明細書

画像符号化装置および画像符号化方法、画像復号装置および画像復号方法、並びに提供媒体

技術分野

本発明は、画像符号化装置および画像符号化方法、画像復号装置および画像復号方法、並びに提供媒体に関する。特に、例えば、動画像データを、光磁気ディスクや磁気テープなどの記録媒体に記録し、これを再生してディスプレイなどに表示したり、テレビ会議システム、テレビ電話システム、放送用機器、マルチメディアデータベース検索システムなどのように、動画像データを伝送路を介して送信側から受信側に伝送し、受信側において、受信された動画像データを表示する場合や、編集して記録する場合などに用いて好適な画像符号化装置および画像符号化方法、画像復号装置および画像復号方法、並びに提供媒体に関する。

背景技術

例えば、テレビ会議システム、テレビ電話システムなどのように、動画像データを遠隔地に伝送するシステムにおいては、伝送路を効率良く利用するため、画像データを、そのライン相関やフレーム間相関を利用して圧縮符号化するようになされている。

動画像の高能率符号化方式として代表的なものとしてMPEG（Moving Picture Experts Group）（蓄積用動画像符号化）方式がある。これはＩＳＯ－ＩＥＣ／ＪＴＣ１／ＳＣ２／ＷＧ１１において議論され、標準案として提案されたものであり、動き補償予測符号化とＤＣＴ（Discrete Cosine Transform）符号化を組み合わせたハイブリッド方式が採用されている。

ＭＰＥＧでは、様々なアプリケーションや機能に対応するために、いくつかのプロファイルおよびレベルが定義されている。最も基本となるのが、メインプロファ

イルメインレベル（ＭＰ＠ＭＬ（Main Profile at Main Level））である。

図１は、ＭＰＥＧ方式におけるＭＰ＠ＭＬのエンコーダの一例の構成を示している。

符号化すべき画像データは、フレームメモリ３１に入力され、一時記憶される。そして、動きベクトル検出器３２は、フレームメモリ３１に記憶された画像データを、例えば、１６画素×１６画素などで構成されるマクロブロック単位で読み出し、その動きベクトルを検出する。

ここで、動きベクトル検出器３２においては、各フレームの画像データを、Ｉピクチャ（フレーム内符号化）、Ｐピクチャ（前方予測符号化）、またはＢピクチャ（両方向予測符号化）のうちのいずれかとして処理する。なお、シーケンシャルに入力される各フレームの画像を、Ｉ，Ｐ，Ｂピクチャのいずれのピクチャとして処理するかは、例えば、予め定められている（例えば、Ｉ，Ｂ，Ｐ，Ｂ，Ｐ，・・・Ｂ，Ｐとして処理される）。

即ち、動きベクトル検出器３２は、フレームメモリ３１に記憶された画像データの中の、予め定められた所定の参照フレームを参照し、その参照フレームと、現在符号化の対象となっているフレームの１６画素×１６ラインの小ブロック（マクロブロック）とをパターンマッチング（ブロックマッチング）することにより、そのマクロブロックの動きベクトルを検出する。

ここで、ＭＰＥＧにおいては、画像の予測モードには、イントラ符号化（フレーム内符号化）、前方予測符号化、後方予測符号化、両方向予測符号化の４種類があり、Ｉピクチャはイントラ符号化され、Ｐピクチャはイントラ符号化または前方予測符号化のいずれかで符号化され、Ｂピクチャはイントラ符号化、前方予測符号化、後方予測符号化、または両方法予測符号化のいずれかで符号化される。

即ち、動きベクトル検出器３２は、Ｉピクチャについては、予測モードとしてイントラ符号化モードを設定する。この場合、動きベクトル検出器３２は、動きベクトルの検出は行わず、予測モード（イントラ予測モード）を、ＶＬＣ（可変長符号化）器３６および動き補償器４２に出力する。

また、動きベクトル検出器３２は、Ｐピクチャについては、前方予測を行い、その動きベクトルを検出する。さらに、動きベクトル検出器３２は、前方予測を行うことにより生じる予測誤差と、符号化対象のマクロブロック（Ｐピクチャのマクロブロック）の、例えば分散とを比較する。その比較の結果、マクロブロックの分散の方が予測誤差より小さい場合、動きベクトル検出器３２は、予測モードとしてイントラ符号化モードを設定し、ＶＬＣ器３６および動き補償器４２に出力する。また、動きベクトル検出器３２は、前方予測を行うことにより生じる予測誤差の方が小さければ、予測モードとして前方予測符号化モードを設定し、検出した動きベクトルとともに、ＶＬＣ器３６および動き補償器４２に出力する。

さらに、動きベクトル検出器３２は、Ｂピクチャについては、前方予測、後方予測、および両方向予測を行い、それぞれの動きベクトルを検出する。そして、動きベクトル検出器３２は、前方予測、後方予測、および両方向予測についての予測誤差の中の最小のもの（以下、適宜、最小予測誤差という）を検出し、その最小予測誤差と、符号化対象のマクロブロック（Ｂピクチャのマクロブロック）の、例えば分散とを比較する。その比較の結果、マクロブロックの分散の方が最小予測誤差より小さい場合、動きベクトル検出器３２は、予測モードとしてイントラ符号化モードを設定し、ＶＬＣ器３６および動き補償器４２に出力する。また、動きベクトル検出器３２は、最小予測誤差の方が小さければ、予測モードとして、その最小予測誤差が得られた予測モードを設定し、対応する動きベクトルとともに、ＶＬＣ器３６および動き補償器４２に出力する。

