、ブラインド音源分離（ＢＳＳ：blind source separation）ソリューションを使用して、１つの周波数における特定の方向での最大応答と、別の周波数における異なる方向での最大応答とを有するフィルタを取得し得る。しかしながら、そのようなアルゴリズムは、遅い収束、極小値への収束、および／またはスケーリングのあいまいさを示すことがある。

【０１１６】

[00155]良好な初期条件を提供するデータ独立型、開ループ手法（たとえば、ＭＶＤＲビームフォーマー）を、音声アクティビティ検出器を使用せずに出力間の相関を最小限に抑える閉ループ方法（たとえば、ＢＳＳ）と組み合わせて、それによって改良されたロバ

ストな分離ソリューションを提供することが望ましいことがある。ＢＳＳ方法は経時的に適応を実行するので、残響環境においてもロバストなソリューションを生成することが期待され得る。

【０１１７】

[00156]ヌルビームを使用してフィルタを初期化する既存のＢＳＳ初期化手法とは対照的に、本明細書で説明するソリューションは、音源ビームを使用してフィルタを初期化し、指定の音源方向に集中する。そのような初期化なしに、ＢＳＳ方法がリアルタイムで有用なソリューションに適応するのを期待することは現実的でないことがある。

【０１１８】

[00157]図１６Ａに、フィルタバンクＢＫ１０と、フィルタ配向モジュールＯＭ１０と、フィルタ更新モジュールＵＭ１０とを含み、マルチチャネル信号（この例では、入力チャネルＭＣＳ１０－１およびＭＣＳ１０－２）を受信するように構成された、装置Ａ１００のブロック図を示す。フィルタバンクＢＫ１０は、マルチチャネル信号に基づく第１の信号に複数の第１の係数を適用して、第１の出力信号ＯＳ１０－１を生成するように構成される。フィルタバンクＢＫ１０はまた、マルチチャネル信号に基づく第２の信号に複数の第２の係数を適用して、第２の出力信号ＯＳ１０－２を生成するように構成される。フィルタ配向モジュールＯＭ１０は、第１の音源方向ＤＡ１０に基づく複数の第１の係数の値の初期セットＣＶ１０を生成し、第１の音源方向ＤＡ１０とは異なる第２の音源方向ＤＡ２０に基づく複数の第２の係数の値の初期セットＣＶ２０を生成するように構成される。フィルタ更新モジュールＵＭ１０は、第１および第２の出力信号からの情報に基づいて、複数の第１および第２の係数の値の初期セットを更新して、値の対応する更新されたセットＵＶ１０およびＵＶ２０を生成するように構成される。

【０１１９】

[00158]音源方向ＤＡ１０およびＤＡ２０の各々が、入力チャネルＭＣＳ１０－１およびＭＣＳ１０－２を生成するマイクロフォンアレイに対する（たとえば、アレイのマイクロフォンの軸に対する）対応する音源の推定方向を示すことが望ましいことがある。図１６Ｂに、マイクロフォンアレイＲ１００と、アレイから（たとえば、入力チャネルＭＣＳ１０－１およびＭＣＳ１０－２を含む）マルチチャネル信号ＭＣＳ１０を受信するように構成された装置Ａ１００のインスタンスとを含む、デバイスＤ１０のブロック図を示す。アレイＲ１００は、図１のアレイ１８と、図１のシステム１４中の装置Ａ１００との中に含まれ得る。

【０１２０】

[00159]図１６Ｃに、点音源ｊから受信された信号成分の、アレイＲ１００のマイクロフォンＭＣ１０およびＭＣ２０の軸に対する到来方向θ_jを示す。アレイの軸は、マイクロフォンの音響的に敏感な面の中心を通る線として定義される。この例では、標示ｄは、マイクロフォンＭＣ１０とマイクロフォンＭＣ２０との間の距離を示す。

【０１２１】

[00160]フィルタ配向モジュールＯＭ１０は、ビームフォーミングアルゴリズムを実行して、それぞれの音源方向ＤＡ１０、ＤＡ２０におけるビームを記述した係数値の初期セットＣＶ１０、ＣＶ２０を発生するように実装され得る。ビームフォーミングアルゴリズムの例としては、ＤＳＢ（遅延和ビームフォーマー（delay-and-sum beamformer））、ＬＣＭＶ（線形制約最小分散（linear constraint minimum variance））、およびＭＶＤＲ（最小分散無ひずみ応答（minimum variance distortionless response））がある。一例では、フィルタ配向モジュールＯＭ１０は、次のようなデータ独立式に従って、各フィルタが他の音源方向においてゼロ応答（またはヌルビーム）を有するように、ビームフォーマーのＮ×Ｍ係数行列Ｗを計算するように実装される。

【０１２２】

$$W(\omega) = D^H(\omega, \theta)\,[D(\omega, \theta)\,D^H(\omega, \theta) + r(\omega) \times I]^{-1}$$

ただし、ｒ（ω）は、非反転性（noninvertibility）を補償するための正則化項である。別の例では、フィルタ配向モジュールＯＭ１０は、次のような式に従って、ＭＶＤＲビー

ムフォーマーのＮ×Ｍ係数行列Ｗを計算するように実装される。
【数９】

$$\mathbf{W} = \frac{\mathbf{\Phi}^{-1}\mathbf{D}(\omega)}{\mathbf{D}^{H}(\omega)\mathbf{\Phi}^{-1}\mathbf{D}(\omega)}. \quad (1)$$

【０１２３】
これらの例では、Ｎは出力チャネルの数を示し、Ｍは入力チャネルの数（たとえば、マイクロフォンの数）を示し、Φは雑音の正規化クロスパワースペクトル密度行列を示し、Ｄ（ω）は、（指向性行列とも呼ばれる）Ｍ×Ｎアレイマニホルド行列を示し、上付き文字Ｈは共役転置関数を示す。通常、ＭはＮ以上である。
【０１２４】
[00161]係数行列Ｗの各行は、フィルタバンクＢＫ１０の対応するフィルタの係数の初期値を定義する。一例では、係数行列Ｗの第１の行は初期値ＣＶ１０を定義し、係数行列Ｗの第２の行は初期値ＣＶ２０を定義する。別の例では、係数行列Ｗの第１の行は初期値ＣＶ２０を定義し、係数行列Ｗの第２の行は初期値ＣＶ１０を定義する。
【０１２５】
[00162]行列Ｄの各列ｊは、次の式として表され得る周波数ωにわたる遠距離場音源ｊの指向性ベクトル（または「ステアリングベクトル」）である。D_{mj}（ω）＝ｅｘｐ（－ｉ×ｃｏｓ（θ_j）×ｐｏｓ（ｍ）×ω／ｃ）この式において、ｉは虚数を示し、ｃは媒質中の音の伝播速度（たとえば、空中では３４０ｍ／秒）を示し、θ_jは、図１６Ｃに示される入射到来角としてマイクロフォンアレイの軸に対する音源jの方向（たとえば、ｊ＝１の方向ＤＡ１０およびｊ＝２の方向ＤＡ２０）を示し、ｐｏｓ（ｍ）は、Ｍ個のマイクロフォンのアレイにおけるｍ番目のマイクロフォンの空間座標を示す。均一なマイクロフォン間間隔ｄをもつマイクロフォンの線形アレイの場合、ファクタｐｏｓ（ｍ）は（ｍ－１）ｄとして表され得る。
【０１２６】
[00163]拡散雑音場の場合、行列Φは次のようなコヒーレンス関数Γを使用して置き換えられ得る。
【数１０】

$$\Gamma_{ij} = \begin{cases} \operatorname{sinc}\left(\frac{\omega d_{ij}}{c}\right), & i \neq j \\ 1, & i = j \end{cases},$$

【０１２７】
ただし、d_{ij}はマイクロフォンｉとマイクロフォンｊとの間の距離を示す。さらなる一例では、行列Φは、（Γ＋λ（ω）Ｉ）に置き換えられ、ただし、λ（ω）は、（たとえば安定性に関する）対角ローディングファクタである。
【０１２８】
[00164]一般に、フィルタバンクＢＫ１０の出力チャネルの数Ｎは、入力チャネルの数Ｍ以下である。図１６Ａは、Ｎの値が２である（すなわち、２つの出力チャネルＯＳ１０－１およびＯＳ１０－２をもつ）装置Ａ１００の実装形態を示しているが、ＮおよびＭは２よりも大きい値（たとえば、３、４、またはそれ以上）を有し得ることを理解されたい。そのような一般的な場合、フィルタバンクＢＫ１０は、Ｎ個のフィルタを含むように実装され、フィルタ配向モジュールＯＭ１０は、これらのフィルタのために初期係数値のＮ個の対応するセットを生成するように実装され、これらの原理のそのような拡張は、明確に企図され、本明細書によって開示される。
【０１２９】

[00165]たとえば、図１７に、ＮとＭの両方の値が４である装置Ａ１００の実装形態Ａ１１０のブロック図を示す。装置Ａ１１０は、４つのフィルタを含むフィルタバンクＢＫ１０の実装形態ＢＫ１２を含み、各フィルタは、入力チャネルＭＣＳ１０－１、ＭＣＳ１０－２、ＭＣＳ１０－３、およびＭＣＳ１０－４の各々をフィルタ処理して、出力信号（またはチャネル）ＯＳ１０－１、ＯＳ１０－２、ＯＳ１０－３、およびＯＳ１０－４のうちの対応する１つを生成するように構成される。装置Ａ１００はまた、フィルタバンクＢＫ１２のフィルタのために係数値の初期セットＣＶ１０、ＣＶ２０、ＣＶ３０、およびＣＶ４０を生成するように構成された、フィルタ配向モジュールＯＭ１０の実装形態ＯＭ１２と、係数値の初期セットを適応させて、値の対応する更新されたセットＵＶ１０、ＵＶ２０、ＵＶ３０、およびＵＶ４０を生成するように構成された、フィルタ適応モジュールＡＭ１０の実装形態ＡＭ１２とを含む。

【０１３０】

[00166]（「ビームパターン」とも呼ばれる）周波数ビン対入射角に関するフィルタバンクＢＫ１０のフィルタの初期応答は、ＭＶＤＲビームフォーミングアルゴリズム（たとえば、上の式（１））に従ってフィルタ配向モジュールＯＭ１０によって発生されたフィルタの係数値によって判断される。この応答は、入射角０（たとえば、マイクロフォンアレイの軸の方向）を中心として対称的であり得る。初期条件の異なるセット（たとえば、所望の音源からの音と、干渉音源からの音との推定到来方向の異なるセット）の下でのこのビームパターンの変形が有され得る。

【０１３１】

[00167]特定の適用例に適していると考えられる指向性とサイドローブ発生との間の折衷に従って選択されたビームフォーマー設計に従って係数値ＣＶ１０およびＣＶ２０を生成するようにフィルタ配向モジュールＯＭ１０を実装することが望ましいことがある。上記の例は、周波数領域ビームフォーマー設計について説明しているが、時間領域ビームフォーマー設計に従って係数値のセットを生成するように構成されたフィルタ配向モジュールＯＭ１０の代替実装形態も、明確に企図され、本明細書によって開示される。

【０１３２】

[00168]フィルタ配向モジュールＯＭ１０は、（たとえば、上記で説明したようにビームフォーミングアルゴリズムを実行することによって）係数値ＣＶ１０およびＣＶ２０を発生させるように、またはストレージから係数値ＣＶ１０およびＣＶ２０を取り出すように実装され得る。たとえば、フィルタ配向モジュールＯＭ１０は、音源方向（たとえば、ＤＡ１０およびＤＡ２０）に従って値（たとえば、ビーム）の事前計算されたセットの中から選択することによって、係数値の初期セットを生成するように実装され得る。そのような係数値の事前計算されたセットをオフラインで計算して、対応する所望の解像度における方向および／または周波数の所望の範囲をカバーし得る（たとえば、０、２０、または３０度から１５０、１６０、または１８０度までの範囲における、５度、１０度、または２０度の各間隔についての係数値の異なるセット）。

【０１３３】

[00169]フィルタ配向モジュールＯＭ１０によって生成される初期係数値（たとえば、ＣＶ１０およびＣＶ２０）は、音源信号の間に所望のレベルの分離をもたらすようにフィルタバンクＢＫ１０を構成するには十分でないことがある。これらの初期値が基づく推定音源方向（たとえば、方向ＤＡ１０およびＤＡ２０）が完全に正確であったとしても、フィルタを一定の方向にステアリングするだけでは、アレイからから遠く離れた音源間の最良の分離、または特定の離れた音源への最良の集中は実現しないことがある。

【０１３４】

[00170]フィルタ更新モジュールＵＭ１０は、第１および第２の出力信号ＯＳ１０－１およびＯＳ１０－２からの情報に基づいて、第１および第２の係数の初期値ＣＶ１０およびＣＶ２０を更新して、値の対応する更新されたセットＵＶ１０およびＵＶ２０を生成するように構成される。たとえば、フィルタ更新モジュールＵＭ１０は、これらの初期係数値によって記述されるビームパターンを適応させるために適応ＢＳＳアルゴリズムを実行

するように実装され得る。

【０１３５】

[00171]ＢＳＳ方法は、Y_j（ω，ｌ）＝Ｗ（ω）X_j（ω，ｌ）などの式に従って様々な音源から、統計的に独立した信号成分を分離し、ただし、X_jは周波数領域における入力（混合）信号のｊ番目のチャネルを示し、Y_jは周波数領域における出力（分離）信号のｊ番目のチャネルを示し、ωは周波数ビンインデックスを示し、ｌは時間フレームインデックスを示し、Ｗはフィルタ係数行列を示す。概して、ＢＳＳ方法は、次のような式による逆混合行列Ｗの経時的適応として記述され得る。

【数１１】

$$\mathbf{W}_{l+r}(\omega) = \mathbf{W}_l(\omega) + \mu[\mathbf{I} - \langle \Phi(Y(\omega,l))Y(\omega,l)^H \rangle]\mathbf{W}_l(\omega), \qquad (2)$$

【０１３６】

ただし、ｒは適応間隔（または更新レート）パラメータを示し、μは適応速度（または学習レート）ファクタを示し、Ｉは恒等行列を示し、上付き文字Ｈは共役転置関数を示し、Φはアクティブ化関数（activation function）を示し、括弧＜・＞は（たとえば、フレームｌからｌ＋Ｌ－１にわたるものであって、Ｌは一般にｒ以下である）時間平均化演算を示す。一例では、μの値は０．１である。式（２）はＢＳＳ学習ルールまたはＢＳＳ適応ルールとも呼ばれる。アクティブ化関数Φは一般に、所望の信号の累積密度関数に近似するように選択され得る非線形有界関数である。そのような方法において使用されるアクティブ化関数Φの例としては、双曲正接関数（hyperbolic tangent function）、シグモイド関数（sigmoid function）、および符号関数（sign function）がある。

【０１３７】

[00172]フィルタ更新モジュールＵＭ１０は、本明細書で説明するＢＳＳ方法に従ってフィルタ配向モジュールＯＭ１０によって生成された係数値（たとえば、ＣＶ１０およびＣＶ２０）を適応させるように実装され得る。そのような場合、出力信号ＯＳ１０－１およびＯＳ１０－２は、周波数領域信号Ｙのチャネル（たとえば、それぞれ第１のチャネルおよび第２のチャネル）であり、係数値ＣＶ１０およびＣＶ２０は、逆混合行列Ｗの対応する行（たとえば、それぞれ第１の行および第２の行）の初期値であり、適応された値は、適応後の逆混合行列Ｗの対応する行（たとえば、それぞれ第１の行および第２の行）によって定義される。

【０１３８】

[00173]周波数領域における適応のためのフィルタ更新モジュールＵＭ１０の典型的な実装形態では、逆混合行列Ｗは有限インパルス応答（ＦＩＲ）多項式行列である。そのような行列は、要素としてＦＩＲフィルタの周波数変換（たとえば、離散フーリエ変換）を有する。時間領域における適応のためのフィルタ更新モジュールＵＭ１０の典型的な実装形態では、逆混合行列ＷはＦＩＲ行列である。そのような行列は要素としてＦＩＲフィルタを有する。そのような場合、係数値の各初期セット（たとえば、ＣＶ１０およびＣＶ２０）は、一般に複数のフィルタを記述することになることを理解されよう。たとえば、係数値の各初期セットは、逆混合行列Ｗの対応する行の要素ごとにフィルタを記述し得る。周波数領域実装形態の場合、係数値の各初期セットは、マルチチャネル信号の周波数ビンごとに、逆混合行列Ｗの対応する行の各要素のフィルタの変換を記述し得る。

【０１３９】

[00174]ＢＳＳ学習ルールは、一般に、出力信号間の相関を減らすように設計される。たとえば、ＢＳＳ学習ルールは、出力信号間の相互情報量を最小限に抑えるように、出力信号の統計的独立性を高めるように、または出力信号のエントロピーを最大にするように選択され得る。一例では、フィルタ更新モジュールＵＭ１０は、独立成分分析（ＩＣＡ：independent component analysis）として知られているＢＳＳ方法を実行するように実装される。そのような場合、フィルタ更新モジュールＵＭ１０は、上記で説明したアクティ

ブ化関数、または、たとえば、次のようなアクティブ化関数を使用するように構成され得る。

【数１２】

$$\Phi(Y_j(\omega,l)) = Y_j(\omega,l) \Big/ |Y_j(\omega,l)|$$

【０１４０】

周知のＩＣＡ実装形態の例としては、Ｉｎｆｏｍａｘ、ＦａｓｔＩＣＡ（www-dot-cis-dot-hut-dot-fi/projects/ica/fasticaでオンライン入手可能）、およびＪＡＤＥ（固有行列の結合近似対角化（Joint Approximate Diagonalization of Eigenmatrices））がある。

【０１４１】

[00175]スケーリングおよび周波数置換は、ＢＳＳにおいて一般に遭遇される２つのあいまいさである。フィルタ配向モジュールＯＭ１０によって生成される初期ビームは置換されないが、そのようなあいまいさは、ＩＣＡの場合に適応中に生じ得る。置換されない解を維持するために、代わりに、周波数ビン間の予想される依存性をモデル化するソースプライアを使用する複素ＩＣＡの一変形である独立ベクトル解析（ＩＶＡ）を使用するようにフィルタ更新モジュールＵＭ１０を構成することが望ましいことがある。この方法では、アクティブ化関数Φは、たとえば、次の式などの多変量アクティブ化関数である。

【数１３】

$$\Phi(Y_j(\omega,l)) = Y_j(\omega,l) \Big/ \left(\sum_{\omega} |Y_j(\omega,l)|^p\right)^{1/p}$$

【０１４２】

ただし、ｐは１以上の整数値（たとえば、１、２、または３）を有する。この関数において、分母の項は、すべての周波数ビンにわたる分離された音源スペクトルに関係する。この場合、置換のあいまいさは解決される。

【０１４３】

[00176]得られた適応係数値によって定義されるビームパターンは、直線ではなく畳み込まれているように見え得る。そのようなパターンは、遠くの音源の分離には一般に不十分である初期係数値ＣＶ１０およびＣＶ２０によって定義されるビームパターンよりも良好な分離をもたらすと予想され得る。たとえば、１０～１２ｄＢから１８～２０ｄＢへの干渉消去の増加が観測されている。適応係数値によって表されるソリューションはまた、マイクロフォン応答（たとえば、利得および／または位相応答）の不整合に対し、開ループビームフォーミングソリューションよりもロバストであると予想され得る。

【０１４４】

[00177]上記の例は、周波数領域におけるフィルタ適応について説明しているが、時間領域における係数値のセットを更新するように構成されたフィルタ更新モジュールＵＭ１０の代替実装形態も、明確に企図され、本明細書によって開示される。時間領域ＢＳＳ方法は、置換のあいまいさの影響を受けないが、一般に、周波数領域ＢＳＳ方法よりも長いフィルタの使用を伴い、実際には扱いにくいことがある。

【０１４５】

[00178]ＢＳＳ方法を使用して適応されたフィルタは概して、良好な分離を達成するが、そのようなアルゴリズムも、特に音源が遠くにある場合に、分離信号にさらなる残響をもたらす傾向がある。特定の到来方向において単位利得を強制する幾何学的制約を追加することによって、適応ＢＳＳソリューションの空間応答を制御することが望ましいことが

ある。ただし、上述のように、単一の到来方向に対してフィルタ応答を調整するのは、残響環境では不十分であり得る。その上、ＢＳＳ適応において（ヌルビーム方向とは反対の）ビーム方向を強制しようとすると、問題が生じかねない。

【０１４６】

[00179]フィルタ更新モジュールＵＭ１０は、方向に対する値の適応されたセットの判断された応答に基づいて、複数の第１の係数の値の適応されたセットと複数の第２の係数の値の適応されたセットとのうちの少なくとも１つを調整するように構成される。この判断された応答は、指定の特性を有する応答に基づき、異なる周波数では異なる値を有し得る。一例では、判断された応答は、最大応答である（たとえば、指定の特性は最大値である）。調整されるべき係数のセットｊごとに、また調整されるべき範囲内の各周波数ωおいて、たとえば、この最大応答R_j（ω）は、次のような式に従って、その周波数における適応されたセットの複数の応答のうちの最大値として表され得る。

【数１４】

$$R_j(\omega) = \max_{\theta=[-\pi,\pi]} \left| W_{j1}(\omega)D_{\theta 1}(\omega) + W_{j2}(\omega)D_{\theta 2}(\omega) + \ldots + W_{jM}(\omega)D_{\theta M}(\omega) \right|, \quad (3)$$

【０１４７】

ただし、Ｗは適応された値の行列（たとえば、ＦＩＲ多項式行列）であり、W_{jm}は、行ｊおよび列ｍにおける行列Ｗの要素を示し、列ベクトルD_{θ}（ω）の各要素ｍは、次の式で表され得る距離θの遠距離場音源から受信される信号に関する周波数ωにおける位相遅延を示す。

【０１４８】

$D_{\theta m}$（ω）＝ｅｘｐ（－ｉ×ｃｏｓ（θ）×ｐｏｓ（ｍ）×ω／ｃ）

別の例では、判断された応答は、最小応答（たとえば、各周波数における適応されたセットの複数の応答の中の最小値）である。

【０１４９】

[00180]一例では、式（３）は、範囲［－π，＋π］においてθの６４個の均一に離間した値について評価される。他の例では、式（３）は、θの異なる数の値（たとえば、１６個または３２個の均一に離間した値、５度または１０度の増分における値など）について、不均一な間隔で（たとえば、横方向の範囲にわたって、縦方向における範囲よりも大きい解像度で、またはその逆）、および／または異なる関心領域（たとえば、［－π，０］、［－π／２，＋π／２］、［－π，＋π／２］）にわたって評価され得る。均一なマイクロフォン間間隔ｄをもつマイクロフォンの線形アレイの場合、係数ｐｏｓ（ｍ）は（ｍ－１）ｄとして表され得、したがって、ベクトルD_{θ}（ω）の各要素ｍは次のように表され得る。$D_{\theta m}$（ω）＝ｅｘｐ（－ｉ×ｃｏｓ（θ）×（ｍ－１）ｄ×ω／ｃ）

式（３）が最大値を有する方向θの値は、周波数ωの値が異なる場合には異なると予想され得る。音源方向（たとえば、ＤＡ１０および／またはＤＡ２０）は、式（３）が評価されるθの値の中に含まれ得、または、代替的に、それらの値とは別個であり得る（たとえば、音源方向が、式（３）が評価されるθの値の隣接するものの間の角度を示す場合）ことに留意されたい。

【０１５０】

[00181]図１８Ａに、フィルタ更新モジュールＵＭ１０の実装形態ＵＭ２０のブロック図を示す。フィルタ更新モジュールＵＭ１０は、出力信号ＯＳ１０－１およびＯＳ１０－２からの情報に基づいて係数値ＣＶ１０および係数値ＣＶ２０を適応させて、値の対応する適応されたセットＡＶ１０およびＡＶ２０を生成するように構成された適応モジュールＡＰＭ１０を含む。たとえば、適応モジュールＡＰＭ１０は、本明細書で説明するＢＳＳ方法のいずれか（たとえば、ＩＣＡ、ＩＶＡ）を実行するように実装され得る。

【０１５１】

[00182]フィルタ更新モジュールＵＭ２０はまた、（たとえば、上記の式（３）による）方向に対する値の適応されたセットＡＶ１０の最大応答に基づいて適応された値ＡＶ１０を調整して、値の更新されたセットＵＶ１０を生成するように構成された調整モジュールＡＪＭ１０を含む。この場合、フィルタ更新モジュールＵＭ２０は、更新された値ＵＶ２０としてそのような調整をせずに適応された値ＡＶ２０を生成するように構成される。（本明細書で開示する構成の範囲はまた、係数値ＣＶ２０が適応も調整もされないという点で、装置Ａ１００とは異なる装置を含むことに留意されたい。そのような構成は、たとえば、信号が残響をほとんどまたはまったく伴わずに直接経路を介して対応する音源から到来する状況において使用され得る。）

[00183]調整モジュールＡＪＭ１０は、値の適応されたセットを、方向に対する各周波数における所望の利得応答（たとえば、最大の単位利得応答）を有するようにセットを正規化することによって調整するように実装され得る。そのような場合、調整モジュールＡＪＭ１０は、係数値の適応されたセットｊ（たとえば、適応された値ＡＶ１０）の各値を、セットの最大応答R_j（ω）で除算して、係数値の対応する更新されたセット（たとえば、更新された値ＵＶ１０）を取得するように実装され得る。

【０１５２】

[00184]所望の利得応答が単位利得応答以外である場合、調整モジュールＡＪＭ１０は、適応された値および／または正規化された値に利得係数を適用することを調整演算が含むように実装され得、ここで、利得係数値の値は周波数とともに変化して、所望の利得応答を記述する（たとえば、音源のピッチ周波数のハーモニックを選好し、および／または干渉物によって支配され得る１つまたは複数の周波数を減衰させる）。判断された応答が最小応答である場合、調整モジュールＡＪＭ１０は、（たとえば、各周波数の）最小応答を減算することによって、または方向に対する各周波数における所望の利得応答（たとえば、最小のゼロの利得応答）を有するようにセットを再マッピングすることによって、適応されたセットを調整するように実装され得る。

【０１５３】

[00185]係数値のセットのうちの２つ以上について、また場合によってはすべてについて（たとえば、少なくとも、定位された音源に関連しているフィルタについて）そのような正規化を実行するように調整モジュールＡＪＭ１０を実装することが望ましいことがある。図１８Ｂに、調整モジュールＡＪＭ１０の実装形態ＡＪＭ１２を含むフィルタ更新モジュールＵＭ２０の実装形態ＵＭ２２のブロック図を示し、ＡＪＭ１２はまた、方向に対する値の適応されたセットＡＶ２０の最大応答に基づいて、適応された値ＡＶ２０を調整して、値の更新されたセットＵＶ２０を生成するように構成される。

【０１５４】

[00186]そのようなそれぞれの調整は、追加の適応フィルタに（たとえば、適応行列Ｗの他の行に）同じ方法で拡張され得ることを理解されたい。たとえば、図１７に示したフィルタ更新モジュールＵＭ１２は、係数値の４つのセットＣＶ１０、ＣＶ２０、ＣＶ３０、およびＣＶ４０を適応させて、値の４つの対応する適応されたセットを生成するように構成された適応モジュールＡＰＭ１０の一実装形態と、値の対応する適応されたセットの最大応答に基づいて、値の更新されたセットＵＶ３０およびＵＶ４０の一方または両方の各々を生成するように構成された調整モジュールＡＪＭ１２の一実装形態とを含むように、フィルタ更新モジュール２２の一実装形態として構成され得る。

【０１５５】

[00187]従来のオーディオ処理ソリューションは、雑音基準の計算と、計算された雑音基準を適用する後処理ステップとを含み得る。本明細書で説明する適応ソリューションは、後処理への依存を弱め、フィルタ適応への依存を強めて、干渉する点音源を除去することによって干渉消去と残響除去とを改善するように実装され得る。残響は、周波数とともに変化する利得応答を有する伝達関数（たとえば、室内応答伝達関数）として考えられ得、減衰する周波数成分もあれば、増幅する周波数成分もある。たとえば、室内のジオメトリは、様々な周波数における信号の相対強度に影響を与えることがあり、いくつかの周波

数が支配的になり得る。ある周波数から別の周波数に変化する方向において（すなわち、各周波数における主要ビームの方向において）所望の利得応答を有するようにフィルタを抑制することによって、本明細書で説明する正規化演算は、異なる周波数において空間中で信号のエネルギーが拡散される度合いの差異を補償することによって、信号を残響除去するのを助け得る。

【０１５６】

[00188]最良の分離および残響除去の結果を達成するために、一部の到来角範囲内で音源から到来するエネルギーを通過させ、他の角度で干渉音源から到来するエネルギーをブロックする空間応答を有するように、フィルタバンクＢＫ１０のフィルタを構成することが望ましいことがある。本明細書で説明するように、ＢＳＳ適応を使用して、フィルタが初期解の近傍でより良い解を見つけることを可能にするように、フィルタ更新モジュールＵＭ１０を構成することが望ましいことがある。ただし、所望の音源に向けられた主要ビームを維持する制約なしに、フィルタ適応は、同様の方向からの干渉音源が（たとえば、干渉音源からのエネルギーを除去する広いヌルビームを作ることによって）主要ビームを損なうのを許容し得る。

