(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2006-60801
(P2006-60801A)
(43)公開日　平成18年3月2日(2006.3.2)

| (51)Int.Cl. | | | F I | | | テーマコード（参考） |
|---|---|---|---|---|---|---|
| H04N | 7/173 | (2006.01) | H04N | 7/173 | 610Z | 5C053 |
| H04N | 5/765 | (2006.01) | H04N | 5/91 | L | 5C059 |
| H04N | 7/26 | (2006.01) | H04N | 7/173 | 640A | 5C164 |
| | | | H04N | 7/13 | Z | |

審査請求 未請求　請求項の数 28　ＯＬ　外国語出願　（全 31 頁）

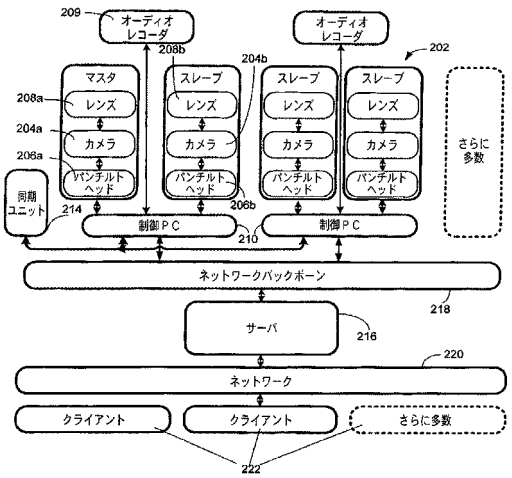
(21)出願番号　特願2005-217367 (P2005-217367)
(22)出願日　平成17年7月27日 (2005.7.27)
(31)優先権主張番号　10/901,477
(32)優先日　平成16年7月27日 (2004.7.27)
(33)優先権主張国　米国 (US)
(31)優先権主張番号　10/923,107
(32)優先日　平成16年8月19日 (2004.8.19)
(33)優先権主張国　米国 (US)

(71)出願人　500046438
マイクロソフト　コーポレーション
アメリカ合衆国　ワシントン州　９８０５２－６３９９　レッドモンド　ワン　マイクロソフト　ウェイ
(74)代理人　100077481
弁理士　谷　義一
(74)代理人　100088915
弁理士　阿部　和夫
(72)発明者　ホワ　カイ
アメリカ合衆国　９８０５２　ワシントン州　レッドモンド　ワン　マイクロソフト　ウェイ　マイクロソフト　コーポレーション内

最終頁に続く

(54)【発明の名称】対話式マルチビュービデオのクライアントサービスのためのシステムおよび方法

(57)【要約】
【課題】新しいタイプのビデオ取込みシステム、ビデオフォーマット、ビデオ圧縮アルゴリズム、サービスを提示する、対話式マルチビュービデオを提供する。
【解決手段】多くのビデオカメラが、関連する様々な位置および方向からイベントを取り込むように割り振られる。取り込まれたビデオは、圧縮され、リアルタイムでサーバに送信される。ユーザは、ユーザがサーバに接続して対話式にマルチビュービデオを受信することを可能にする新しいタイプのサービスに加入することができる。従来の再生制御に加えて、ユーザは、カメラ位置および配向の制御を操作すること、視聴方向を選択すること、スイープ効果やフリーズ回転効果などの特殊効果を楽しむことなどができる。対話式マルチビュービデオは、イベント視聴におけるまったく新しい体験を提供する。
【選択図】図２



【特許請求の範囲】
【請求項１】
ビデオ視聴サービスをユーザに提供するためのコンピュータ実施方法であって、
複数のカメラを使用して同じイベントの様々な画像ストリームを様々なビューから同時に取り込むプロセス動作と、
前記取り込まれた画像の特定ストリームをクライアントが対話式に要求するプロセス動作と、
前記取り込まれた画像の前記要求された特定ストリームをサーバから前記クライアントに提供するプロセス動作と
を備えることを特徴とするコンピュータ実施方法。
【請求項２】
取り込まれた画像の前記特定ストリームはリアルタイムで提供されることを特徴とする請求項１に記載のコンピュータ実施方法。
【請求項３】
前記クライアントは少なくとも１つのカメラの視点をリアルタイムで制御することができることを特徴とする請求項２に記載のコンピュータ実施方法。
【請求項４】
前記クライアントは少なくとも１つのカメラのパンチルト値を制御することができることを特徴とする請求項３に記載のコンピュータ実施方法。
【請求項５】
前記取り込まれたビデオストリームは後で視聴するために記憶されることを特徴とする請求項１に記載のコンピュータ実施方法。
【請求項６】
前記ユーザは、前記ビデオストリームが正しいテンポで継続している間に、所望の視点を提供する様々なカメラからの前記ビデオストリームにアクセスすることによって、あるカメラ視点からの前記取り込まれたビデオストリームの視聴から、別のビデオ視点からの別の前記取り込まれたビデオストリームの視聴に切り換えることができることを特徴とする請求項５に記載のコンピュータ実施方法。
【請求項７】
前記ユーザは、前記ビデオストリームが正しいテンポで継続している間に、所望の視点を提供する隣接カメラからの前記ビデオストリームにアクセスすることによって、隣接カメラ視点からの前記取り込まれたビデオストリームの視聴にスイープすることができることを特徴とする請求項５に記載のコンピュータ実施方法。
【請求項８】
時が静止し、前記カメラ視点が所与の点の周りで回転することを特徴とする請求項５に記載のコンピュータ実施方法。
【請求項９】
前記ユーザは前に視聴または作成されたビデオシーケンスを再生することができることを特徴とする請求項５に記載のコンピュータ実施方法。
【請求項１０】
ユーザはオンデマンドで再生できるビューおよび特殊効果のセットのスクリプトを作成することができることを特徴とする請求項５に記載のコンピュータ実施方法。
【請求項１１】
前記ユーザは前記スクリプトを他のユーザに送信することができ、前記他のユーザは、前記スクリプトが起動されると、スクリプトされたのと同じビデオイベントを見ることができることを特徴とする請求項１０に記載のコンピュータ実施方法。
【請求項１２】
サーバが、前記取り込まれたビデオからビデオストリームを生成し、前記クライアントの要求に基づいて前記ビデオストリームを前記クライアントに送信することを特徴とする請求項１に記載のコンピュータ実施方法。

【請求項１３】
１つのクライアントに対して２つの通信チャネルがあることを特徴とする請求項１２に記載のコンピュータ実施方法。
【請求項１４】
一方の通信チャネルは、レイテンシを低減するためにオーディオ／ビデオデータの送信に使用されることを特徴とする請求項１３に記載のコンピュータ実施方法。
【請求項１５】
前記通信チャネルはユーザデータグラムプロトコル（ＵＤＰ）チャネルであることを特徴とする請求項１４に記載のコンピュータ実施方法。
【請求項１６】
一方の通信チャネルはコマンドおよび制御データの送信に使用されることを特徴とする請求項１３に記載のコンピュータ実施方法。
【請求項１７】
前記通信チャネルは伝送制御プロトコル（ＴＣＰ）チャネルであることを特徴とする請求項１６に記載のコンピュータ実施方法。
【請求項１８】
ユーザがビデオを見るためのコンピュータ実施方法であって、
複数のカメラを使用して同じイベント空間の様々な画像ストリームを様々な視点から同時に取り込むプロセス動作と、
前記取り込まれた様々な画像ストリームのうちユーザから要求されたストリームを、ユーザによって見られるようにコンピュータで受信するプロセス動作と
を備えることを特徴とするコンピュータ実施方法。
【請求項１９】
前記取り込まれた様々な画像ストリームのうちユーザから要求されたストリームは、前記画像が取り込まれているときにリアルタイムで前記コンピュータに提供されることを特徴とする請求項１８に記載のコンピュータ実施方法。
【請求項２０】
取り込まれた様々な画像ストリームは記憶媒体に記憶され、前記取り込まれた様々な画像ストリームのうちユーザから要求されたストリームは、前記記憶媒体から前記コンピュータに提供されることを特徴とする請求項１８に記載のコンピュータ実施方法。
【請求項２１】
前記取り込まれた様々な画像ストリームはサーバに記憶されることを特徴とする請求項２０に記載のコンピュータ実施方法。
【請求項２２】
前記取り込まれた様々な画像ストリームは前記ユーザのコンピュータに記憶されることを特徴とする請求項２１に記載のコンピュータ実施方法。
【請求項２３】
前記ユーザは、前記様々な画像ストリームが取り込まれる際の視点を制御することができることを特徴とする請求項１８に記載のコンピュータ実施方法。
【請求項２４】
ユーザが見るためのビデオを提供するためのシステムであって、
汎用コンピューティングデバイスと、
前記汎用コンピューティングデバイスによって実行可能なプログラムモジュールを含むコンピュータプログラムとを備え、
前記コンピューティングデバイスは、前記コンピュータプログラムの前記プログラムモジュールにより、
複数のカメラによって様々な視点から同時に取り込まれた同じイベントの複数のビデオストリームを入力するステップと、
前記入力された複数のビデオストリームのうちユーザから要求されたストリームを、クライアント位置にいるユーザによって見られるようにサーバから前記クライアントに提供

するステップと
を行うように命令される
ことを特徴とするシステム。
【請求項２５】
前記複数のビデオストリームのうちユーザから要求されたストリームは、リアルタイムで前記クライアントに提供されることを特徴とする請求項２４に記載のシステム。
【請求項２６】
前記ビデオストリームが正しいテンポで継続する間に、あるカメラ視点から取り込まれたビデオストリームから、別のカメラ視点から取り込まれた少なくとも１つのビデオストリームに切り換わるビデオストリームが、前記ユーザに送信されることを特徴とする請求項２４に記載のシステム。
【請求項２７】
前記ビデオストリームが正しいテンポで継続する間に、隣接カメラ視点からのビデオストリームが前記ユーザに送信されることを特徴とする請求項２４に記載のシステム。
【請求項２８】
時が静止し、所与の点の周りで回転するカメラ視点からのビデオストリームが前記ユーザに送信されることを特徴とする請求項２４に記載のシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、新しいタイプのビデオ視聴サービスを含む対話式マルチビュービデオのためのシステムおよび方法を対象とする。
【背景技術】
【０００２】
現在一般に使用されているビデオ形式は、いわゆるシングルビュービデオである。これは、１台のビデオカメラから取り込まれた１つのビデオクリップ、または連続的な複数の期間を使用して連結された複数のビデオクリップからなる。任意の時間インスタンスに対して、イベントのビューは１つしかない。この種のビデオ形式は、テレビジョン（ＴＶ）やパーソナルコンピュータ（ＰＣ）やその他のデバイスで、ビデオストリーミングや放送や通信に広く使用されている。
【０００３】
従来のマルチメディアサービス（従来のＴＶ、ビデオオンデマンド、ビデオストリーミング、デジタルビデオディスク（ＤＶＤ）など）を見直してみると、いくつかの制限が存在する。例えば、従来のマルチメディアサービスでは、任意の時間インスタンスにおけるイベントに対して、ビデオストリームは１つしかない。加えて、従来のマルチメディアサービスでは、任意の時間インスタンスにおける視聴方向は、番組編集者によって選択される。ユーザは受動的な位置にあり、カメラアングルまたは視点を変更することはできない。さらに、ユーザは、録画されてユーザに提供されたものを見ることができるだけであり、視聴アングルを選択する能力はない。
【０００４】
従来のシングルビュービデオの拡張であるアイビジョン（ＥｙｅＶｉｓｉｏｎ）（非特許文献１参照）は、カーネギーメロン大学コンピュータビジョン教授の金出武雄氏によって共同開発されたスポーツ放送システムである。アイビジョンは、スーパーボウル２００１で３０台のカムコーダを利用して試合を撮影した。３０台のカムコーダから取り込まれたビデオはすべてビデオルーティング切換え装置に入力され、編集済みビデオがＴＶ視聴者に放映された。しかし、アイビジョンシステムは、１つの編集済みビデオしかユーザに提供せず、ユーザは、視聴方向を選択することやカメラ制御を実施することはできない。また、アイビジョンシステムはＴＶ視聴者だけに役立ち、他のマルチメディアフォーマットでは利用不可能である。
【０００５】

アイビジョンに加えて、別のマルチメディアデバイスである３Ｄレコーダが、自由視点ビデオを録画再生するために設計された（非特許文献２参照）。これは、最初に２Ｄビデオを取り込み、次いで背景から前景を抽出する。ソースコーディングを適用して、３Ｄ前景オブジェクト（例えば人間）を生み出す。しかし、アイビジョンと同様にこの３Ｄレコーダでも、ユーザがカメラを制御することはできない。加えて、この３Ｄビデオレコーダによって利用される処理では、背景から前景を分類する必要があり、これはかなりの計算資産を要する。

【０００６】

マルチビュービデオに対する需要の増大に伴って、最近、標準化の取組みが行われている（非特許文献３、４参照）。ＭＰＥＧ界は、２００１年１２月から、３ＤＡＶ（３Ｄオーディオビジュアル）技術の探究に取り組んでいる。３Ｄビデオという用語に関しては、非常に多岐にわたる多くの応用および技術が論じられてきた。これらの応用はどれも、動的な実際の視聴覚状況で、または実際に取り込まれた画像から再構築された３Ｄオブジェクトを含む動的な状況で、ユーザが自分の視点および／または方向を選択することが可能だという意味での対話性に焦点を合わせてはいなかった。応用シナリオに関して、マルチビュービデオは、最も不完全で非効率的で利用不可能な要素を含む、最も難題を呈するシナリオであることがわかっている。この領域は、近い将来、最も多くの標準化労力を必要とする。さらに、対話性を扱った標準化の取組みはなかった。