動き補償器４２は、動きベクトル検出器３２から予測モードと動きベクトルの両方を受信すると、その予測モードおよび動きベクトルにしたがって、フレームメモリ４１に記憶されている、符号化され、かつ既に局所復号された画像データを読み出し、この読み出された画像データを、予測画像データとして、演算器３３および４０に供給する。

演算器３３は、動きベクトル検出器３２がフレームメモリ３１から読み出した画像データと同一のマクロブロックをフレームメモリ３１から読み出し、そのマクロ

ブロックと、動き補償器４２からの予測画像との差分を演算する。この差分値は、ＤＣＴ器３４に供給される。

一方、動き補償器４２は、動きベクトル検出器３２から予測モードのみを受信した場合、即ち、予測モードがイントラ符号化モードである場合には、予測画像を出力しない。この場合、演算器３３（演算器４０も同様）は、特に処理を行わず、フレームメモリ３１から読み出したマクロブロックを、そのままＤＣＴ器３４に出力する。

ＤＣＴ器３４では、演算器３３の出力データに対して、ＤＣＴ処理が施され、その結果得られるＤＣＴ係数が、量子化器３５に供給される。量子化器３５では、バッファ３７のデータ蓄積量（バッファ３７に記憶されているデータの量）（バッファフィードバック）に対応して量子化ステップ（量子化スケール）が設定され、その量子化ステップで、ＤＣＴ器３４からのＤＣＴ係数が量子化される。この量子化されたＤＣＴ係数（以下、適宜、量子化係数という）は、設定された量子化ステップとともに、ＶＬＣ器３６に供給される。

ＶＬＣ器３６では、量子化器３５より供給される量子化係数が、例えばハフマン符号などの可変長符号に変換され、バッファ３７に出力される。さらに、ＶＬＣ器３６は、量子化器３５からの量子化ステップ、動きベクトル検出器３２からの予測モード（イントラ符号化（画像内予測符号化）、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれが設定されたかを示すモード）および動きベクトルも可変長符号化し、その結果得られる符号化データを、バッフ３７に出力する。

バッファ３７は、ＶＬＣ器３６からの符号化データを一時蓄積することにより、そのデータ量を平滑化し、符号化ビットストリームとして、例えば、伝送路に出力し、または記録媒体に記録する。

また、バッファ３７は、そのデータ蓄積量を量子化器３５に出力しており、量子化器３５は、このバッファ３７からのデータ蓄積量にしたがって量子化ステップを設定する。即ち、量子化器３５は、バッファ３７がオーバーフローしそうなとき、

量子化ステップを大きくし、これにより、量子化係数のデータ量を低下させる。また、量子化器３５は、バッファ３７がアンダーフローしそうなとき、量子化ステップを小さくし、これにより、量子化係数のデータ量を増大させる。このようにして、バッファ３７のオーバフローとアンダフローを防止するようになっている。

量子化器３５が出力する量子化係数と量子化ステップは、ＶＬＣ器３６だけでなく、逆量子化器３８にも供給されるようになされている。逆量子化器３８では、量子化器３５からの量子化係数が、同じく量子化器３５からの量子化ステップにしたがって逆量子化され、これによりＤＣＴ係数に変換される。このＤＣＴ係数は、ＩＤＣＴ器（逆ＤＣＴ器）３９に供給される。ＩＤＣＴ器３９では、ＤＣＴ係数が逆ＤＣＴ処理され、その処理の結果得られるデータが、演算器４０に供給される。

演算器４０には、ＩＤＣＴ器３９の出力データの他、上述したように、動き補償器４２から、演算器３３に供給されている予測画像と同一のデータが供給されている。演算器４０は、ＩＤＣＴ器３９の出力データ（予測残差（差分データ））と、動き補償器４２からの予測画像データとを加算することで、元の画像データを局所復号し、この局所復号された画像データ（局所復号画像データ）が出力される（但し、予測モードがイントラ符号化である場合には、ＩＤＣＴ器３９の出力データは、演算器４０をスルーして、そのまま、局所復号画像データとして、フレームメモリ４１に供給される）。なお、この復号画像データは、受信側において得られる復号画像データと同一のものである。

演算器４０において得られた復号画像データ（局所復号画像データ）は、フレームメモリ４１に供給されて記憶され、その後、インター符号化（前方予測符号化、後方予測符号化、量方向予測符号化）される画像に対する参照画像データ（参照フレーム）として用いられる。