【０１５７】

[00189]フィルタ更新モジュールＵＭ１０は、制約付きＢＳＳを介して適応ヌルビームフォーミングを使用して、音源定位解からの大きい逸脱を防ぐ一方、小さい定位誤差を訂正することができるように構成され得る。しかしながら、フィルタが異なる音源に方向を変えるのを防ぐフィルタ更新ルールに関する空間制約を課すことが望ましいこともある。たとえば、フィルタを適応させるプロセスが、干渉音源の到来方向にヌル制約を含めることが望ましいことがある。そのような制約は、ビームパターンが低周波数において当該干渉方向にそれの配向を変えるのを防ぐことが望ましいことがある。

【０１５８】

[00190]ＢＳＳ逆混合行列の一部のみを適応させるようにフィルタ更新モジュールＵＭ１０を実装する（たとえば、適応モジュールＡＰＭ１０を実装する）ことが望ましいことがある。たとえば、フィルタバンクＢＫ１０のフィルタのうちの１つまたは複数を固定することが望ましいことがある。そのような制約は、（たとえば、上記の式（２）に示した）フィルタ適応プロセスが係数行列Ｗの対応する行を変えるのを防止することによって実装され得る。

【０１５９】

[00191]一例では、そのような制約は、固定されるべき各フィルタに対応する（たとえば、フィルタ配向モジュールＯＭ１０によって生成された）係数値の初期セットを維持するために、適応プロセスの開始時から適用される。そのような実装形態は、たとえば、静止した干渉物にビームパターンが向けられているフィルタにとって適切であり得る。別の例では、そのような制約は、係数値の適応されたセットのさらなる適応を防止するために（たとえば、フィルタが収束したことが検出されたときに）後で適用される。そのような実装形態は、たとえば、安定した残響環境における静止した干渉物にビームパターンが向けられているフィルタにとって適切であり得る。フィルタ係数値の正規化されたセットが固定されると、セットが固定されている間は調整モジュールＡＪＭ１０はそれらの値の調整を実行する必要がないが、調整モジュールＡＪＭ１０は係数値の他のセットを（たとえば、調整モジュールＡＪＭ１０によるそれらの適応に応答して）調整し続け得ることに留意されたい。

【０１６０】

[00192]代替または追加として、周波数範囲の一部分のみでフィルタのうちの１つまたは複数を適応させるようにフィルタ更新モジュールＵＭ１０を実装する（たとえば、適応モジュールＡＰＭ１０を実装する）ことが望ましいことがある。フィルタのそのような固定化は、当該範囲から外れた周波数に（たとえば、上記の式（２）中のωの値に）対応するフィルタ係数値を適応させないことによって達成され得る。

【０１６１】

[00193]有用な情報を含んでいる周波数範囲でのみ、フィルタのうちの１つまたは複数（場合によってはすべて）の各々を適応させ、別の周波数範囲ではフィルタを固定することが望ましいことがある。適応されるべき周波数範囲は、マイクロフォンアレイから話者までの予想される距離、マイクロフォン間の距離（例：たとえば空間エイリアシングにより、空間フィルタ処理がいずれにせよ失敗する周波数でフィルタを適応させるのを回避するため）、部屋のジオメトリ、および／または室内のデバイスの配置などのファクタに基づき得る。たとえば、入力信号は、特定の周波数範囲（たとえば、高周波数範囲）にわたって、その範囲で正しいＢＳＳ学習をサポートするのに十分な情報を含んでいないことがある。そのような場合、適応なしにこの範囲で初期の（または場合によっては直近の）フィルタ係数値を使用し続けることが望ましいことがある。

【０１６２】

[00194]音源がアレイから３～４メートル以上離れているとき、一般的に、音源によって放出される高周波エネルギーで、マイクロフォンに到達するものはほとんどない。そのような場合、フィルタ適応を適切にサポートする情報は、高周波数範囲ではほとんど得られないことがあるので、高周波数でフィルタを固定し、低周波数でのみそれらを適応させることが望ましいことがある。

【０１６３】

[00195]追加または代替として、どの周波数を適応させるべきかの決定は、周波数帯域において現在利用可能なエネルギーの量、および／またはマイクロフォンアレイから現在の話者までの推定距離などのファクタに従って、実行時間中に変わり得、フィルタごとに異なり得る。たとえば、ある時間には最高２ｋＨｚ（あるいは３ｋＨｚまたは５ｋＨｚ）の周波数でフィルタを適応させ、別の時間には最高４ｋＨｚ（あるいは５ｋＨｚ、８ｋＨｚ、または１０ｋＨｚ）の周波数でフィルタを適応させることが望ましいことがある。特定の周波数のために固定され、すでに調整されている（たとえば、正規化されている）フィルタ係数値を調整モジュールＡＪＭ１０が調整する必要はないが、調整モジュールＡＪＭ１０は他の周波数で係数値を（たとえば、適応モジュールＡＰＭ１０によるそれらの適応に応答して）調整し続け得ることに留意されたい。

【０１６４】

[00196]フィルタバンクＢＫ１０は、更新された係数値（たとえば、ＵＶ１０およびＵＶ２０）をマルチチャネル信号の対応するチャネルに適用する。更新された係数値は、（たとえば、調整モジュールＡＪＭ１０による）本明細書で説明する調整後の（たとえば、適応モジュールＡＰＭ１０によって適応された）逆混合行列Ｗの対応する行の値であるが、そのような値が本明細書で説明するように固定されている場合は除く。係数値の各更新されたセットは一般に、複数のフィルタを記述することになる。たとえば、係数値の各更新されたセットは、逆混合行列Ｗの対応する行の要素ごとにフィルタを記述し得る。

【０１６５】

[00197]概して、各推定音源方向（たとえば、ＤＡ１０および／またはＤＡ２０）は、測定、計算、予測、予想、および／または選択され得、所望の音源、干渉音源、または反射からの音の到来方向を示し得る。フィルタ配向モジュールＯＭ１０は、別のモジュールまたはデバイスから（たとえば、音源定位モジュールから）推定音源方向を受信するように構成され得る。そのようなモジュールまたはデバイスは、（たとえば、顔および／または動き検出を実行することによる）カメラからの画像情報および／または超音波反射からの測距情報に基づいて推定音源方向を生成するように構成され得る。そのようなモジュールまたはデバイスはまた、音源の数を推定するように、および／または動いている１つまたは複数の音源を追跡するように構成され得る。図１９Ａに、そのような画像情報をキャプチャするために使用され得るカメラＣＭ１０をもつアレイＲ１００の４マイクロフォン実装形態Ｒ１０４の構成の一例の上面図を示す。

【０１６６】

[00198]代替的に、装置Ａ１００は、マルチチャネル信号ＭＣＳ１０内の情報および／またはフィルタバンクＢＫ１０によって生成される出力信号内の情報に基づいて、推定音

源方向（たとえば、ＤＡ１０およびＤＡ２０）を計算するように構成された方向推定モジュールＤＭ１０を含むように実装され得る。そのような場合、方向推定モジュールＤＭ１０はまた、上記で説明したように画像情報および／または測距情報に基づいて推定音源方向を計算するように実装され得る。たとえば、方向推定モジュールＤＭ１０は、マルチチャネル信号ＭＣＳ１０に適用される、一般化相互相関（ＧＣＣ：generalized cross-correlation）アルゴリズム、またはビームフォーマーアルゴリズムを使用して音源ＤＯＡを推定するように実装され得る。

【０１６７】

[00199]図２０に、マルチチャネル信号ＭＣＳ１０内の情報に基づいて推定音源方向ＤＡ１０およびＤＡ２０を計算するように構成された方向推定モジュールＤＭ１０のインスタンスを含む装置Ａ１００の実装形態Ａ１２０のブロック図を示す。この場合、方向推定モジュールＤＭ１０およびフィルタバンクＢＫ１０は、同じ領域中で動作する（たとえば、周波数領域信号としてマルチチャネル信号ＭＣＳ１０を受信し、処理する）ように実装される。図２１に、装置Ａ１２０およびＡ２００の実装形態Ａ２２０のブロック図を示し、ここでは、方向推定モジュールＤＭ１０は、変換モジュールＸＭ２０から周波数領域においてマルチチャネル信号ＭＣＳ１０からの情報を受信するように構成される。

【０１６８】

[00200]一例では、方向推定モジュールＤＭ１０は、位相変換を使用したステアード応答パワー（ＳＲＰ－ＰＨＡＴ：steered response power using the phase transform）アルゴリズムを使用して、マルチチャネル信号ＭＣＳ１０内の情報に基づいて推定音源方向を計算するように実装される。ＳＲＰ－ＰＨＡＴアルゴリズムは、最尤音源定位から得られるものであり、出力信号の相関が最大となる時間遅延を判断する。相互相関は、各ビンにおいて電力によって正規化され、それにより、より良いロバストネスが与えられる。残響環境では、ＳＲＰ－ＰＨＡＴは、競合する音源定位方法よりも良い結果をもたらすことが予想され得る。

【０１６９】

[00201]ＳＲＰ－ＰＨＡＴアルゴリズムは、周波数領域における受信信号ベクトルＸ（すなわち、マルチチャネル信号ＭＣＳ１０）

$X(\omega) = [X_1(\omega), \ldots, X_p(\omega)]^T = S(\omega)G(\omega) + S(\omega)H(\omega) + N(\omega)$

で表され得、

ただし、Ｓは音源信号ベクトルを示し、利得行列Ｇ、室内伝達関数ベクトルＨ、および雑音ベクトルＮは次のように表され得る。

【数１５】

$$\mathbf{X}(\omega) = [X_1(\omega), \ldots, X_p(\omega)]^T,$$

$$\mathbf{G}(\omega) = [\alpha_1(\omega)e^{-j\omega\tau_1}, \ldots, \alpha_p(\omega)e^{-j\omega\tau_p}]^T,$$

$$\mathbf{H}(\omega) = [H_1(\omega), \ldots, H_p(\omega)]^T,$$

$$\mathbf{N}(\omega) = [N_1(\omega), \ldots, N_p(\omega)]^T.$$

【０１７０】

これらの式において、Ｐはセンサーの数（すなわち、入力チャネルの数）を示し、αは利得ファクタを示し、τは音源からの伝搬時間を示す。

【０１７１】

[00202]この例では、複合雑音ベクトル$N^c(\omega) = S(\omega)H(\omega) + N(\omega)$は、以下のゼロ平均、周波数独立、結合ガウス分布（zero-mean, frequency-independent, join

t Gaussian distribution）を有すると仮定され得る。
【数１６】

$$p(\mathbf{N}^c(\omega)) = \rho \exp\left\{-\frac{1}{2}[\mathbf{N}^c(\omega)]^H \mathbf{Q}^{-1}(\omega)\mathbf{N}^c(\omega)\right\},$$

【０１７２】
ただし、Ｑ（ω）は共分散行列であり、ρは定数である。音源方向は、次の式を最大化することによって推定され得る。
【数１７】

$$J_2 = \int_\omega \frac{[\mathbf{G}^H(\omega)\mathbf{Q}^{-1}(\omega)\mathbf{X}(\omega)]^H \mathbf{G}^H(\omega)\mathbf{Q}^{-1}(\omega)\mathbf{X}(\omega)}{\mathbf{G}^H(\omega)\mathbf{Q}^{-1}(\omega)\mathbf{G}(\omega)} d\omega.$$

【０１７３】
Ｎ（ω）＝０であるとの仮定の下で、この式は次のように書き直され得る。
【数１８】

$$J_2 = \frac{1}{\gamma P}\int \left|\sum_{i=1}^{P} \frac{X_i(\omega)e^{j\omega\tau_i}}{|X_i(\omega)|}\right|^2 d\omega, \qquad (4)$$

【０１７４】
ただし、０＜γ＜１は設計定数であり、式（４）の右辺を最大化する時間遅延Γ$_i$は音源の到来方向を示す。
【０１７５】
[00203]図２２に、周波数ωの範囲にわたる異なる２音源シナリオのＤＯＡ推定にＳＲＰ－ＰＨＡＴのそのような実装形態を使用した結果によるプロットの例を示す。これらのプロットでは、ｙ軸は
【数１９】

$$\left|\sum_{i=1}^{P} \frac{X_i(\omega)e^{j\omega\tau_i}}{|X_i(\omega)|}\right|^2$$

【０１７６】
の値を示し、ｘ軸は、アレイ軸に対する推定音源到来方向θ$_i$（
【数２０】

$$\theta_i \left(= \cos^{-1}\left(\tau_i c / d\right)\right)$$

【０１７７】
）を示す。各プロットにおいて、各線は範囲内の異なる周波数に対応し、各プロットはマイクロフォンアレイの縦方向を中心として対称的である（すなわち、θ＝０）。左上のプロットは、アレイから４メートルの距離にある２つの音源のヒストグラムを示している。右上のプロットは、アレイから４メートルの距離にある２つの近接した音源のヒストグラ

ムを示している。左下のプロットは、アレイから２．５メートルの距離にある２つの音源のヒストグラムを示している。右下のプロットは、アレイから２．５メートルの距離にある２つの近接した音源のヒストグラムを示している。これらのプロットの各々は、推定音源方向を、全周波数にわたる単一のピークとしてではなく、重心によって特徴づけられ得る角度範囲として示すことがわかるであろう。

【０１７８】

[00204]別の例では、方向推定モジュールＤＭ１０は、ブラインド音源分離（ＢＳＳ）アルゴリズムを使用して、マルチチャネル信号ＭＣＳ１０内の情報に基づいて推定音源方向を計算するように実装される。ＢＳＳ方法は、干渉音源からのエネルギーを除去する信頼できるヌルビームを発生する傾向があり、これらのヌルビームの方向は、対応する音源の到来方向を示すために使用され得る。方向推定モジュールＤＭ１０のそのような実装形態は、次のような式に従って、マイクロフォンｊおよびｊ’のアレイの軸に対する周波数ｆにおける音源ｉの到来方向（ＤＯＡ）を計算するように実装され得る。

【数２１】

$$\hat{\theta}_{i,jj'}(f) = \cos^{-1}\left(\frac{\arg\left([W^{-1}]_{ji}/[W^{-1}]_{j'i}\right)}{2\pi f c^{-1}\left\|p_j - p_{j'}\right\|}\right), \qquad (5)$$

【０１７９】

ただし、Ｗは逆混合行列を示し、p_jおよび$p_{j'}$は、それぞれマイクロフォンｊおよびｊ’の空間的座標を示す。この場合、本明細書で説明するようにフィルタ更新モジュールＵＭ１０によって更新されるフィルタとは別個に方向推定モジュールＤＭ１０のＢＳＳフィルタ（たとえば、逆混合行列Ｗ）を実装することが望ましいことがある。

【０１８０】

[00205]図２３に、４つのヒストグラムのセットの一例を示し、各ヒストグラムは、４行逆混合行列Ｗの対応するインスタンスの（アレイ軸に対する）各入射角に式（５）がマッピングする周波数ビンの数を示し、ただし、Ｗは、マルチチャネル信号ＭＣＳ１０内の情報に基づいており、本明細書で説明するＩＶＡ適応ルールに従って方向推定モジュールＤＭ１０の一実装形態によって計算される。この例では、入力マルチチャネル信号は、約４０～６０度の角度だけ分離された２つのアクティブな音源からのエネルギーを含んでいる。左上のプロットは、（音源１の方法を示す）ＩＶＡ出力１のヒストグラムを示しており、右上のプロットは、（音源２の方法を示す）ＩＶＡ出力２のヒストグラムを示している。これらのプロットの各々は、推定音源方向を、全周波数にわたる単一のピークとしてではなく、重心によって特徴づけられ得る角度範囲として示すことがわかるであろう。下のプロットは、ＩＶＡ出力３および４のヒストグラムを示しており、これらは、両方の音源からのエネルギーをブロックし、残響からのエネルギーを含んでいる。

【０１８１】

[00206]別の例では、方向推定モジュールＤＭ１０は、複数の異なる周波数成分の各々についてマルチチャネル信号ＭＣＳ１０のチャネル間の位相差に基づいて推定音源方向を計算するように実装される。（たとえば、図１９Ｂに示された平面波面の仮定が有効になるように）遠距離場に点音源が１つあり、残響がない理想的な場合、位相差と周波数との比は周波数に対して一定である。図１５Ｂに示されたモデルを参照すると、方向推定モジュールＤＭ１０のそのような実装形態は、量

【数２２】

$$\frac{c\Delta\varphi_i}{d2\pi f_i}$$

【０１８２】

の（アークコサインとも呼ばれる）逆コサインとして音源方向θ_iを計算するように構成され得、ただし、ｃは音速（約３４０ｍ／秒）を示し、ｄはマイクロフォン間の距離を示し、$\Delta\phi_i$は２つのマイクロフォンチャネルの対応する位相推定間のラジアンの差分を示し、f_iは、位相推定が対応する周波数成分（たとえば、対応するＦＦＴサンプルの周波数、あるいは対応するサブバンドの中心周波数またはエッジ周波数）である。

【０１８３】

画像中のオブジェクト深さ判断

[00207]以下で、画像からオブジェクト深さ情報を判断するための例示的な構成について説明する。第１の構成では、画像中のオブジェクトの推定深さを判断するために、マルチカメラ画像視差技法が使用される。第２の構成では、画像シーン中のオブジェクト範囲を推定するために単一カメラ自動フォーカス技法が使用され得る。ＳＩＦＴキーポイント探索は、推定キーポイント深さ情報を含むことによってよりロバストにされ得る。

【０１８４】

[00208]図２４は、画像またはビデオキャプチャ中にシーン中のオブジェクトの視差を検出するように構成された画像キャプチャデバイス１３５０の特定の構成の図である。画像キャプチャデバイス１３５０は、画像処理モジュール１３５６に結合された画像センサーペア１３５２を含む。画像処理モジュール１３５６は外部メモリ１３６２に結合される。画像処理モジュール１３５６は、同期およびインターフェースモジュール１３５４と、画像処理機能モジュール１３５８と、視差検出モジュール１３４２と、符号化モジュール１３６０とを含む。

【０１８５】

[00209]画像センサーペア１３５２は、画像データ１３７０を画像処理モジュール１３５６に与えるように構成される。単一のシーンに対応する第１の画像と第２の画像とを使用してオブジェクト深さ判断が実行され得る。第１の画像は、第１のセンサー（たとえば、右センサー）によるシーンの第１の画像キャプチャに対応し得、第２の画像は、第２のセンサー（たとえば、左センサー）によるシーンの第２の画像キャプチャに対応し得、第２の画像キャプチャは、図２４に示すセンサーペア１３５２などによって、第１の画像キャプチャと実質的に同時である。

【０１８６】

[00210]同期およびインターフェースモジュール１３５４は、データ１３７２を画像処理機能モジュール１３５８に与えるように構成される。画像処理機能モジュール１３５８は、処理された画像データ１３８０を視差検出モジュール１３４２に与えるように構成される。符号化モジュール１３６０は、画像／ビデオデータ１３８２を受信し、オブジェクト深さデータで符号化された画像／ビデオデータ１３８４を発生するように構成される。

【０１８７】

[00211]視差検出モジュール１３４２は、画像センサーペア１３５２によってキャプチャされたシーン内のオブジェクトに対応する視差値を判断するように構成され得る。特定の構成では、視差検出モジュール１３４２は、シーン固有オブジェクト検出またはキーポイント検出および視差判断機能を組み込む。

【０１８８】

[00212]画像センサーペア１３５２は、代表的な図では、右センサー（すなわち、閲覧者の右眼によって知覚されるシーンに関連する画像をキャプチャする第１のセンサー）と、左センサー（すなわち、閲覧者の左眼によって知覚されるシーンに関連する画像をキャプチャする第２のセンサー）とを含むセンサーのペアとして示されている。画像データ１３７０は、左センサーによって生成された左画像データと、右センサーによって生成された右画像データとを含む。各センサーは、水平方向に延在する感光性構成要素の行と、垂直方向に延在する感光性構成要素の列とを有するものとして示されている。左センサーと右センサーは、水平方向に沿って互いに距離ｄにおいて実質的に位置合わせされる。本明細書で使用する画像データ内の「水平」方向は、右画像データ中のオブジェクトのロケーションと、左画像データ中の同じオブジェクトのロケーションとの間の変位の方向である

。

【０１８９】

[00213]図２５は、図２４のシステム中に含まれ得る画像処理システム１４４０の特定の実施形態の図である。処理システム１４４０は、入力画像データ１４０４を受信し、出力画像データ１４２８を発生するように構成される。処理システム１４４０は、較正入力１４５０を介して受信されるカメラ較正パラメータ１４０６に応答し得る。

【０１９０】

[00214]画像処理システム１４４０は、微細ジオメトリ補正モジュール１４１０と、キーポイント検出モジュール１４１２と、キーポイント整合モジュール１４１４と、深さ計算モジュール１４１６とを含む。

【０１９１】

[00215]ジオメトリ補正モジュール１４１０は、データ経路１４７０を介して入力画像データ１４０４を受信し、補正された画像データ１４５４を発生するように構成される。ジオメトリ補正モジュール１４１０は、カメラ較正パラメータ１４０６からのデータを使用し得、入力画像データ１４０４を調整して、画像データ１４０４のレンダリングに悪影響を及ぼし得る不整合、収差、または他の較正状態について訂正し得る。例示のために、ジオメトリ補正モジュール１４１０は、較正パラメータ１４０６について調整するために、任意のグリッド上で画像データ１４０４のリサンプリングを効果的に実行し得る。

【０１９２】

[00216]処理システム１４４０がコンピューティングデバイス中に実装され得る構成では、カメラ較正パラメータ１４０６は、画像／ビデオデータファイルのヘッダ中でなど、入力画像データ１４０４とともに受信され得る。処理システム１４４０が図２４の画像キャプチャデバイス１３５０などの画像キャプチャデバイス中に実装される構成では、カメラ較正パラメータ１４０６は、画像キャプチャデバイスの画像センサーペアに対応し得、微細ジオメトリ補正モジュール１４１０にとってアクセス可能なメモリに記憶され得る。

【０１９３】

[00217]キーポイント検出モジュール１４１２は、補正された画像データ１４５４を受信し、キーポイントロケーションデータ１４５６を発生するように構成される。キーポイント検出モジュール１４１２は、補正された画像データ１４５４中の特徴的なポイントを識別するように構成される。たとえば、特徴的なポイントは、シーン中のオブジェクトの垂直エッジ、または水平方向において高周波成分を有するそのシーンの他のポイントに対応し得る。画像データ中のそのような特徴的な要素を本明細書では「キーポイント」または「オブジェクト」と呼ぶが、そのような識別された要素は、個々のピクセル、ピクセルのグループ、分数ピクセル部分、他の画像成分、またはそれらの任意の組合せに対応し得ることを理解されたい。たとえば、キーポイントは、受信された画像データのサブサンプリングされたルーマ成分をもつピクセルに対応し得、垂直エッジ検出フィルタを使用して検出され得る。

【０１９４】

[00218]キーポイント整合モジュール１４１４は、キーポイントロケーションデータ１４５４を受信し、識別されたキーポイントに対応する視差データ１４５８を発生するように構成される。キーポイント整合モジュール１４１４は、探索範囲内でキーポイントの周りを探索し、視差ベクトルの信頼性測度を生成するように構成され得る。

【０１９５】

[00219]深さ計算モジュール１４１６は、視差データ１４５８を受信し、センサー１３５２からのキーポイントの推定距離を示すレンジデータ１４６０を発生するように構成される。

【０１９６】

[00220]処理システム１４４０の動作中に、レンジ評価プロセスが実行される。画像データ１４０４をキャプチャした２つのセンサー間の相対位置を推定し、補正するように設計された較正手順が、オフラインで（たとえば、デバイスのエンドユーザへの配信より前

に）実行され得るが、ジオメトリ補正は画像データ１４０４のフレームごとに実行され得る。

【０１９７】

[00221]処理は、（たとえば、キーポイント検出モジュール１４１２において）キーポイント検出を続ける。視差を確実に推定するために使用され得る画像のオブジェクトまたはピクセル（キーポイント）のセットが選択される。推定視差における高い信頼性が達成され得るが、シーン中のすべての領域またはオブジェクトが使用されるとは限らない。キーポイントのセットの選択は、適切な（１つまたは複数の）解像度を生成するために、画像サブサンプリングを含み得る。（たとえば、垂直方向の特徴に対応する水平周波数のみを探すために）画像高域フィルタを適用し、その後、フィルタを適用することによって発生した結果の平方値または絶対値を取り得る。所定のしきい値を超える結果は、潜在的キーポイントとして識別され得る。一部の局所近傍内の最良のキーポイント（たとえば、所定の領域内にあるすべてのキーポイントの最大フィルタ結果に対応するキーポイント）を選択するために、潜在的キーポイントに対してキーポイントプルーニングプロセスが実行され得る。

【０１９８】

[00222]検出されたキーポイントを使用して、（たとえば、キーポイント整合モジュール１４１４において）キーポイント整合が実行され得る。第１の画像（たとえば左画像または右画像）中のキーポイントと、第２の画像（たとえば左画像および右画像のうちの他方）中の対応するエリアとの間の対応が判断され得る。信頼性推定値が生成され得、それは、キーポイント選択とともに視差推定精度を著しく改善し得る。左画像中のキーポイントと右画像中のキーポイントとの間の整合がどれくらい近接しているかの判断を可能にするために、整合は、正規化された相互共分散（cross-covariance）を使用して実行され得る。信頼性測度はこの正規化された相互共分散に基づき得る。特定の実施形態では、第１の画像中のキーポイントに対応する第２の画像中のキーポイントの位置を特定するための探索範囲は、センサー較正のための画像補正がすでに実行されているので、水平のみであり、探索範囲は、第１の画像中のキーポイントの周りの一定の範囲のみをカバーするように調整される。視差値はこれらの比較から計算される。

【０１９９】

[00223]図２６Ａおよび図２６Ｂは、知覚されたオブジェクト深さと相関させられたオブジェクト視差の例示的な実施形態の図である。オブジェクト深さ判断は、異なる画像を各眼１５０４、１５０６にダイレクトすることに依拠する。目的は、オブジェクト視差（水平シフト）が深さと相関させられるように、左および右（Ｌ／Ｒ）画像から深さの錯覚を再生成することである。図２６Ａは、ディスプレイ表面１５２４を越えて知覚されるオブジェクト１５３０に対応する正の視差１５５０を示している。視差１５５０は、左画像中のオブジェクトのロケーション１５２０と、右画像中のオブジェクトのロケーション１５２２との間の距離を示す。観察者は、左画像中のオブジェクト１５３０の画像と、右画像中のオブジェクト１５３０の画像とを融合させて、左眼１５０４の見通し線１５６０と、右眼１５０６の見通し線１５６２との交点においてオブジェクト１５３０を知覚することになる。

【０２００】

[00224]図２６Ｂは、ディスプレイ表面１５２４の前で知覚されるオブジェクト１５３０に対応する負の視差１５５０を示している。視差１５５０は、左画像中のオブジェクトのロケーション１５２０と、右画像中のオブジェクトのロケーション１５２２との間の距離を示す。観察者は、左画像中のオブジェクト１５３０の画像と、右画像中のオブジェクト１５３０の画像とを融合させて、左眼１５０４の見通し線１５６０と、右眼１５０６の見通し線１５６２との交点において、ディスプレイ表面１５３４の前でオブジェクト１５３０を知覚することになる。

【０２０１】

[00225]２つの眼から見えるオブジェクト変位は、視覚野によって深さとして解釈され

る。２つのキャプチャされた画像間の視差はシーンに依存することになる。シーン深さを感知することを使用すると、画像中のキーポイント探索を特定の深さでのまたはその近くでのオブジェクトのみに狭めることができ、したがって、オブジェクト認識の信頼性を高めることができる。

【０２０２】

[00226]深さ計算モジュール６０２によって実行されるシーンレンジ推定は、左画像と右画像との間のスパース動きベクトル推定として一般化され得る。シーンレンジ評価プロセスはキー（特徴的な）ポイント識別を含むことができる。水平シフトのみが存在する（および測定される）ので、垂直変化は必要とされない。水平変化（何らかの垂直成分をもつエッジ）が使用される。いくつかの構成では、キーポイントは異なる解像度で検出され得る。オブジェクトレンジ推定プロセスはまた、キーポイント整合を含むことができる。光源レベル非依存になるために、およびロバストな視差信頼性メトリックを生成するために、キーポイント整合は、正規化された相互共分散を使用して実行され得る。その結果、キーポイントを異なる解像度で整合させることは不要になり得る。