【０００７】

いくつかの文献に上述のような従来の技術に関連した技術内容が開示されている（例えば、非特許文献１～５参照）。

【０００８】

【非特許文献１】http://www.ri.cmu.edu/projects/project449.html

【非特許文献２】Ｓ．Ｗｕｒｍｌｉｎ、Ｅ．Ｌａｍｂｏｒａｙ、Ｏ．Ｇ．Ｓｔａａｄｔ、Ｍ．Ｈ．Ｇｒｏｓｓ、「３Ｄビデオレコーダ」、パシフィックグラフィックス０２議事録、３２５～３３４ページ、２００２年１０月９～１１日

【非特許文献３】ＩＳＯ／ＩＥＣ　ＪＴＣ１／ＳＣ２９／ＷＧ１１　Ｎ５８７７、「３ＤＡＶの応用および要件」、２００３年７月

【非特許文献４】ＩＳＯ／ＩＥＣ　ＪＴＣ１／ＳＣ２９／ＷＧ１１　Ｎ５８７８、「３ＤＡＶの探究に関する報告」、２００３年７月

【非特許文献５】Ｚ．Ｚｈａｎｇ、「カメラ較正のためのフレキシブルな新技法」、パターン分析および機械知能に関するＩＥＥＥ会報、２２（１１）：１３３０～１３３４、２０００年

【発明の開示】

【発明が解決しようとする課題】

【０００９】

したがって、所与のインスタンスにおいて多くのビデオストリームを有し、ユーザが視聴方向の選択およびカメラ制御に関与することを可能にするビデオを、効率的に取り込み視聴するためのシステムおよび方法が必要とされている。このシステムおよび方法は、その較正が高精度であるべきであり、効率的な圧縮技法を可能にすべきである。さらに、これらの圧縮技法は、様々な視聴体験の提示を容易にすべきである。ハードウェアも相対的に安価であるのが最適である。このようなシステムは、視聴者が様々な視聴体験に参加できるようにすべきであり、特殊効果を可能にすべきである。加えて、このシステムおよび方法は、計算が効率的であるべきであり、大量の画像オーディオデータならびにユーザ対話の処理に対して頑強であるべきである。

【００１０】

本発明は、このような状況に鑑みてなされたもので、その目的とするところは、新しいタイプのビデオ視聴サービスを含む、対話式マルチビュービデオのクライアントサービスのためのシステムおよび方法を提供することにある。

【課題を解決するための手段】

【００１１】
カメラの使用がより一般的になり、コンピュータ処理力がより強力になり、ネットワーク帯域幅がより広くなるにつれて、ユーザは、これらの利点を利用してよりリッチなマルチメディア体験を追及することを望む。さらに、手術やスポーツ選手権イベントなどいくつかの重要なイベントを、異なる視点およびアングルから包括的に取り込むことが非常に望ましい。
【００１２】
前に論じたシングルビュービデオ形式の自然な拡張が、本発明のマルチビュービデオ形式である。マルチビュービデオでは、あるイベントまたはイベント空間の複数のビデオが、様々な視点およびアングルで同時に取り込まれる。これらのマルチビュービデオは、圧縮され、送信され、記憶され、最後にユーザに送達される。本発明のマルチビュービデオの重要な特徴の１つは、ユーザがビデオの取込みを制御でき、様々な方向からのイベント視聴を選択できることである。
【００１３】
本発明は、新しいタイプのビデオ取込みシステムおよび方法、新しいタイプのビデオフォーマット、新しいタイプのオンラインおよびオフラインビデオ圧縮手順、新しいタイプのビデオサービスを含む、対話式マルチビュービデオのためのシステムおよび方法を対象とする。
【００１４】
この新しいタイプのビデオ取込みシステムは、ビデオカメラ、制御ＰＣ、サーバ、ネットワークコンポーネント、クライアントからなる。オーディオコンポーネントを使用して、関連する任意のオーディオを取り込むこともできる。複数のカメラ、一実施形態では何十台または何百台ものビデオカメラが、マスター－スレーブ構成で、イベント位置でのイベント取込みに割り振られる。これらのカメラは、１つまたは複数の制御ＰＣによって制御される。イベント空間におけるイベントは、これらのカメラによって様々な視点および方向から同時に取り込まれる。次いで、これらの取り込まれたビデオは、制御ＰＣ中で圧縮され、リアルタイムで１つまたは複数のサーバに送信される。次いで圧縮ビデオは、エンドユーザにリアルタイムで送達することもでき、あるいはビデオ間の空間相関および時間相関を利用してさらに圧縮することもできる。
【００１５】
本発明の一実施形態では、自動のパターンなし較正ツールを利用して、複数のカメラを較正する。画像点とパターン点との対応を用いるパターンベースの方法とは対照的に、パターンなし較正方法は、様々なビューからの画像点間の対応に基づく。
【００１６】
１つまたは複数のサーバは、制御ＰＣからビデオを受信し、次いでこれらをマルチビュービデオまたはビデオビームの形式に保存する。ビデオビームは、様々な視聴方向から同時に撮った同じイベントのビデオストリームのセットからなり、ユーザが任意の瞬間に視聴方向を選択することを可能にする。本発明の対話式マルチビュービデオの記憶方式は、大量のビデオデータと、複数のユーザが同時にビデオビームを効率的に検索することとをサポートする。本発明の一実施形態では、索引ファイル方式を生み出して検索を高速化する。この核となる技法は、索引ファイルを使用して、任意の時間インスタンスにおけるオーディオビデオビットストリームの検索を容易にすることである。
【００１７】
従来のアルゴリズムを使用することもできるが、新規なオンラインおよびオフライン圧縮手順が、本発明の対話式ビデオシステムおよび方法と共に利用される。オンライン圧縮手順は、リアルタイムのマルチビュービデオ取込みのために設計されている。この出力は、オンラインサービスのために直接使用することもでき、あるいは将来の処理（例えばオフライン圧縮および／または再生）のためにディスクに保存することもできる。オフライン圧縮アルゴリズムは、トランスコーディングプロセスで採用されて、符号化済みビットストリームがずっと効率的に圧縮される。その後、出力ビットストリームは記憶およびオ

フラインサービスのためにディスクに保存される。
【００１８】
ユーザは、１つまたは複数のサーバに接続して、ユーザがマルチビュービデオを対話式に受信できるようにする新しいタイプのサービスに加入することができる。従来の再生制御に加えて、ユーザは、カメラ位置および配向の制御を操作すること、視聴方向を選択すること、スイープ効果やフリーズ回転効果などの特殊効果を楽しむことなどができる。対話式マルチビュービデオは、イベント視聴におけるまったく新しい体験を提供する。
【００１９】
対話式マルチビュービデオは、メディアストリーミング、放送、通信で一般に使用されている現在のシングルビュービデオの自然な拡張である。対話式マルチビュービデオは、技術開発および顧客需要の傾向に合致する。対話式マルチビュービデオは、メディアプレーヤ、メッセージングシステム、ミーティングシステムなど、様々なメディア用途に対して強力な影響を有するであろう。
【００２０】
本発明の対話式マルチビュービデオシステムは、多くの利点を有する。この対話式マルチビュービデオシステムは、ビデオストリームの選択およびカメラの制御をユーザに提供し、これによりユーザは、任意の時間インスタンスにおける視聴方向を選択することができる。本発明のこの対話式マルチビュービデオシステムでは、従来のシステムとは異なり、前景と背景のオブジェクトを分類する必要はない。加えて、この対話式マルチビュービデオシステムにより、従来のビデオシステムよりも効率的な符号化が採用され、特殊効果の表現を容易にするよりリッチな機能が備わる。
【００２１】
上述した利益に加えて、本発明の他の利点も、添付の図面と共に後続の詳細な記述を読めば明らかになるであろう。
【００２２】
本発明の具体的な特徴、態様、利点は、以下の記述、添付の特許請求の範囲、添付の図面を考慮すればよりよく理解されるようになるであろう。
【発明を実施するための最良の形態】
【００２３】
以下、図面を参照して本発明を適用できる実施形態を詳細に説明する。
【００２４】
本発明の好ましい実施形態に関する以下の記述では、本明細書の一部をなす添付の図面を参照する。図面には、例示として、本発明を実施することのできる具体的な実施形態を示す。本発明の範囲を逸脱することなく、他の実施形態を利用することもでき、構造上の変更を加えることもできることを理解されたい。
【００２５】
１．０　例示的な動作環境
図１に、本発明を実施することのできる適したコンピューティングシステム環境の例１００を示す。コンピューティングシステム環境１００は、適したコンピューティング環境の一例にすぎず、本発明の使用または機能の範囲についてどんな限定を意味するものでもない。またコンピューティング環境１００は、この例示的な動作環境１００に示すコンポーネントのいずれか１つまたは組合せに関してどんな依存や要件を有するものとも解釈すべきではない。
【００２６】
本発明は、その他多くの汎用または専用コンピューティングシステム環境または構成でも機能する。本発明で使用するのに適するであろう周知のコンピューティングシステム、環境、および／または構成の例には、限定しないがパーソナルコンピュータ、サーバコンピュータ、ハンドヘルドデバイスまたはラップトップデバイス、マルチプロセッサシステム、マイクロプロセッサベースのシステム、セットトップボックス、プログラム可能な民生用電子機器、ネットワークＰＣ、ミニコンピュータ、メインフレームコンピュータや、

これらのシステムまたはデバイスのいずれかを含む分散コンピューティング環境などが含まれる。

【００２７】

本発明は、プログラムモジュールなど、コンピュータによって実行されるコンピュータ実行可能命令の一般的なコンテキストで述べることができる。一般に、プログラムモジュールは、特定のタスクを実施するか特定の抽象データ型を実現するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。本発明は分散コンピューティング環境で実施することもでき、その場合、タスクは通信ネットワークを介してリンクされたリモート処理デバイスによって実施される。分散コンピューティング環境では、プログラムモジュールは、メモリ記憶デバイスを含めたローカルとリモートの両方のコンピュータ記憶媒体に位置することができる。

【００２８】

図１を参照すると、本発明を実施するための例示的なシステムは、コンピュータ１１０の形の汎用コンピューティングデバイスを含む。コンピュータ１１０のコンポーネントには、限定しないがプロセッサ１２０と、システムメモリ１３０と、システムメモリを含めた様々なシステムコンポーネントをプロセッサ１２０に結合するシステムバス１２１とを含めることができる。システムバス１２１は、様々なバスアーキテクチャのいずれかを用いた、メモリバスまたはメモリコントローラ、周辺バス、ローカルバスを含めて、いくつかのタイプのバス構造のいずれかとすることができる。限定ではなく例として、このようなアーキテクチャには、ＩＳＡ（Ｉｎｄｕｓｔｒｙ Ｓｔａｎｄａｒｄ Ａｒｃｈｉｔｅｃｔｕｒｅ）バス、ＭＣＡ（Ｍｉｃｒｏ Ｃｈａｎｎｅｌ Ａｒｃｈｉｔｅｃｔｕｒｅ）バス、ＥＩＳＡ（Ｅｎｈａｎｃｅｄ ＩＳＡ）バス、ＶＥＳＡ（Ｖｉｄｅｏ Ｅｌｅｃｔｒｏｎｉｃｓ Ｓｔａｎｄａｒｄｓ Ａｓｓｏｃｉａｔｉｏｎ）ローカルバス、ＰＣＩ（Ｐｅｒｉｐｈｅｒａｌ Ｃｏｍｐｏｎｅｎｔ Ｉｎｔｅｒｃｏｎｎｅｃｔ）バス（メザニンバスとも呼ばれる）が含まれる。

【００２９】

コンピュータ１１０は通常、様々なコンピュータ可読媒体を備える。コンピュータ可読媒体は、コンピュータ１１０からアクセスできる任意の利用可能な媒体とすることができ、揮発性と不揮発性の媒体、取外し可能と取外し不可能の媒体の両方がこれに含まれる。限定ではなく例として、コンピュータ可読媒体には、コンピュータ記憶媒体および通信媒体を含めることができる。コンピュータ記憶媒体には、コンピュータ可読命令、データ構造、プログラムモジュール、その他のデータなどの情報を記憶するための任意の方法または技術で実現された、揮発性と不揮発性、取外し可能と取外し不可能の両方の媒体が含まれる。コンピュータ記憶媒体には、限定しないがＲＡＭ、ＲＯＭ、ＥＥＰＲＯＭ、フラッシュメモリまたはその他のメモリ技術、ＣＤ－ＲＯＭ、デジタル多用途ディスク（ＤＶＤ）またはその他の光ディスク記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置またはその他の磁気記憶デバイスが含まれ、あるいは、所望の情報を記憶するのに使用できコンピュータ１１０からアクセスできるその他の任意の媒体が含まれる。通信媒体は通常、コンピュータ可読命令、データ構造、プログラムモジュール、またはその他のデータを、搬送波やその他のトランスポート機構などの被変調データ信号に組み入れるものであり、任意の情報送達媒体がこれに含まれる。用語「被変調データ信号」は、信号中の情報が符号化される形で１つまたは複数の特性が設定または変更される信号を意味する。限定ではなく例として、通信媒体には、有線ネットワークや直接有線接続などの有線媒体と、音響、無線周波数、赤外線などの無線媒体およびその他の無線媒体とが含まれる。以上のいずれかの組合せもコンピュータ可読媒体の範囲に含めるべきである。