次に、図２は、図１のエンコーダから出力される符号化データを復号する、ＭＰＥＧにおけるＭＰ@ＭＬのデコーダの一例の構成を示している。

伝送路を介して伝送されてきた符号化ビットストリーム（符号化データ）が図示せぬ受信装置で受信され、または記録媒体に記録された符号化ビットストリーム

（符号化データ）が図示せぬ再生装置で再生され、バッファ１０１に供給されて記憶される。

ＩＶＬＣ器（逆ＶＬＣ器（可変長復号器））１０２は、バッファ１０１に記憶された符号化データを読み出し、可変長復号することにより、その符号化データを、マクロブロック単位で、動きベクトル、予測モード、量子化ステップ、および量子化係数に分離する。これらのデータのうち、動きベクトルおよび予測モードは動き補償器１０７に供給され、量子化ステップおよびマクロブロックの量子化係数は逆量子化器１０３に供給される。

逆量子化器１０３は、ＩＶＬＣ器１０２より供給されたマクロブロックの量子化係数を、同じくＩＶＬＣ器１０２より供給された量子化ステップにしたがって逆量子化し、その結果得られるＤＣＴ係数を、ＩＤＣＴ器１０４に出力する。ＩＤＣＴ器１０４は、逆量子化器１０３からのマクロブロックのＤＣＴ係数を逆ＤＣＴし、演算器１０５に供給する。

演算器１０５には、ＩＤＣＴ器１０４の出力データの他、動き補償器１０７の出力データも供給されている。即ち、動き補償器１０７は、フレームメモリ１０６に記憶されている、既に復号された画像データを、図１の動き補償器４２における場合と同様に、ＩＶＬＣ器１０２からの動きベクトルおよび予測モードにしたがって読み出し、予測画像データとして、演算器１０５に供給する。演算器１０５は、ＩＤＣＴ器１０４の出力データ（予測残差（差分値））と、動き補償器１０７からの予測画像データとを加算することで、元の画像データを復号する。この復号画像データは、フレームメモリ１０６に供給されて記憶される。なお、ＩＤＣＴ器１０４の出力データが、イントラ符号化されたものである場合には、その出力データは、演算器１０５をスルーして、復号画像データとして、そのままフレームメモリ１０６に供給されて記憶される。

フレームメモリ１０６に記憶された復号画像データは、その後に復号される画像データの参照画像データとして用いられる。さらに、復号画像データは、出力再生画像として、例えば、図示せぬディスプレイなどに供給されて表示される。

なお、ＭＰＥＧ１および２では、Ｂピクチャは、参照画像データとして用いられないため、エンコーダまたはデコーダのそれぞれにおいて、フレームメモリ４１（図１）または１０６（図２）には記憶されない。

以上の図１、図２に示したエンコーダ、デコーダは、ＭＰＥＧ１／２の規格に準拠したものであるが、現在、画像を構成する物体などのオブジェクトのシーケンスであるＶＯ（Video Object）単位で符号化を行う方式につき、ＩＳＯ－ＩＥＣ／ＪＴＣ１／ＳＣ２９／ＷＧ１１において、ＭＰＥＧ（Moving Picture Experts Group）４として標準化作業が進められている。

ところで、ＭＰＥＧ４については、主として、通信の分野で利用されるものとして、標準化作業が進められていたため、ＭＰＥＧ１／２において規定されているＧＯＰ（Group Of Picture）は、ＭＰＥＧ４では規定されておらず、従って、ＭＰＥＧ４が蓄積メディアに利用された場合には、効率的なランダムアクセスが困難になることが予想される。

発明の開示

本発明は、このような状況に鑑みてなされたものであり、効率的なランダムアクセスをすることができるようにするものである。

請求の範囲第１項に記載の画像符号化装置は、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化する符号化手段を備えることを特徴とする。

請求の範囲第６項に記載の画像符号化方法は、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化する符号化ステップを備えることを特徴とする。

請求の範囲第１１項に記載の画像復号装置は、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られた符号化ビットストリームを復号する復号手段を備えることを特徴とする。

請求の範囲第１６項に記載の画像復号方法は、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られた符号化ビットストリームを復号する復号ステップを備えることを特徴とする。

請求の範囲第２１項に記載の提供媒体は、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られる符号化ビットストリームを提供することを特徴とする。

請求の範囲第２６項に記載の画像符号化装置は、秒精度の時刻を表す秒精度時刻情報を生成する秒精度時刻情報生成手段と、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成する詳細時間情報生成手段とを備えることを特徴とする。

請求の範囲第２９項に記載の画像符号化方法は、秒精度の時刻を表す秒精度時刻情報を生成する秒精度時刻情報生成ステップと、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成する詳細時間情報生成ステップとを備えることを特徴とする。

請求の範囲第３２項に記載の画像復号装置は、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を、秒精度時刻情報および詳細時間情報に基づいて求める表示時刻算出手段を備えることを特徴とする。

請求の範囲第３５項に記載の画像復号方法は、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を、秒精度時刻情報および詳細時間情報に基づいて求める表示時刻算出ステップを備えることを特徴とする。

請求の範囲第３８項に記載の提供媒体は、秒精度の時刻を表す秒精度時刻情報を生成し、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成し、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時

刻を表す情報として、秒精度時刻情報および詳細時間情報を、対応するＩ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加することにより得られる符号化ストリームを提供することを特徴とする。