【０２０３】

オーディオシーン分解

[00227]音響分解サブシステム２２は、シーンから記録されたオーディオ信号を分解するために、このセクションで説明する技法を採用することができる。本明細書で開示するものは、楽音（note）のペンデンシ（pendency）にわたる楽音のスペクトルの変化に関係する情報を含む基底関数インベントリと、スパース復元技法とを使用する、オーディオ信号の分解である。そのような分解は、信号の分析、符号化、再生、および／または合成をサポートするために使用され得る。本明細書では、調波楽器（すなわち、非打楽器）および打楽器からの混合音を含むオーディオ信号の定量分析の例を示す。

【０２０４】

[00228]開示する技法は、キャプチャされたオーディオ信号を一連のセグメントとして処理するように構成され得る。典型的なセグメント長は約５または１０ミリ秒から約４０または５０ミリ秒にわたり、セグメントは、重複しても（たとえば、隣接するセグメントが２５％または５０％だけ重複する）、重複しなくてもよい。１つの特定の例では、信号は、１０ミリ秒の長さをそれぞれ有する一連の重複しないセグメントまたは「フレーム」に分割される。また、そのような方法によって処理されるセグメントは、異なる演算によって処理されるより大きいセグメントのセグメント（すなわち、「サブフレーム」）であり得、またはその逆も同様である。

【０２０５】

[00229]２つ以上の楽器および／またはボーカル信号の混合から個々のノート／ピッチプロファイルを抽出するために音楽シーンを分解することが望ましいことがある。潜在的な使用事例としては、複数のマイクロフォンを用いてコンサート／ビデオゲームシーンをテープに記録すること、空間／スパース復元処理を用いて楽器とボーカルとを分解すること、ピッチ／ノートプロファイルを抽出すること、補正ピッチ／ノートプロファイルを用いて個々の音源を部分的にまたは完全にアップミックスすることがある。そのような動作は、音楽アプリケーション（たとえば、ＱｕａｌｃｏｍｍのＱＵＳＩＣアプリケーション、Ｒｏｃｋ　ＢａｎｄまたはＧｕｉｔａｒ　Ｈｅｒｏなどのビデオゲーム）の機能をマルチプレーヤ／シンガーシナリオに拡張するために使用され得る。

【０２０６】

[00230]（たとえば、図３４に示すように）同時に２人以上のボーカリストがアクティブであり、および／または複数の楽器がプレイされるシナリオを音楽アプリケーションが処理することを可能にすることが望ましいことがある。そのような機能は、現実的な音楽テープ記録シナリオ（マルチピッチシーン）をサポートするために望ましいことがある。ユーザは、各音源を別々に編集および再合成する能力を希望し得るが、サウンドトラックを生成することは、それらの音源を同時に記録することを伴い得る。

【０２０７】

[00231]本開示では、複数の音源が同時にアクティブになり得る音楽アプリケーションのための使用事例を可能にするために使用され得る方法について説明する。そのような方法は、基底関数インベントリベースのスパース復元（たとえば、スパース分解）技法を使用してオーディオ混合信号を分析するように構成され得る。

【０２０８】

[00232]基底関数のセットについて（たとえば、効率的なスパース復元アルゴリズムを使用して）アクティブ化係数の最もスパースなベクトルを見つけることによって混合信号スペクトルを音源成分に分解することが望ましいことがある。基底関数のセットは、図２の画像／ビデオ処理ブロック５４によってシーン中に存在すると示された特定のタイプの楽器に減少させられ得る。アクティブ化係数ベクトルを（たとえば、基底関数のセットとともに）使用して、混合信号を再構成するか、または混合信号の（たとえば、１つまたは複数の選択された楽器からの）選択された部分を再構成し得る。また、（たとえば、大きさおよび時間サポートに従って）スパース係数ベクトルを後処理することが望ましいことがある。

【０２０９】

[00233]図２７Ａに、オーディオ信号を分解する方法Ｍ１００のフローチャートを示す。方法Ｍ１００は、オーディオ信号のフレームからの情報に基づいて、周波数範囲にわたる対応する信号表現を計算するタスクＴ１００を含む。方法Ｍ１００は、タスクＴ１００によって計算された信号表現と、複数の基底関数とに基づいて、アクティブ化係数のベクトルを計算するタスクＴ２００をも含み、アクティブ化係数の各々は、複数の基底関数のうちの異なる１つに対応する。

【０２１０】

[00234]タスクＴ１００は、信号表現を周波数領域ベクトルとして計算するように実装され得る。そのようなベクトルの各要素は、メルまたはバーク尺度（mel or Bark scale）に従って取得され得る、サブバンドのセットのうち対応する１つのサブバンドのエネルギーを示し得る。しかしながら、そのようなベクトルは、一般に、高速フーリエ変換（ＦＦＴ）、または短時間フーリエ変換（ＳＴＦＴ）など、離散フーリエ変換（ＤＦＴ）を使用して計算される。そのようなベクトルは、たとえば、６４、１２８、２５６、５１２、または１０２４ビンの長さを有し得る。一例では、オーディオ信号は、８ｋＨｚのサンプリングレートを有し、０～４ｋＨｚ帯域は、長さ３２ミリ秒の各フレームについて２５６ビンの周波数領域ベクトルによって表される。別の例では、信号表現は、オーディオ信号の重複セグメントにわたる修正離散コサイン変換（ＭＤＣＴ）を使用して計算される。

【０２１１】

[00235]さらなる一例では、タスクＴ１００は、フレームの短期電力スペクトルを表すケプストラム係数（たとえば、メル周波数ケプストラム係数またはＭＦＣＣ）のベクトルとして信号表現を計算するように実装される。この場合、タスクＴ１００は、フレームのＤＦＴ周波数領域ベクトルの大きさにメル尺度フィルタバンクを適用することと、フィルタ出力の対数をとることと、対数値のＤＣＴをとることとによって、そのようなベクトルを計算するように実装され得る。そのような手順は、たとえば、「STQ: DSR - Front-end feature extraction algorithm; compression algorithm」と題する、ＥＴＳＩドキュメントＥＳ　２０１　１０８に記載されているオーロラ規格（欧州通信規格協会、２０００年）において説明されている。

【０２１２】

[00236]楽器は、一般に、明確な音色を有する。楽器の音色は、それのスペクトルエンベロープ（たとえば、周波数範囲にわたるエネルギーの分布）によって記述され得るので、異なる楽器の音色の範囲は、個々の楽器のスペクトルエンベロープを符号化する基底関数のインベントリを使用してモデル化され得る。

【０２１３】

[00237]各基底関数は、周波数範囲にわたる対応する信号表現を備える。これらの信号表現の各々は、タスクＴ１００によって計算された信号表現と同じ形態を有することが望

ましいことがある。たとえば、各基底関数は、長さ６４、１２８、２５６、５１２、または１０２４ビンの周波数領域ベクトルであり得る。代替的に、各基底関数は、ＭＦＣＣのベクトルなどのケプストラム領域ベクトルであり得る。さらなる一例では、各基底関数はウェーブレット領域ベクトルである。

【０２１４】

[00238]基底関数インベントリＡは、各楽器ｎ（たとえば、ピアノ、フルート、ギター、ドラムなど）の基底関数のセットA_nを含み得る。たとえば、楽器の音色は、概して、各楽器ｎの基底関数のセットA_nが、一般に、楽器ごとに異なり得るある所望のピッチ範囲にわたる各ピッチについて少なくとも１つの基底関数を含むようなピッチ従属である。たとえば、半音階スケールにチューニングされた楽器に対応する基底関数のセットは、オクターブ当たり１２ピッチの各々の異なる基底関数を含み得る。ピアノの基底関数のセットは、ピアノの各キーについて異なる基底関数を含み、合計で８８個の基底関数を含み得る。別の例では、各楽器の基底関数のセットは、５オクターブ（たとえば、５６ピッチ）または６オクターブ（たとえば、６７ピッチ）など、所望のピッチ範囲内の各ピッチについて異なる基底関数を含む。基底関数のこれらのセットA_nは独立であり得、または２つ以上のセットが１つまたは複数の基底関数を共有し得る。

【０２１５】

[00239]セットの各基底関数は、楽器の音色を異なる対応するピッチで符号化し得る。音楽信号のコンテキストでは、人間ボイスは、インベントリが１つまたは複数の人間ボイスモデルの各々の基底関数のセットを含み得るような楽器と見なされ得る。

【０２１６】

[00240]基底関数のインベントリは、アドホック記録された個々の楽器記録から学習された一般的な楽器ピッチデータベースに基づき得、および／または（たとえば、独立成分分析（ＩＣＡ）、期待値最大化（ＥＭ：expectation-maximization）などの分離方式を使用した）混合の分離されたストリームに基づき得る。

【０２１７】

[00241]オーディオを処理するための基底関数のセットの選択は、図２の画像／ビデオ処理ブロック５４によって与えられる楽器候補のリストに基づき得る。たとえば、基底関数のセットは、画像／ビデオ処理ブロック５４のオブジェクト認識プロセスによってシーン中で識別される楽器のみに制限され得る。

【０２１８】

[00242]タスクＴ１００によって計算された信号表現と、インベントリＡからの基底関数の複数Ｂとに基づいて、タスクＴ２００はアクティブ化係数のベクトルを計算する。このベクトルの各係数は、基底関数の複数Ｂのうちの異なる１つに対応する。たとえば、タスクＴ２００は、基底関数の複数Ｂに従って、ベクトルが信号表現のための最も有望なモデルを示すように、ベクトルを計算するように構成され得る。図３２に、そのようなモデル$Bf=y$を示し、ここで、基底関数の複数Ｂは、Ｂ個の列が個々の基底関数であり、ｆが基底関数アクティブ化係数の列ベクトルであり、ｙが、記録された混合信号のフレーム（たとえば、スペクトログラム周波数ベクトルの形態の５、１０、または２０ミリ秒フレーム）の列ベクトルであるような行列である。

【０２１９】

[00243]タスクＴ２００は、線形プログラミング問題を解くことによって、オーディオ信号の各フレームのアクティブ化係数ベクトルを復元するように構成され得る。そのような問題を解くために使用され得る方法の例としては、非負値行列因子分解（ＮＮＭＦ：nonnegative matrix factorization）がある。ＮＮＭＦに基づくシングルチャネル基準方法は、（たとえば、以下で説明するように）期待値最大化（ＥＭ）更新ルールを使用して、基底関数とアクティブ化係数とを同時に計算するように構成され得る。

【０２２０】

[00244]既知または部分的に既知の基底関数空間における最もスパースなアクティブ化係数ベクトルを見つけることによって、オーディオ混合信号を（１つまたは複数の人間ボ

イスを含み得る）個々の楽器に分解することが望ましいことがある。たとえば、タスクＴ２００は、既知の楽器基底関数のセットを使用して、（たとえば、効率的なスパース復元アルゴリズムを使用して）基底関数インベントリにおける最もスパースなアクティブ化係数ベクトルを見つけることによって、入力信号表現を音源成分（たとえば、１つまたは複数の個々の楽器）に分解するように構成され得る。

【０２２１】

[00245]劣決定系（underdetermined system）の連立一次方程式（すなわち、式よりも多い未知数を有する系）の最小Ｌ１ノルム解は、しばしばその系の最もスパースな解でもあることが知られている。Ｌ１ノルムの最小化によるスパース復元は、以下のように実行され得る。

【０２２２】

[00246]ターゲットベクトルf_0は、Ｋ＜Ｎ個の非０成分を有する長さＮのスパースベクトルであり（すなわち、「Ｋスパース」であり）、射影行列（すなわち、基底関数行列）Ａは、サイズ約Ｋのセットについてインコヒーレント（ランダム様）であると仮定する。信号ｙ＝Ａf_oであることがわかる。次いで、Ａｆ＝ｙｚを条件として

【数２３】

$$\min \|f\|_{l_2}$$

【０２２３】

を解くこと（ただし、||ｆ||$_{l1}$は

【数２４】

$$\sum_{i=1}^{n} |f_i|$$

【０２２４】

として定義される）により、f_oが正確に復元される。その上、扱いやすいプログラムを解くことによって、Ｍ≧Ｋ・ｌｏｇＮ個のインコヒーレント測定値からf_oを復元することができる。測定値の数Ｍは、アクティブな成分の数にほぼ等しい。

【０２２５】

[00247]１つの手法は、圧縮感知（compressive sensing）からのスパース復元アルゴリズムを使用することである。（「圧縮感知（compressed sensing）」とも呼ばれる）圧縮感知の一例では、信号復元Φｘ＝ｙであり、ｙは、長さＭの観測信号ベクトルであり、ｘは、ｙの凝縮（condensed）表現である、Ｋ＜Ｎ個の非０成分を有する長さＮのスパースベクトル（すなわち、「Ｋスパースモデル」）であり、Φは、サイズＭ×Ｎのランダム射影行列である。ランダム射影Φはフルランクでないが、それは、高確率でスパース／圧縮可能信号モデルについて可逆である（すなわち、それは不良設定逆問題（ill-posed inverse problem）を解く）。

【０２２６】

[00248]アクティブ化係数ベクトルｆは、対応する基底関数セットA_nのアクティブ化係数を含む各楽器ｎのサブベクトルf_nを含むと見なされ得る。これらの楽器固有のアクティブ化サブベクトルは独立して（たとえば、後処理動作において）処理され得る。たとえば、１つまたは複数のスパーシティ制約（たとえば、ベクトル要素の少なくとも半分が０であること、楽器固有のサブベクトル中の非０要素の数が最大値を超えないことなど）を強制することが望ましいことがある。アクティブ化係数ベクトルの処理は、各フレームについて各非０アクティブ化係数のインデックス番号を符号化すること、各非０アクティブ

化係数のインデックスと値とを符号化すること、またはスパースベクトル全体を符号化することを含み得る。そのような情報は、示されたアクティブな基底関数を使用して混合信号を再生するため、または混合信号の特定の部分のみ（たとえば、特定の楽器によってプレイされるノートのみ）を再生するために、（たとえば、別の時間および／またはロケーションにおいて）使用され得る。

【０２２７】

[00249]楽器によって生成されるオーディオ信号は、ノートと呼ばれる一連のイベントとしてモデル化され得る。ノートをプレイする調波楽器の音は、たとえば、（アタックとも呼ばれる）オンセット段階、（サスティーンとも呼ばれる）定常段階、および（リリースとも呼ばれる）オフセット段階の、時間的に異なる領域に分割され得る。ノートの時間エンベロープの別の記述（ＡＤＳＲ）は、アタックとサスティーンとの間の追加のディケイ（decay）段階を含む。このコンテキストでは、ノートの持続時間は、アタック段階の開始からリリース段階の終了まで（または、同じ弦上の別のノートの開始など、そのノートを終了する別のイベントまで）の間隔として定義され得る。ノートは単一のピッチを有すると仮定されるが、インベントリは、単一のアタックと（たとえば、ビブラートまたはポルタメントなどのピッチベンディング効果によって生成される）複数のピッチとを有するノートをモデル化するようにも実装され得る。いくつかの楽器（たとえば、ピアノ、ギター、またはハープ）は、コードと呼ばれるイベントにおいて一度に２つ以上のノートを生成し得る。

【０２２８】

[00250]異なる楽器によって生成されるノートはサスティーン段階中に同様の音色を有し得るので、そのような期間中にどの楽器がプレイしているかを識別することは困難であり得る。しかしながら、ノートの音色は、段階ごとに変化することが予想され得る。たとえば、アクティブな楽器を識別することは、サスティーン段階中よりもアタックまたはリリース段階中に容易であり得る。

【０２２９】

[00251]アクティブ化係数ベクトルが適切な基底関数を示す可能性を高めるために、基底関数間の差分を最大にすることが望ましいことがある。たとえば、基底関数が時間に対するノートのスペクトルの変化に関係する情報を含むことが望ましいことがある。

【０２３０】

[00252]時間に対する音色の変化に基づいて基底関数を選択することが望ましいことがある。そのような手法は、ノートの音色のそのような時間領域展開に関係する情報を基底関数インベントリに符号化することを含み得る。たとえば、特定の楽器ｎの基底関数のセットA_nは、２つ以上の対応する信号表現の各々がノートの展開における異なる時間（たとえば、アタック段階の時間、サスティーン段階の時間、およびリリース段階の時間）に対応するように、各ピッチにおいてこれらの信号表現を含み得る。これらの基底関数は、ノートをプレイする楽器の記録の対応するフレームから抽出され得る。

【０２３１】

[00253]図２７Ｃに、一般的構成による、オーディオ信号を分解するための装置ＭＦ１００のブロック図を示す。装置ＭＦ１００は、（たとえば、タスクＴ１００に関して本明細書で説明したように）オーディオ信号のフレームからの情報に基づいて、周波数範囲にわたる対応する信号表現を計算するための手段Ｆ１００を含む。装置ＭＦ１００は、（たとえば、タスクＴ２００に関して本明細書で説明したように）手段Ｆ１００によって計算された信号表現と、複数の基底関数とに基づいて、アクティブ化係数の各々が複数の基底関数のうちの異なる１つに対応する、アクティブ化係数のベクトルを計算するための手段Ｆ２００をも含む。

【０２３２】

[00254]図２７Ｄに、変換モジュール２１００と係数ベクトル計算器２２００とを含む、別の一般的構成による、オーディオ信号を分解するための装置Ａ１００のブロック図を示す。変換モジュール２１００は、（たとえば、タスクＴ１００に関して本明細書で説明

したように）オーディオ信号のフレームからの情報に基づいて、周波数範囲にわたる対応する信号表現を計算するように構成される。係数ベクトル計算器２２００は、（たとえば、タスクＴ２００に関して本明細書で説明したように）変換モジュール２１００によって計算された信号表現と、複数の基底関数とに基づいて、アクティブ化係数の各々が複数の基底関数のうちの異なる１つに対応する、アクティブ化係数のベクトルを計算するように構成される。

【０２３３】

[00255]図２７Ｂに、基底関数インベントリが各ピッチにおける各楽器のための複数の信号表現を含む、方法Ｍ１００の実装形態Ｍ２００のフローチャートを示す。これらの複数の信号表現の各々は、周波数範囲にわたるエネルギーの複数の異なる分布（たとえば、複数の異なる音色）を記述する。インベントリはまた、異なる時間関係モダリティのために異なる複数の信号表現を含むように構成され得る。１つのそのような例では、インベントリは、各ピッチにおける弾かれた（bowed）弦の複数の信号表現と、各ピッチにおけるはじかれた（plucked）（たとえば、ピッツィカートの）弦の異なる複数の信号表現とを含む。

【０２３４】

[00256]方法Ｍ２００は、タスクＴ１００の複数のインスタンス（この例では、タスクＴ１００ＡおよびＴ１００Ｂ）を含み、各インスタンスは、オーディオ信号の対応する異なるフレームからの情報に基づいて、周波数範囲にわたる対応する信号表現を計算する。様々な信号表現は連結され得、同様に、各基底関数は複数の信号表現の連結であり得る。この例では、タスクＴ２００は、混合フレームの連結を各ピッチにおける信号表現の連結に整合させる。図３３に、混合信号ｙのフレームｐ１とフレームｐ２とが整合のために連結された、図３２のモデルＢｆ＝ｙの変形Ｂ’ｆ＝ｙの一例を示す。

【０２３５】

[00257]インベントリは、各ピッチにおける複数の信号表現がトレーニング信号の連続するフレームからとられるように構築され得る。他の実装形態では、各ピッチにおける複数の信号表現が、時間的により大きい窓にわたる（たとえば、連続するフレームではなく時間的に分離されたフレームを含む）ことが望ましいことがある。たとえば、各ピッチにおける複数の信号表現が、アタック段階と、サスティーン段階と、リリース段階との中の少なくとも２つからの信号表現を含むことが望ましいことがある。ノートの時間領域展開に関するより多くの情報を含むことによって、異なるノートの基底関数のセット間の差分が増加され得る。

【０２３６】

[00258]図２８Ａに、セグメントの高周波を強調するタスクＴ３００を含む方法Ｍ１００の実装形態Ｍ３００のフローチャートを示す。この例では、タスクＴ１００は、事前強調の後にセグメントの信号表現を計算するように構成される。図２９Ａに、タスクＴ３００の複数のインスタンスＴ３００Ａ、Ｔ３００Ｂを含む、方法Ｍ２００の実装形態Ｍ４００のフローチャートを示す。一例では、事前強調タスクＴ３００は、２００Ｈｚを上回るエネルギーと総エネルギーとの比を増加させる。

【０２３７】

[00259]図２８Ｂに、変換モジュール２１００の上流でオーディオ信号に対して高周波強調を実行するように構成された事前強調フィルタ２３００（たとえば、１次高域フィルタなどの高域フィルタ）を含む装置Ａ１００の実装形態Ａ３００のブロック図を示す。図２８Ｃに、事前強調フィルタ２３００が変換係数に対して高周波事前強調を実行するように構成された、装置Ａ１００の別の実装形態Ａ３１０のブロック図を示す。これらの場合、また、基底関数の複数Ｂに対して高周波事前強調（たとえば、高域フィルタ処理）を実行することが望ましいことがある。

【０２３８】

[00260]楽音は、ビブラートおよび／またはトレモロなどのカラーレーション効果を含み得る。ビブラートは、一般に、４または５から７、８、１０、または１２ヘルツまでの

範囲内にある変調レートをもつ周波数変調である。ビブラートによるピッチ変化は、シンガーの場合には、０．６から２半音の間で変動し得、管弦楽器の場合には、概して＋／－０．５半音よりも少ない（たとえば、弦楽器の場合には、０．２から０．３５半音の間である）。トレモロは、一般に同様の変調レートを有する振幅変調である。

【０２３９】

[00261]基底関数インベントリにおいてそのような効果をモデル化することは困難であり得る。そのような効果の存在を検出することが望ましいことがある。たとえば、ビブラートの存在は、４～８Ｈｚの範囲内の周波数領域ピークによって示され得る。また、そのような特性は、再生中に効果を復元するために使用され得るので、検出された効果のレベルの測度を（たとえば、このピークのエネルギーとして）記録することが望ましいことがある。トレモロの検出および定量化では、同様の処理が時間領域において実行され得る。効果が検出され、場合によっては定量化された後、ビブラートの場合には時間に対して周波数を平滑化することによって、またはトレモロの場合には時間に対して振幅を平滑化することによって変調を除去することが望ましいことがある。

【０２４０】

[00262]図３０Ｂに、変調レベル計算器（ＭＬＣ：modulation level calculator）を含む装置Ａ１００の実装形態Ａ７００のブロック図を示す。ＭＬＣは、上記で説明したように、オーディオ信号のセグメント中の検出された変調の測度（たとえば、時間または周波数領域における検出された変調ピークのエネルギー）を計算し、場合によっては記録するように構成される。

【０２４１】

[00263]本開示では、複数の音源が同時にアクティブになり得る音楽アプリケーションのための使用事例を可能にするために使用され得る方法について説明する。そのような事例では、可能な場合、アクティブ化係数ベクトルを計算する前に音源を分離することが望ましいことがある。この目的を達成するために、マルチチャネル技法とシングルチャネル技法との組合せが提案される。

【０２４２】

[00264]図２９Ｂに、信号を空間クラスタに分離するタスクＴ５００を含む方法Ｍ１００の実装形態Ｍ５００のフローチャートを示す。タスクＴ５００は、音源をできる限り多くの空間クラスタに隔離するように構成され得る。一例では、タスクＴ５００は、記録された音響シナリオをできる限り多くの空間クラスタに分離するために、マルチマイクロフォン処理を使用する。そのような処理は、マイクロフォン信号間の利得差および／または位相差に基づき得、そのような差は、周波数帯域全体にわたって評価されるか、あるいは複数の異なる周波数サブバンドまたは周波数ビンの各々において評価され得る。

【０２４３】

[00265]空間分離方法のみでは、所望の分離レベルを達成するには不十分であり得る。たとえば、いくつかの音源は、マイクロフォンアレイに対して近接しすぎているか、または場合によっては準最適に構成されることがある（たとえば、複数のバイオリン奏者および／または調波楽器が１つのコーナーに位置し得、打楽器奏者が通常後方に位置する）。典型的な音楽バンドシナリオでは、音源は（たとえば、図３４に示すように）互いに近接して位置するかまたは他の音源のさらに後ろに位置し得るので、空間情報のみを使用して、バンドに対して同じ概略的な方向にあるマイクロフォンのアレイによってキャプチャされた信号を処理すると、音源のすべてを互いから区別することができないことがある。タスクＴ１００およびＴ２００は、（たとえば、図３４に示すように）個々の楽器を分離するために、本明細書で説明するシングルチャネル基底関数インベントリベースのスパース復元（たとえば、スパース分解）技法を使用して個々の空間クラスタを分析し得る。

【０２４４】

[00266]計算しやすさのために、基底関数の複数Ｂは、基底関数のインベントリＡよりもかなり小さいことが望ましいことがある。大きいインベントリから開始して、所与の分離タスクのためのインベントリを狭めることが望ましいことがある。基底関数Ｂのセット

の選択は、記録されたシーン中の楽器の視覚的認識に基づいて低減され得る。たとえば、Ｂ個の基底関数は、図２の画像／ビデオ処理ブロック５４によって与えられる楽器候補のリストに対応するもの、または図６のシステム５００によって識別されるものに制限され得る。

【０２４５】

[00267]別の例では、そのような低減はまた、セグメントが打楽器からの音を含むのか調波楽器からの音を含むのかを判断することと、整合のためにインベントリから基底関数の適切な複数Ｂを選択することとによって実行され得る。打楽器は、調波音の場合の水平線とは反対にインパルス様のスペクトログラム（たとえば、垂直線）を有する傾向がある。

【０２４６】

[00268]調波楽器は、一般に、ある基本ピッチおよび関連する音色と、この調波パターンの対応する高周波拡張とによって、スペクトログラムにおいて特徴づけられ得る。したがって、別の例では、スペクトルの高周波レプリカは、低周波スペクトルに基づいて予測され得るので、これらのスペクトルのより低いオクターブのみを分析することによって計算タスクを低減することが望ましいことがある。整合の後に、アクティブな基底関数を高周波に外挿し、混合信号から減算して、残差信号を取得し得、残差信号は、符号化されおよび／またはさらに分解され得る。

【０２４７】

[00269]そのような低減はまた、グラフィカルユーザインターフェースにおけるユーザ選択を通して実行され、ならびに／あるいは、第１のスパース復元ラン（sparse recovery run）または最尤適合に基づく、可能性が最も高い楽器および／またはピッチの事前分類によって実行され得る。たとえば、スパース復元演算の第１のランを実行して、復元されたスパース係数の第１のセットを取得し得、この第１のセットに基づいて、適用可能なノート基底関数がスパース復元演算の別のランのために狭められ得る。

【０２４８】

[00270]１つの低減手法は、いくつかのピッチ間隔においてスパーシティスコアを測定することによって、いくつかの楽器ノートの存在を検出することを含む。そのような手法は、初期ピッチ推定値に基づいて、１つまたは複数の基底関数のスペクトル形状を改善することと、方法Ｍ１００において、改善された基底関数を複数Ｂとして使用することとを含み得る。

【０２４９】

[00271]低減手法は、対応する基底関数に射影された音楽信号のスパーシティスコアを測定することによってピッチを識別するように構成され得る。最良のピッチスコアが与えられれば、基底関数の振幅形状は、楽器ノートを識別するために最適化され得る。アクティブな基底関数の低減されたセットは、次いで、方法Ｍ１００において複数Ｂとして使用され得る。

【０２５０】

[00272]図３０Ａに、基底関数の第１ランインベントリ低減を含む方法Ｍ１００の実装形態Ｍ６００のフローチャートを示す。方法Ｍ６００は、（たとえば、メルまたはバーク尺度の場合のように、隣接する要素間の周波数距離が周波数とともに増加する）非線形周波数領域におけるセグメントの信号表現を計算するタスクＴ６００を含む。一例では、タスクＴ６００は、定Ｑ変換（constant-Q transform）を使用して非線形信号表現を計算するように構成される。方法Ｍ６００はまた、非線形信号表現と、複数の同様に非線形の基底関数とに基づいて、アクティブ化係数の第２のベクトルを計算するタスクＴ７００を含む。第２のアクティブ化係数ベクトルからの（たとえば、アクティブなピッチ範囲を示し得るアクティブ化された基底関数の識別情報からの）情報に基づいて、タスクＴ８００は、タスクＴ２００において使用する基底関数の複数Ｂを選択する。また、方法Ｍ２００、Ｍ３００、およびＭ４００は、そのようなタスクＴ６００、Ｔ７００、およびＴ８００を含むように実装され得ることに明確に留意されたい。

【０２５１】
[00273]図３１に、基底関数のより大きいセットから（たとえば、インベントリから）複数の基底関数を選択するように構成されたインベントリ低減モジュール（ＩＲＭ：inventory reduction module）を含む装置Ａ１００の実装形態Ａ８００のブロック図を示す。モジュールＩＲＭは、（たとえば、定Ｑ変換に従って）非線形周波数領域におけるセグメントの信号表現を計算するように構成された第２の変換モジュール２１１０を含む。モジュールＩＲＭは、本明細書で説明するように、非線形周波数領域における計算された信号表現と、第２の複数の基底関数とに基づいて、アクティブ化係数の第２のベクトルを計算するように構成された第２の係数ベクトル計算器２２１０をも含む。モジュールＩＲＭは、本明細書で説明するように、第２のアクティブ化係数ベクトルからの情報に基づいて、基底関数のインベントリの中から複数の基底関数を選択するように構成された基底関数セレクタをも含む。