【００３０】

システムメモリ１３０は、読取り専用メモリ（ＲＯＭ）１３１やランダムアクセスメモリ（ＲＡＭ）１３２など、揮発性および／または不揮発性メモリの形のコンピュータ記憶媒体を含む。ＲＯＭ１３１には通常、起動中などにコンピュータ１１０内の要素間で情報を転送するのを助ける基本ルーチンを含むＢＩＯＳ（ｂａｓｉｃ ｉｎｐｕｔ／ｏｕｔｐ

ｕｔ　ｓｙｓｔｅｍ）１３３が記憶されている。ＲＡＭ１３２は通常、プロセッサ１２０がすぐにアクセス可能な、かつ／またはプロセッサ１２０が現在作用している、データおよび／またはプログラムモジュールを含む。限定ではなく例として、図１には、オペレーティングシステム１３４、アプリケーションプログラム１３５、その他のプログラムモジュール１３６、プログラムデータ１３７を示す。

【００３１】

コンピュータ１１０は、その他の取外し可能／取外し不可能、揮発性／不揮発性コンピュータ記憶媒体を備えることもできる。例にすぎないが図１には、ノンリムーバブル不揮発性の磁気媒体に対して読み書きするハードディスクドライブ１４１と、リムーバブル不揮発性の磁気ディスク１５２に対して読み書きする磁気ディスクドライブ１５１と、ＣＤ　ＲＯＭやその他の光媒体などリムーバブル不揮発性の光ディスク１５６に対して読み書きする光ディスクドライブ１５５を示す。この例示的な動作環境で使用できるその他の取外し可能／取外し不可能、揮発性／不揮発性コンピュータ記憶媒体には、限定しないが磁気テープカセット、フラッシュメモリカード、デジタル多用途ディスク、デジタルビデオテープ、固体ＲＡＭ、固体ＲＯＭなどが含まれる。ハードディスクドライブ１４１は通常、インタフェース１４０などの取外し不可能メモリインタフェースを介してシステムバス１２１に接続され、磁気ディスクドライブ１５１および光ディスクドライブ１５５は通常、インタフェース１５０などの取外し可能メモリインタフェースでシステムバス１２１に接続される。

【００３２】

以上に論じ図１に示したドライブおよびそれらに関連するコンピュータ記憶媒体は、コンピュータ可読命令、データ構造、プログラムモジュール、その他のデータの記憶域をコンピュータ１１０に提供する。例えば図１には、ハードディスクドライブ１４１がオペレーティングシステム１４４、アプリケーションプログラム１４５、その他のプログラムモジュール１４６、プログラムデータ１４７を記憶しているのが示されている。これらのコンポーネントは、オペレーティングシステム１３４、アプリケーションプログラム１３５、その他のプログラムモジュール１３６、プログラムデータ１３７と同じものとすることもでき、異なるものとすることもできることに留意されたい。ここでは、オペレーティングシステム１４４、アプリケーションプログラム１４５、その他のプログラムモジュール１４６、プログラムデータ１４７が少なくとも異なるコピーであることを示すために、異なる番号を付けてある。ユーザは、キーボード１６２、マウスやトラックボールやタッチパッドと一般に呼ばれるポインティングデバイス１６１などの入力デバイスを介して、コンピュータ１１０にコマンドおよび情報を入力することができる。その他の入力デバイス（図示せず）には、マイクロホン、ジョイスティック、ゲームパッド、衛星受信アンテナ、スキャナなどを含めることができる。これらおよびその他の入力デバイスは、システムバス１２１に結合されたユーザ入力インタフェース１６０を介してプロセッサ１２０に接続されることが多いが、パラレルポート、ゲームポート、ユニバーサルシリアルバス（ＵＳＢ）など、その他のインタフェースおよびバス構造で接続されてもよい。モニタ１９１または他のタイプの表示デバイスも、ビデオインタフェース１９０などのインタフェースを介してシステムバス１２１に接続される。モニタに加えて、コンピュータは通常、スピーカ１９７やプリンタ１９６など、その他の周辺出力デバイスも備えることができ、これらは出力周辺インタフェース１９５を介して接続することができる。本発明にとって特に重要なのは、一連の画像１９３を取り込むことのできるカメラ１９２（デジタル／電子のスチルまたはビデオカメラ、あるいはフィルム／写真スキャナなど）も、パーソナルコンピュータ１１０への入力デバイスとして含めることができることである。さらに、１つのカメラだけが示してあるが、複数のカメラをパーソナルコンピュータ１１０への入力デバイスとして含めることもできる。１つまたは複数のカメラからの画像１９３は、適切なカメラインタフェース１９４を介してコンピュータ１１０に入力される。このインタフェース１９４はシステムバス１２１に接続され、それにより、画像がＲＡＭ１３２に、またはコンピュータ１１０に関連するその他のデータ記憶デバイスの１つにルーティングされて

記憶されるようにする。ただし画像データは、カメラ１９２の使用を必要とせずに、前述のコンピュータ可読媒体のいずれかからコンピュータ１１０に入力されてもよいことに留意されたい。オーディオデータを取り込むために、オーディオレコーダ１９８もオーディオインタフェースデバイス１９９を介してコンピュータに接続することができる。

【００３３】

コンピュータ１１０は、リモートコンピュータ１８０など１つまたは複数のリモートコンピュータへの論理接続を用いて、ネットワーク化された環境で動作することができる。リモートコンピュータ１８０は、パーソナルコンピュータ、サーバ、ルータ、ネットワークＰＣ、ピアデバイス、またはその他の一般的なネットワークノードとすることができ、通常はコンピュータ１１０に関して上述した要素の多くまたはすべてを備えるが、図１にはメモリ記憶デバイス１８１だけが示してある。図１に示す論理接続は、ローカルエリアネットワーク（ＬＡＮ）１７１およびワイドエリアネットワーク（ＷＡＮ）１７３を含むが、その他のネットワークを含むこともできる。このようなネットワーキング環境は、オフィス、企業全体のコンピュータネットワーク、イントラネット、インターネットでよく見られる。

【００３４】

ＬＡＮネットワーキング環境で使用されるときは、コンピュータ１１０は、ネットワークインタフェースまたはアダプタ１７０を介してＬＡＮ１７１に接続される。ＷＡＮネットワーキング環境で使用されるときは、コンピュータ１１０は通常、インターネットなどのＷＡＮ１７３を介した通信を確立するためのモデム１７２またはその他の手段を備える。モデム１７２は内蔵でも外付けでもよく、ユーザ入力インタフェース１６０またはその他の適切な機構を介してシステムバス１２１に接続することができる。ネットワーク化された環境では、コンピュータ１１０に関して示したプログラムモジュールまたはその一部をリモートのメモリ記憶デバイスに記憶することができる。限定ではなく例として、図１には、リモートアプリケーションプログラム１８５がメモリデバイス１８１上にあるのが示されている。図示のネットワーク接続は例示的なものであり、コンピュータ間で通信リンクを確立するための他の手段を使用してもよいことは理解されるであろう。

【００３５】

例示的な動作環境について論じたが、この記述セクションの残りの部分は、本発明を組み入れたプログラムモジュールに関する記述に充てる。

【００３６】

２．０　対話式マルチビュービデオのためのシステムおよび方法

以下の各セクションで、本発明によるシステムおよび方法について詳細に述べる。この対話式マルチビュービデオのシステムは、３つの主要部分、すなわち取込みコンポーネント、サーバコンポーネント、クライアントコンポーネントからなる。

【００３７】

２．１　取込みコンポーネント

本発明の対話式マルチビューカメラシステムの取込みコンポーネント２０２は、カメラ（例えばビデオカメラ）、レンズ、パンチルトヘッド、制御ＰＣ、同期ユニットを備える。図２に示すように、本発明の一実施形態では、２つのビデオカメラ２０４ａ、２０４ｂ（それぞれ、それ自体のパンチルトヘッド２０６ａ、２０６ｂおよびレンズ（例えばズームレンズ）２０８ａ、２０８ｂを有する）が、１つの制御ＰＣ２１０および１３９４ポート（図示せず）にそれぞれ接続される。各カメラは、それ自体のＩＤ番号を有する。制御ＰＣ２１０は、パンチルトヘッド２０６およびレンズ２０８を例えばＲＳ２３２インタフェースを介して制御することによって、カメラの視点およびアングルを変更することができる。同期ユニット２１４は、１つまたは複数のＰＣ２１０に、好ましくはそれらの１３９４ポートまたは他の適した手段を介してリンクされる。本システムの取込みコンポーネントは、特定位置の任意のオーディオを録音するオーディオ録音機器２０９を備えることもできる。

【００３８】

同期ユニット２１４は、すべてのカメラが同じ瞬間にトリガして撮影するようにするために使用される。それにより制御ＰＣは、各カメラから同時にビデオを取り込むことができる。これらすべてのカメラから、１つがマスタカメラとして選択され、残りはスレーブカメラと呼ばれる。マスタカメラはカメラマンによって制御され、スレーブカメラは、マスタカメラと同じ関心点を向くように駆動させることができる。これは、いわゆるマスター－スレーブ追跡プロセスによって実現される。通常、カメラマンは人間である。しかし場合によっては、マスタカメラは、実際のカメラマンからのコマンドなしにオブジェクト追跡アルゴリズムによって制御することもできる。

【００３９】

制御コマンドが、マスタカメラの制御ＰＣに入力される。パンチルトパラメータが計算され、他の制御ＰＣに送信されて、すべてのスレーブカメラが駆動される。取り込まれたビデオは、制御ＰＣによって受信され、圧縮され、サーバに送信される。本発明の一実施形態では、各ビデオは、６４０×４８０のサイズで、毎秒３０フレームのフレームレートで取り込まれる。本発明の一実施形態で使用される詳細なオンライン圧縮手順については、セクション３．１で提示する。

【００４０】

２．１．１　カメラ較正

マスター－スレーブ追跡の前に、カメラを較正すべきである。本発明のマルチビュービデオシステムでは、内部パラメータと、外部パラメータと、ハンド－アイ関係とを決定する較正プロセスを利用する。図３に、このプロセスの全体的なフローチャートを示す。最初に内部カメラパラメータを計算し（プロセス動作３０２）、続いて外部カメラパラメータを決定する（プロセス動作３０４）。次いで、ハンド－アイパラメータを決定する（プロセス動作３０６）。最後に、決定した内部、外部、ハンド－アイパラメータを使用して、すべてのカメラの外部パラメータを共通座標系で調整することによってカメラを較正する。これらのすべてのパラメータと、マスタカメラのパンチルトパラメータとが与えられれば、スレーブカメラをマスタカメラと同じ関心点に向けるスレーブカメラのパンチルトパラメータを、効率的に計算し調整することができる。

【００４１】

内部パラメータは、基本的なピンホールカメラモデルを使用して定義される。内部パラメータは、カメラの内部構造だけに依存する。内部パラメータには、１つの画像ピクセルの焦点距離と幅との比率、１つの画像ピクセルの焦点距離と高さとの比率、主点のｘ座標、主点のｙ座標が含まれる。外部パラメータは、カメラの内部構造に依存しない。外部パラメータは、既知の世界基準系に対するカメラ基準系の位置および配向を定義する。外部パラメータには通常、回転行列および３Ｄ平行移動ベクトルが含まれる。ハンド－アイ関係パラメータには、各カメラのパンチルトヘッドに対するカメラの位置および配向が含まれる。

【００４２】

本発明のマルチビュー対話式ビデオシステムおよび方法では、パターンベースの較正とパターンなし較正の２つの較正方法を採用する。パターンベースの較正は、好ましくは地面またはその他の適した基準面に配置された、大きな較正パターンを使用して実現され、パターンなし較正は、地面からもたらされる情報を利用する。以下、この２つの方法をより詳細に述べる。

【００４３】

２．１．２　パターンベースの較正

本発明の一実施形態では、平面ベースのアルゴリズム（非特許文献５参照）を、その精度および単純さゆえに使用して、内部パラメータを較正する。内部パラメータの変化はごくわずかなので、このような較正は何週間かに１度実施するだけでよい。すべてのカメラの外部パラメータは、共通の世界座標系で、好ましくはパターン面の座標系で較正する。次いで、各カメラのハンド－アイ関係も、３つ以上のパンチルト位置でその外部パラメータから較正する。