請求の範囲第１項に記載の画像符号化装置においては、符号化手段が、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化するようになされている。

請求の範囲第６項に記載の画像符号化方法においては、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化するようになされている。

請求の範囲第１１項に記載の画像復号装置においては、復号手段が、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られた符号化ビットストリームを復号するようになされている。

請求の範囲第１６項に記載の画像復号方法においては、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られた符号化ビットストリームを復号するようになされている。

請求の範囲第２１項に記載の提供媒体においては、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られる符号化ビットストリームを提供するようになされている。

請求の範囲第２６項に記載の画像符号化装置においては、秒精度時刻情報生成手段は、秒精度の時刻を表す秒精度時刻情報を生成し、詳細時間情報生成手段は、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成するようになされている。

請求の範囲第２９項に記載の画像符号化方法においては、秒精度の時刻を表す秒精度時刻情報を生成し、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度よ

り細かい精度で表す詳細時間情報を生成するようになされている。

請求の範囲第３２項に記載の画像復号装置においては、表示時刻算出手段が、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を、秒精度時刻情報および詳細時間情報に基づいて求めるようになされている。

請求の範囲第３５項に記載の画像復号方法においては、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を、秒精度時刻情報および詳細時間情報に基づいて求めるようになされている。

請求の範囲第３８項に記載の提供媒体においては、秒精度の時刻を表す秒精度時刻情報を生成し、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成し、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時刻を表す情報として、秒精度時刻情報および詳細時間情報を、対応するＩ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加することにより得られる符号化ストリームを提供するようになされている。

図面の簡単な説明

図１は、従来のエンコーダの一例の構成を示すブロック図である。

図２は、従来のデコーダの一例の構成を示すブロック図である。

図３は、本発明を適用したエンコーダの一実施の形態の構成例を示すブロック図である。

図４は、時刻によって、ＶＯの位置、大きさが変化することを説明するための図である。

図５は、図３のＶＯＰ符号化部3_1～3_Nの構成例を示すブロック図である。

図６は、空間スケーラビリティを説明するための図である。

図７は、空間スケーラビリティを説明するための図である。

図８は、空間スケーラビリティを説明するための図である。

図９は、空間スケーラビリティを説明するための図である。

図１０は、ＶＯＰのサイズデータおよびオフセットデータの決定方法を説明するための図である。

図１１は、図５の下位レイヤ符号化部２５の構成例を示すブロック図である。

図１２は、図５の上位レイヤ符号化部２３の構成例を示すブロック図である。

図１３は、空間スケーラビリティを説明するための図である。

図１４は、時間スケーラビリティを説明するための図である。

図１５は、本発明を適用したデコーダの一実施の形態の構成例を示すブロック図である。

図１６は、図１５のＶＯＰ復号部72_1～72_Nの他の構成例を示すブロック図である。

図１７は、図１６の下位レイヤ復号部９５の構成例を示すブロック図である。

図１８は、図１６の上位レイヤ復号部９３の構成例を示すブロック図である。

図１９は、スケーラブル符号化によって得られるビットストリームのシンタクスを示す図である。

図２０は、ＶＳのシンタクスを示す図である。

図２１は、ＶＯのシンタクスを示す図である。

図２２は、ＶＯＬのシンタクスを示す図である。

図２３は、ＶＯＰのシンタクスを示す図である。

図２４は、modulo_time_base と VOP_time_increment との関係を示す図である。

図２５は、本発明によるビットストリームのシンタクスを示す図である。

図２６は、ＧＯＶのシンタクスを示す図である。

図２７は、 time_code の構成を示す図表である。

図２８は、ＧＯＶ層の time_code、並びにＧＯＶの先頭のＩ－ＶＯＰの modulo_time_base と VOP_time_increment の符号化方法を示す図である。

図２９は、ＧＯＶ層の time_code、並びにＧＯＶの先頭のＩ－ＶＯＰよりも前に位置するＢ－ＶＯＰの modulo_time_base と VOP_time_increment の符号化方法を示す図である。

図３０は、modulo_time_base と VOP_time_increment の定義を変更しない場合のそれらの関係を示す図である。

図３１は、Ｂ－ＶＯＰの modulo_time_base と VOP_time_increment の第１の方法による符号化処理を示す図である。

図３２は、Ｉ／Ｐ－ＶＯＰの modulo_time_base と VOP_time_increment の第１および第２の方法による符号化処理を示すフローチャートである。

図３３は、Ｂ－ＶＯＰの modulo_time_base と VOP_time_increment の第１の方法による符号化処理を示すフローチャートである。

図３４は、第１および第２の方法により符号化したＩ／Ｐ－ＶＯＰの modulo_time_base と VOP_time_increment の復号処理を示すフローチャートである。

図３５は、第１の方法により符号化したＢ－ＶＯＰの modulo_time_base と VOP_time_increment の復号処理を示すフローチャートである。

図３６は、Ｂ－ＶＯＰの modulo_time_base と VOP_time_increment の第２の方法による符号化処理を示す図である。

図３７は、Ｂ－ＶＯＰの modulo_time_base と VOP_time_increment の第２の方法による符号化処理を示すフローチャートである。

図３８は、第２の方法により符号化したＢ－ＶＯＰの modulo_time_base と VOP_time_increment の復号処理を示すフローチャートである。