【０２５２】
[00274]方法Ｍ１００は、オンセット検出（たとえば、楽音のオンセットを検出すること）と、調波楽器スパース係数を改善するための後処理とを含むことが望ましいことがある。アクティブ化係数ベクトルｆは、楽器固有の基底関数セットB_nのアクティブ化係数を含む、各楽器ｎの対応するサブベクトルf_nを含むと見なされ得、これらのサブベクトルは独立して処理され得る。

【０２５３】
[00275]一般的なオンセット検出方法はスペクトルの大きさ（たとえば、エネルギー差）に基づき得る。たとえば、そのような方法は、スペクトルエネルギーおよび／またはピークスロープに基づいてピークを見つけることを含み得る。

【０２５４】
[00276]また、各個々の楽器のオンセットを検出することが望ましいことがある。たとえば、調波楽器の中のオンセット検出の方法は、時間的な対応する係数差に基づき得る。１つのそのような例では、調波楽器ｎのオンセット検出は、現在のフレームの楽器ｎの係数ベクトル（サブベクトルf_n）の最大大きさの要素のインデックスが、前のフレームの楽器ｎの係数ベクトルの最大大きさの要素のインデックスに等しくない場合にトリガされる。そのような動作は、各楽器について反復され得る。

【０２５５】
[00277]調波楽器のスパース係数ベクトルの後処理を実行することが望ましいことがある。たとえば、調波楽器では、大きい大きさを有し、および／または指定された基準を満たす（たとえば、十分に鋭い）アタックプロファイルを有する、対応するサブベクトルの係数を保持すること、ならびに／あるいは残差係数を除去（たとえば、ゼロアウト）することが望ましいことがある。

【０２５６】
[00278]各調波楽器について、支配的な大きさと許容できるアタック時間とを有する係数が保持され、残差係数がゼロ化されるように、（たとえば、オンセット検出が示されたときに）各オンセットフレームにおいて係数ベクトルを後処理することが望ましいことがある。アタック時間は、時間に対する平均大きさなどの基準に従って評価され得る。１つのそのような例では、係数の現在の平均値が係数の過去の平均値よりも小さい場合（たとえば、フレーム（ｔ－５）からフレーム（ｔ＋４）までなど、現在の窓にわたる係数の値の和が、フレーム（ｔ－１５）からフレーム（ｔ－６）までなど、過去の窓にわたる係数の値の和よりも小さい場合）、現在のフレームｔの楽器の各係数はゼロアウトされる（すなわち、アタック時間は許容できない）。各オンセットフレームにおける調波楽器のための係数ベクトルのそのような後処理は、最大大きさをもつ係数を保持し、他の係数をゼロアウトすることをも含み得る。各非オンセットフレームにおける各調波楽器について、前のフレーム中の値が０でなかった係数のみを保持し、ベクトルの他の係数をゼロアウトするように係数ベクトルを後処理することが望ましいことがある。

【０２５７】

[00279]上述のように、ＥＭアルゴリズムは、初期基底関数行列を発生するため、および／または（たとえば、アクティブ化係数ベクトルに基づいて）基底関数行列を更新するために使用され得る。ＥＭ手法のための更新ルールの例について次に説明する。スペクトログラムV_{ft}が与えられれば、各時間フレームについてスペクトル基底ベクトルＰ（ｆ｜ｚ）と重みベクトルP_t（ｚ）とを推定することが望まれる。これらの分布から行列分解が得られる。

【０２５８】

[00280]ＥＭアルゴリズムは、以下のように適用される。最初に、重みベクトルP_t（ｚ）とスペクトル基底ベクトルＰ（ｆ｜ｚ）とをランダムに初期化する。次いで、収束するまで後続のステップ間を反復する。１）予想（Ｅ）ステップ　－　スペクトル基底ベクトルＰ（ｆ｜ｚ）と重みベクトルP_t（ｚ）とが与えられれば、後の分布P_t（ｚ｜ｆ）を推定する。この推定は、以下のように表され得る。

【数２５】

$$P_t(z|f) = \frac{P_t(f|z)P(z)}{\sum_z P_t(f|z)P(z)}.$$

【０２５９】

[00281]２）最大化（Ｍ）ステップ　－　後の分布P_t（Ｚ｜ｆ）が与えられれば、重みベクトルP_t（ｚ）とスペクトル基底ベクトルＰ（ｆ｜ｚ）とを推定する。重みベクトルの推定は、以下のように表され得る。

【数２６】

$$P_t(z) = \frac{\sum_f V_{ft} P_t(z|f)}{\sum_z \sum_f V_{ft} P_t(z|f)}.$$

【０２６０】

スペクトル基底ベクトルの推定は、以下のように表され得る。

【数２７】

$$P(f|z) = \frac{\sum_t V_{ft} P_t(z|f)}{\sum_t \sum_f V_{ft} P_t(z|f)}.$$

【０２６１】

[00282]本明細書で開示するシステムおよび方法は、コンピュータ、ゲーミングコンソール、またはセルラーフォン、携帯情報端末（ＰＤＡ）、スマートフォンなどのハンドヘルドデバイスなどを含む、任意の好適な視聴覚システム中に含まれ得る。本明細書で説明した構成要素の主な機能は、概してデジタル処理領域において実装される。しかしながら、これらの構成要素は、代替的に、好適なアナログ構成要素を使用するアナログ領域において実装されるか、またはアナログ電子構成要素とデジタル電子構成要素との任意の好適な組合せにおいて実装され得る。

【０２６２】

[00283]音響信号を受信するように構成された２つ以上のマイクロフォンのアレイと、１つまたは複数のカメラとを有するポータブル視聴覚感知デバイス内で、本明細書で説明したシステムおよび方法を実装することが望ましいことがある。そのようなアレイを含むように実装され得、オーディオ記録および／または音声通信適用例のために使用され得るポータブルオーディオ感知デバイスの例としては、電話ハンドセット（たとえば、セルラー電話ハンドセット）、ハンドヘルドオーディオおよび／またはビデオレコーダ、携帯情報端末（ＰＤＡ）または他のハンドヘルドコンピューティングデバイス、およびノートブ

ックコンピュータ、ラップトップコンピュータ、ネットブックコンピュータ、タブレットコンピュータ、あるいは他のポータブルコンピューティングデバイスがある。ポータブルコンピューティングデバイスの種類は現在、ラップトップコンピュータ、ノートブックコンピュータ、ネットブックコンピュータ、ウルトラポータブルコンピュータ、タブレットコンピュータ、モバイルインターネットデバイス、スマートブック、およびスマートフォンなどの名称を有するデバイスを含む。そのようなデバイスは、ディスプレイスクリーンを含む上部パネルと、キーボードを含み得る下部パネルとを有し得、それらの２つのパネルは、クラムシェルまたは他のヒンジ結合関係で接続され得る。そのようなデバイスは、上面上にタッチスクリーンディスプレイを含むタブレットコンピュータとして同様に実装され得る。そのような方法を実行するように構築され得、オーディオ記録および／または音声通信適用例のために使用され得るオーディオ感知デバイスの他の例としては、テレビジョンディスプレイ、セットトップボックス、ならびにオーディオおよび／またはビデオ会議デバイスがある。

【０２６３】

[00284]本明細書で開示するシステムおよび方法は、リアルタイムで視聴覚情報を処理するように、ならびに以前に記録された視聴覚情報を処理するように実装され得る。

【０２６４】

[00285]本明細書で説明したシステム、装置、デバイスおよびそれらのそれぞれの構成要素の機能、ならびに方法ステップおよびモジュールは、ハードウェアで実装されるか、ハードウェアによって実行されるソフトウェア／ファームウェアで実装されるか、またはそれらの任意の好適な組合せで実装され得る。ソフトウェア／ファームウェアは、マイクロプロセッサ、ＤＳＰ、埋込みコントローラまたは知的財産（ＩＰ：intellectual property）コアなど、１つまたは複数のデジタル回路によって実行可能な命令のセット（たとえば、プログラミングコードセグメント）を有するプログラムであり得る。ソフトウェア／ファームウェアで実装される場合、機能は、命令またはコードとして１つまたは複数のコンピュータ可読媒体上に記憶され得る。コンピュータ可読媒体はコンピュータ記憶媒体を含み得る。記憶媒体は、コンピュータによってアクセスされ得る任意の利用可能な媒体であり得る。限定ではなく例として、そのようなコンピュータ可読媒体は、ＲＡＭ、ＲＯＭ、ＥＥＰＲＯＭ（登録商標）、ＣＤ－ＲＯＭまたは他の光ディスクストレージ、磁気ディスクストレージまたは他の磁気ストレージデバイス、あるいは命令またはデータ構造の形態の所望のプログラムコードを搬送または記憶するために使用され得、コンピュータによってアクセスされ得る、任意の他の媒体を備えることができる。本明細書で使用するディスク（disk）およびディスク（disc）は、コンパクトディスク（disc）（ＣＤ）、レーザディスク（登録商標）（disc）、光ディスク（disc）、デジタル多用途ディスク（disc）（ＤＶＤ）、フロッピー（登録商標）ディスク（disk）およびｂｌｕ－ｒａｙ（登録商標）ディスク（disc）を含み、ディスク（disk）は、通常、データを磁気的に再生し、ディスク（disc）は、データをレーザで光学的に再生する。上記の組合せもコンピュータ可読媒体の範囲内に含まれるべきである。

【０２６５】

[00286]視聴覚シーン分析システムおよび方法のいくつかの例が開示された。これらのシステムおよび方法は例であり、可能な結合は本明細書で説明したものに限定されない。その上、これらの例に対する様々な変更が可能であり、本明細書で提示した原理は他のシステムにも同様に適用され得る。たとえば、本明細書で開示する原理は、パーソナルコンピュータ、エンターテインメントカウンセル、ビデオゲームなどのデバイスに適用され得る。さらに、様々な構成要素および／または方法ステップ／ブロックは、特許請求の範囲から逸脱することなく、明確に開示したもの以外の構成で実装され得る。

【０２６６】

[00287]したがって、これらの教示に鑑みて、他の実施形態および変更形態は当業者に容易に行われる。したがって、以下の特許請求の範囲は、上記の明細書および添付の図面とともに閲覧されたとき、すべてのそのような実施形態および変更形態を包含するもので

ある。

【図１】

図１

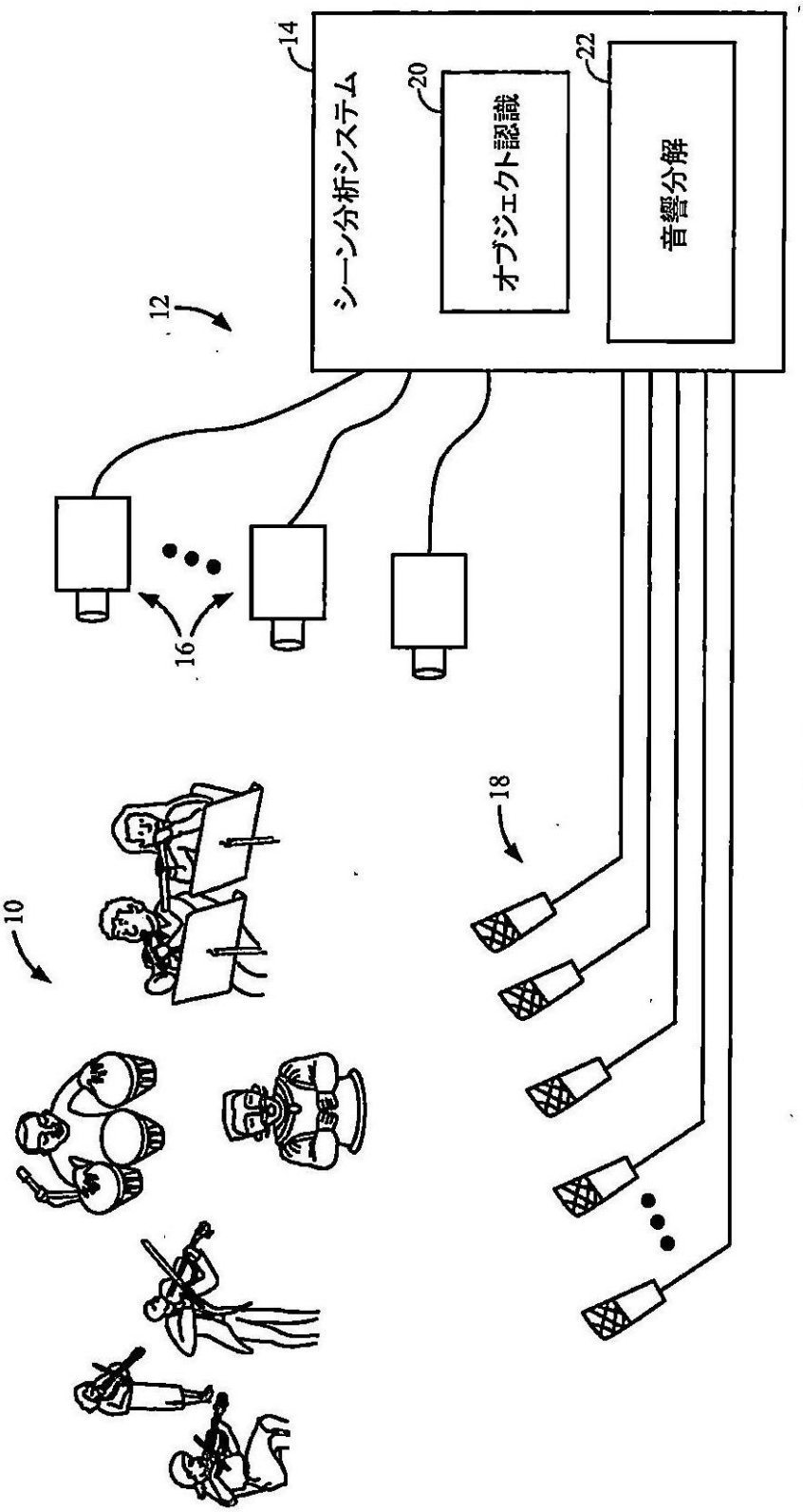


FIG. 1

【図２】

図 2

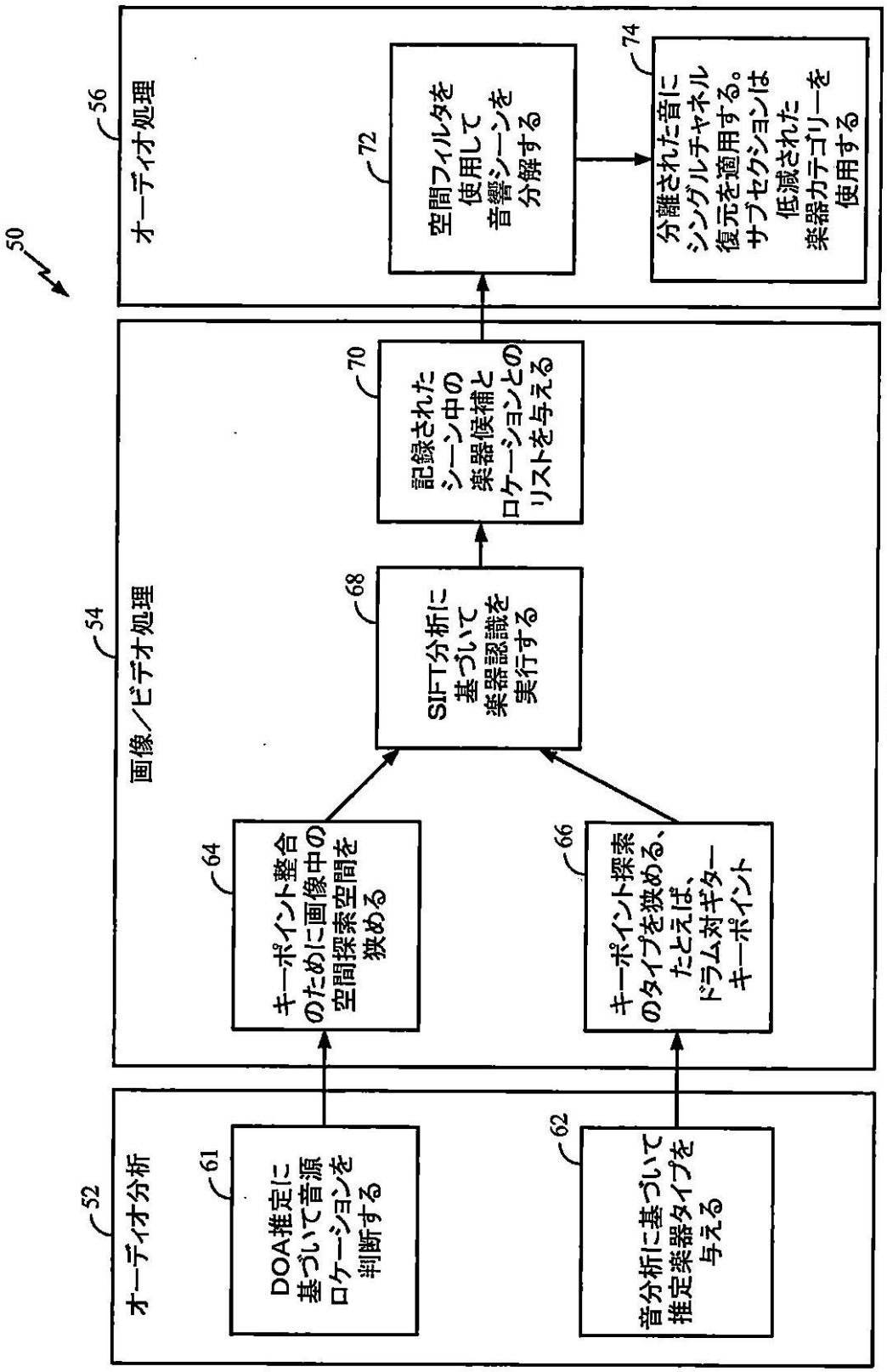

50
52
オーディオ分析
61
DOA推定に基づいて音源ロケーションを判断する
62
音分析に基づいて推定楽器タイプを与える
54
画像／ビデオ処理
64
キーポイント整合のために画像中の空間探索空間を狭める
66
キーポイント探索のタイプを狭める、たとえば、ドラム対ギターキーポイント
68
SIFT分析に基づいて楽器認識を実行する
70
記録されたシーン中の楽器候補とロケーションとのリストを与える
56
オーディオ処理
72
空間フィルタを使用して音響シーンを分解する
74
分離された音にシングルチャネル復元を適用する。サブセクションは低減された楽器カテゴリーを使用する

FIG. 2

【図３】

図３

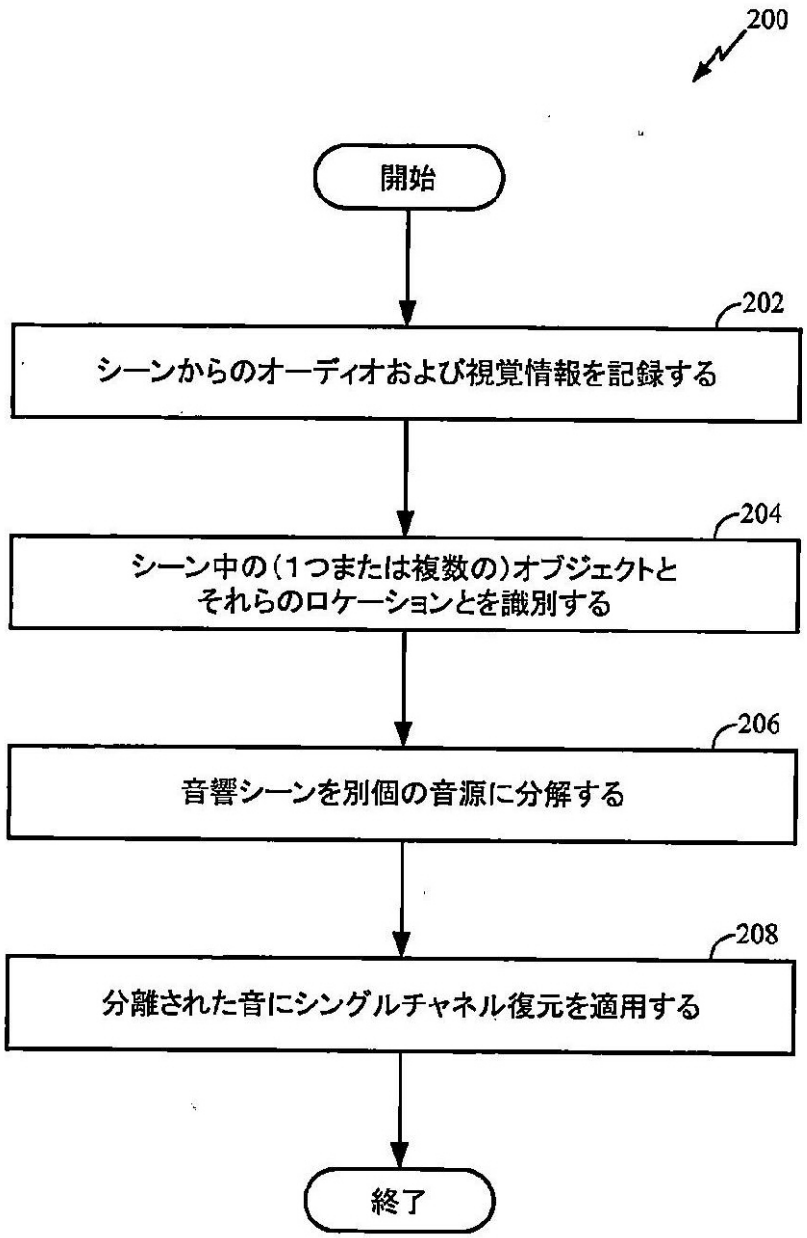

200
開始
202
シーンからのオーディオおよび視覚情報を記録する
204
シーン中の（１つまたは複数の）オブジェクトと
それらのロケーションとを識別する
206
音響シーンを別個の音源に分解する
208
分離された音にシングルチャネル復元を適用する
終了

FIG. 3

【図４】

図４

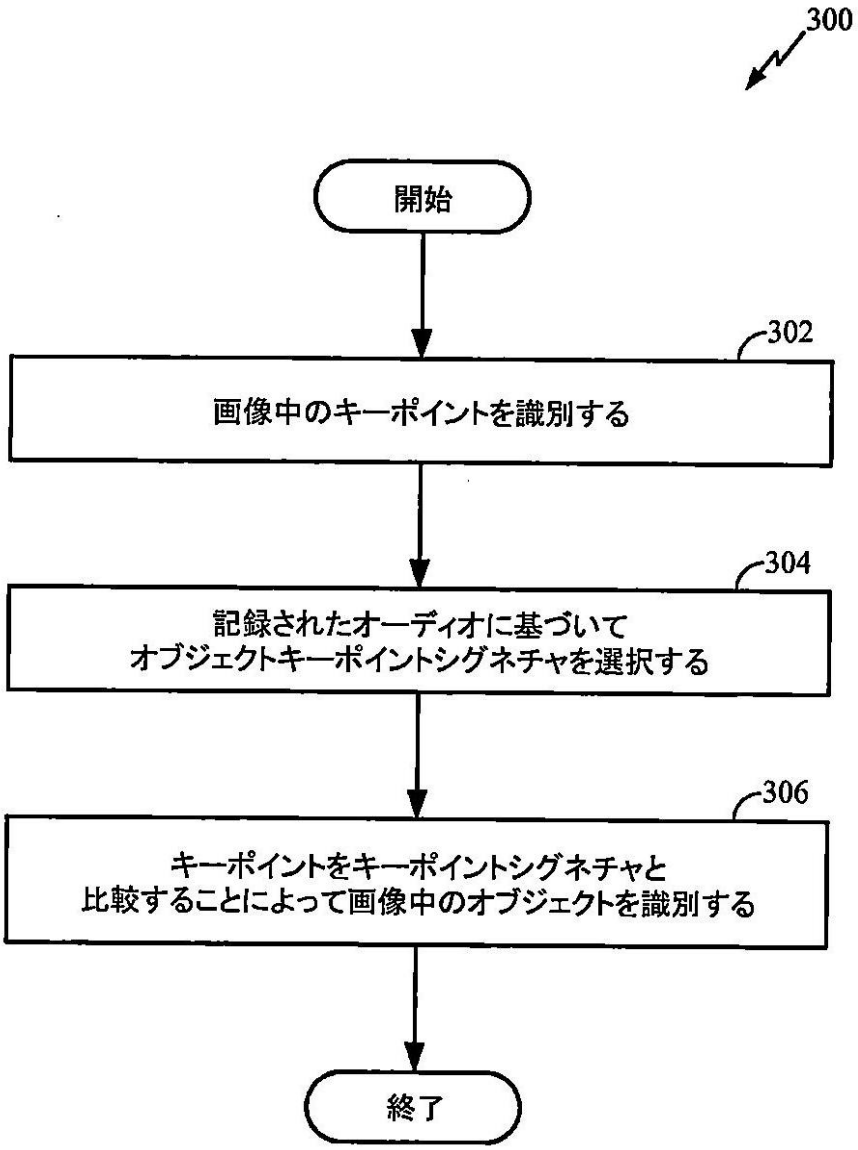

300
開始
302
画像中のキーポイントを識別する
304
記録されたオーディオに基づいて
オブジェクトキーポイントシグネチャを選択する
306
キーポイントをキーポイントシグネチャと
比較することによって画像中のオブジェクトを識別する
終了