【００４４】
パターンベースの方法は、精密にわかっている幾何形状を有する平面パターンの画像を使用する。パターンベースの較正を自動にするために、本発明の一実施形態では、図４Ａに示す特別な較正パターンを設計した。この較正パターンは、３種類の色（赤、緑、青）を使用して、すべてのコーナ点の位置を符号化している。様々なパンチルト運動が行われているカメラによってパターンの画像を取り込み、次いで、色で符号化された位置と共にパターンのコーナを検出するように、自動手順を設計した。
【００４５】
図４Ｂに、パターンベースの較正を単純化した流れ図を示す。パターンを地面または他の適した基準系の上に配置し、パターンのコーナ、および場合によっては他の基準点を、既知の位置に配置する（プロセス動作４０２）。次いで、すべてのカメラが較正パターンの画像を取り込む（プロセス動作４０４）。画像から取り込まれた特徴点と、既知の座標にある基準パターン点との対応を見つけて使用することにより、従来の技法を用いて外部カメラパラメータを精密に推定することができる（プロセス動作４０６）。正確な較正を得るために、基準パターンは、精密に製造すべきであり、較正に使用される画像の大部分を占めるべきである。さらに、大規模なシステムでは、高精度な大きい基準パターンをセットアップするのは些細な作業ではなく、特別な機器を必要とする。この不便を回避するために、パターンなし較正方法を開発した。これについて以下に述べる。
【００４６】
２．１．３　パターンなし較正
２．１．３．１　パターンなし較正手順の概観
本発明の一実施形態では、自動のパターンなし較正ツールを利用する。画像点とパターン点との対応を使用してカメラの外部パラメータを決定するパターンベースの方法とは対照的に、パターンなし較正方法は、様々なカメラからの画像点の間の対応に基づく。図５に、本発明の対話式マルチビュービデオシステムのパターンなし較正手順の全体的な流れ図を提供する。最初に、プロセス動作５０２に示すように、マスタカメラとスレーブカメラの両方の各画像中で特徴点を抽出する。これらの特徴点を使用して、各画像中の特徴をマスタカメラの画像にマッピングする画像間ホモグラフィのセットを推定する（プロセス動作５０４）。次いで、プロセス動作５０６および５０８に示すように、好ましくは特異値分解（ＳＶＤ）演算を使用して、これらのホモグラフィに基づいて外部パラメータの線形解を得ることができる。ＳＶＤは、行列の固有値および固有ベクトルを見つけるのに使用することのできる古典的な数学演算である。本発明で使用される方法では、ＳＶＤを使用して、特徴点のホモグラフィとその転置との積行列の、固有値およびそれらに対応する固有ベクトルを見つける。得られたこれらの固有成分に基づいて、カメラの外部パラメータを、一次方程式のセットに対する最小２乗解として推定することができる。この後、プロセス動作５１０に示すように、外部カメラパラメータのバンドル調整を適用して、すべての特徴対応の再投影誤差の合計を最小化することによって外部カメラパラメータを精緻化する。推定された外部パラメータを使用して、マスタ画像（例えばマスタカメラによって撮られたもの）中の特徴をスレーブ画像（例えばスレーブカメラによって撮られたもの）上に投影することができる。用語「再投影誤差」は、スレーブ画像上に投影された特徴と、それらに対応するマスタ画像中の特徴との間の誤差を指す。投影誤差の合計を使用することは、較正されたパラメータの精度を評価するための好都合な方法である。本発明の一実施形態では、推定されたパラメータは、レベンベルグ－マーカート（ＬＭ、Ｌｅｖｅｎｂｅｒｇ－Ｍａｒｑｕａｒｄｔ）法を使用して投影誤差の合計を最小化することによって精緻化される。
【００４７】
２．１．３．２　ホモグラフィ推定
本発明のパターンなし較正技法は、以下のようにより具体的に述べることができる。ほとんどの環境では常に、優勢な面、通常は地面がある。このようなシナリオで複数のカメラをセットアップするとき、各カメラは、優勢な面などの共通面の画像を形成する。例え

ば、地面を見る別々の位置にある２つのカメラ（一方はマスタカメラ、他方はスレーブ）からの２つの画像が、以下の式で定義される３×３ホモグラフィＨによってリンクされる。

【００４８】

【数１】

$$\mathbf{H} \fallingdotseq \mathbf{A}_2(\mathbf{R} + \frac{\mathbf{t}\mathbf{n}^T}{d})\mathbf{A}_1^{-1}$$

【００４９】

上式で、$A_1$および$A_2$は、それぞれマスタカメラおよびスレーブカメラの内部行列である。記号　は、０でないスケールまで等しいことを示す。というのは、ホモグラフィはスケールまでしか推定できないからである。Ｒおよびｔは、マスタの基準座標系におけるスレーブカメラの外部パラメータ（回転および平行移動）であり、ｎは地面の単位法線ベクトルである。

【００５０】

２つの画像（同一直線上にない）の間で５つ以上の点対応がある場合、ホモグラフィを推定できる様々な従来技法がある。例えば、ホモグラフィは、直接線形変換（ＤＬＴ、Ｄｉｒｅｃｔ　Ｌｉｎｅａｒ　Ｔｒａｎｓｆｏｒｍ）という名称の基本的なコンピュータビジョンアルゴリズムによって推定することができる。本発明の一実施形態では、ランダムサンプルコンセンサス（ＲＡＮＳＡＣ、Ｒａｎｄｏｍ　Ｓａｍｐｌｅ　Ｃｏｎｓｅｎｓｕｓ）技法を利用してホモグラフィを推定する。この方法は、次の５つのステップからなる。

【００５１】

１．特徴点を検出する。一実施形態では、コーナ検出演算子を使用して２つの画像から特徴を検出する。

【００５２】

２．特徴点の周りの画像間強度類似性を利用して、対応する特徴セットの仮説を得る。

【００５３】

３．ＲＡＮＳＡＣアルゴリズムによってホモグラフィを初期化する。

【００５４】

４．レベンベルグ－マーカートアルゴリズムによって、ホモグラフィを精緻化して、対応するすべての特徴対における再投影誤差を最小化する。

【００５５】

５．推定されたホモグラフィを使用して、より多くの対応する特徴対を見つける。ここで、ステップ４および５を何回か反復してホモグラフィを改善することができる。

【００５６】

ホモグラフィが得られた後は、以下のプロセスによってカメラの外部パラメータを線形推定することができる。

【００５７】

２．１．３．３　外部パラメータの決定

ホモグラフィＨについて、

【００５８】

【数２】

$$\mathbf{A}_2^{-1}\mathbf{H}\mathbf{A}_1$$

【００５９】

をＭで示し、Ｍの固有値を$v_j$で示す（ｊ＝１，２，３）。Ｈの特性に従って、ｎに関する３つの式を確立することができる。

【００６０】

【数３】

$$\begin{cases} v_1^T n = 0 \\ v_2^T n = sign(b_2)|b_2|(a_2+1) \\ v_3^T n = sign(b_3)|b_3|(a_3+1) \end{cases}$$

【００６１】

上式で、$b_j$および$a_j$は２つの中間変数であり、$|b_j|$および$a_j$の値はＭの固有値から得られる。このことは、１つの画像間ホモグラフィから、未知の符号のｎに関する３つの式が得られることを意味する。マスタカメラを含むｍ＋１個のカメラによって取り込まれた平面場面の画像がｍ＋１個ある場合、マスタ画像から他の画像へのｍ個のホモグラフィを推定することができる。次いで、各Ｍからの固有値および固有ベクトルをさらに決定することができる。これらに基づいて、上記の制約は、３ｍ個の一次方程式のセットを構成することができる。これは、法線ベクトルｎを推定するための可能な方法の１つを提示する。実際には、初期化ステップによってｎの初期値を得ることができ、次いで上式における符号を決定することができる。これに基づいて、さらにｎを推定することができる。本発明の一実施形態によれば、１つのホモグラフィから２つの可能な解が得られるので、ボーティング（ｖｏｔｉｎｇ）ベースの初期化ステップを採用して$b_j$の符号を決定する。

【００６２】

より具体的には、全体的な手順は次のように述べることができる。

【００６３】

ステップ１。画像を獲得し、特徴点を検出し、従来の方法でまたは前述のようにホモグラフィＨを推定する。

【００６４】

ステップ２。標準的なＳＶＤ分解演算によって$M^TM$の固有値および固有ベクトルを計算する。

【００６５】

ステップ３。ボーティング法によって法線ベクトルｎの初期値を推定する。

【００６６】

ステップ４。式における符号を決定し、次いでベクトルｎを精緻化する。

【００６７】

ステップ５。平行移動ｔ（スケールまで）および回転Ｒを推定する。

【００６８】

ステップ６。すべての特徴対応の再投影誤差の合計を最小化することによって、外部カメラパラメータをバンドル調整する。

【００６９】

２．２　サーバコンポーネント

サーバは、対話式マルチビュービデオシステム中で最も強力なユニットである。サーバは、大量のビデオデータの伝送および記憶を管理し、多くのクライアントにサービスを提供する。図２に示すように、サーバ２１６は２つのネットワーク２１８、２２０に接続される。ネットワーク２１８は、例えば広帯域ネットワークバックボーンなどであり、圧縮ビデオを制御ＰＣ２１０からサーバ２１６に送達できるようサーバ２１６と制御ＰＣ２１０とを接続するために採用される。本発明の一実施形態では、本発明のマルチビュービデオシステムは、１ＧＢネットワークを使用してサーバ２１６とすべての制御ＰＣ２１０とを接続する。外部ネットワーク２２０（例えばＬＡＮ、ＷＡＮ、さらにはインターネット）を使用して、サーバ２１６をクライアント２２２に接続する。本発明の一実施形態では、クライアント２２２は、１０／１００ＭＢまたはそれ以上のネットワークを介してサーバ２１６に接続される。本発明の別の実施形態では、クライアント２２２は、インターネットを介してサーバ２１６に接続される。

【００７０】

２．２．１　マルチビュービデオフォーマット

サーバ２１６は、制御ＰＣ２１０からビデオを受信し、次いでこれらをマルチビュービデオまたはビデオビームの形式に保存する。ビデオビームは、同時に撮った同じイベントまたはイベント空間のビデオと好ましくはオーディオとのストリームのセットからなる。本発明の対話式マルチビュービデオの記憶方式は、大量のビデオデータと、ビデオビームの効率的な検索とをサポートする。本発明の一実施形態では、索引構造を生み出して検索を高速化する。本発明のマルチビュービデオは、膨大なビデオビームを維持することができ、また、多数のユーザが同時にビームにアクセスするのをサポートすることができる。この核となる技法は、索引を使用して、任意の時間インスタンスにおけるオーディオビデオビットストリームの検索を容易にすることである。図６Ａおよび６Ｂに、これらの索引構造の例を示す。図６Ａには、ビデオビットストリーム６０２のフォーマットを示し、図６Ｂには、ビデオビットストリームに対応するオーディオビットストリーム６０４のフォーマットを示す。実際のビデオオーディオデータは、索引ファイルと共にサーバに記憶されることが多い。これらはまた、オフライン再生のためにクライアントでローカルに記憶してもよい。例えば、ビデオビームをＤＶＤディスクに記憶し、クライアント位置にある任意のＰＣで再生することができる。

【００７１】

マルチビュービデオのサイズは非常に大きい場合があるので、本発明の一実施形態では、６４ビットポインタを使用して任意の圧縮マルチビュービデオフレームの開始点を表す。一方、任意の圧縮オーディオフレームの開始点を表すには、３２ビットポインタを使用すれば十分である。さらに、ビデオビットストリームを突き止める時間消費を短縮し、ビデオ索引ファイルのサイズを縮小するために、６４ビットポインタを３２ビットの高アドレスポインタと３２ビットの低アドレスポインタとに分割する。フラグ（例えば「ｂＣｒｏｓｓ４Ｇ」という名称のフラグ）を使用して、高アドレスポインタ中の移行があるか否かを表す。フラグが「真」にセットされている場合は、低アドレスをチェックすべきである。その場合、現在の低アドレスの値が前の低アドレスの値よりも小さければ、現在のポインタから残りのポインタについては高アドレスを１つ増加させるべきである。

【００７２】

オーディオとビデオの索引は、異なるファイルに別々に保存される。ビデオ索引ファイルは階層構造に構成される。第１の層は多くのフィールド６０６（例えば「ＶｉｄｅｏＩｎｄｅｘＩｎｆｏＨｅａｄｅｒ」フィールド）からなり、各フィールドは、タイムスタンプと、ビデオ索引データのオフセットと、３２ビット高アドレスと、高アドレスポインタへの移行があるか否かを示すフラグ（例えば「ｂＣｒｏｓｓ４Ｇ」フラグ）と、その瞬間に利用されたカメラの数とを含む。第２の層は、図６Ａに示すように、第１の層６０８によってポイントされているのと同じタイムスタンプを有する詳細なビデオ索引データ６１０（例えば「ＶｉｄｅｏＩｎｄｅｘ」フィールド）を含む。第２の層の各フィールドは、カメラＩＤと、このフレームの符号化タイプと、３２ビット低アドレスポインタとからなる。あるタイムスタンプに対する「ＶｉｄｅｏＩｎｄｅｘ」フィールドの数は、「ＶｉｄｅｏＩｎｄｅｘＩｎｆｏＨｅａｄｅｒ」フィールド中の「ｂｙＣａｍｅｒａＮｕｍ」で表されるカメラ総数と等しいことに留意されたい。また、異なるタイムスタンプにおけるカメラの数は異なる可能性があることにも留意されたい。