図３９は、modulo_time_base について説明するための図である。

図４０は、本発明を適用したエンコーダおよびデコーダの他の実施の形態の構成例を示すブロック図である。

発明を実施するための最良の形態

以下図面について本発明の実施例を詳述する。その前に、特許請求の範囲に記載の発明の各手段と以下の実施の形態との対応関係を明らかにするために、各手段の後の括弧内に、対応する実施の形態（但し、一例）を付加して、本発明の特徴を記述すると、次のようになる。

即ち、請求の範囲第１項に記載の画像符号化装置は、画像を符号化し、その結果得られる符号化ビットストリームを出力する画像符号化装置であって、画像を受信する受信手段（例えば、図１１や図１２に示すフレームメモリ３１など）と、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化する符号化手段（例えば、図１１や図１２に示すＶＬＣ器３６など）とを備えることを特徴とする。

請求の範囲第４項に記載の画像符号化装置は、イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、符号化開始秒精度絶対時刻を基準とする秒精度の時刻を表す秒精度時刻情報を生成する秒精度時刻情報生成手段（例えば、図３２に示すプログラムの処理ステップＳ３～Ｓ７、および図３７に示すプログラムの処理ステップＳ４３～Ｓ４７など）と、オブジェクトのグループに含まれるＩ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成する詳細時間情報生成手段（例えば、図３２に示すプログラムの処理ステップＳ８、および図３７に示すプログラムの処理ステップＳ４８など）と、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時刻を表す情報として、秒精度時刻情報および詳細時間情報を、対応するＩ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加する付加手段（例えば、図１１や図１２に示すＶＬＣ器３６など）とをさらに備えることを特徴とする。

請求の範囲第１１項に記載の画像復号装置は、画像を復号する画像復号装置であって、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られた符号化ビットストリームを受信する受信手段（例えば、図１７や図１８に示すバッファ１０１など）と、符号

化ビットストリームを復号する復号手段（例えば、図１７や図１８に示すＩＶＬＣ器１０２など）とを備えることを特徴とする。

請求の範囲第１３項に記載の画像復号装置は、オブジェクトのグループには、その符号化を開始した、画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻が含められており、オブジェクトのグループに含まれるオブジェクトの表示時刻を、符号化開始秒精度絶対時刻に基づいて求める表示時刻算出手段（例えば、図３４に示すプログラムの処理ステップＳ２２～Ｓ２７および図３８に示すプログラムの処理ステップＳ５２～Ｓ５７など）をさらに備えることを特徴とする。

請求の範囲第２６項に記載の画像符号化装置は、画像を、その画像を構成するオブジェクトごとに符号化し、その結果得られる符号化ビットストリームを出力する画像符号化装置であって、イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、秒精度の時刻を表す秒精度時刻情報を生成する秒精度時刻情報生成手段（例えば、図３２に示すプログラムの処理ステップＳ３～Ｓ７、および図３７に示すプログラムの処理ステップＳ４３～Ｓ４７など）と、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成する詳細時間情報生成手段（例えば、図３２に示すプログラムの処理ステップＳ８、および図３７に示すプログラムの処理ステップＳ４８など）と、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時刻を表す情報として、秒精度時刻情報および詳細時間情報を、対応するＩ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加する付加手段（例えば、図１１や図１２に示すＶＬＣ器３６など）とを備えることを特徴とする。

請求の範囲第３２項に記載の画像復号装置は、イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとととともに、秒精度の時刻を表す秒精度時刻情報と、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報とが、表示時刻を表す情報として、対応するＩ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加されているとき、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を、秒精度時刻情報および詳細時間情報に基づいて求める表示時刻算出手段（例えば、図３４に示すプログラムの処理ステップＳ２２～Ｓ２７、および図３８に示すプログラムの処理ステップＳ５２～Ｓ５７など）と、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰを、対応する表示時刻にしたがって復号する復号手段（例えば、図１７や図１８に示すＩＶＬＣ器１０２など）とを備えることを特徴とする。

なお、勿論この記載は、各手段を上記したものに限定することを意味するものではない。

図３は、本発明を適用したエンコーダの一実施の形態の構成例を示している。

符号化すべき画像（動画像）データは、ＶＯ（Video Object）構成部１に入力され、ＶＯ構成部１では、そこに入力される画像を構成するオブジェクトごとに、そのシーケンスであるＶＯが構成され、ＶＯＰ構成部2_1～2_Nに出力される。即ち、ＶＯ構成部１においてＮ個のＶＯ＃１～ＶＯ＃Ｎが構成された場合、そのＮ個のＶＯ＃１～ＶＯ＃Ｎは、ＶＯＰ構成部2_1～2_Nにそれぞれ出力される。

具体的には、例えば、符号化すべき画像データが、独立した背景Ｆ１のシーケンスと前景Ｆ２のシーケンスとから構成される場合、ＶＯ構成部１は、例えば、前景Ｆ２のシーケンスを、ＶＯ＃１として、ＶＯＰ構成部2_1に出力するとともに、背