FIG. 4

【図５Ａ】

図 5A

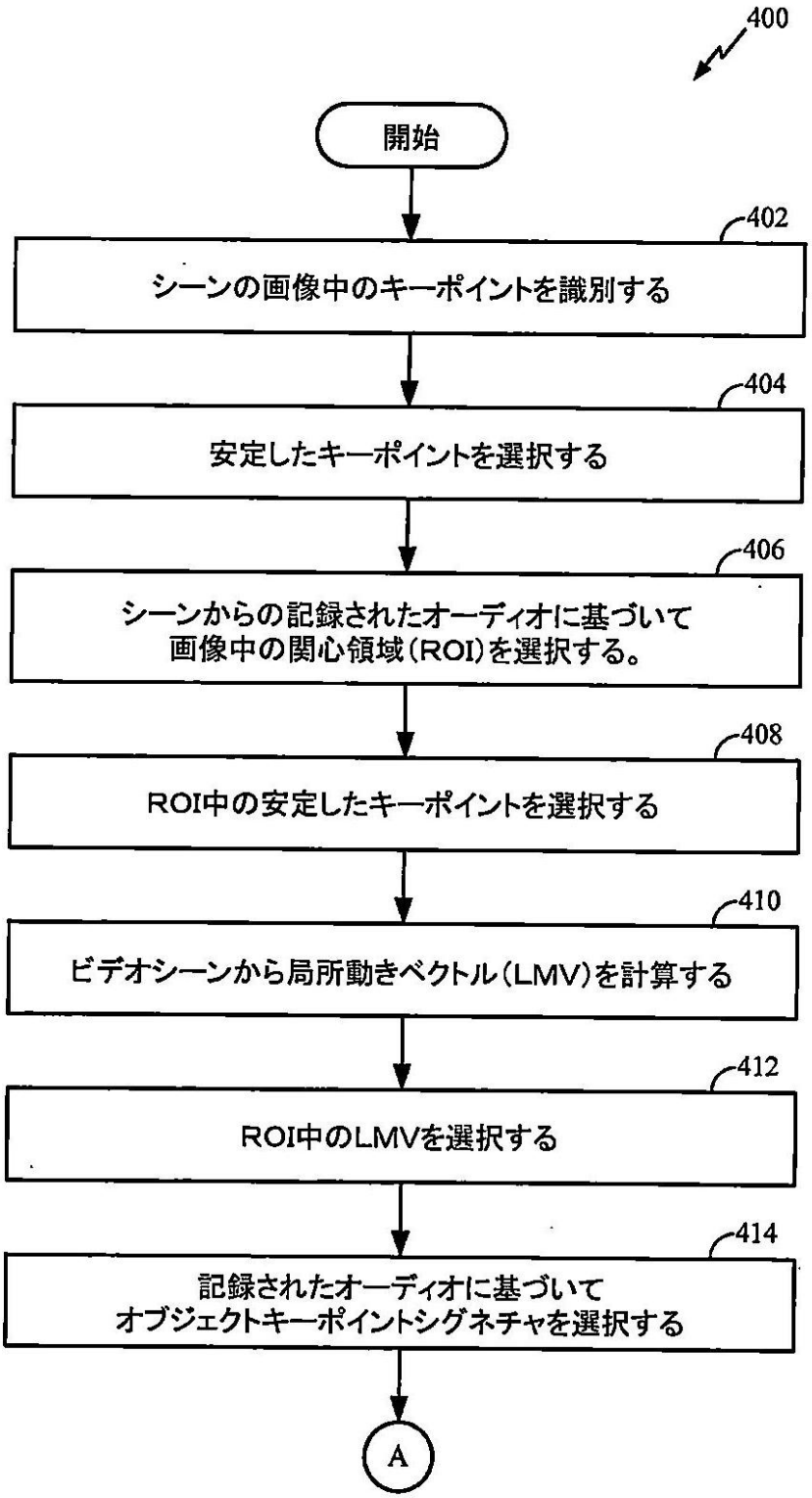


FIG. 5A

【図５Ｂ】

図 5B

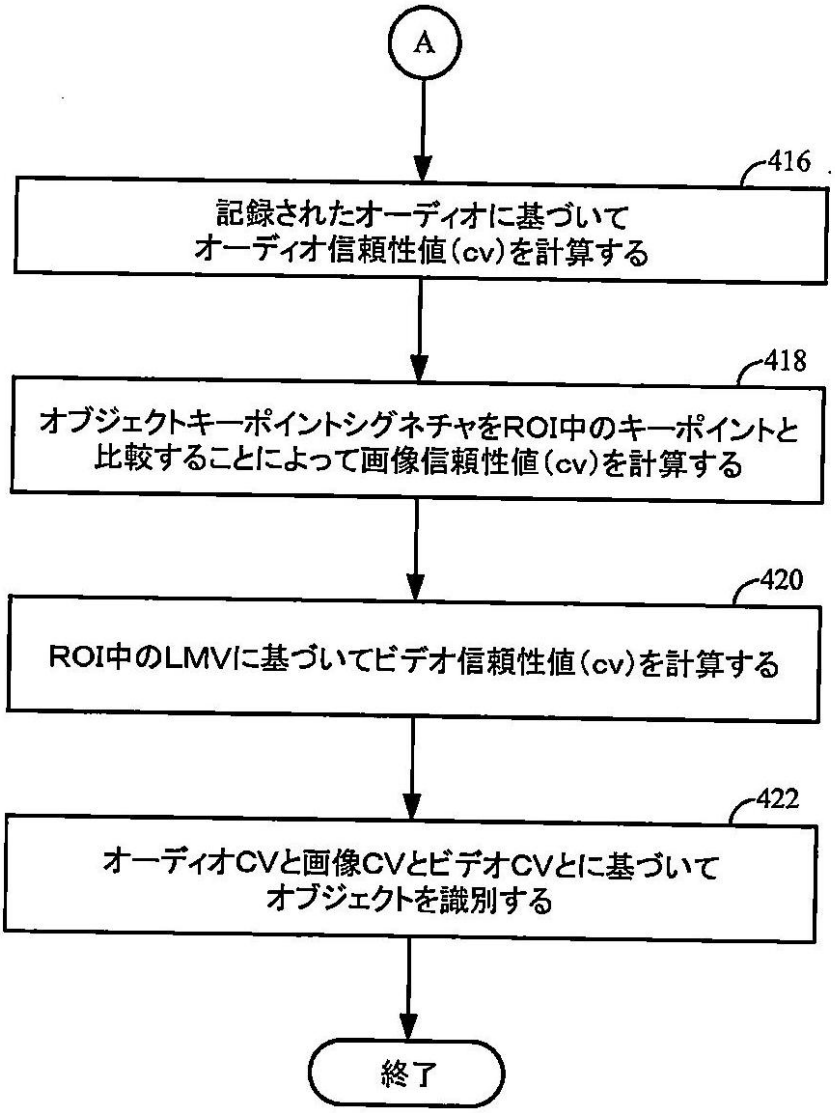


FIG. 5B

【図６】

図 6

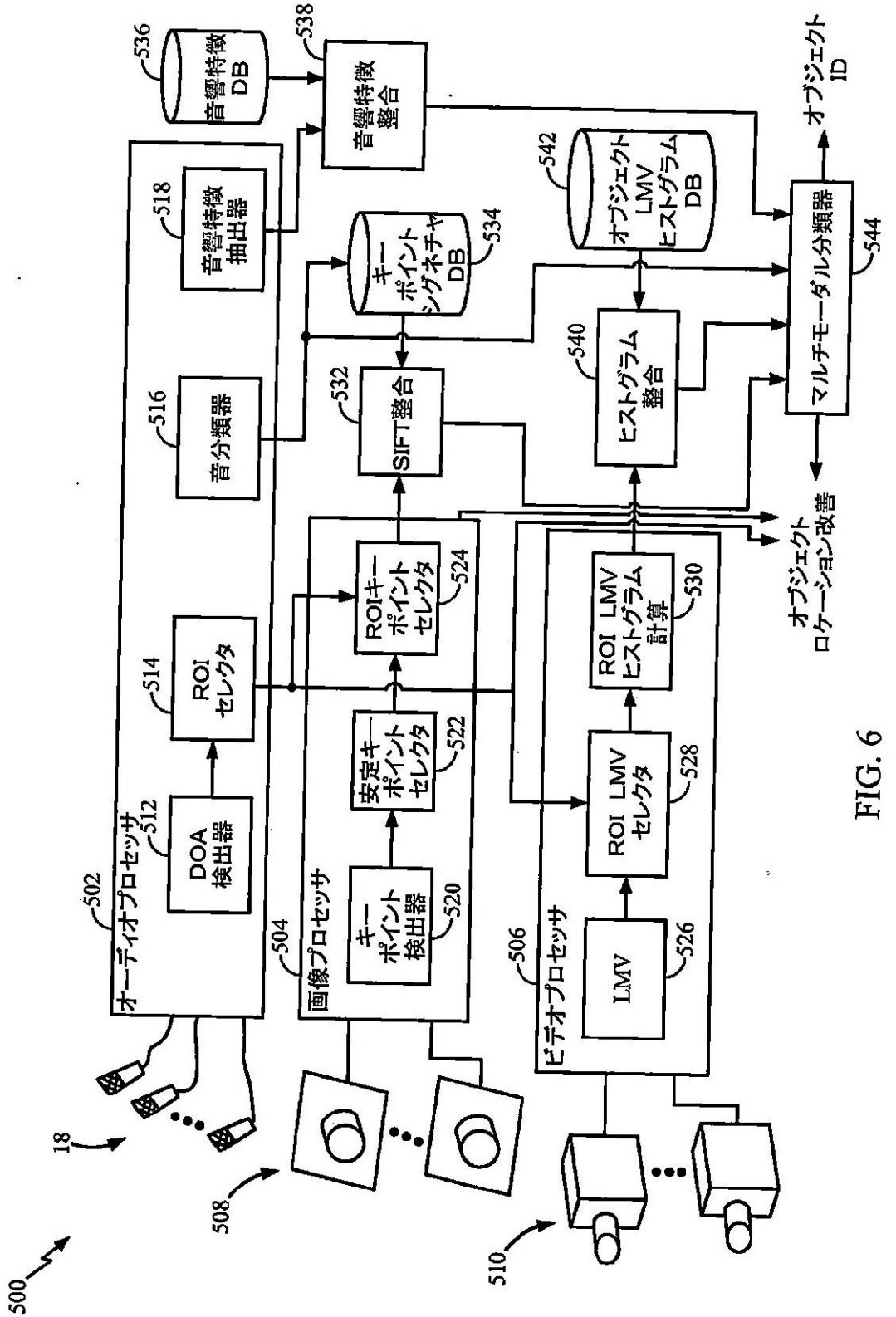

500
18
508
510
オーディオプロセッサ
502
DOA
検出器
512
ROI
セレクタ
514
音分類器
516
音響特徴
抽出器
518
画像プロセッサ
504
キー
ポイント
検出器
520
安定キー
ポイント
セレクタ
522
ROIキー
ポイント
セレクタ
524
SIFT整合
532
音響特徴
DB
536
音響特徴
整合
538
キー
ポイント
シグネチャ
DB
534
ビデオプロセッサ
506
LMV
526
ROI LMV
セレクタ
528
ROI LMV
ヒストグラム
計算
530
オブジェクト
LMV
ヒストグラム
DB
542
ヒストグラム
整合
540
マルチモーダル分類器
544
オブジェクト
ID
オブジェクト
ロケーション改善

FIG. 6

【図７】

図７

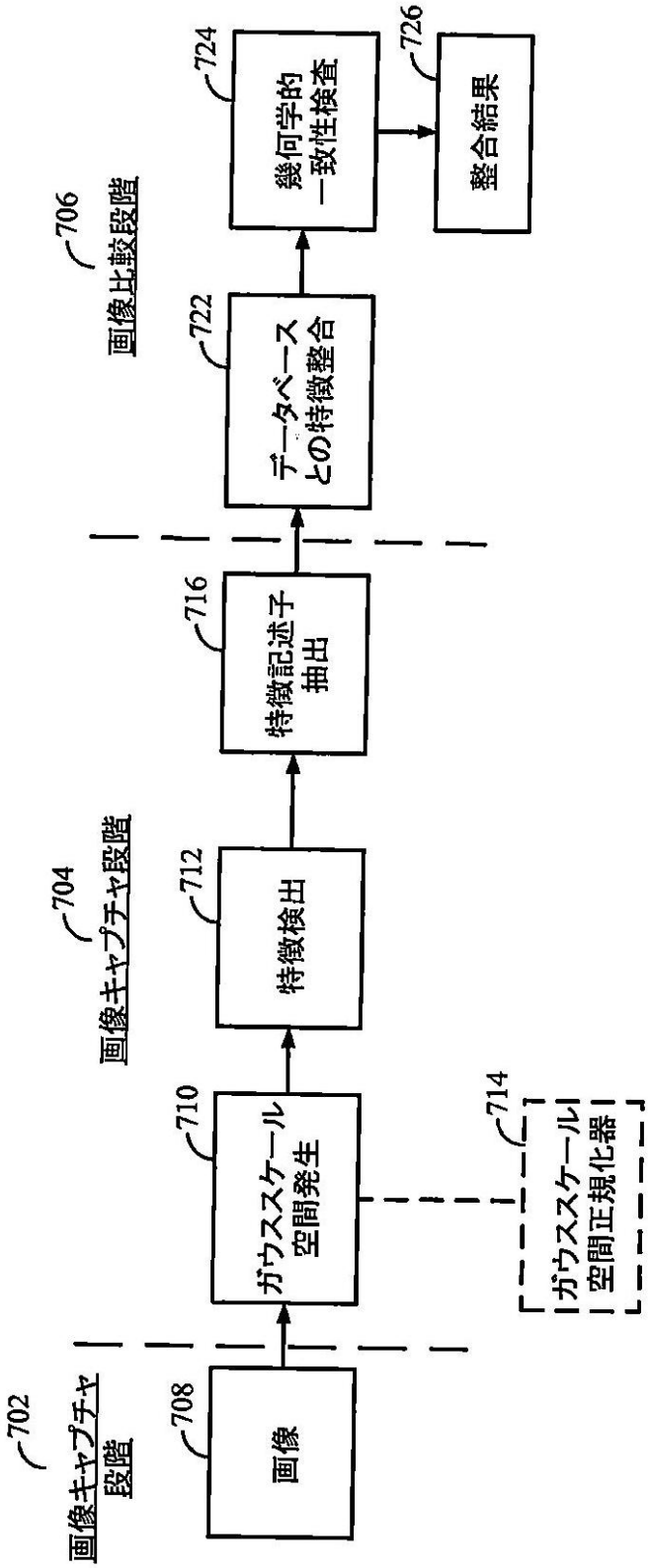

702
画像キャプチャ段階
708
画像
710
ガウススケール空間発生
712
特徴検出
704
画像キャプチャ段階
716
特徴記述子抽出
722
データベースとの特徴整合
706
画像比較段階
724
幾何学的一致性検査
726
整合結果
714
ガウススケール空間正規化器

FIG. 7

【図８】

図 8

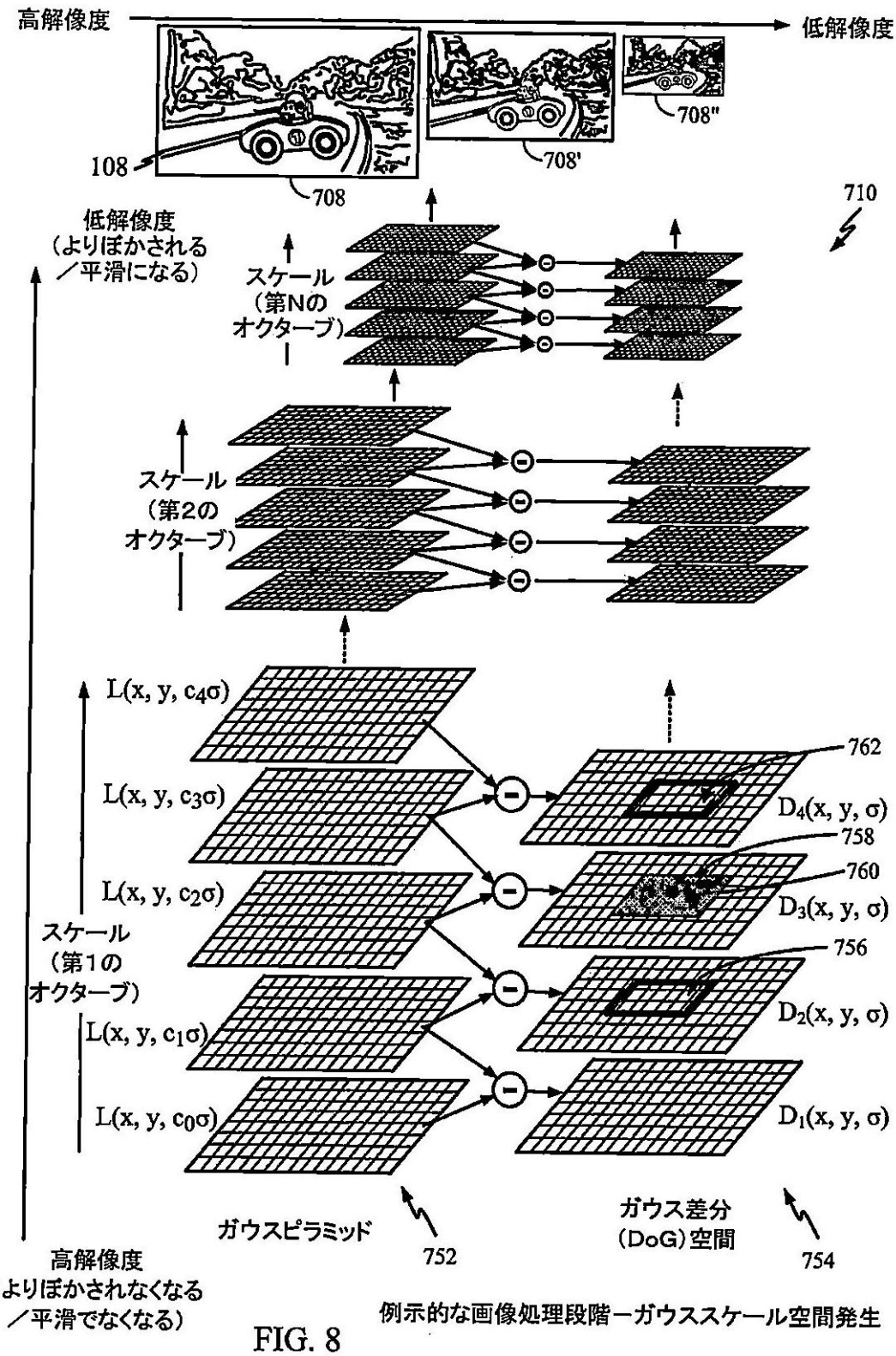


FIG. 8

例示的な画像処理段階－ガウススケール空間発生

【図９】

図 9

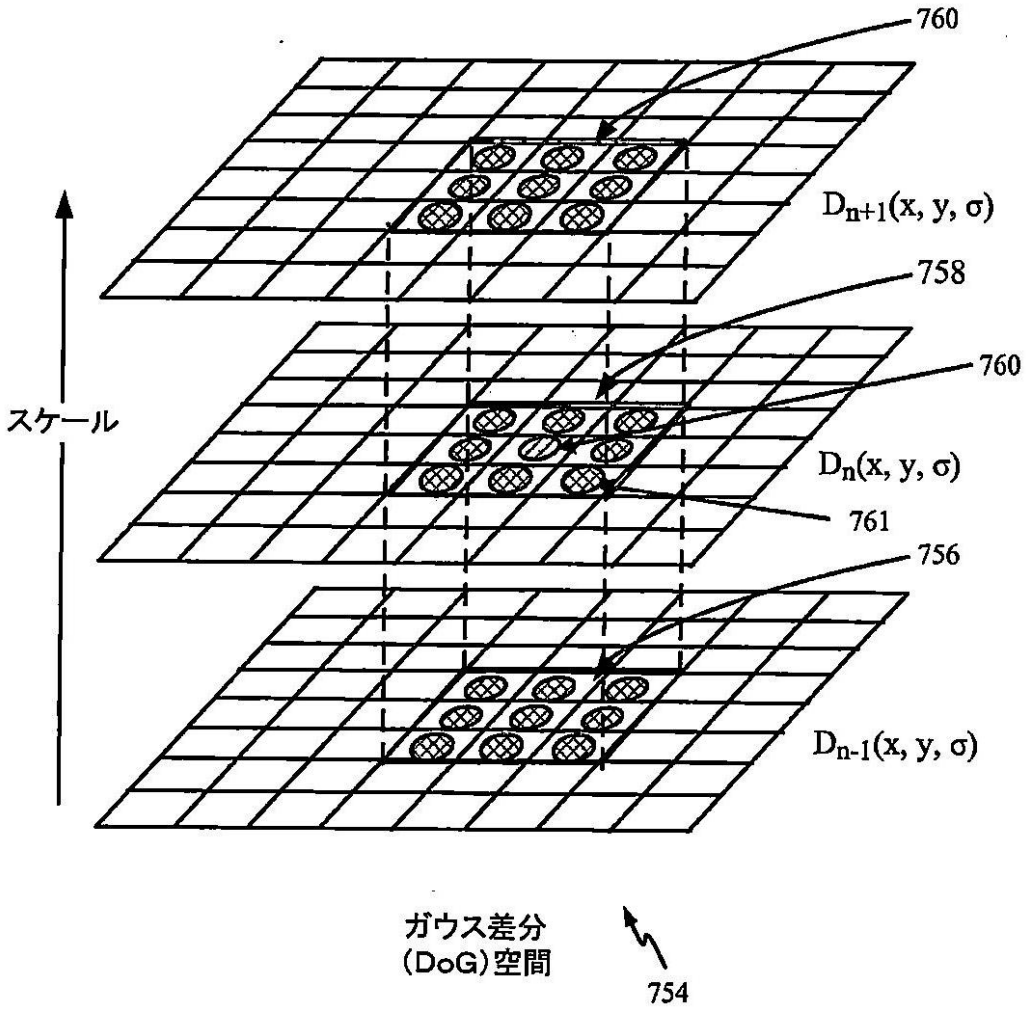

760
$D_{n+1}(x, y, \sigma)$
758
760
スケール
$D_n(x, y, \sigma)$
761
756
$D_{n-1}(x, y, \sigma)$
ガウス差分
(DoG)空間
754