【００７３】

ビデオ索引の構造の例を以下に示す。

【００７４】

【数４】

```
// first layer

Struct VideoIndexInfoHeader
{
        DWORD       dwTimeStamp; // time stamp of multi-view video
        DWORD       dwOffset;    // 32-bits offset of the VideoIndexHeader
        DWORD       dwOffsetHigh; // the high-address of the offset
        BOOL        bCross4G;    // indicate whether the offsets have the same
dwOffsetHigh or not
        BYTE        byCameraNum; // total number of cameras at that time stamp
}

//second layer
Struct VideoIndex
{
        BYTE        byCameraID;  // the ID of camera, maximum 255
        BYTE        byFrameType; // coding type of the video frame
        DWORD       dwOffsetLow; // the low-address pointer
}
```

【００７５】

オーディオ索引ファイル６０４も階層構造に構成される。第１の層は多くのフィールド６１４（例えば「ａｕｄｉＩｎｄｅｘＩｎｆｏＨｅａｄｅｒ」）からなり、各フィールドは、タイムスタンプと、オーディオ索引データのオフセットと、その瞬間のオーディオ録音の数とを含む。第２の層６１６は、図６Ｂに示すように、同じタイムスタンプを有する詳細なオーディオ索引データ（例えば「ＡｕｄｉｏＩｎｄｅｘ」フィールド）を含む。あるタイムスタンプに対する「ＡｕｄｉｏＩｎｄｅｘ」フィールドの数は、「ＡｕｄｉｏＩｎｄｅｘＩｎｆｏＨｅａｄｅｒ」フィールド中の「ＡｕｄｉｏＮｕｍ」で表されるオーディオストリーム総数と等しいことに留意されたい。また、異なるタイムスタンプにおけるオーディオストリームの数は異なる可能性があることにも留意されたい。

【００７６】

オーディオ索引の構造の例を以下に示す。

【００７７】

【数５】

```
// first layer

Struct AudioIndexInfoHeader
{
        DWORD       dwTimeStamp; // time stamp of multi-view video
        DWORD       dwOffset;    // 32-bits offset of the AudioIndexHeader
        BYTE        byAudioNum   // total number of audios at that time stamp
}

//second layer
Struct AudioIndex

{
        BYTE        byAudioID;   // the ID of audio, maximum 255
        DWORD       dwOffset;    // the 32-bits pointer
}
```

【００７８】

２．３　クライアントコンポーネント

受信したビデオビームは、オンライン対話サービスのために直接使用することもでき、あるいはオフライン処理のために保存することもできる。本発明によるシステムおよび方法の一実施形態のコンテキストでは、オンラインは、視聴されるビデオビームがリアルタイムで取り込まれることを意味する。オフラインは、ビデオビームが取り込まれて記憶媒体に記憶されたことを意味する。オフライン再生には２つのタイプがある。一方のタイプは、例えばビデオオンデマンド（ＶＯＤ）で行われるように、ビデオビームがサーバで記憶され、クライアントがそれをストリーミングプロセスによって再生するものである。このモードでは、サーバはストリーミングサーバとして働く。したがって、このタイプのオフライン再生は「ストリーミングサービス」と呼ばれる。他方のタイプのオフライン再生は、ビデオビームがローカルディスクまたは別の位置に記憶されているときに行われる。このモードでは、クライアントは、ビデオビームをサーバの助けなしに再生することができる。

【００７９】

オンライン対話式サービスの場合、サーバはクライアントからのユーザコマンドに応答する。本発明の例示的な一実施形態でサポートされるコマンドには、ＶＣＲなど通常のメディアプレーヤにおける従来のコマンドに加えて、切換え、スイープ、フリーズおよび回転、履歴閲覧が含まれる。ユーザコマンドに従って、サーバは、取り込まれたビデオからビデオストリームを生成し、次いでこれをクライアントに送信する。本発明の一実施形態では、１つのクライアントに対して２つの通信チャネルがある。一方はユーザデータグラムプロトコル（ＵＤＰ）チャネルであり、これは、レイテンシ（latency）を短縮するためにオーディオ／ビデオデータの送信に使用される。他方は伝送制御プロトコル（ＴＣＰ）チャネルであり、これは、正確さを保証するために、取込みカメラを制御するためのコマンドおよび制御データの送信に使用される。オフライン処理の場合、ビデオビームをトランスコードしてデータ量をさらに削減する。詳細なオフライン圧縮手順はセクション３．２に提示する。以下、クライアントコンポーネントの詳細について論じる。

【００８０】

２．３．１　オンラインサービス

オンラインサービスでは、クライアントはＬＡＮ、ＷＡＮ、さらにはインターネットにおいてサーバにリモート接続することができる。クライアントとサーバの間の接続が確立されると、ユーザはクライアント部分で、通常のメディアプレーヤにおけるような従来のコマンドに申し込むことができ、また、対話式マルチビューにおける固有のコマンド（例えば切換え、スイープ、フリーズおよび回転、履歴閲覧など）を発行する能力にも申し込むことができる。

【００８１】

クライアントは、自分のコマンドをサーバに送信する。ユーザのコマンドに応答して、サーバは、期待されるビデオをユーザのコマンドに従って生成し、各クライアントにそれぞれ送信する。一言で言えば、ユーザは、マルチビュービデオを対話式で再生することができる。場合によっては、ユーザは、カメラＩＤやパンチルト値などのパラメータをクライアントに入力することもできる。クライアントは、これらのパラメータをサーバに送信し、次いで制御ＰＣに送信して、取込みカメラを制御することができる。

【００８２】

２．３．２　オフラインサービス

オフライン再生では、クライアントは、ローカルディスクまたは別の位置に記憶されたマルチビュービデオビームを直接開いて再生することができる。通常のビデオプレーヤにおける従来の効果（例えば再生、早送り、巻戻し、一時停止、停止など）に加えて、ユーザは、例えば、異なるビデオストリーム間での切換え、スイープ効果、フリーズおよび回転効果を含めて、いくつかの凝った特殊効果を体験することもできる。これらの特殊効果に関する簡単な記述を以下に提供する。

【００８３】
ストリーミングモードでは、クライアントは、オンラインモードと同様にＬＡＮ、ＷＡＮ、さらにはインターネットを介してサーバにリモート接続することができる。このモードでは、サーバコンポーネントは、クライアントの接続およびビデオビームを管理するストリーミングサーバとして働き、ユーザは、自分のコマンドをサーバに申し込んで、ビデオビームから自分の望むコンテンツを選択し、様々なビデオ効果（例えば切換え、スイープ、フリーズおよび回転、履歴閲覧、スクリプト）を見ることができる。このモードは、現在のビデオオンデマンド（ＶｏＤ）システムの拡張である。ストリーミングサービスとオンラインサービスの主な違いは、ストリーミングモードではビデオビームが取り込まれてサーバコンポーネントに記憶済みであり、リアルタイムで取り込まれるのではないことである。ストリーミングサービスは、以下に挙げるユーザコマンドすべてをサポートする。

【００８４】
切換え効果：切換え効果は、ビデオが正しいテンポで継続している間にユーザがあるカメラ視点と別のカメラ視点との間で切り換えることができるものである。これは、所望の視点を提供する様々なカメラからのビデオストリームにアクセスすることを含む。一例は、ユーザが第２のカメラの視点から続けて第５のカメラの視点に切り換えることである。

【００８５】
スイープ効果：スイープ効果は、時間がそれまでどおり進んでいる間に隣接カメラビューにスイープするものである。これによりユーザは、様々な視点からイベントを見ることができる。一例として、合計８つの視点があるとすると、ユーザは第１の視点から開始し、継続的に第２の視点、第３の視点に切り換えて第８の視点まで同様にしてから、第８の視点を見る。

【００８６】
フリーズおよび回転効果：フリーズおよび回転効果では、時が静止し、カメラビューが所与の点の周りで回転する。一例として、合計８つの視点があるとすると、ユーザは第１の視点から開始し、継続的に第２、第３の視点に切り換え、以下同様にして第８の視点まで行ったり来たりする。

【００８７】
履歴効果：履歴効果では、ユーザは、前に視聴または作成されたビデオシーケンスを再生することができる。

【００８８】
スクリプト：ユーザは、オンデマンドで再生することのできるビューおよび特殊効果のセットのスクリプトを作成することもできる。ユーザはまた、このスクリプトを他のユーザに送信することもでき、他のユーザは、スクリプトが起動されると、スクリプトされたのと同じビデオイベントを見ることになる。

【００８９】
スイープ、切換え、フリーズおよび回転の効果は、オンラインモードでも利用可能にすることができる。

【００９０】
３．０　圧縮手順
本発明の対話式マルチビュービデオシステムおよび方法と共に、オンラインとオフラインの両方の圧縮手順を使用することができる。オンライン圧縮手順は、リアルタイムのマルチビュービデオ取込みのために設計されている。この出力は、オンラインサービスのために直接使用することもでき、あるいは将来の処理のために（例えばオフラインでさらに圧縮したり後で再生したりするために）ディスクに保存することもできる。オフライン圧縮手順は、トランスコーディングプロセスで採用されて、符号化済みビットストリームがずっと効率的に圧縮される。その後、出力ビットストリームは記憶およびオフラインサービスのためにディスクに保存される。

【００９１】

以下のセクションで、特定の新規なオンラインおよびオフライン圧縮手順について述べるが、本発明のシステムおよび方法はこれらのタイプの圧縮に限定されないことに留意されたい。従来の圧縮アルゴリズムを使用することもできる。

【００９２】

３．１　オンライン圧縮

概して、従来のシングルビュービデオ符号化と同様、本発明の対話式マルチビュービデオシステムの一実施形態で使用されるオンライン圧縮では、各ビデオビューをＩＰＰＰフレームのフォーマットで符号化することができる。

【００９３】

背景として、通常のビデオ圧縮は、インターフレーム（Ｐフレーム）圧縮とイントラフレーム（Ｉフレーム）圧縮の２つの基本的な圧縮技法を利用する。インターフレーム圧縮は、フレーム間であり、連続するピクチャのデータ冗長性（例えば時間冗長性）を最小限に抑えるようになっている。イントラフレーム圧縮は、個々のフレーム内で行われ、各ピクチャ中のデータの重複（例えば空間冗長性）を最小限に抑えるようになっている。従来のビデオ符号化では、イントラピクチャフレームは本質的にＪＰＥＧフォーマットでソース画像を符号化する（いくつかの違いはあるが）。通常、ピクセルのブロックは、離散コサイン変換（ＤＣＴ）にかけられ、マクロブロックベースで量子化される。イントラピクチャフレームは、他のどんなフレームにも依存せず、ランダムアクセスのための「ジャンプイン」ポイントとして使用される。インターフレームは、予測フレーム（Ｐフレーム）と呼ばれることもあり、前のＩまたはＰフレームを使用して現在のフレームの内容を「予測」し、次いで予測と実際のフレーム内容との差を圧縮する。予測は、前のフレーム中で現在のマクロブロックの位置に近い領域を見つけようとすることによって行われ、この領域は類似するピクセルを含む。前の予測領域を（通常は半ピクセル精度で）現在のマクロブロックに移動させる動きベクトルを計算する。動きベクトルは、動きがない場合は論理的には０ベクトルとすることができ、これは当然、非常に効率的に符号化される。予測ピクセルとそれらの実際の値との差を計算し、ＤＣＴ変換し、係数を量子化する（ＩフレームのＤＣＴ係数よりも粗く）。十分に類似するピクセルグループを前のフレーム中で見つけられなかった場合は、Ｐフレームは単純に、マクロブロックをＩフレームであるかのように空間符号化することができる。

【００９４】

従来のビデオ符号化と同様、本発明のオンライン圧縮アルゴリズムには、「Ｉ」フレームと「Ｐ」フレームの２つのタイプのフレームがある。各「Ｉ」フレームの圧縮は、そのフレームの相関だけに基づくが、「Ｐ」フレームの圧縮は、そのフレームとその前フレームとの相関に基づく。基本的に、「Ｐ」フレームの圧縮効率は、「Ｉ」フレームよりもずっと高い。「Ｉ」フレームは、効率的な圧縮をもたらすことはできないが、誤差に対して非常に頑強である。さらに、各「Ｉ」フレームは他のフレームに依存しないので、Ｉフレームへのアクセスは容易である。この理由で、通常のビデオエンコーダは、フレームを定期的に「Ｉ」フレームとして圧縮する。

【００９５】

しかし、本発明の対話式マルチビュービデオシステムのオンライン圧縮の、従来方式との大きな違いは、予測符号化を高速化するために導入される独特な「静的」モードにある。静的モードを得るには、元画像と基準画像との差を計算する必要がある。計算複雑度をさらに低減するために、この静的モードを使用するか否かを、すべてのビュー間で合同で判定する。この合同判定ではまず、あるビューの静的領域を検出する。次いで、それらに対応する、近隣ビューが重なった領域は、静的である可能性が高いと考えられる。最後に、非常に簡単なチェックにかけて、この判定を確認する（本発明の一実施形態では、ピクセルのわずかな部分だけを使用して、元画像と基準画像との差を計算する）。静的モードでは、関係するマクロブロック（ＭＢ）は従来のインターモードと同様に符号化されるが、それに対応する基準画像（時間予測のために次のフレームによって使用される）は、単にその前の再構築済み画像からコピーされる。この結果、このＭＢの基準画像の作成には