景Ｆ１のシーケンスを、ＶＯ＃２として、ＶＯＰ構成部２$_2$に出力する。

なお、ＶＯ構成部１は、符号化すべき画像データが、例えば、背景Ｆ１と前景Ｆ２とを、既に合成したものである場合、所定のアルゴリズムにしたがって、画像を領域分割することにより、背景Ｆ１と前景Ｆ２とを取り出し、それぞれのシーケンスとしてのＶＯを、対応するＶＯＰ構成部２ｎ（但し、ｎ＝１，２，・・・，Ｎ）に出力する。

ＶＯＰ構成部２ｎは、ＶＯ構成部１の出力から、ＶＯＰ（VO Plane）を構成する。即ち、例えば、各フレームから物体を抽出し、その物体を囲む、例えば、最小の長方形（以下、適宜、最小長方形という）をＶＯＰとする。なお、このとき、ＶＯＰ構成部２ｎは、その横および縦の画素数が、例えば、１６の倍数となるようにＶＯＰを構成する。ＶＯ構成部２ｎは、ＶＯＰを構成すると、そのＶＯＰを、ＶＯＰ符号化部３ｎに出力する。

さらに、ＶＯＰ構成部２ｎは、ＶＯＰの大きさ（例えば、横および縦の長さ）を表すサイズデータ（VOP size）と、フレームにおける、そのＶＯＰの位置（例えば、フレームの最も左上を原点とするときの座標）を表すオフセットデータ（VOP offset）とを検出し、これらのデータも、ＶＯＰ符号化部３ｎに供給する。

ＶＯＰ符号化部３ｎは、ＶＯＰ構成部２ｎの出力を、例えば、ＭＰＥＧや、H.２６３などの規格に準拠した方式で符号化し、その結果得られるビットストリームを、多重化部４に出力する。多重化部４は、ＶＯＰ符号化部３$_1$～３$_N$からのビットストリームを多重化し、その結果得られる多重化データを、例えば、地上波や、衛星回線、ＣＡＴＶ網その他の伝送路５を介して伝送し、または、例えば、磁気ディスク、光磁気ディスク、光ディスク、磁気テープその他の記録媒体６に記録する。

ここで、ＶＯおよびＶＯＰについて説明する。

ＶＯは、ある合成画像のシーケンスが存在する場合の、その合成画像を構成する各オブジェクト（物体）のシーケンスであり、ＶＯＰは、ある時刻におけるＶＯを意味する。即ち、例えば、いま、画像Ｆ１およびＦ２を合成して構成される合成画像Ｆ３がある場合、画像Ｆ１またはＦ２が時系列に並んだものが、それぞれＶＯで

あり、ある時刻における画像Ｆ１またはＦ２が、それぞれＶＯＰである。従って、ＶＯは、異なる時刻の、同一オブジェクトのＶＯＰの集合ということができる。

なお、例えば、画像Ｆ１を背景とするとともに、画像Ｆ２を前景とすると、合成画像Ｆ３は、画像Ｆ２を抜くためのキー信号を用いて、画像Ｆ１およびＦ２を合成することによって得られるが、この場合における画像Ｆ２のＶＯＰには、その画像Ｆ２を構成する画像データ（輝度信号および色差信号）の他、適宜、そのキー信号も含まれるものとする。

画像フレーム（画枠）のシーケンスは、その大きさおよび位置のいずれも変化しないが、ＶＯは、大きさや位置が変化する場合がある。即ち、同一のＶＯを構成するＶＯＰであっても、時刻によって、その大きさや位置が異なる場合がある。

具体的には、図４は、背景である画像Ｆ１と、前景である画像Ｆ２とからなる合成画像を示している。

画像Ｆ１は、例えば、ある自然の風景を撮影したものであり、その画像全体のシーケンスが１つのＶＯ（ＶＯ＃０とする）とされている。また、画像Ｆ２は、例えば、人が歩いている様子を撮影したものであり、その人を囲む最小の長方形のシーケンスが１つのＶＯ（ＶＯ＃１とする）とされている。

この場合、ＶＯ＃０は風景の画像であるから、基本的に、通常の画像のフレームと同様に、その位置および大きさの両方とも変化しない。これに対して、ＶＯ＃１は人の画像であるから、人物が左右に移動したり、また、図面において手前側または奥側に移動することにより、その大きさや位置が変化する。従って、図４は、同一時刻におけるＶＯ＃０およびＶＯ＃１を表しているが、ＶＯの位置や大きさは、時間の経過にともなって変化することがある。

そこで、図３のＶＯＰ符号化部３ｎは、その出力するビットストリームに、ＶＯＰを符号化したデータの他、所定の絶対座標系におけるＶＯＰの位置（座標）および大きさに関する情報も含めるようになされている。なお、図４においては、ＶＯ＃０を構成する、ある時刻のＶＯＰ（画像Ｆ１）の位置を示すベクトルをＯＳＴ０と、その時刻と同一時刻における、ＶＯ＃１のＶＯＰ（画像Ｆ２）の位置を表すベ