FIG. 9

【図１０】

図 10

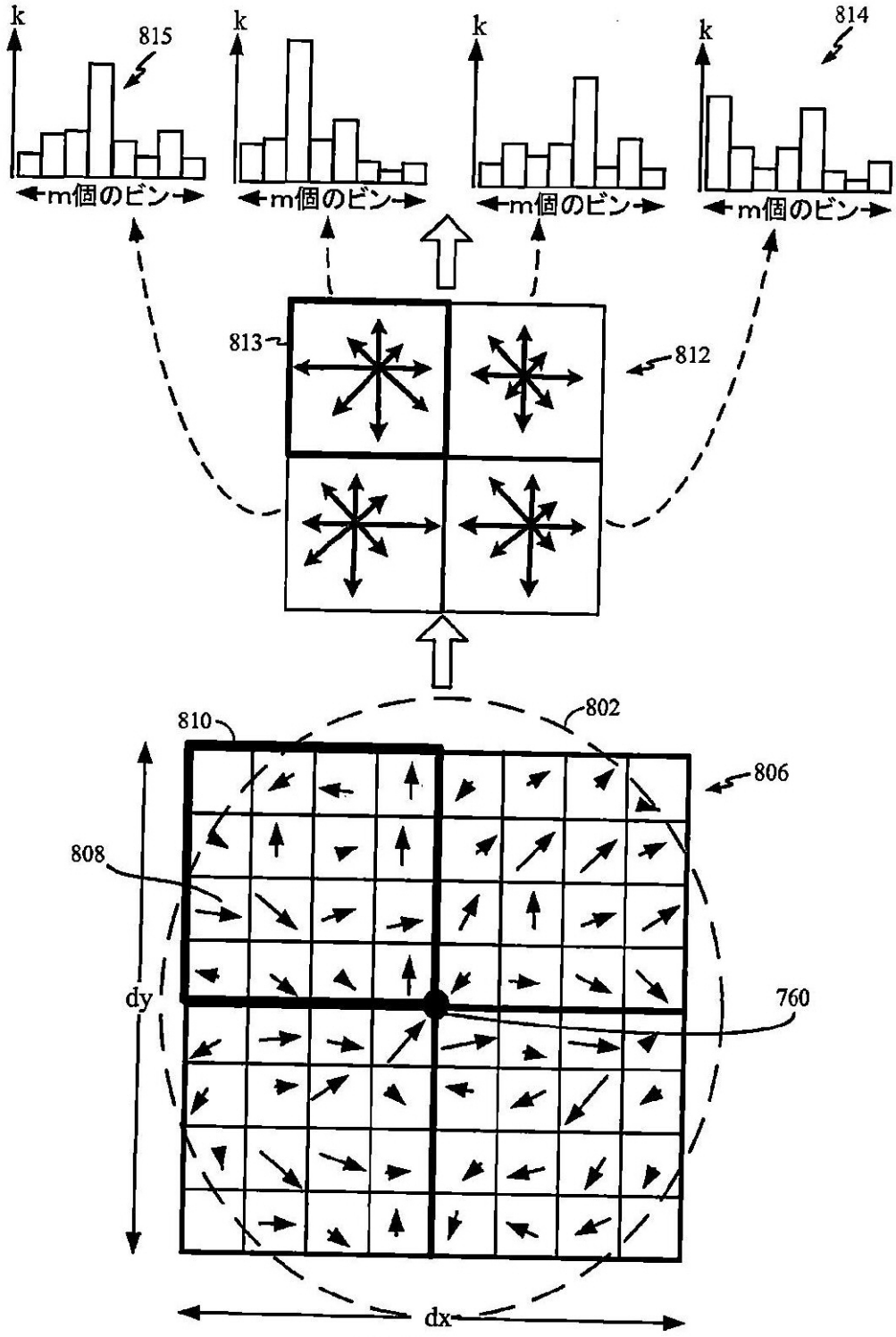


FIG. 10

【図１１】

図 11

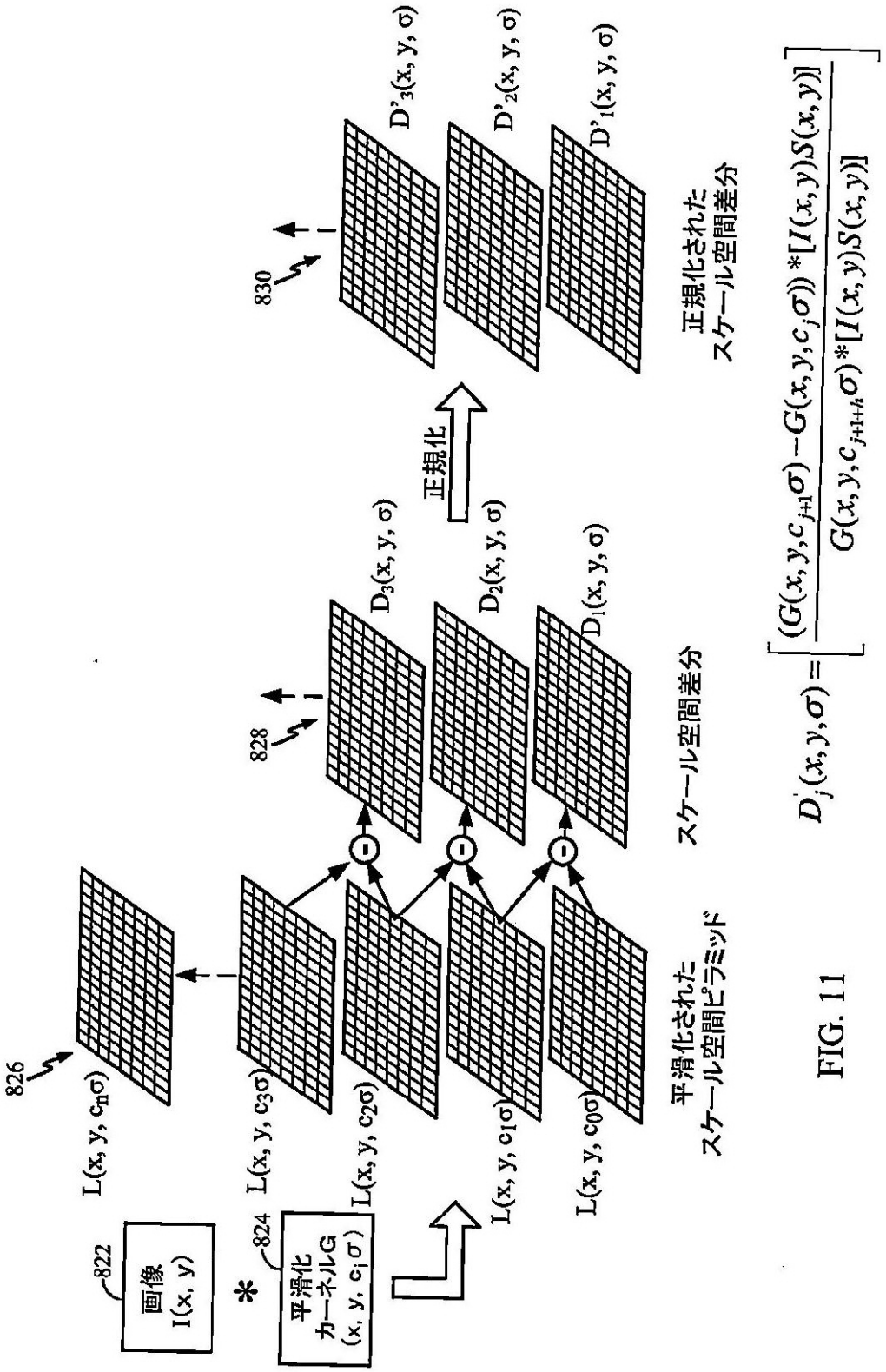


FIG. 11

【図１２】

図 12

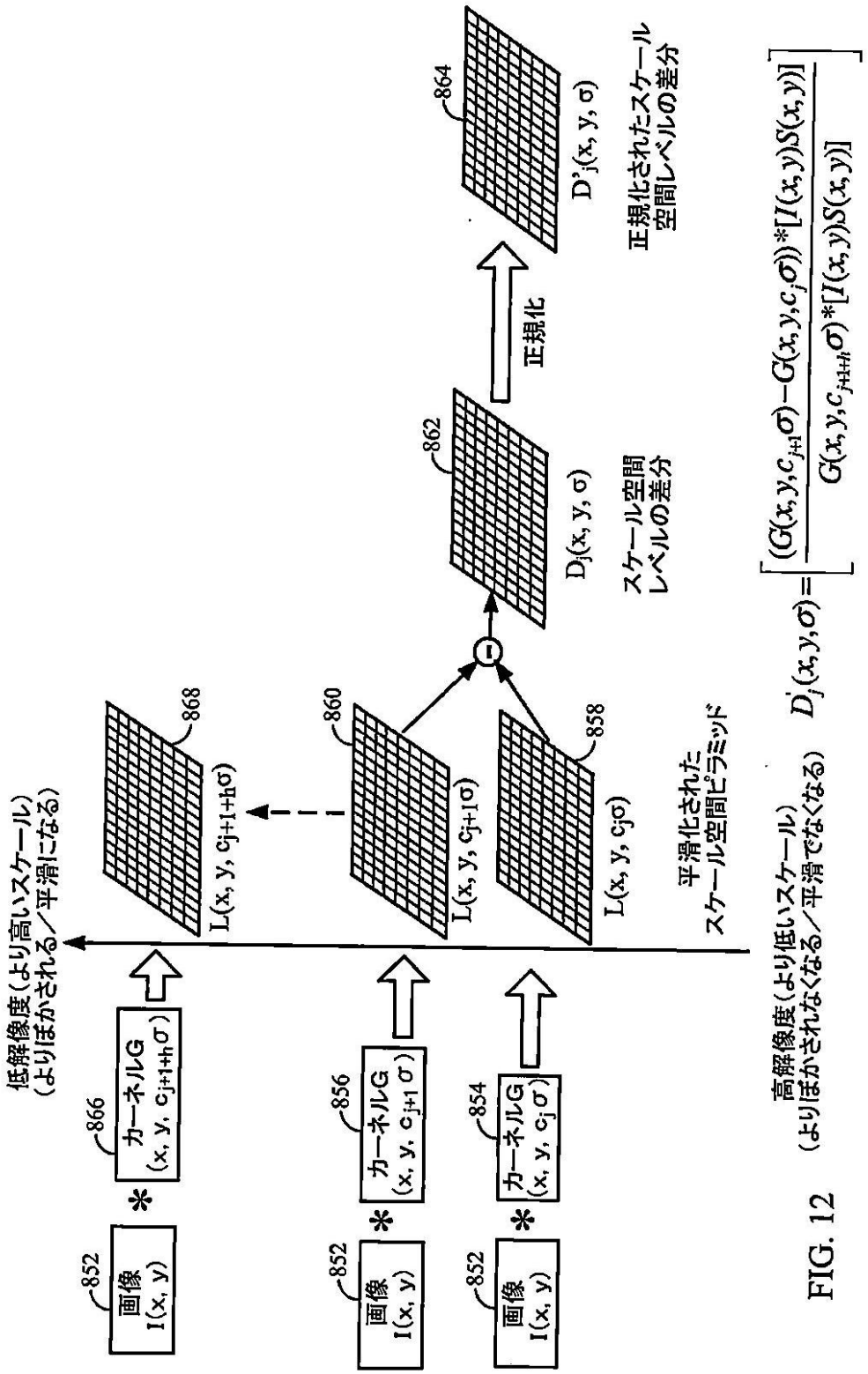


FIG. 12

【図１３】

図 13

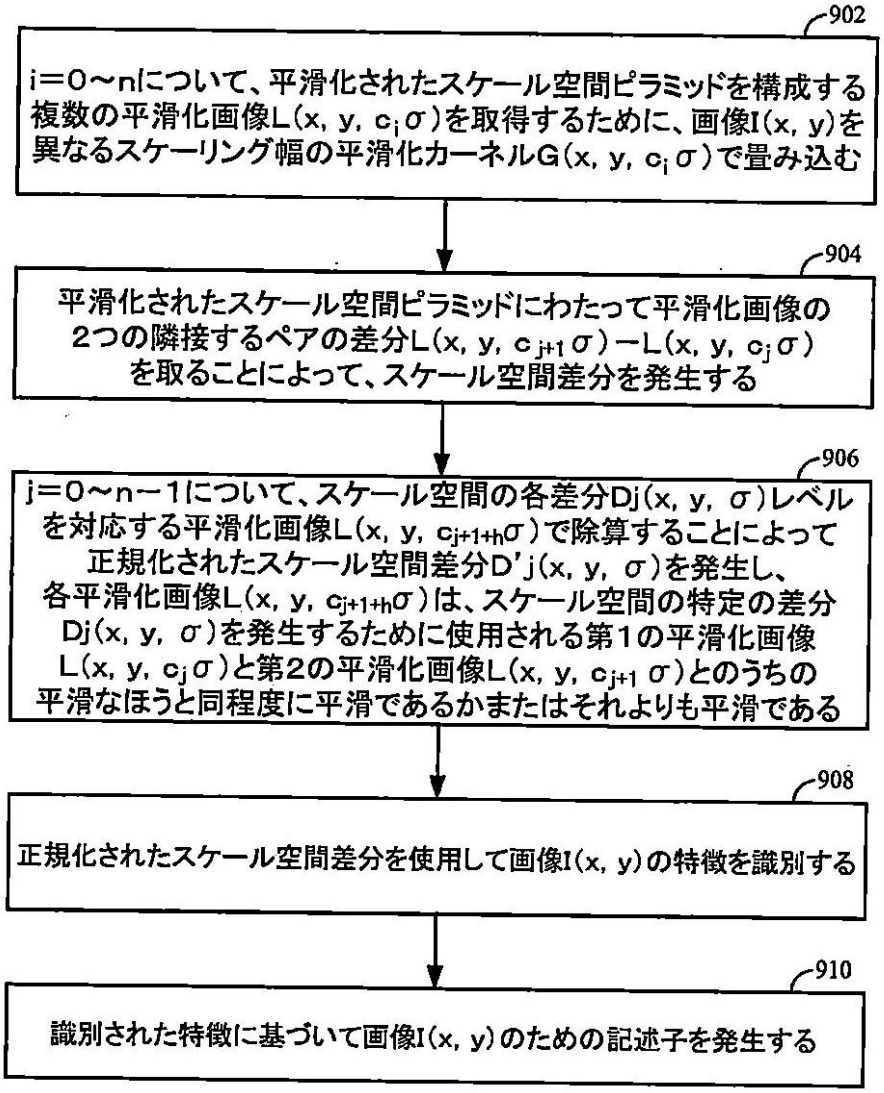


FIG. 13

【図１４】

図 14

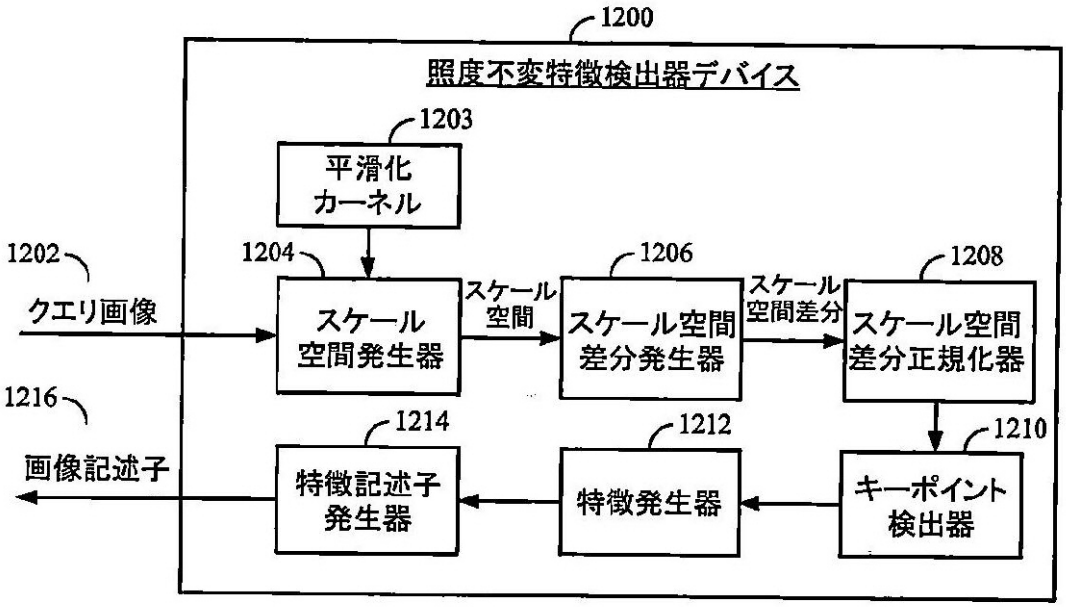


FIG. 14

【図１５】

図 15

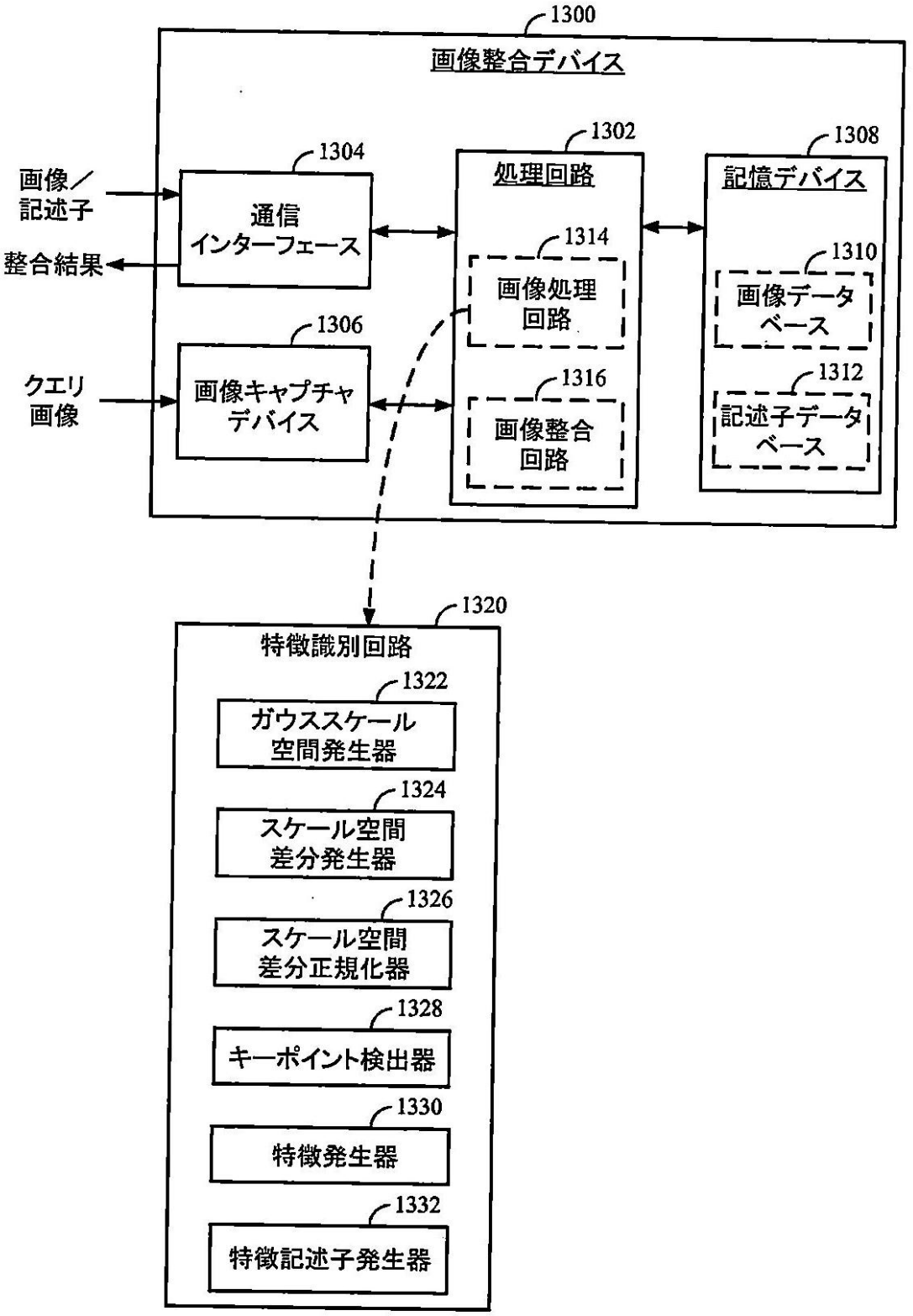


FIG. 15

【図１６Ａ】

図 16A

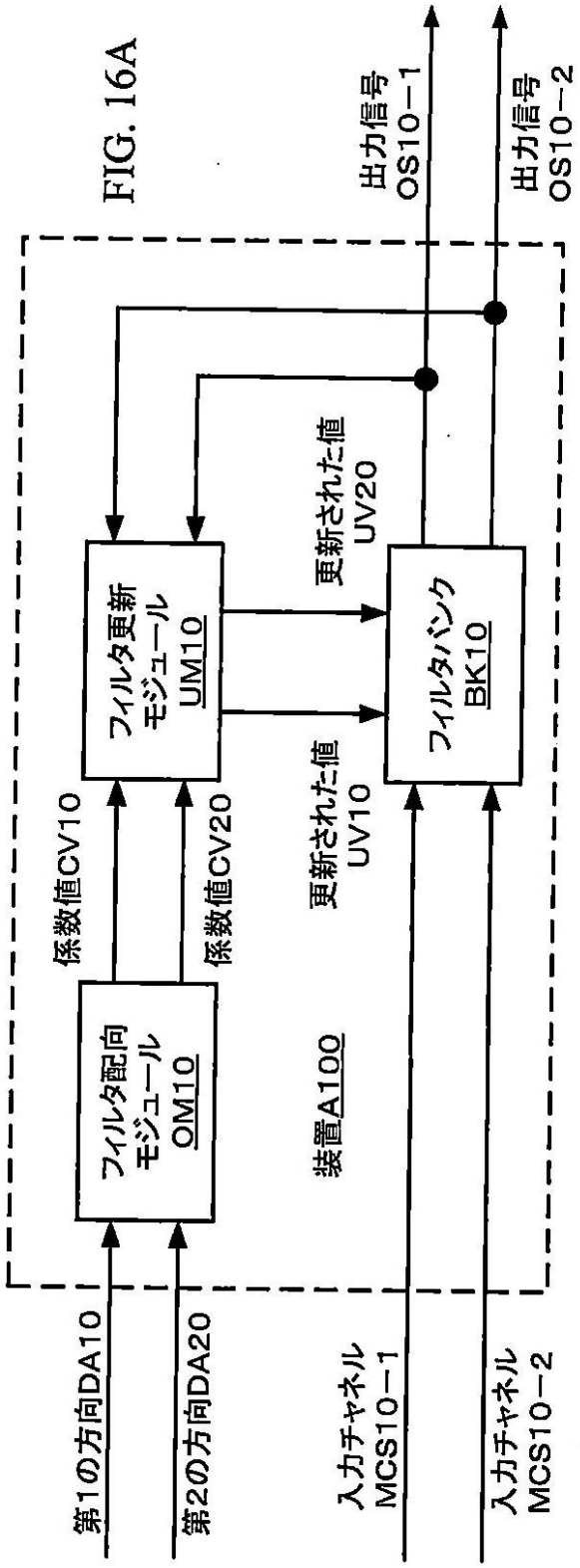


FIG. 16A
出力信号
OS10−1
出力信号
OS10−2
フィルタ更新
モジュール
UM10
更新された値
UV20
フィルタバンク
BK10
係数値CV10
係数値CV20
更新された値
UV10
フィルタ配向
モジュール
OM10
装置A100
第1の方向DA10
第2の方向DA20
入力チャネル
MCS10−1
入力チャネル
MCS10−2

【図１６Ｂ】

図 16B

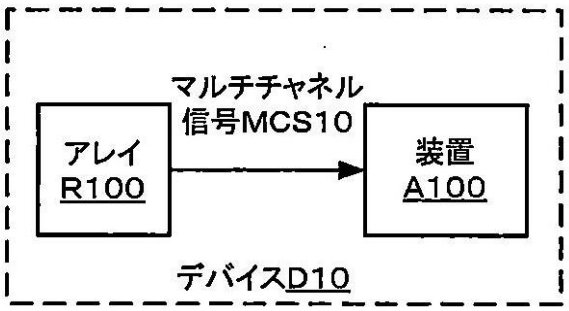
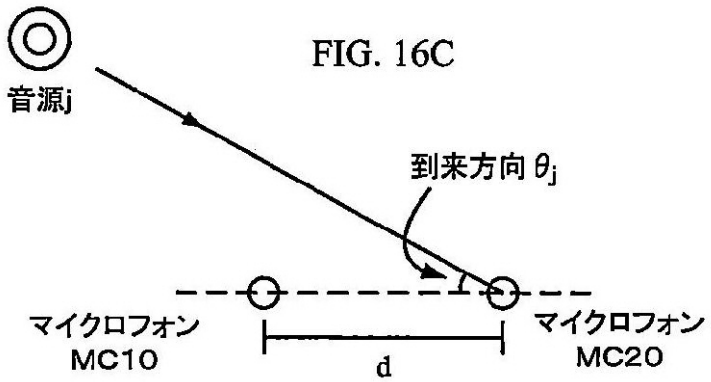


FIG. 16B

【図１６Ｃ】

図 16C

FIG. 16C



【図１７】

図 17

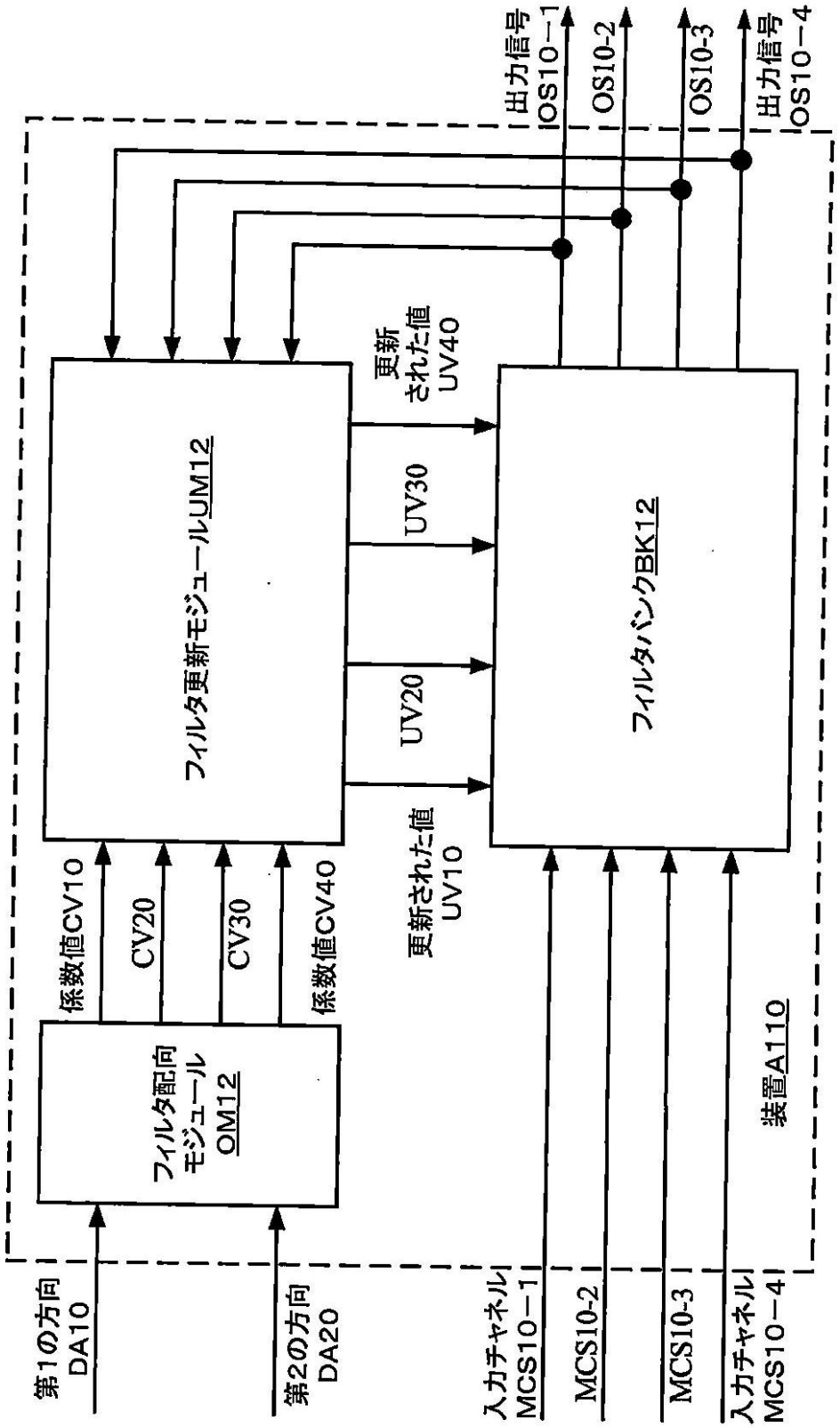
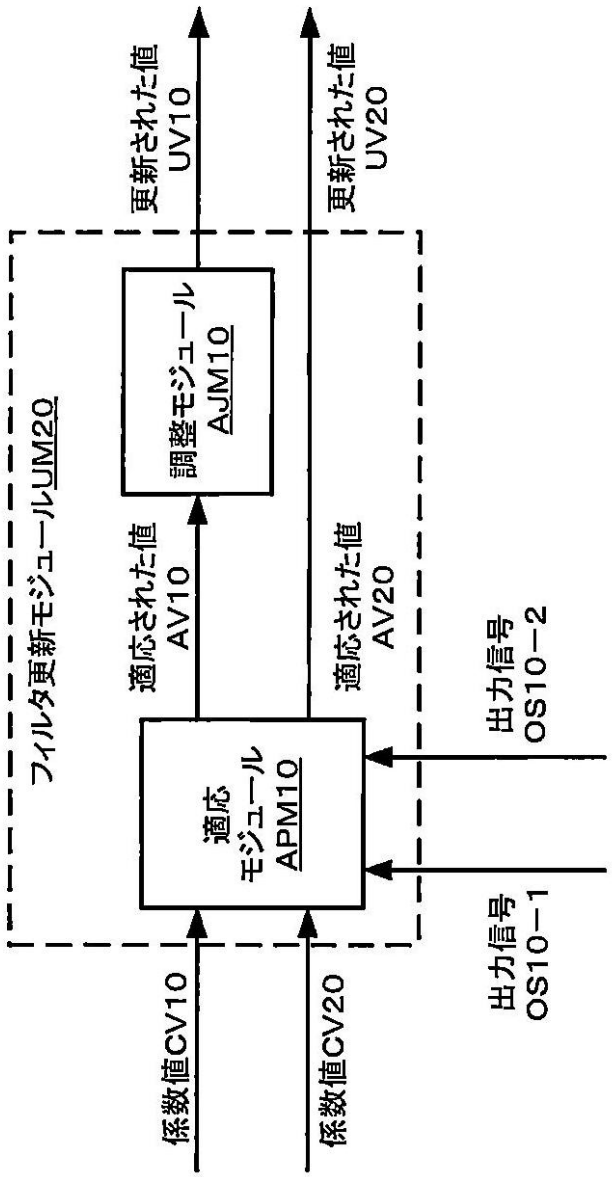