、逆量子化、逆ＤＣＴ、動き補償のどれも必要ない。
【００９６】
この新しい符号化モードに加えて、合同動き推定（ＭＥ）も適用して、ＭＥの複雑度を低減する。この新しいＭＥでは、あるビューに対してまず従来のＭＥを適用する。次いで、このビューについて見つかったＭＶに基づいて３Ｄ　ＭＶを作成する。その後、３ＤＭＶを近隣ビューに投影して、それら自体のＭＶを予測する。予測されたＭＶに基づいて、これらのビューの検索範囲を縮小することができ、したがって複雑度をかなり低減することができる。例えば、従来のシングルビュービデオ符号化では、エンコーダは通常、あるマクロブロックの動きベクトルを見つけるのに３２×３２の領域内を検索しなければならない。しかし、本発明によるシステムおよび方法のマルチビュービデオ符号化では、３Ｄ動きが得られてこれがあるビューに投影されると、このビューの検索範囲を絞り込むことができ（例えば８×８ピクセルに）、したがってこのビューの動きベクトルを見つける計算はかなり低減される。一方、このことはまた、異なるビューの動きベクトルが相関することも意味する。したがって、これらの動きベクトルはさらに圧縮することができる。本発明の一実施形態では、真の動きベクトルＶと、他のビューから得られた予測ベクトル
【００９７】
【数６】


【００９８】
との差だけを符号化する。
【００９９】
図７に、１つのカメラについての、本発明のオンライン符号化方式の全体的な例示的フローチャートを示す。この例では、システムは３つのビデオカメラを有し、各ビデオカメラは毎秒３０フレームでビデオを取り込むと仮定する。したがって、フレームサイズは６４０×４８０ピクセルである。そのため、毎秒３×３０フレームを圧縮する必要がある。最初に、単一のカメラによって取り込まれたフレームの圧縮について考え、次いで複数ビデオの場合について論じる。
【０１００】
図７のプロセス動作７０２に示すように、フレームを符号化する際は、どんなタイプのフレームであるかにかかわらず、最初にフレームをブロックに、好ましくはマクロブロック（ＭＢ）に分割する。１つのＭＢのサイズは１６×１６ピクセルである。すなわち、上の例では１フレームあたり６４０×４８０／１６／１６ＭＢが得られる。次いで、各フレームを所定の符号化タイプに従って圧縮する。各「Ｉ」フレームについては、すべてのＭＢをイントラモードで符号化する（プロセス動作７０４、７０８）。一方、「Ｐ」フレームについては、各ＭＢを符号化するときに３つの符号化モードを選択することができる。モード決定はＭＢベースである。言い換えれば、１つの「Ｐ」フレーム中で、異なるＭＢが異なる符号化モードを有することができる。どのモードを使用するか決定するために、エンコーダはまず、各ＭＢについて動き推定演算を実施して、現在のフレームとその前のフレームとの類似を計算する（プロセス動作７１０）。差が非常に大きい場合、これはこのＭＢに相関がほとんどないことを示し、この場合はイントラモードを選択する（プロセス動作７１２および７１４）。差が非常に小さい場合は「静的」モードを選択する（プロセス動作７１６、７１８）。残りの場合は「インター」モードを選択する（プロセス動作７２０）。これは、１つのビデオストリームのみからの入力についてのモード決定である。
【０１０１】
以下は、オンライン圧縮の場合の３つの符号化モードに関する記述である。図１１Ａ、１１Ｂ、１１Ｃに、前述のモードの符号化アーキテクチャ（それぞれインターモード、イントラモード、静的モード）を示す。
【０１０２】

１）イントラモード：図８に示すように、まず各ＭＢ中の係数を、変換または「Ｔ」モジュールによって変換して、それらの空間的相関を除去する（プロセス動作８０２）。その後、変換された係数を「Ｑ」モジュールによって量子化する（プロセス動作８０４）。（量子化プロセスの単純な例は次のとおりである。２つの係数６７および１６があり、量子化レベルが６４であると仮定する。量子化後、第１の係数は６４になり、第２の係数は０になる。量子化の目的は、係数を容易に符号化できるように係数の不確定性を除去することであることがわかる。当然、量子化後にはいくらかの情報が失われる。）量子化された係数を符号化する（例えば「エントロピー符号化」モジュールを使用して）（プロセス動作８０６）。最後に、圧縮されたビットストリームを得る（プロセス動作８０８）。

【０１０３】

２）インターモード：図９に示すように、まず現在のＭＢおよび前の基準フレームを入力する（プロセス動作９０２）。次いで、「フレームバッファ」に保存されている前の基準フレームに対して「動き推定」プロセスを実施して、現在のＭＢに最も類似する領域を見つける（動き推定プロセスは通常、図７に示したようにモード決定プロセスによって現在のＭＢに対して実施され、したがってここで再び実施する必要はないことに留意されたい）。その後、プロセス動作９０６に示すように、動き補償（ＭＣ）モジュールにより、動き補償操作を適用して、見つかった領域を「フレームバッファ」からコピーする。今や２つのＭＢがあり、一方は元のフレームからのＭＢ、他方は「ＭＣ」モジュールからのＭＢである。この２つのＭＢは類似しているが、これらの間にはなおいくらかの差がある。これらの間の差は残差と呼ばれるが、次いでこの差を「Ｔ」モジュールによって変換し、「Ｑ」モジュールによって量子化する（プロセス動作９０８および９１０）。最後に、量子化された結果を「エントロピー符号化」モジュールによって符号化する（プロセス動作９１２）。また、基準画像を次のフレームのために更新する必要がある。これは、逆量子化モジュール（「Ｑ－１」）および逆変換モジュール（「Ｔ－１」）によって達成され（プロセス動作９１４および９１６に示す）、次いで、これらの動作の結果として回復された残差を、動き補償された結果に加える（プロセス動作９１８）。この後には、エンコーダは、デコーダと同じ基準画像を有する。

【０１０４】

３）静的モード：静的モードは、本発明のシステムおよび方法によって利用される新しいモードである。この最初の部分は、インターモードの最初の部分と非常によく似ている。しかし、第２の部分、すなわち基準フレームの作成には大きな違いがある。この新しいモードでは、新しい基準は前の基準からコピーされるだけであり、一方、前のインターモードでは、逆量子化、逆変換、残差加算が必要である。この結果、多量の計算を省くことができる。図１０に、静的モード処理の流れ図を示す。図１０に示すように、まず現在のＭＢおよび前の基準フレームを入力する（処理動作１００２）。次いで、「フレームバッファ」に保存された前の基準フレームに対して「動き推定」プロセスを実施して、現在のＭＢに最も類似する領域を見つける（プロセス動作１００４）。（動き推定プロセスは通常、図７に示したようにモード決定プロセスによって実施され、したがってここで再び実施する必要はないことに留意されたい）。その後、プロセス動作１００６に示すように、「ＭＣ」モジュール（すなわち動き補償）を適用して、見つかった領域を「フレームバッファ」からコピーする。この場合２つのＭＢがあり、一方は元のフレームからのＭＢ、他方は「ＭＣ」モジュールからの結果である。次いで、この２つのＭＢ間の差を「Ｔ」モジュールによって変換し、「Ｑ」モジュールによって量子化する（プロセス動作１００８および１０１０）、最後に、量子化された結果を「エントロピー符号化」モジュールによって符号化する（プロセス動作１０１２）。新しい基準フレームは、単に、動き補償されたＭＢをコピーすることによって得られる（プロセス動作１０１４）。この静的モードでは、ＭＢは実際に静的である必要はなく、動きを含んでいてもよいことを指摘しておくのは重要である。さらに、ＭＢをインターモードで符号化するか静的モードで符号化するかを決定するモード決定しきい値が非常に大きくなるときは、ほとんどのインターモードＭＢは静的モードとして符号化されることになる。その場合、複雑度はかなり低減することが

できるが、性能はやや犠牲になる。本発明の一実施形態では、このモード決定しきい値を制御して、複雑度と性能との間の適切なトレードオフを達成する。

【０１０５】

復号プロセスは、符号化プロセスのちょうど逆である。例えば、まず、圧縮済みビットストリームをエントロピーデコーダに入力して、量子化済み係数（ならびに、各ＭＢの符号化モードなど他の必要な情報）を得る。次いで、各ＭＢにつき、それらの符号化モードに従って、量子化済み係数に逆量子化や逆変換などを施す。

【０１０６】

次に、複数カメラの場合のモード決定はどうなるであろうか。３つのカメラの場合と図１２Ａおよび１２Ｂを再び参照する。第１のカメラからのビデオは、前に提示したのとまったく同様にモード決定を実施する（プロセス動作１２０２～１２２２）。その後、画像領域のエピポーラ幾何および類似を用いて、第１のカメラと残りの２つのカメラとの対応を確立することを試みる（プロセス動作１２２４）。対応に基づいて、第２および第３のカメラの符号化モードを推定する（プロセス動作１２２６）。推定は常に正しいとは限らないので、得られたこれらの符号化モード、さらには動きベクトルを、精緻化する必要がある。これは、第２のモード決定プロセスによって、より低い計算コストで達成される（プロセス動作１２２８）。次いで、得られた符号化モードに基づいて各ＭＢを符号化する（プロセス動作１２３０）。シングルビューの場合のモード決定と同様、この第２の決定プロセスでも、元のＭＢと動き補償済みＭＢとの差を計算する。ただし、ピクセルのわずかな部分の差だけを計算する。この結果、複雑さの多くが低減される。

【０１０７】

マルチビューの場合も、シングルビューの場合と同様に各ビューを独立して復号する。ＭＶを近隣ビューから予測する場合は、最初に近隣ビューのＭＶを復号すべきである。

【０１０８】

３．２　オフライン圧縮

オフライン圧縮を使用して、ビデオデータストリームをさらに圧縮することができる。図１３および１４に示すように、オフライン圧縮の鍵となる考え方は、すべてのビューを３Ｄマッピングに分解することであり、この３Ｄマッピングは、３Ｄ環境における特徴点のグループからなる。図１３のプロセス動作１３０２に示すように、各特徴点を、その３Ｄ座標（ｘ，ｙ，ｚ）および対応する色成分（Ｙ，Ｕ，Ｖ）で表す。作成されたマッピングは、各ビュー中のすべてのピクセルを再構築することのできる特徴点の最小限のセットである。ＤＣＴやＤＷＴなどの変換ベースの分解とは異なり、この種の分解は、マルチビュービデオを非相関化するには最も効率的である。明らかに、ビューの数が増加したときは、新しい特徴点（すなわち新しい情報）だけを記録すればよく、他の特徴点は既存のマッピングから見つけることができる。

【０１０９】

３Ｄマッピングを作成した後、プロセス動作１３０４に示すように、得られた特徴点を変換して、これらの間の相関をさらに分解する。変換結果を量子化し、「ベース層」ビットストリームとして符号化する（プロセス動作１３０６、１３０８）。逆量子化された特徴点を再び各ビューにマッピングして、予測ビュー画像を形成する（プロセス動作１３１０）。予測画像は元の画像に近いが、これらの間にはなおいくらかの差がある。この差は、プロセス動作１３１２、１３１４に示すように、各ビューの「エンハンスメント層」として独立して符号化する（エンハンスメント層のビットストリームをスケーラブルに符号化して、ネットワーク適応能力を改善することができる）。さらに、２種類の層を符号化するとき、時間相関も利用する。これは、時間領域では、マッピング情報の静的部分およびエンハンスメント残差は不変だからである。動きのある部分も、３Ｄ動き構造によってやはり圧縮することができる。

【０１１０】

図１４に、オフライン圧縮の例示的な符号化構造を示す。この構造は、３Ｄマッピング作成モジュール１４０２、変換モジュール１４０４、量子化モジュール１４０６、逆変換

モジュール１４０８、逆量子化モジュール１４１０、逆マッピングモジュール１４１２、エントロピー符号化モジュール１４１４、ならびにビューバッファ１４１６を備える。表現をわかりやすくするために、この例では２つのビューだけを考える。ｉ番目に取り込まれたビューについて、すべてのビュー画像およびカメラ位置を「３Ｄマッピング作成」モジュールに入力して、特徴点セット$M_i$を抽出する。次いで、前に構築された特徴点セット

【０１１１】

【数７】

$$\hat{M}_{i-1}$$

【０１１２】

からマッピング情報$M_i$を予測して、その時間相関を除去する。予測残差

【０１１３】

【数８】

$$M_i - \hat{M}_{i-1}$$

【０１１４】

を変換し量子化する（ここではＤＣＴまたは離散ウェーブレット変換（ＤＷＴ）あるいはその他の変換を採用することができる）。最後に、エントロピー符号化を適用してベース層ビットストリームを生成する。次いで、再構築されたマッピング情報

【０１１５】

【数９】

$$\hat{M}_i$$

【０１１６】

を、カメラの位置と共に「逆マッピング」モジュールに入力する。その後、各ビューの予測画像が得られる。時間予測によって、予測画像と元画像との差をさらに非相関化する。残差を変換し量子化する（ここではＤＣＴまたは離散ウェーブレット変換（ＤＷＴ）あるいはその他の変換を採用することができる）。最後に、エントロピー符号化を適用してエンハンスメント層ビットストリームを生成する。（この例では、２つのエンハンスメント層ビットストリーム（各ビューにつき１つのビットストリーム）が得られる。）