クトルをＯＳＴ１と、それぞれ表してある。

次に、図５は、スケーラビリティを実現する、図３のＶＯＰ符号化部３ｎの構成例を示している。即ち、ＭＰＥＧでは、異なる画像サイズやフレームレートに対応するスケーラビリティを実現するスケーラブル符号化方式が導入されており、図５に示したＶＯＰ符号化部３ｎでは、そのようなスケーラビリティを実現することができるようになされている。

ＶＯＰ構成部２ｎからのＶＯＰ（画像データ）、並びにそのサイズデータ（VOP size）、およびオフセットデータ（VOP offset）は、いずれも画像階層化部２１に供給される。

画像階層化部２１は、ＶＯＰから、１以上の階層の画像データを生成する（ＶＯＰの１以上の階層化を行う）。即ち、例えば、空間スケーラビリティの符号化を行う場合においては、画像階層化部２１は、そこに入力される画像データを、そのまま上位レイヤ（上位階層）の画像データとして出力するとともに、それらの画像データを構成する画素数を間引くことなどにより縮小し（解像度を低下させ）、これを下位レイヤ（下位階層）の画像データとして出力する。

なお、入力されたＶＯＰを下位レイヤのデータとするとともに、そのＶＯＰの解像度を、何らかの手法で高くし（画素数を多くし）、これを、上位レイヤのデータとすることなども可能である。

また、階層数は、１とすることが可能であるが、この場合、スケーラビリティは実現されない。なお、この場合、ＶＯＰ符号化部３ｎは、例えば、下位レイヤ符号化部２５だけで構成されることになる。

さらに、階層数は、３以上とすることも可能であるが、ここでは、簡単のために、２階層の場合について説明を行う。

画像階層化部２１は、例えば、時間スケーラビリティ（テンポラルスケーラビリティ）の符号化を行う場合、時刻に応じて、画像データを、下位レイヤまたは上位レイヤのデータとして、例えば、交互に出力する。即ち、例えば、画像階層化部２１は、そこに、あるＶＯを構成するＶＯＰが、ＶＯＰ０，ＶＯＰ１，ＶＯＰ２，Ｖ

ＯＰ３，・・・の順で入力されたとした場合、ＶＯＰ０，ＶＯＰ２，ＶＯＰ４，ＶＯＰ６，・・・を、下位レイヤのデータとして、また、ＶＯＰ１，ＶＯＰ３，ＶＯＰ５，ＶＯＰ７，・・・を、上位レイヤデータとして出力する。なお、時間スケーラビリティの場合は、このようにＶＯＰが間引かれたものが、下位レイヤおよび上位レイヤのデータとされるだけで、画像データの拡大または縮小（解像度の変換）は行われない（但し、行うようにすることも可能である）。

また、画像階層化部２１は、例えば、ＳＮＲ（Signal to Noise Ratio）スケーラビリティの符号化を行う場合、入力された画像データを、そのまま上位レイヤまたは下位レイヤのデータそれぞれとして出力する。即ち、この場合、下位レイヤ並びに上位レイヤの画像データは、同一のデータとなる。

ここで、ＶＯＰごとに符号化を行う場合の空間スケーラビリティについては、例えば、次のような３種類が考えられる。

即ち、例えば、いま、ＶＯＰとして、図４に示したような画像Ｆ１およびＦ２でなる合成画像が入力されたとすると、第１の空間スケーラビリティは、図６に示すように、入力されたＶＯＰ全体（図６（Ａ））を上位レイヤ（Enhancement Layer）とするとともに、そのＶＯＰ全体を縮小したもの（図６（Ｂ））を下位レイヤ（Base Layer）とするものである。

また、第２の空間スケーラビリティは、図７に示すように、入力されたＶＯＰを構成する一部の物体（図７（Ａ）（ここでは、画像Ｆ２に相当する部分））を抜き出して（なお、このような抜き出しは、例えば、ＶＯＰ構成部２ｎにおける場合と同様にして行われ、従って、これにより抜き出された物体も、１つのＶＯＰと考えることができる）、上位レイヤとするとともに、そのＶＯＰ全体を縮小したもの（図７（Ｂ））を下位レイヤとするものである。

さらに、第３の空間スケーラビリティは、図８および図９に示すように、入力されたＶＯＰを構成する物体（ＶＯＰ）を抜き出して、その物体ごとに、上位レイヤおよび下位レイヤを生成するものである。なお、図８は、図４のＶＯＰを構成する背景（画像Ｆ１）から上位レイヤおよび下位レイヤを生成した場合を示しており、

また、図９は、図４のＶＯＰを構成する前景（画像Ｆ２）から上位レイヤおよび下位レイヤを生成した場合を示している。

以上のようなスケーラビリティのうちのいずれを用いるかは予め決められており、画像階層化部２１は、その予め決められたスケーラビリティによる符号化を行うことができるように、ＶＯＰの階層化を行う。

さらに、画像階層化部２１は、そこに入力されるＶＯＰのサイズデータおよびオフセットデータ（それぞれを、以下、適宜、初期サイズデータ、初期オフセットデータという）から、生成した下位レイヤおよび上位レイヤのＶＯＰの所定の絶対座標系における位置を表すオフセットデータと、その大きさを示すサイズデータとを計算（決定）する。