FIG. 17

【図１８Ａ】

図 18A

FIG. 18A



【図１８Ｂ】

図 18B

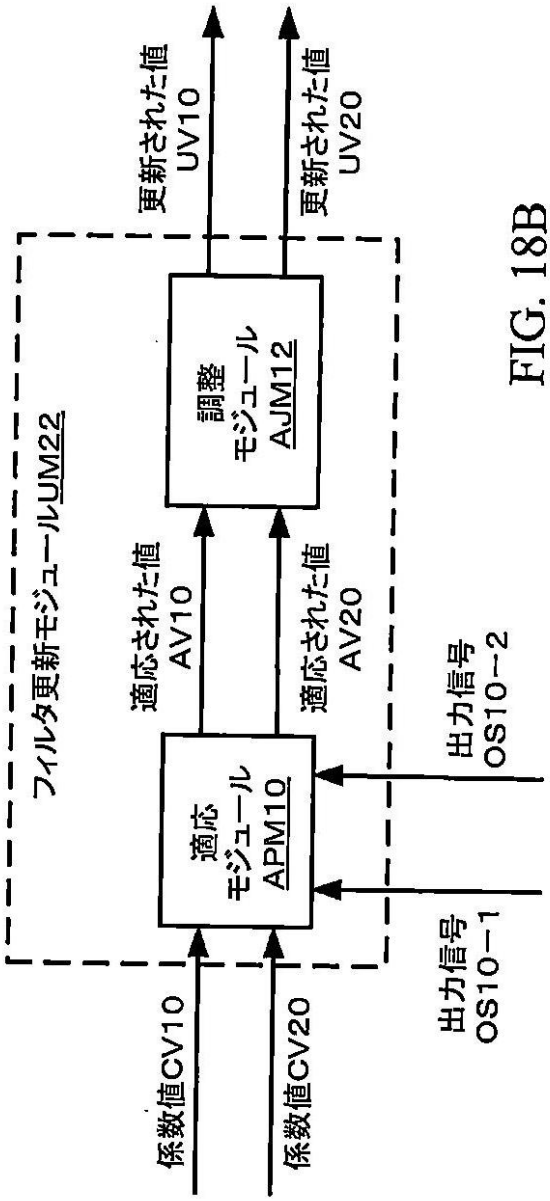


FIG. 18B

【図１９Ａ】

図 19A

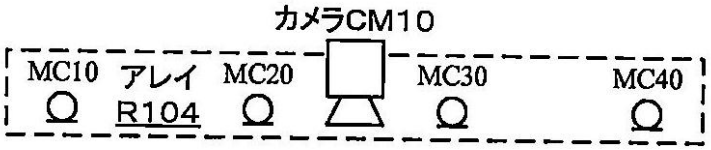


FIG. 19A

【図１９Ｂ】

図 19B

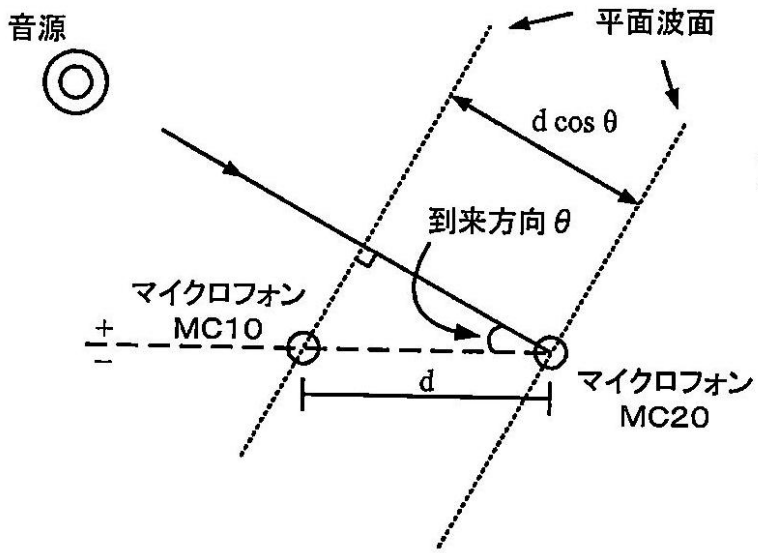


FIG. 19B

【図２０】

図 20

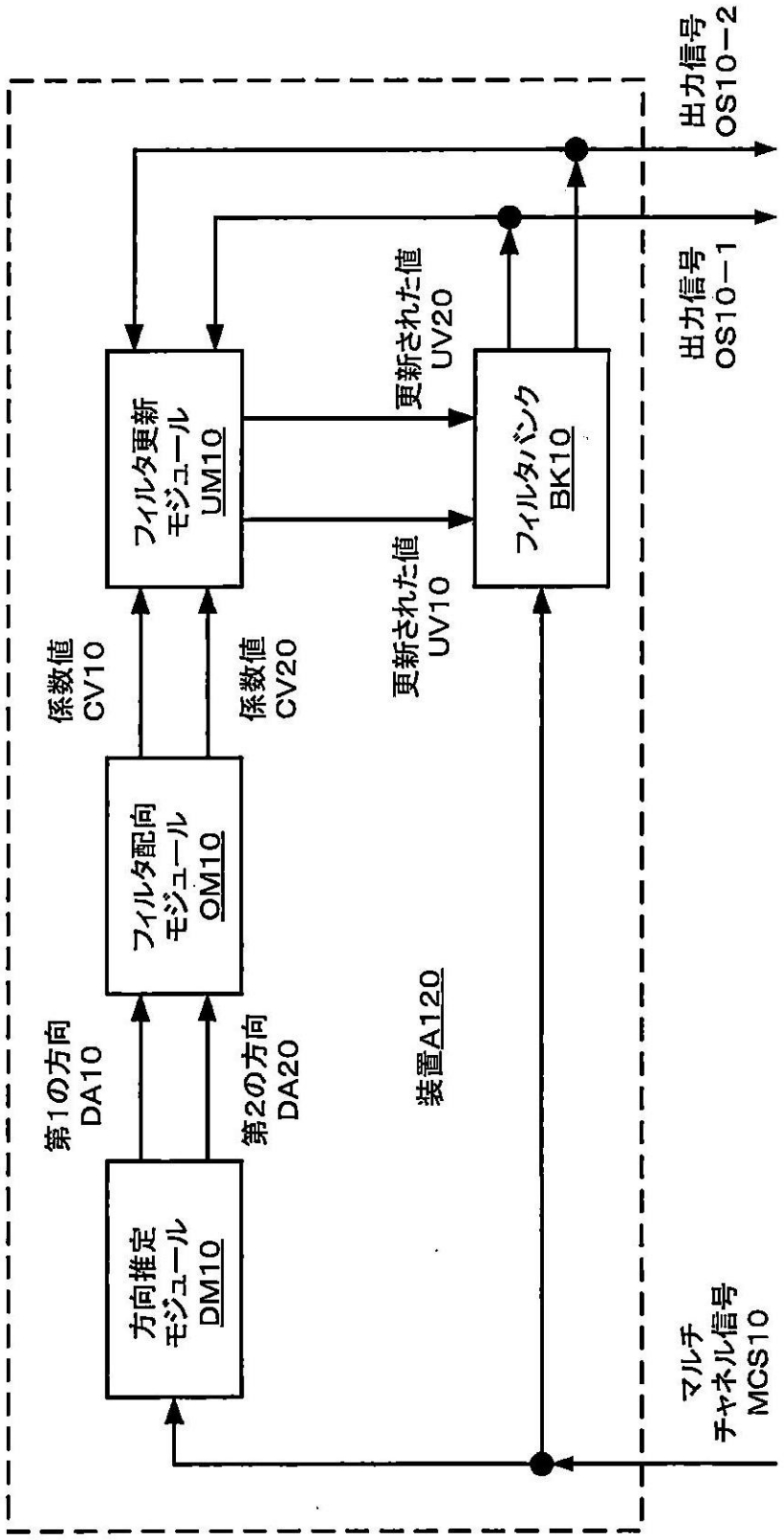


FIG. 20

【図２１】

図 21

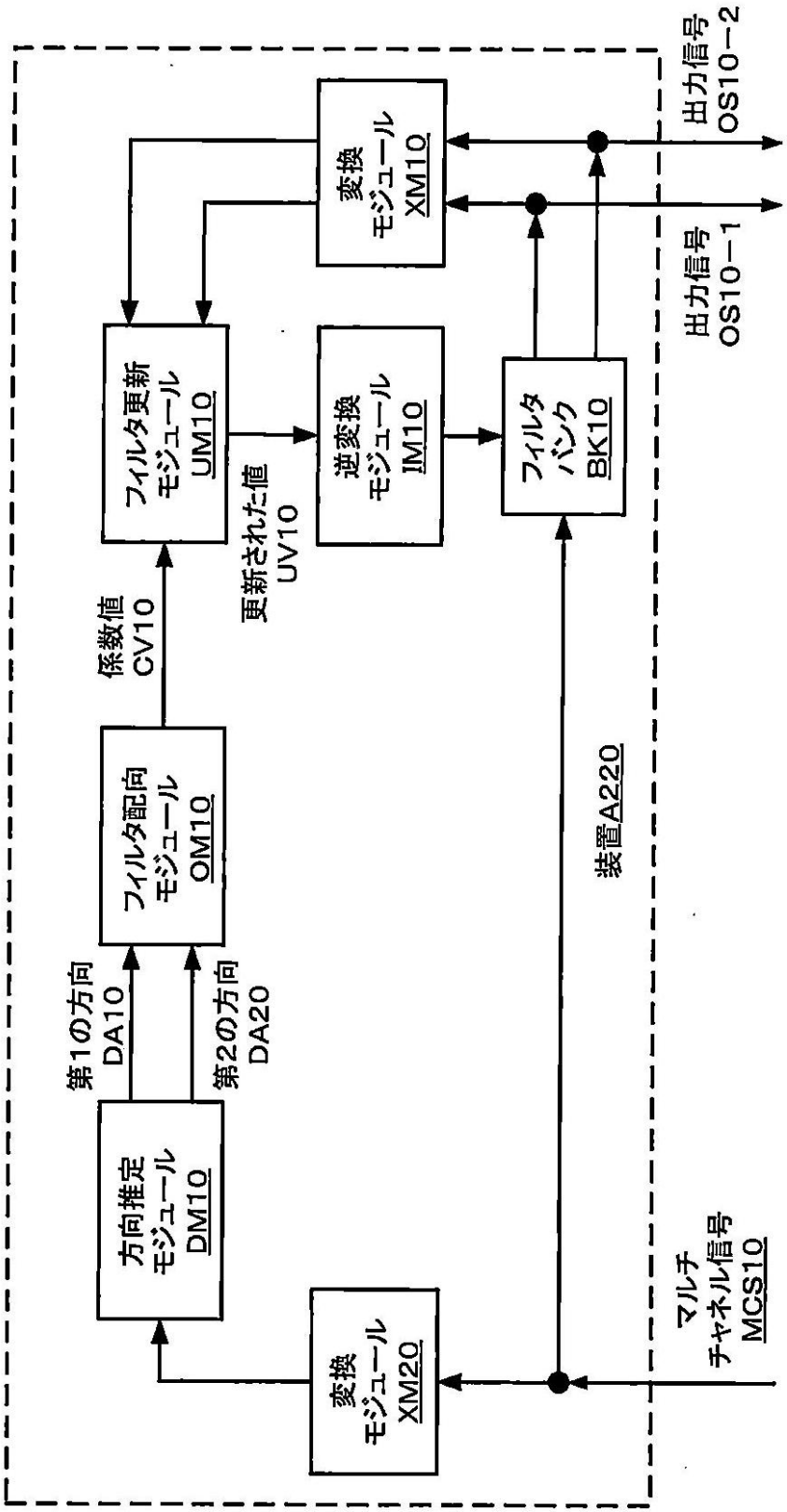


FIG. 21

【図２２】

図 22

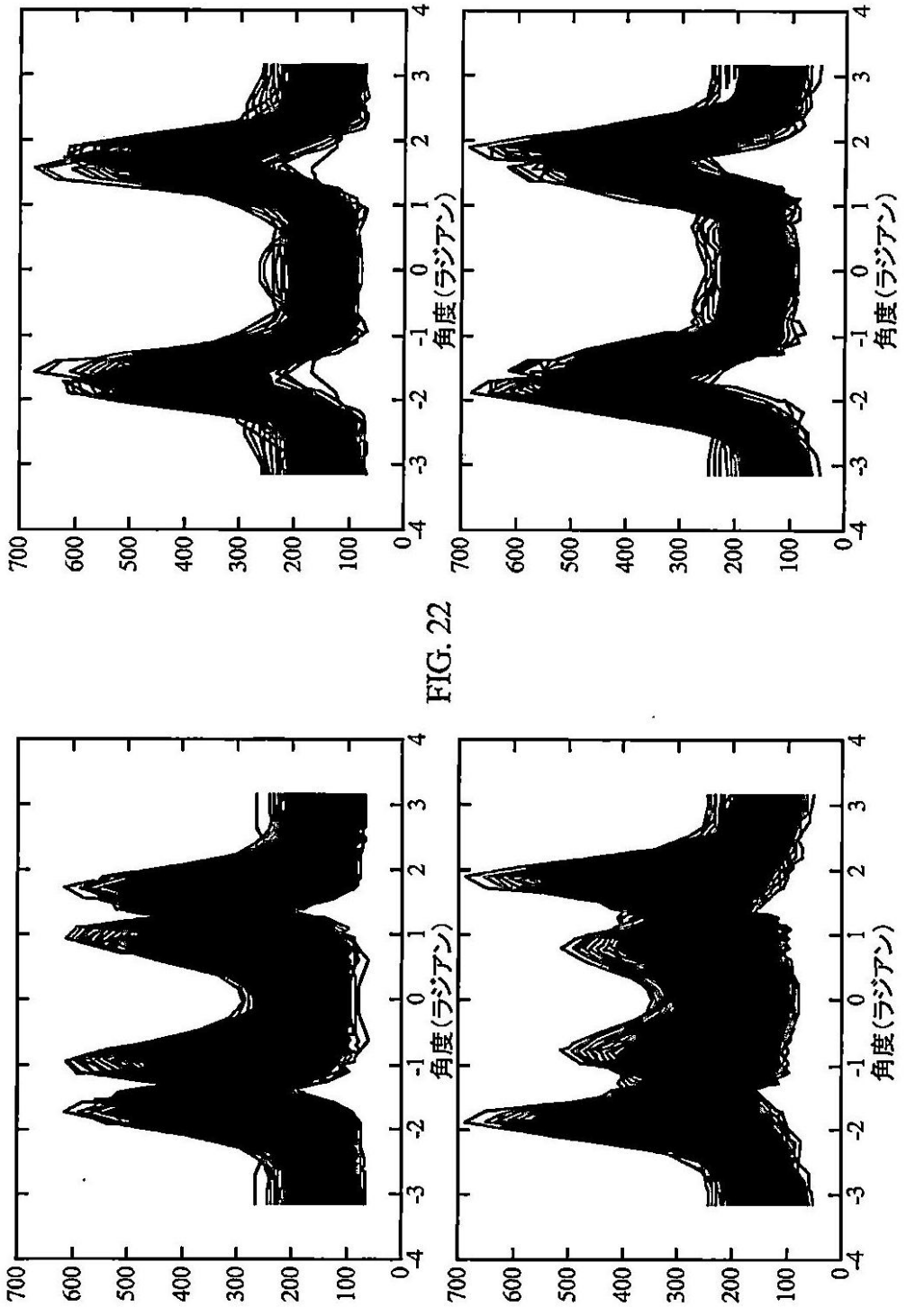


FIG. 22

【図２３】

図 23

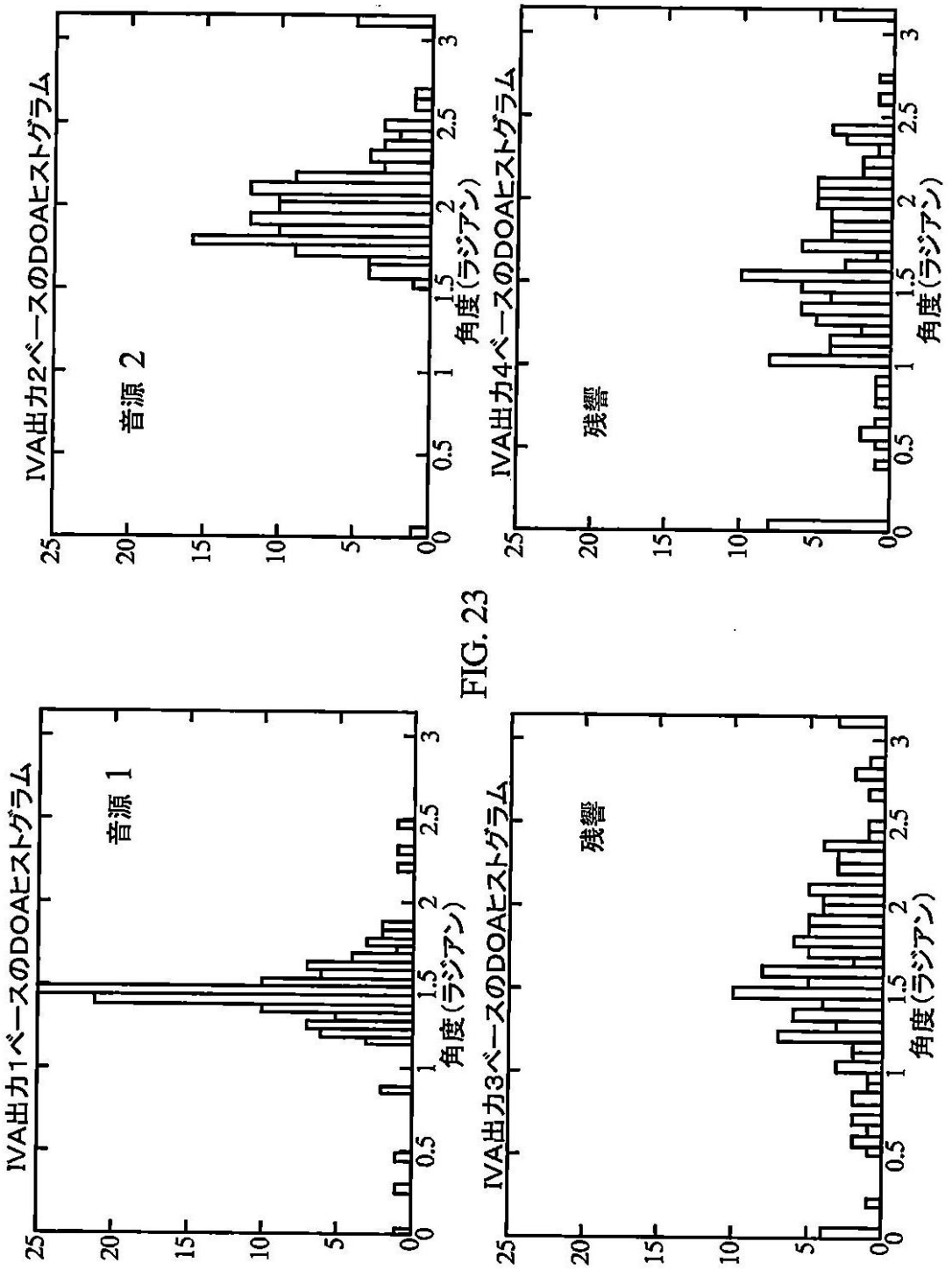


FIG. 23

【図２４】

図 24

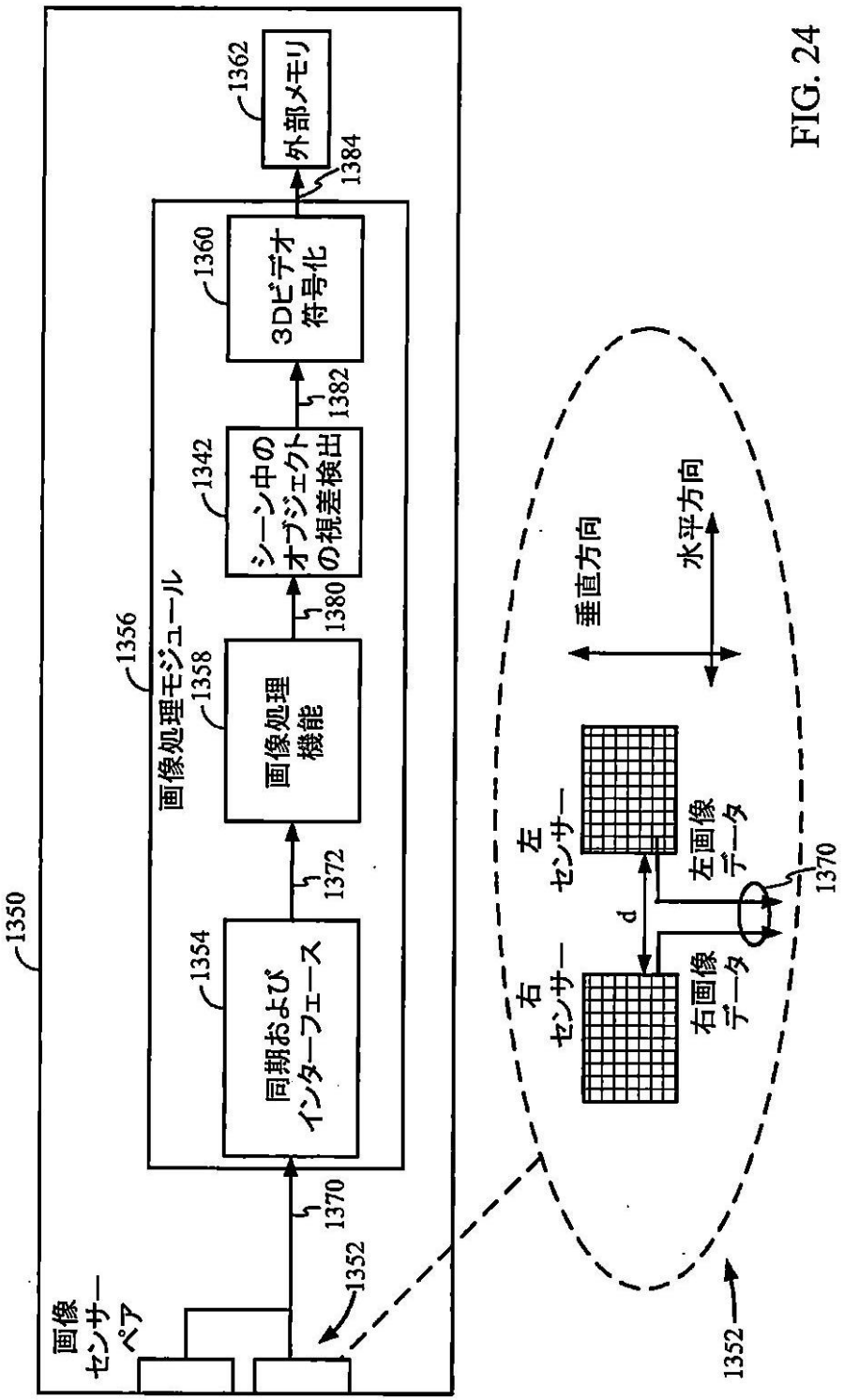


FIG. 24

【図２５】

図 25

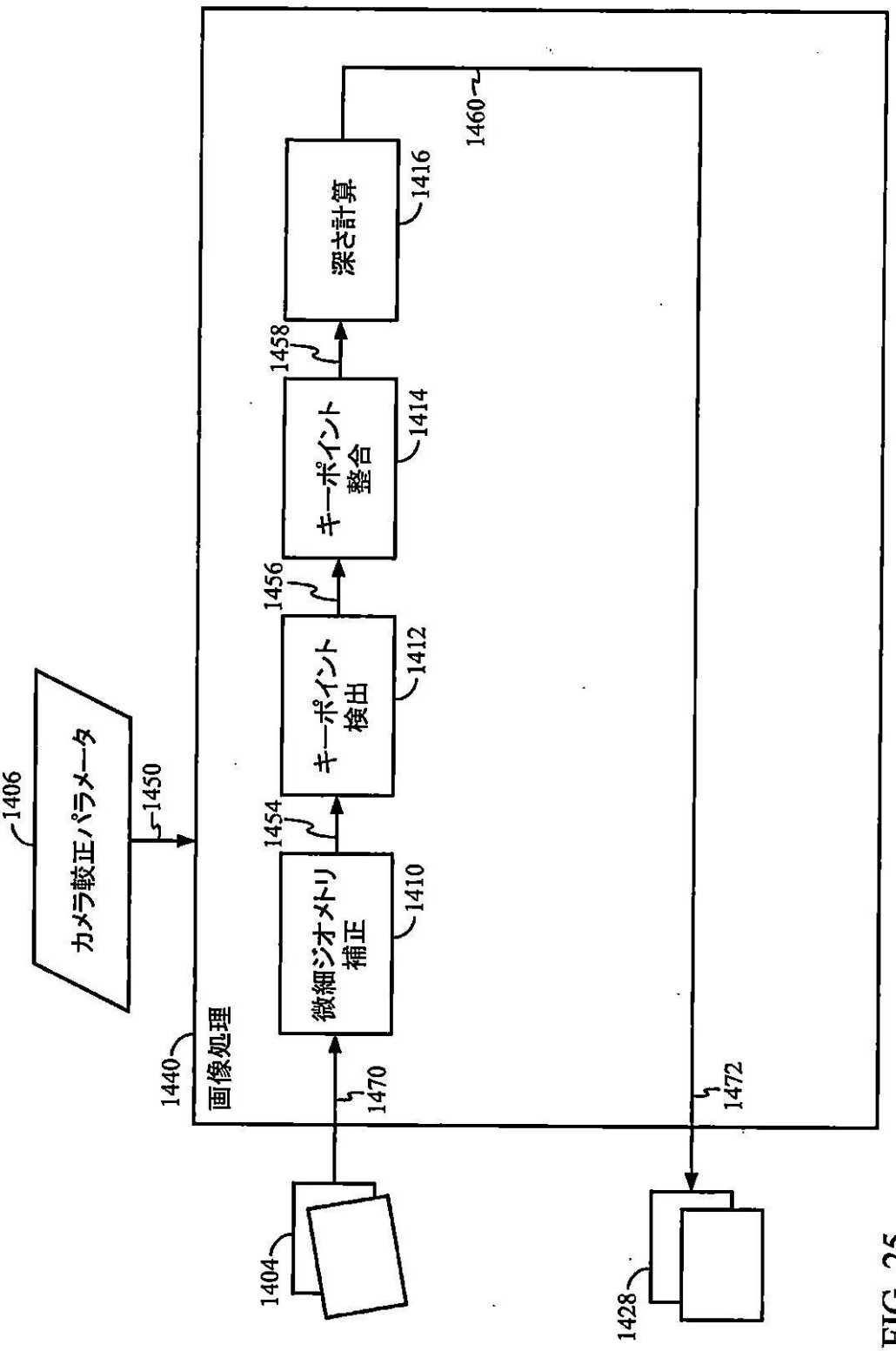


FIG. 25

【図２６Ａ】

図 26A

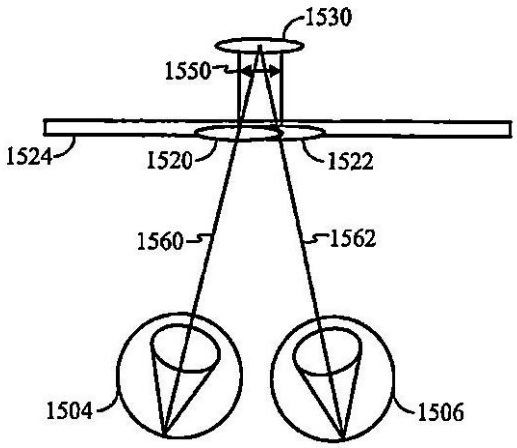


FIG. 26A

【図２６Ｂ】

図 26B

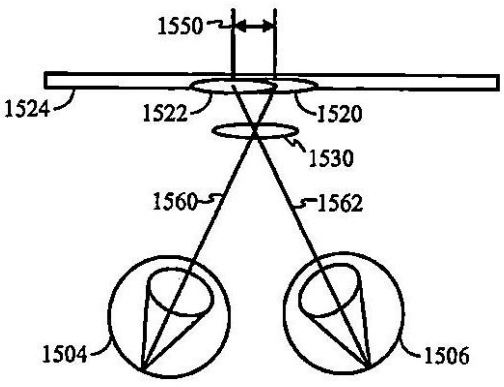


FIG. 26B

【図２７Ａ】

図 27A

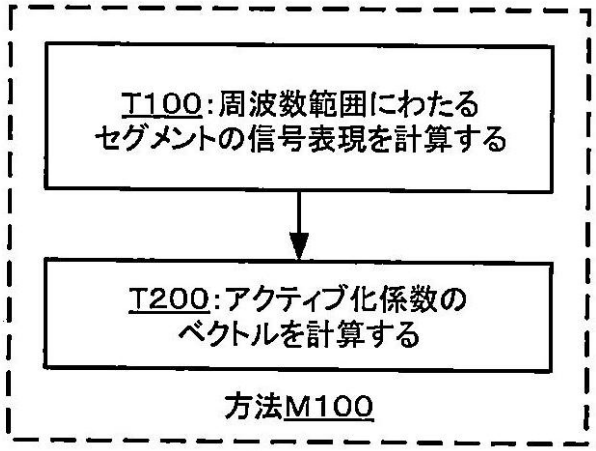


FIG. 27A

【図２７Ｂ】

図 27B

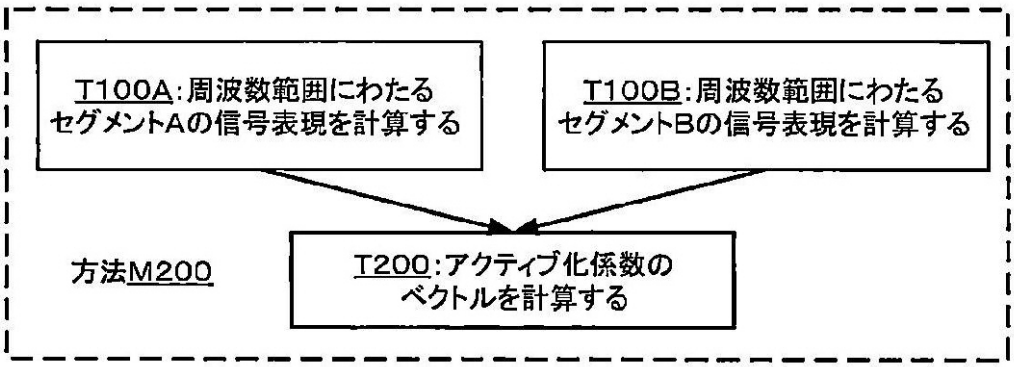


FIG. 27B

【図２７Ｃ】

図 27C

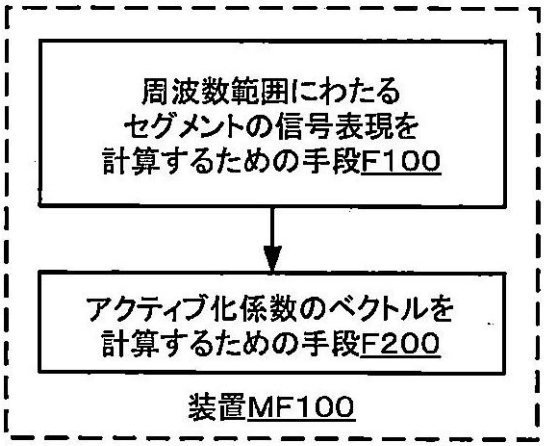


FIG. 27C

【図２７Ｄ】

図 27D

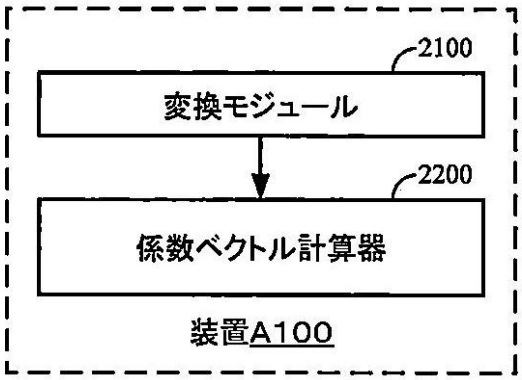


FIG. 27D

【図２８Ａ】

図 28A

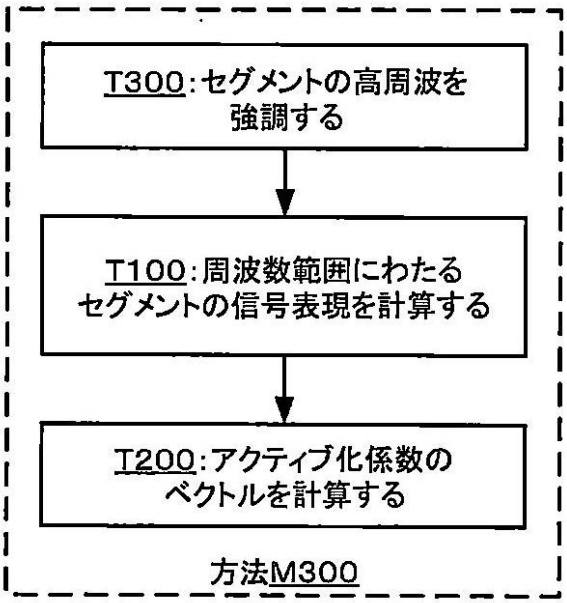
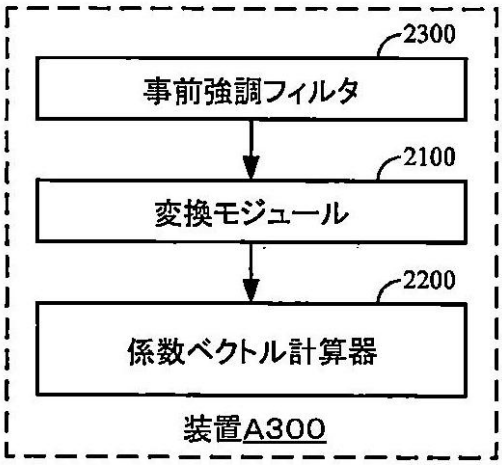