復号プロセスは次のとおりである。あるビューを再構築したいと仮定する。まずベース層を、エントロピー復号、逆量子化、逆変換など（例えばこの層の符号化プロセスの逆）によって復号する。その後、このビューのエンハンスメント層を、エントロピー復号、逆量子化、逆変換などによって復号する。最後に、得られた共通の特徴点（ベース層からの）をこのビューに逆マッピングする。得られた画像と、エンハンスメント層の復号結果とが、このビューの再構築画像を形成する。

【０１１７】

本発明に関する以上の記述は、例示および記述のために提示したものである。これは、網羅的でもなく、開示した厳密な形に本発明を限定するものでもない。前述の教示に鑑みて、多くの修正および変形が可能である。本発明の範囲は、この詳細な記述によってではなく、本明細書に添付された特許請求の範囲によって限定されるものとする。

【図面の簡単な説明】

【０１１８】

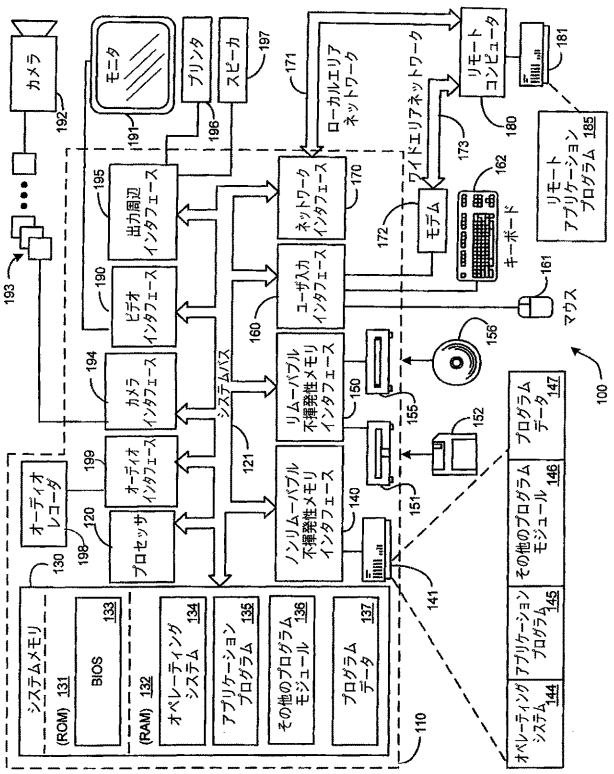
【図１】本発明を実施するための例示的なシステムを構成する汎用コンピューティングデバイスを示す図である。

【図２】本発明による対話式マルチビュービデオシステムを単純化したブロック図である。

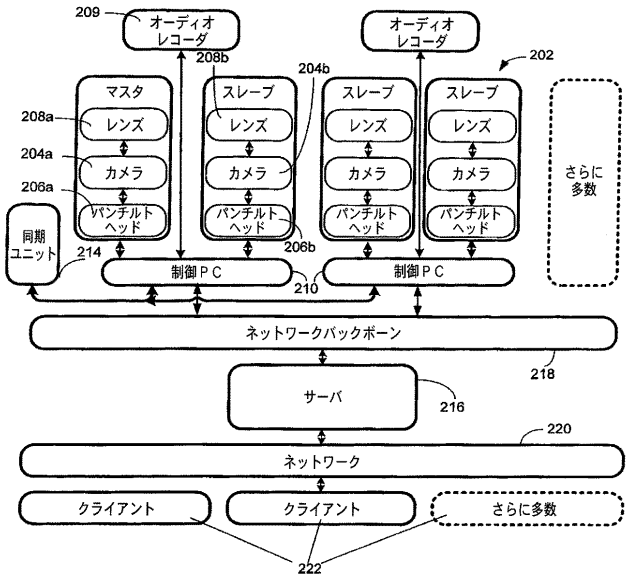
【図３】本発明の対話式マルチビュービデオシステム中で利用される全体的な較正手順を単純化した流れ図である。

【図４Ａ】本発明によるシステムおよび方法の一実施形態で使用される例示的な較正パターンの画像の図である。
【図４Ｂ】本発明の対話式マルチビュービデオシステム中で利用されるパターンベースの較正の流れ図である。
【図５】本発明の対話式マルチビュービデオシステム中で利用されるパターンなし較正の流れ図である。
【図６Ａ】本発明の対話式マルチビュービデオシステム中で使用されるビデオ索引テーブルの図である。
【図６Ｂ】本発明の対話式マルチビュービデオシステム中で使用されるオーディオ索引テーブルの図である。
【図７】本発明の一実施形態の、１つのカメラについてのオンライン圧縮方式を示す流れ図である。
【図８】本発明の一実施形態のイントラモード符号化を示す流れ図である。
【図９】本発明の一実施形態のインターモード符号化を示す流れ図である。
【図１０】本発明の一実施形態の静的モード符号化を示す流れ図である。
【図１１Ａ】本発明の一実施形態のインターモード符号化アーキテクチャの概略図である。
【図１１Ｂ】本発明の一実施形態のイントラモード符号化アーキテクチャの概略図である。
【図１１Ｃ】本発明の一実施形態の静的モード符号化アーキテクチャの概略図である。
【図１２Ａ】複数のカメラのビットストリームを符号化するための符号化論理を示す流れ図である。
【図１２Ｂ】複数のカメラのビットストリームを符号化するための符号化論理を示す流れ図である。
【図１３】本発明の一実施形態のオフライン圧縮方式を示す流れ図である。
【図１４】本発明の一実施形態のオフライン圧縮システムのアーキテクチャの図である。
【符号の説明】
【０１１９】
２０４ａ　カメラ
２０４ｂ　カメラ
２０６ａ　パンチルトヘッド
２０６ｂ　パンチルトヘッド
２０８ａ　レンズ
２０８ｂ　レンズ
２０９　オーディオレコーダ
２１０　制御ＰＣ
２１４　同期ユニット
２１６　サーバ
２１８　ネットワークバックボーン
２２０　ネットワーク
２２２　クライアント

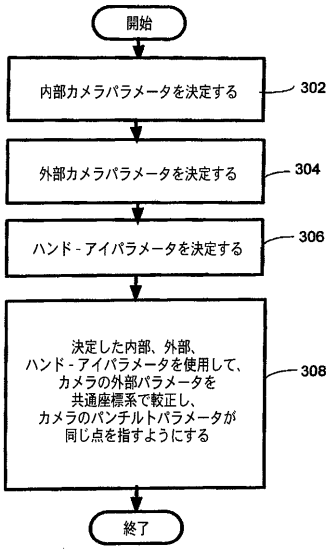
【図１】


カメラ
モニタ
プリンタ
スピーカ
ローカルエリアネットワーク
ワイドエリアネットワーク
リモートコンピュータ
リモートアプリケーションプログラム
出力周辺インタフェース
ビデオインタフェース
カメラインタフェース
オーディオインタフェース
ネットワークインタフェース
モデム
キーボード
ユーザ入力インタフェース
マウス
リムーバブル不揮発性メモリインタフェース
ノンリムーバブル不揮発性メモリインタフェース
システムバス
オーディオレコーダ
プロセッサ
システムメモリ
(ROM) 131
BIOS 133
(RAM) 132
オペレーティングシステム 134
アプリケーションプログラム 135
その他のプログラムモジュール 136
プログラムデータ 137
プログラムデータ 147
その他のプログラムモジュール 146
アプリケーションプログラム 145
オペレーティングシステム 144

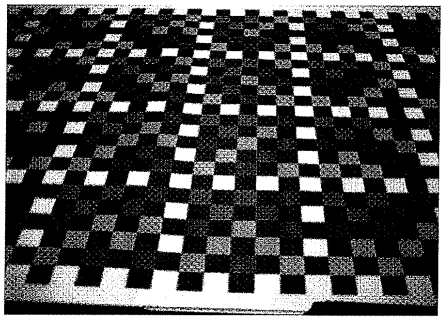
【図２】


オーディオレコーダ
オーディオレコーダ
マスタ
スレーブ
スレーブ
スレーブ
レンズ
カメラ
パンチルトヘッド
さらに多数
同期ユニット
制御ＰＣ
制御ＰＣ
ネットワークバックボーン
サーバ
ネットワーク
クライアント
クライアント
さらに多数

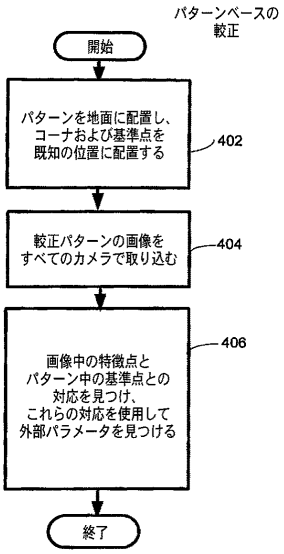
【図３】


開始
内部カメラパラメータを決定する
外部カメラパラメータを決定する
ハンド - アイパラメータを決定する
決定した内部、外部、
ハンド - アイパラメータを使用して、
カメラの外部パラメータを
共通座標系で較正し、
カメラのパンチルトパラメータが
同じ点を指すようにする
終了