FIG. 28A

【図２８Ｂ】

図 28B

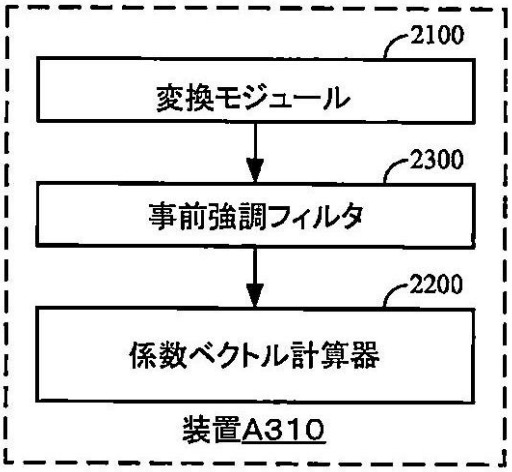
FIG. 28B



【図２８Ｃ】

図 28C

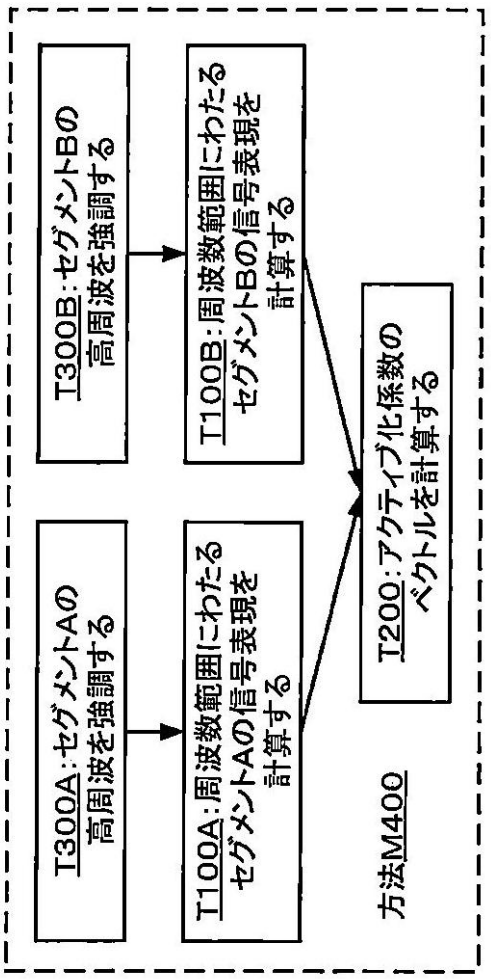
FIG. 28C



【図２９Ａ】

図 29A

FIG. 29A



【図２９Ｂ】

図 29B

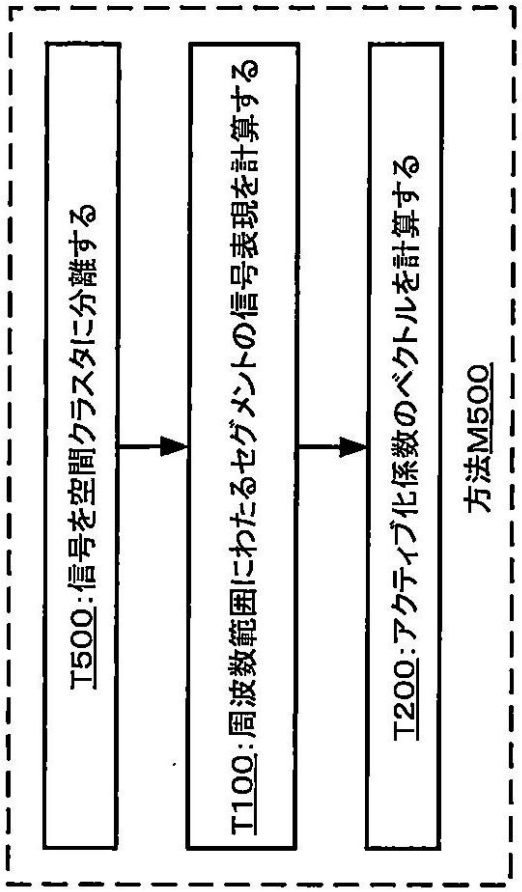


FIG. 29B

【図３０Ａ】

図 30A

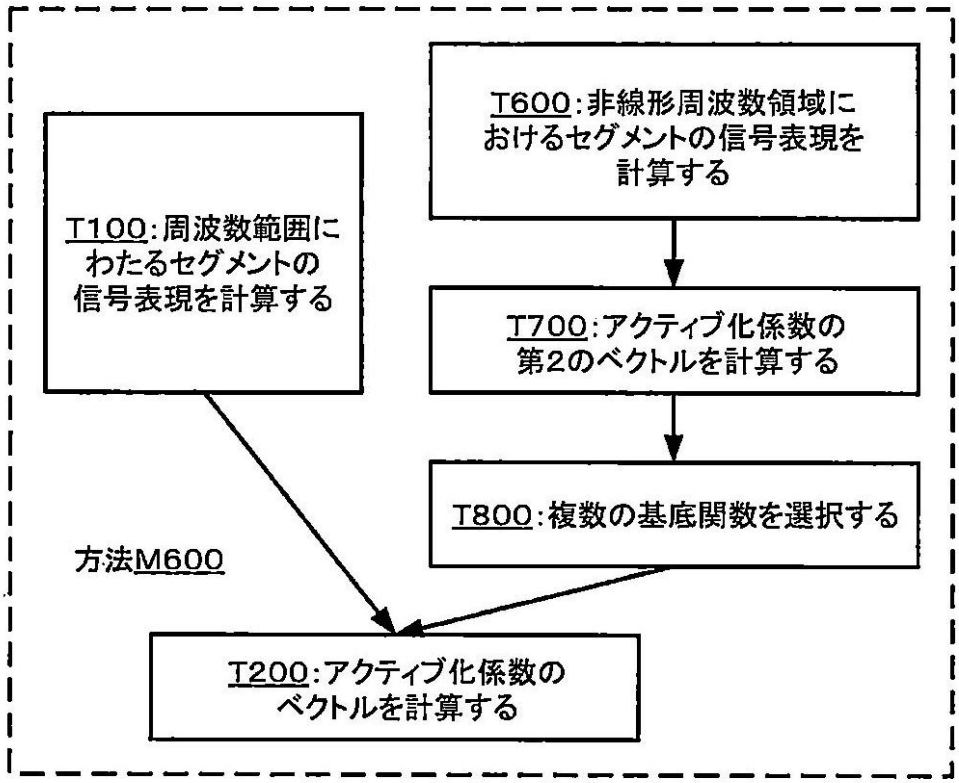


FIG. 30A

【図３０Ｂ】

図 30B

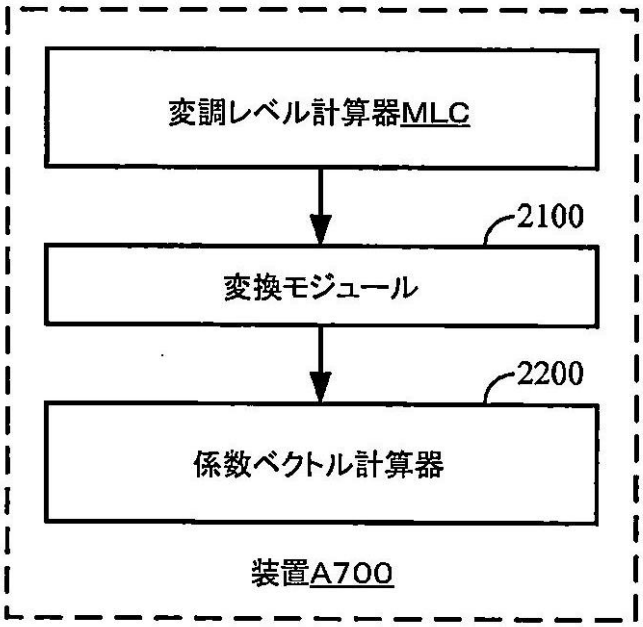


FIG. 30B

【図３１】

図 31

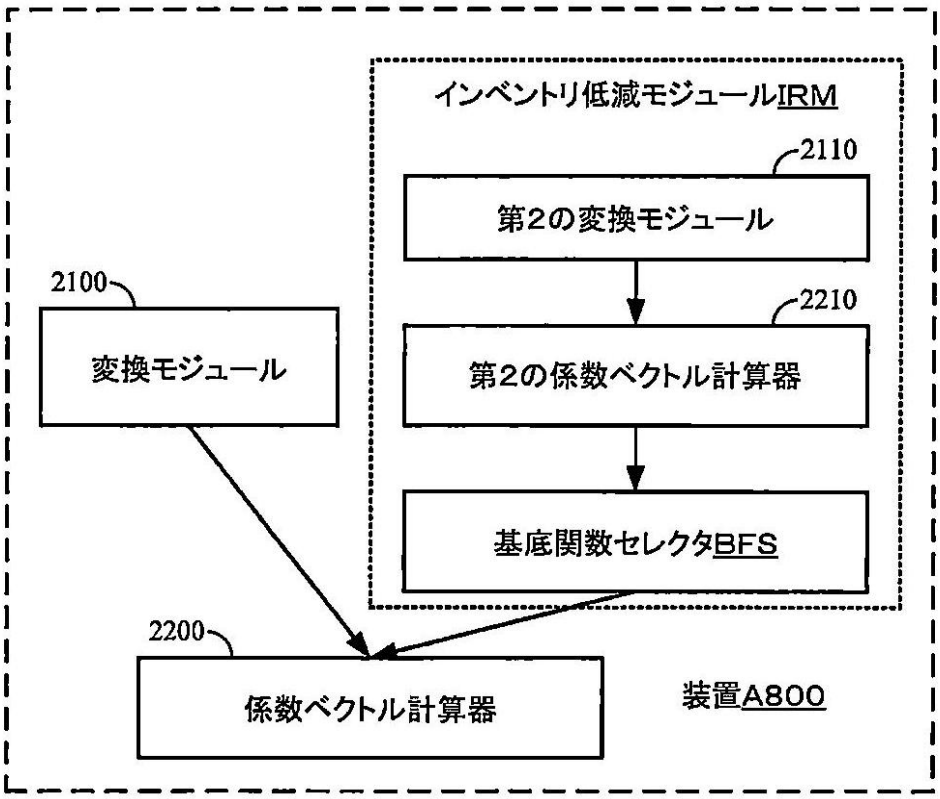


FIG. 31

【図３２】

図 32

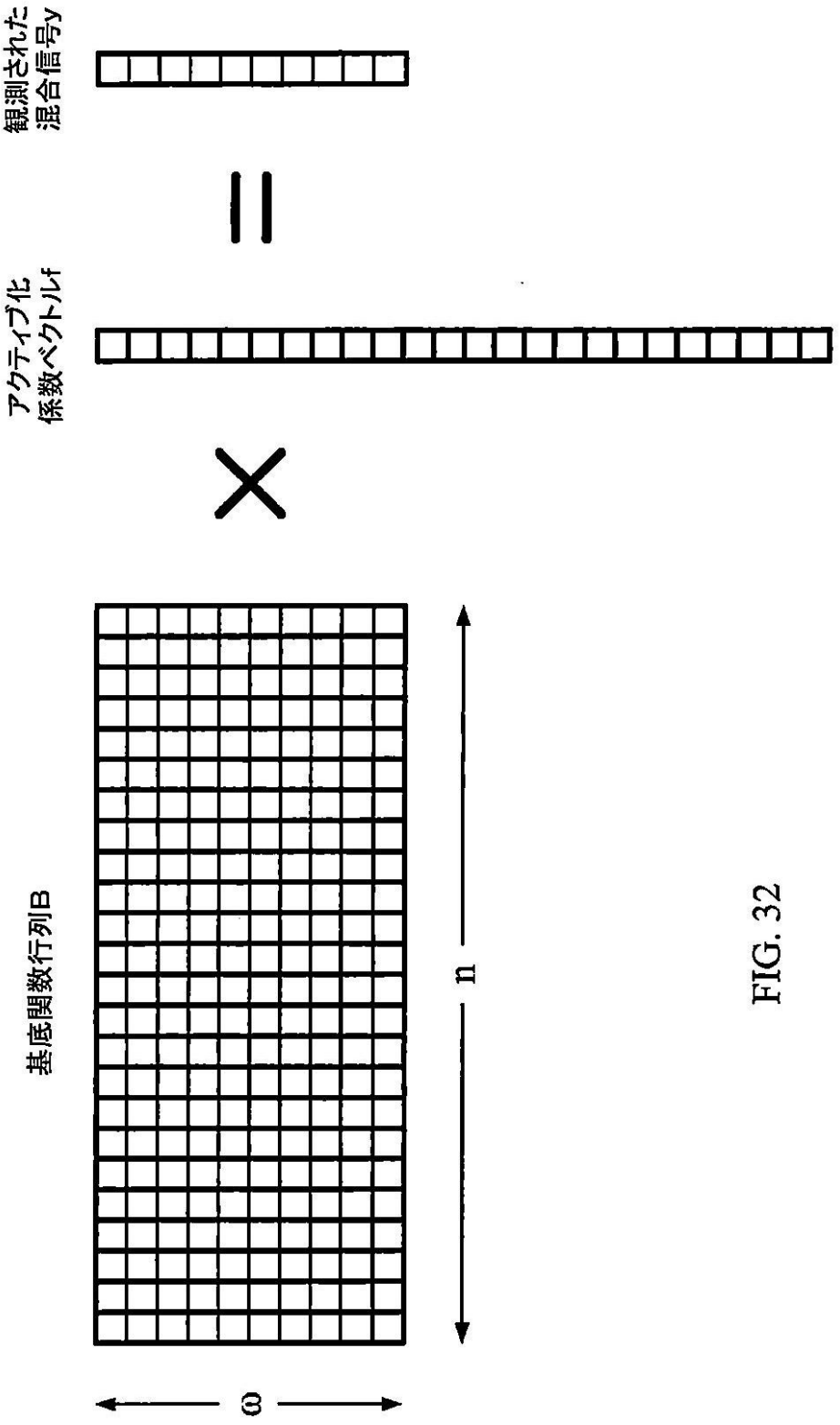


FIG. 32

【図３３】

図 33

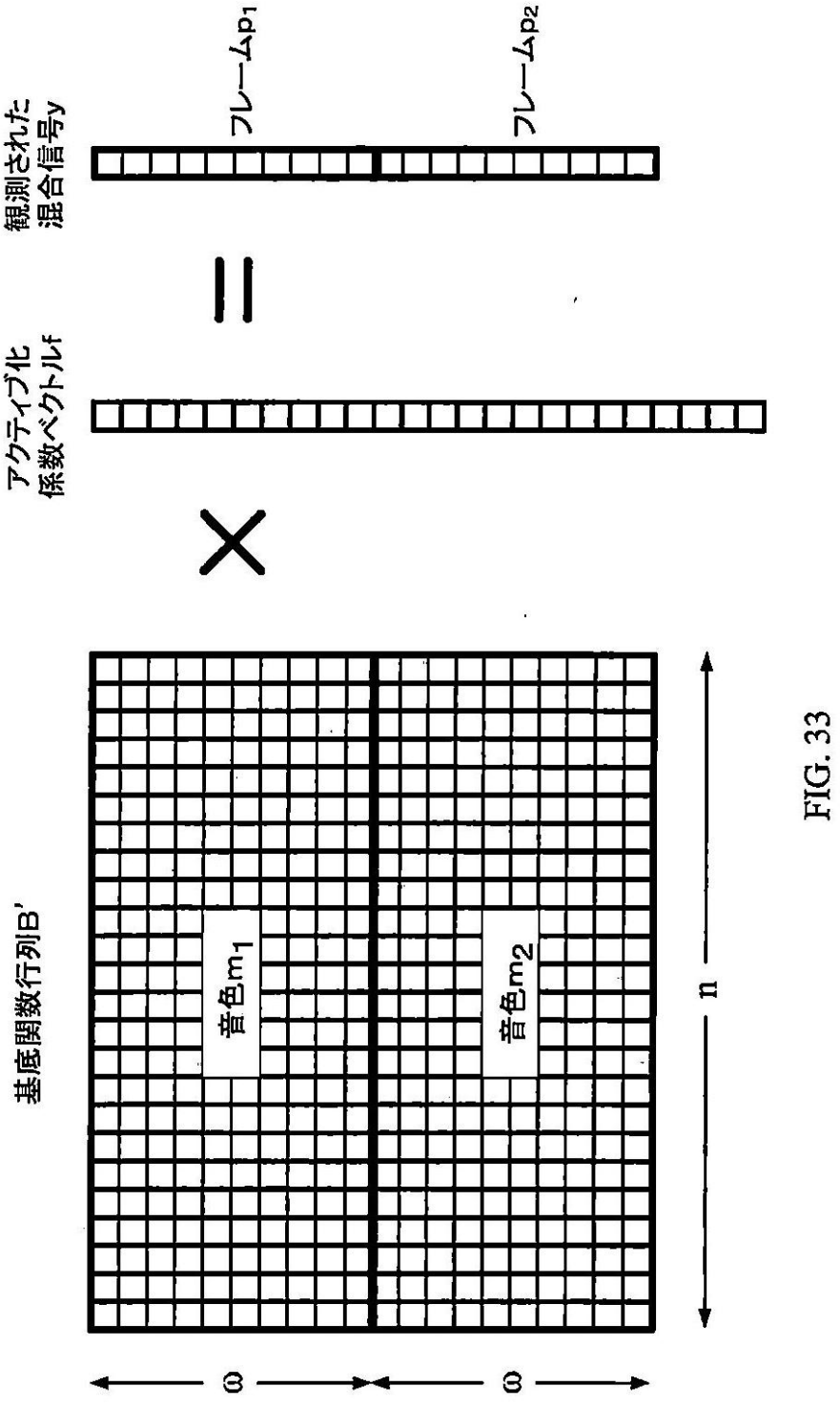


FIG. 33

【図３４】

図 34

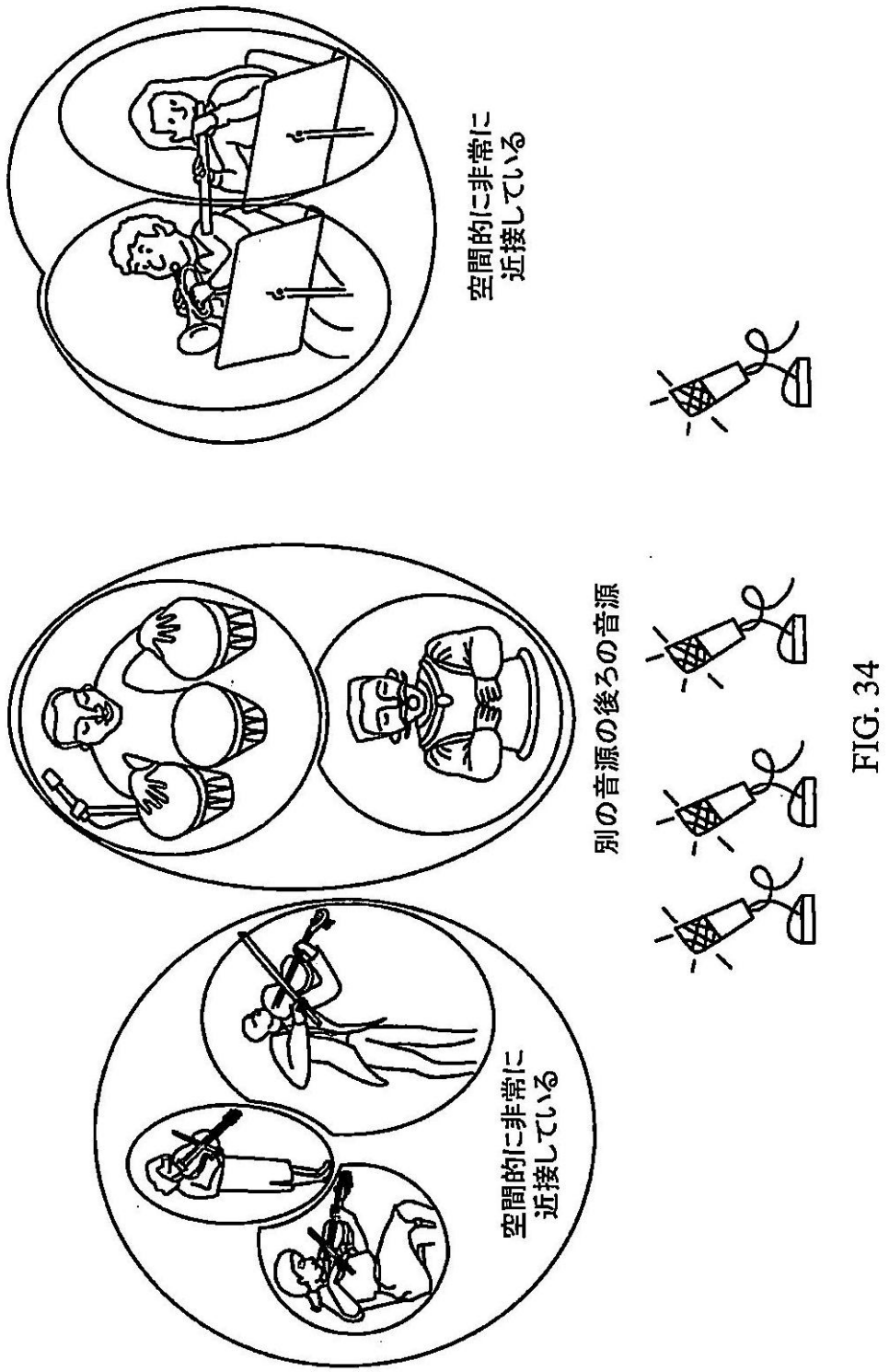


FIG. 34

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/029558

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06K9/62 G06K9/00 G06T7/20 H04R3/00
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)
G06K G06T H04R

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, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	STAFFAN EKVALL ET AL: "Integrating Active Mobile Robot Object Recognition and SLAM in Natural Environments", INTELLIGENT ROBOTS AND SYSTEMS, 2006 IEEE/RSJ INTERNATIONAL CONFERENCE ON, IEEE, PI, 1 October 2006 (2006-10-01), pages 5792-5797, XP031006097, ISBN: 978-1-4244-0258-8 sect III, III.B, IV, V.B ----- -/--	1-40

[X] Further documents are listed in the continuation of Box C. [X] 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	Date of mailing of the international search report
24 June 2013	01/07/2013

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Darolti, Cristina

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/029558

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VERMAAK J ET AL: "Sequential Monte Carlo fusion of sound and vision for speaker tracking", PROCEEDINGS OF THE EIGHT IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION. (ICCV). VANCOUVER, BRITISH COLUMBIA, CANADA, JULY 7 - 14, 2001; [INTERNATIONAL CONFERENCE ON COMPUTER VISION], LOS ALAMITOS, CA : IEEE COMP. SOC, US, vol. 1, 7 July 2001 (2001-07-07), pages 741-746, XP010554056, DOI: 10.1109/ICCV.2001.937600 ISBN: 978-0-7695-1143-6 abstract -----	1-40
A	STROBEL ET AL: "Joint Audio-Video Object Localization and Tracking", IEEE SIGNAL PROCESSING MAGAZINE, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, 1 January 2001 (2001-01-01), pages 22-31, XP002358730, ISSN: 1053-5888, DOI: 10.1109/79.911196 sect. Object Localization; page 24 -----	1-40
A	LO D ET AL: "Multimodal talker localization in video conferencing environments", HAPTIC, AUDIO AND VISUAL ENVIRONMENTS AND THEIR APPLICATIONS, 2004. HA VE 2004. PROCEEDINGS. THE 3RD IEEE INTERNATIONAL WORKSHOP ON OTTAWA, ONT., CANADA 2-3 OCT. 2004, PISCATAWAY, NJ, USA,IEEE, US, 2 October 2004 (2004-10-02), pages 195-200, XP010765318, DOI: 10.1109/HAVE.2004.1391905 ISBN: 978-0-7803-8817-8 sect. III; figures 1-4 -----	1-40
X	US 2003/103647 A1 (RUI YONG [US] ET AL) 5 June 2003 (2003-06-05) paragraphs [0071], [0073], [0105], [0106], [0163] -----	1-5, 11-15, 21-25, 31-35
X	EP 1 643 769 A1 (SAMSUNG ELECTRONICS CO LTD [KR]) 5 April 2006 (2006-04-05) paragraph [0054] - paragraph [0057]; figures 4A-C -----	1-5, 11-15, 21-25, 31-35
	-/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/029558

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SOCIAL ROBOT ED - ANONYMOUS: "Attending to Learn and Learning to Attend for a", HUMANOID ROBOTS, 2006 6TH IEEE-RAS INTERNATIONAL CONFERENCE ON, IEEE, PI, 1 December 2006 (2006-12-01), pages 618-623, XP031053086, ISBN: 978-1-4244-0199-4 page 68; figure 4.1 & Lijin Aryananda: "A Few Days of A Robot's Life in the Human's World: Toward Incremental Individual Recognition", PhD Thesis, 3 April 2007 (2007-04-03), XP055060888, Retrieved from the Internet: URL:http://dspace.mit.edu/handle/1721.1/37144 [retrieved on 2013-04-24] sect.II; figure 1	1-40
X	BENJAMIN FRANSEN ET AL: "Using vision, acoustics, and natural language for disambiguation", HUMAN-ROBOT INTERACTION (HRI), 2007 2ND ACM/IEEE INTERNATIONAL CONFERENCE ON, IEEE, 9 March 2007 (2007-03-09), pages 73-80, XP032211851, ISBN: 978-1-59593-617-2 sect.1, 4.9, 4.10, 5.2, 5.3, 6.1, 8, 9; figures 1, 6,7	1-5, 7-15, 17-25, 27-35, 37-40
X	GERALD FRIEDLAND ET AL: "Visual speaker localization aided by acoustic models", PROCEEDINGS OF THE SEVENTEEN ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, MM '09, 1 January 2009 (2009-01-01), page 195, XP055060737, New York, New York, USA DOI: 10.1145/1631272.1631301 ISBN: 978-1-60-558608-3 sect.4.3, 4.4, 5, 6, 7; figure 6	1-3,7-10
A	Ming-Yu Chen ET AL: "MoSIFT: Recognizing Human Actions in Surveillance Videos", CMU-CS Report, 24 September 2009 (2009-09-24), XP055067839, Retrieved from the Internet: URL:http://www.cs.cmu.edu/~mychen/publication/ChenMoSIFTCMU09.pdf [retrieved on 2013-06-21] sect.1 par.4-5, sect.2, 2.2; figure 2	1-40

-/--

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/029558

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	& Ming-Yu Chen: "Long Term Activity Analysis in Surveillance Video Archives", CMU PhD Thesis, 12 September 2010 (2010-09-12), XP055067853, Retrieved from the Internet: URL:http://www.lti.cs.cmu.edu/research/thesis/2010/mingyu_chen.pdf [retrieved on 2013-06-21] chapter 1, e.g. sect.1.2, 1.3, 1.6.3, 1.6.4	
X	Elie El-Khoury: "Unsupervised video indexing based on audiovisual characterization of persons", These Doctorat Uni Toulouse, 3 June 2010 (2010-06-03), XP55067931, Retrieved from the Internet: URL:http://thesesups.ups-tlse.fr/1025/1/El-Khoury_Elie.pdf [retrieved on 2013-06-24] sect.7	1-3,7-10
X	INOUE N ET AL: "High-Level Feature Extraction Using SIFT GMMs and Audio Models", PATTERN RECOGNITION (ICPR), 2010 20TH INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 23 August 2010 (2010-08-23), pages 3220-3223, XP031772106, ISBN: 978-1-4244-7542-1 sect.I, II.A, III.A	1-3,7-10
X	TALANTZIS F ET AL: "Audioâ Visual Active Speaker Tracking in Cluttered Indoors Environments", IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS. PART B:CYBERNETICS, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 38, no. 3, 1 June 2008 (2008-06-01), pages 799-807, XP011344943, ISSN: 1083-4419, DOI: 10.1109/TSMCB.2008.922063 sect.II, II.C, III.B, IV	1-10

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/029558

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☒ No protest accompanied the payment of additional search fees.

International Application No. PCT/US2013/029558

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-5, 11-15, 21-25, 31-35

Claims 3-5 referring to the features of determining a DOA using an array of microphones to select a portion of the scene based on audio.

---

2. claims: 6, 16, 26, 36

Claim 6 referring to the feature of computing local motion vectors in a video and using them together with the keypoints for recognition

---

3. claims: 7-10, 17-20, 27-30, 37-40

1) Claims 7-8 referring to the features of computing MFCC acoustic recognition features using them together with the keypoints for recognition.
2) Claims 9-10 referring to the features of determining range information (by multicamera disparity) and analysing the keypoints based on range.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2013/029558

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2003103647 A1	05-06-2003	AT 397354 T	15-06-2008
		AT 551676 T	15-04-2012
		CN 1423487 A	11-06-2003
		CN 101093541 A	26-12-2007
		EP 1330128 A2	23-07-2003
		EP 1838104 A2	26-09-2007
		EP 1942679 A2	09-07-2008
		EP 1944975 A2	16-07-2008
		JP 4142420 B2	03-09-2008
		JP 4536789 B2	01-09-2010
		JP 4607984 B2	05-01-2011
		JP 4642093 B2	02-03-2011
		JP 2003216951 A	31-07-2003
		JP 2008204479 A	04-09-2008
		JP 2008243214 A	09-10-2008
		JP 2008243215 A	09-10-2008
		KR 20030045624 A	11-06-2003
		TW I222031 B	11-10-2004
		US 2003103647 A1	05-06-2003
		US 2005129278 A1	16-06-2005
		US 2005147278 A1	07-07-2005
		US 2005188013 A1	25-08-2005
		US 2005210103 A1	22-09-2005
EP 1643769 A1	05-04-2006	NONE	

Form PCT/ISA/210 (patent family annex) (April 2005)

フロントページの続き

(51)Int.Cl.		F I			テーマコード（参考）
Ｇ１０Ｌ　15/28	*(2013.01)*	Ｇ１０Ｌ	15/28	４００	
Ｇ１０Ｌ　15/24	*(2013.01)*	Ｇ１０Ｌ	15/24	Ｚ	

(81)指定国　AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,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,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,CZ,DE,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IS,JP,KE,KG,KM,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LT,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,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC

(74)代理人　100153051
弁理士　河野　直樹
(74)代理人　100140176
弁理士　砂川　克
(74)代理人　100158805
弁理士　井関　守三
(74)代理人　100179062
弁理士　井上　正
(74)代理人　100124394
弁理士　佐藤　立志
(74)代理人　100112807
弁理士　岡田　貴志
(74)代理人　100111073
弁理士　堀内　美保子
(72)発明者　ビッサー、エリック
アメリカ合衆国、カリフォルニア州　９２１２１、サン・ディエゴ、モアハウス・ドライブ　５７７５
(72)発明者　ワン、ヒイン
アメリカ合衆国、カリフォルニア州　９２１２１、サン・ディエゴ、モアハウス・ドライブ　５７７５
(72)発明者　シディクイ、ハシブ・エー.
アメリカ合衆国、カリフォルニア州　９２１２１、サン・ディエゴ、モアハウス・ドライブ　５７７５
(72)発明者　キム、レーホン
アメリカ合衆国、カリフォルニア州　９２１２１、サン・ディエゴ、モアハウス・ドライブ　５７７５

Ｆターム(参考) 5L096 AA09 CA05 FA35 FA67 HA03 HA04 HA07 JA16