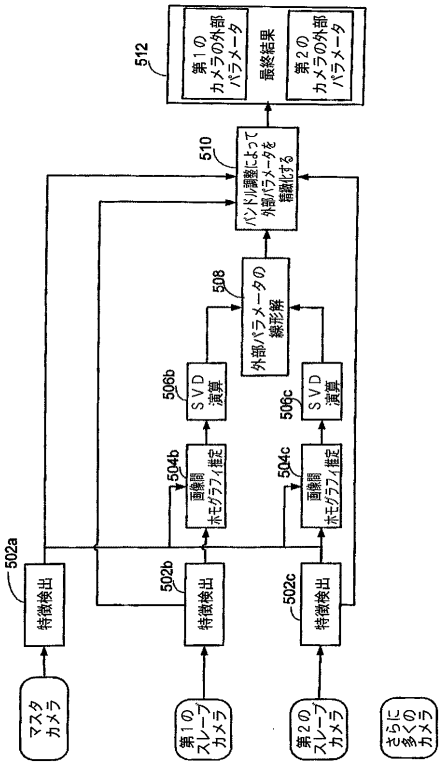
【図４Ａ】



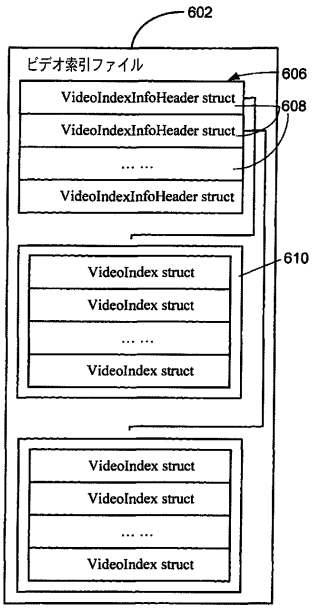
【図４Ｂ】



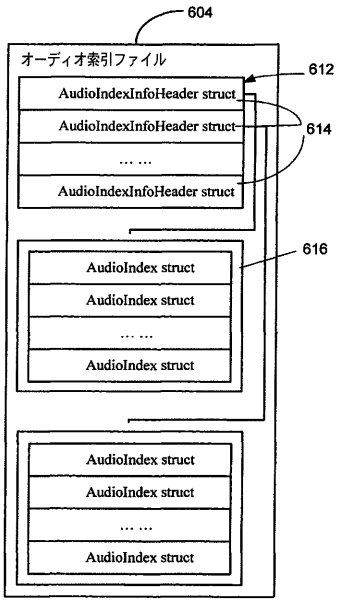

【図５】



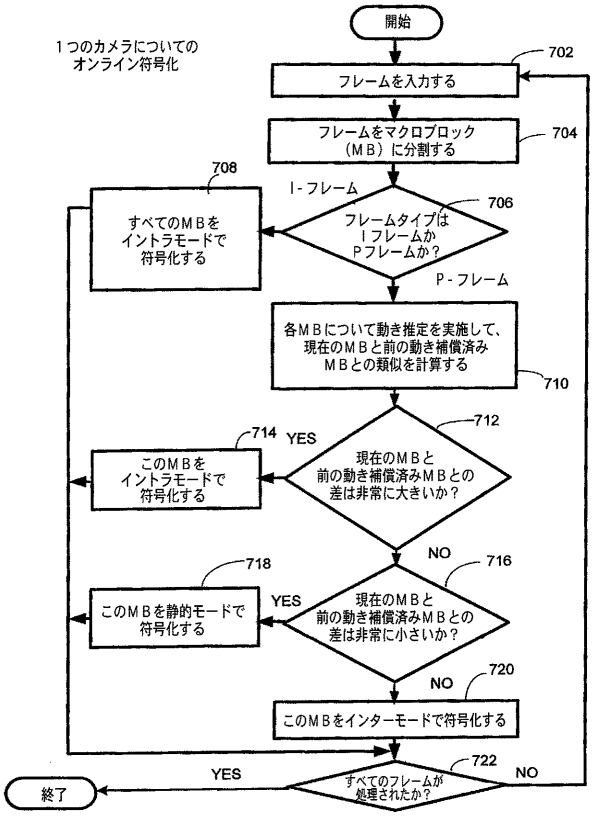

【図６Ａ】



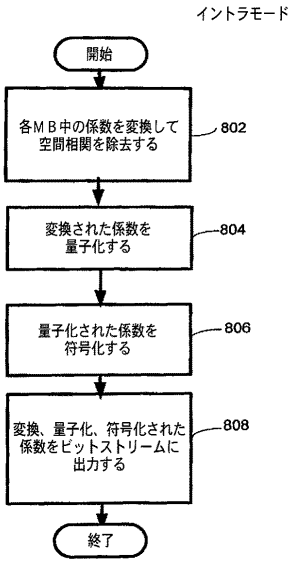

【図６Ｂ】



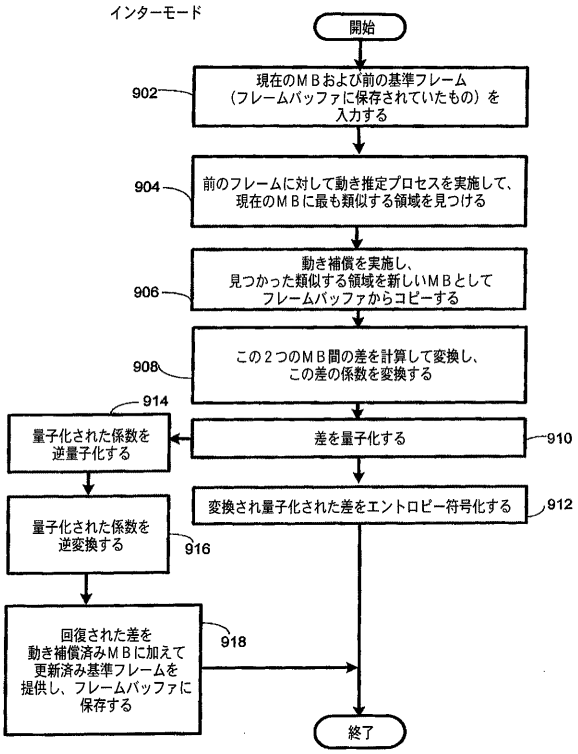
【図７】


1つのカメラについての
オンライン符号化
開始
702
フレームを入力する
704
フレームをマクロブロック
（ＭＢ）に分割する
708
I-フレーム
706
すべてのＭＢを
イントラモードで
符号化する
フレームタイプは
Ｉフレームか
Ｐフレームか？
P-フレーム
各ＭＢについて動き推定を実施して、
現在のＭＢと前の動き補償済み
ＭＢとの類似を計算する
710
712
714
YES
このＭＢを
イントラモードで
符号化する
現在のＭＢと
前の動き補償済みＭＢとの
差は非常に大きいか？
NO
716
718
このＭＢを静的モードで
符号化する
YES
現在のＭＢと
前の動き補償済みＭＢとの
差は非常に小さいか？
NO
720
このＭＢをインターモードで符号化する
722
YES
NO
終了
すべてのフレームが
処理されたか？

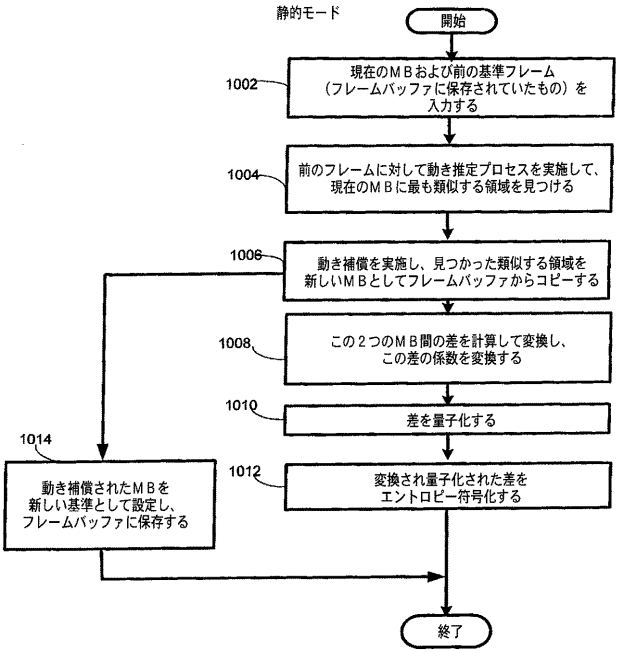
【図８】


イントラモード
開始
各ＭＢ中の係数を変換して
空間相関を除去する
802
変換された係数を
量子化する
804
量子化された係数を
符号化する
806
変換、量子化、符号化された
係数をビットストリームに
出力する
808
終了

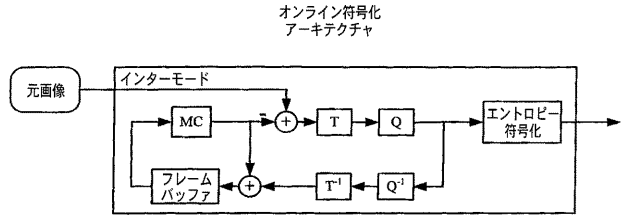
【図９】


インターモード
開始
902
現在のＭＢおよび前の基準フレーム
（フレームバッファに保存されていたもの）を
入力する
904
前のフレームに対して動き推定プロセスを実施して、
現在のＭＢに最も類似する領域を見つける
906
動き補償を実施し、
見つかった類似する領域を新しいＭＢとして
フレームバッファからコピーする
908
この２つのＭＢ間の差を計算して変換し、
この差の係数を変換する
914
量子化された係数を
逆量子化する
差を量子化する
910
912
変換され量子化された差をエントロピー符号化する
量子化された係数を
逆変換する
916
回復された差を
動き補償済みＭＢに加えて
更新済み基準フレームを
提供し、フレームバッファに
保存する
918
終了

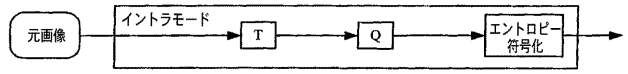
【図１０】


静的モード
開始
1002
現在のＭＢおよび前の基準フレーム
（フレームバッファに保存されていたもの）を
入力する
1004
前のフレームに対して動き推定プロセスを実施して、
現在のＭＢに最も類似する領域を見つける
1006
動き補償を実施し、見つかった類似する領域を
新しいＭＢとしてフレームバッファからコピーする
1008
この２つのＭＢ間の差を計算して変換し、
この差の係数を変換する
1010
差を量子化する
1014
1012
動き補償されたＭＢを
新しい基準として設定し、
フレームバッファに保存する
変換され量子化された差を
エントロピー符号化する
終了

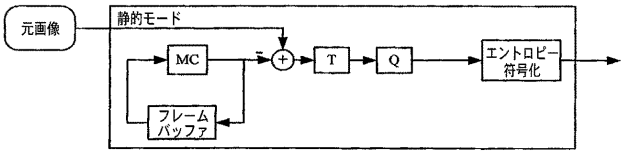
【図１１Ａ】




【図１１Ｂ】

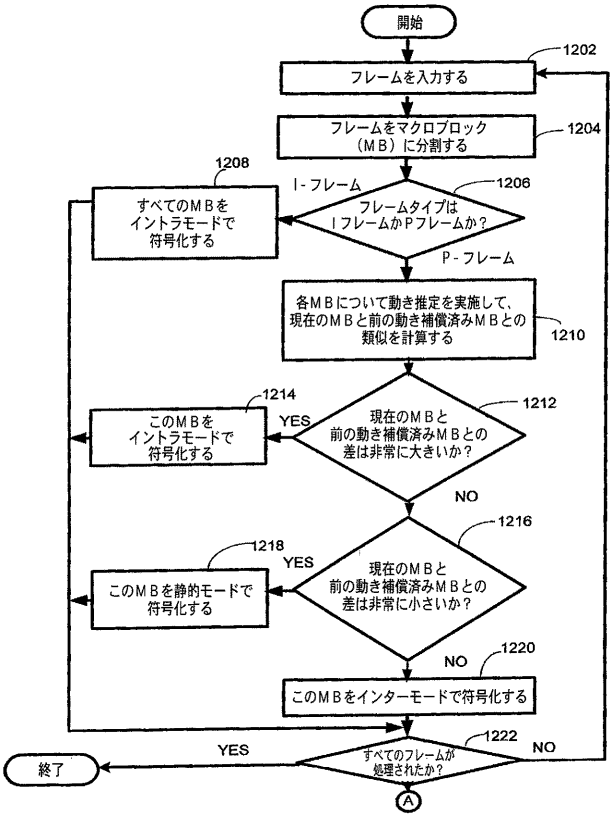



【図１１Ｃ】

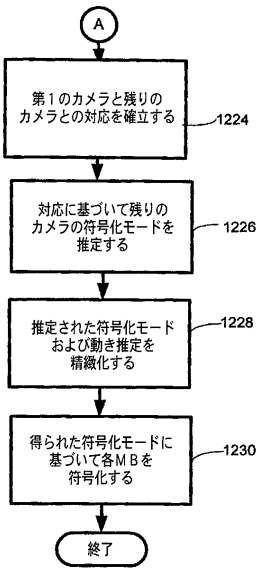



MC：動き補償
T：変換（例えば離散コサイン変換（ＤＣＴ））
Q：量子化
T-1：逆変換
Q-1：逆量子化

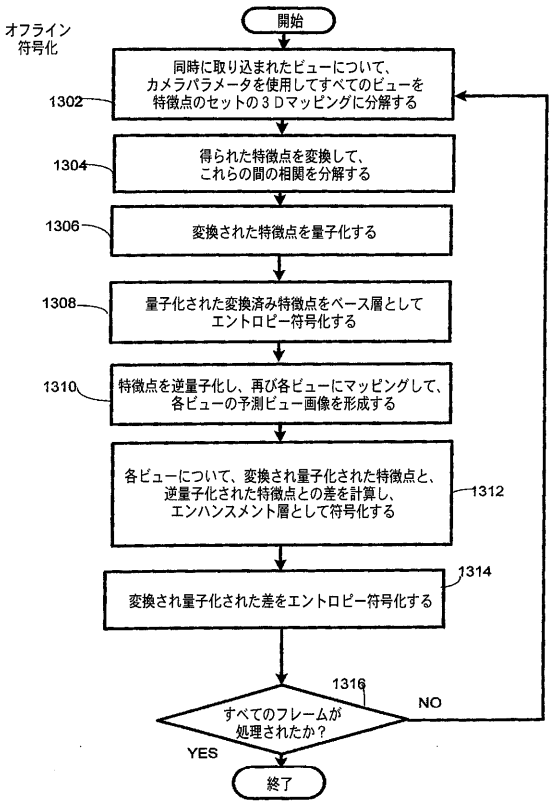
【図１２Ａ】



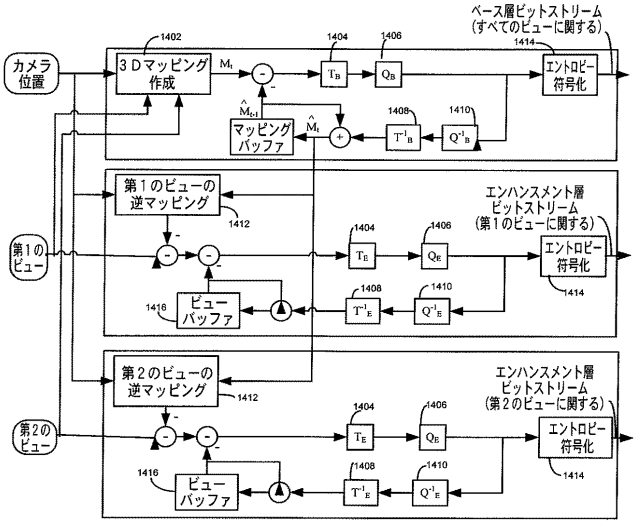

【図１２Ｂ】




【図１３】



【図１４】


ベース層ビットストリーム
(すべてのビューに関する)
カメラ
位置
3Dマッピング
作成
マッピング
バッファ
エントロピー
符号化
第1のビュー
第1のビューの
逆マッピング
エンハンスメント層
ビットストリーム
(第1のビューに関する)
ビュー
バッファ
第2のビュー
第2のビューの
逆マッピング
エンハンスメント層
ビットストリーム
(第2のビューに関する)

フロントページの続き

(72)発明者　チャンクワン　ルー
アメリカ合衆国　９８０５２　ワシントン州　レッドモンド　ワン　マイクロソフト　ウェイ　マイクロソフト　コーポレーション内
(72)発明者　チャン　リー
アメリカ合衆国　９８０５２　ワシントン州　レッドモンド　ワン　マイクロソフト　ウェイ　マイクロソフト　コーポレーション内
(72)発明者　イーホワ　スー
アメリカ合衆国　９８０５２　ワシントン州　レッドモンド　ワン　マイクロソフト　ウェイ　マイクロソフト　コーポレーション内

Ｆターム(参考) 5C053 FA23 GB08 GB11 GB37 HA29 JA12 JA21 LA01 LA11 LA14
5C059 MA00 PP04 SS08
5C164 FA07 MA02S MC02S PA33 SA26P SB02S SB27S SB41P TB35S TC13S
YA11

【外国語明細書】
2006060801000001.pdf