ここで、下位レイヤ並びに上位レイヤのＶＯＰのオフセットデータ（位置情報）およびサイズデータの決定方法について、例えば、上述の第２のスケーラビリティ（図７）を行う場合を例に説明する。

この場合、下位レイヤのオフセットデータＦＰＯＳ＿Ｂは、例えば、図１０（Ａ）に示すように、下位レイヤの画像データを、その解像度および上位レイヤの解像度の違いに基づいて拡大（アップサンプリング）したときに、即ち、下位レイヤの画像を、上位レイヤの画像の大きさと一致するような拡大率（上位レイヤの画像を縮小して下位レイヤの画像を生成したときの、その縮小率の逆数）（以下、適宜、倍率ＦＲという）で拡大したときに、その拡大画像の絶対座標系におけるオフセットデータが、初期オフセットデータと一致するように決定される。また、下位レイヤのサイズデータＦＳＺ＿Ｂも同様に、下位レイヤの画像を倍率ＦＲで拡大したときに得られる拡大画像のサイズデータが初期サイズデータと一致するように決定される。即ち、オフセットデータＦＰＯＳ＿ＢまたはサイズデータＦＳＺ＿Ｂは、それぞれのＦＲ倍か、初期オフセットデータまたは初期サイズデータと一致するように決定される。

一方、上位レイヤのオフセットデータＦＰＯＳ＿Ｅは、例えば、図１０（Ｂ）に示すように、入力されたＶＯＰから抜き出した物体を囲む最小長方形（ＶＯＰ）の、

例えば、左上の頂点の座標が、初期オフセットデータに基づいて求められ、この値に決定される。また、上位レイヤのサイズデータＦＰＯＳ＿Ｅは、入力されたＶＯＰから抜き出した物体を囲む最小長方形の、例えば横および縦の長さに決定される。

従って、この場合、下位レイヤのオフセットデータＦＰＯＳ＿Ｂおよびサイズデータ ＦＰＯＳ＿Ｂを、倍率ＦＲにしたがって変換し（変換後のオフセットデータＦＰＯＳ＿ＢまたはサイズデータＦＰＯＳ＿Ｂを、それぞれ、変換オフセットデータＦＰＯＳ＿Ｂまたは変換サイズデータＦＰＯＳ＿Ｂという）、絶対座標系において、変換オフセットデータＦＰＯＳ＿Ｂに対応する位置に、変換サイズデータＦＳＺ＿Ｂに対応する大きさの画枠を考え、そこに、下位レイヤの画像データをＦＲ倍だけした拡大画像を配置するとともに（図１０（Ａ））、その絶対座標系において、上位レイヤのオフセットデータＦＰＯＳ＿ＥおよびサイズデータＦＰＯＳ＿Ｅにしたがって、上位レイヤの画像を同様に配置すると（図１０（Ｂ））、拡大画像を構成する各画素と、上位レイヤの画像を構成する各画素とは、対応するものどうしが同一の位置に配置されることになる。即ち、この場合、例えば、図１０において、上位レイヤの画像である人の部分と、拡大画像の中の人の部分とは、同一の位置に配置されることになる。

第１および第３のスケーラビリティにおける場合も、同様にして、下位レイヤの拡大画像および上位レイヤの画像を構成する、対応する画素どうしが、絶対座標系において同一の位置に配置されるように、オフセットデータＦＰＯＳ＿ＢおよびＦＰＯＳ＿Ｅ、並びにサイズデータＦＳＺ＿ＢおよびＦＳＺ＿Ｅが決定される。

図５に戻り、画像階層化部２１において生成された上位レイヤの画像データ、オフセットデータＦＰＯＳ＿Ｅ、およびサイズデータＦＳＺ＿Ｅは、遅延回路２２で、後述する下位レイヤ符号化部２５における処理時間だけ遅延され、上位レイヤ符号化部２３に供給される。また、下位レイヤの画像データ、オフセットデータＦＰＯＳ＿Ｂ、およびサイズデータＦＳＺ＿Ｂは、下位レイヤ符号化部２５に供給される。また、倍率ＦＲは、遅延回路２２を介して、上位レイヤ符号化部２３および解像度変換部２４に供給される。

下位レイヤ符号化部２５では、下位レイヤの画像データが符号化され、その結果得られる符号化データ（ビットストリーム）に、オフセットデータＦＰＯＳ＿Ｂおよびサイズデータ ＦＳＺ＿Ｂが含められ、多重化部２６に供給される。

また、下位レイヤ符号化部２５は、符号化データを局所復号し、その結果局所復号結果である下位レイヤの画像データを、解像度変換部２４に出力する。解像度変換部２４は、下位レイヤ符号化部２５からの下位レイヤの画像データを、倍率ＦＲにしたがって拡大（または縮小）することにより、元の大きさに戻し、これにより得られる拡大画像を、上位レイヤ符号化部２３に出力する。

一方、上位レイヤ符号化部２３では、上位レイヤの画像データが符号化され、その結果得られる符号化データ（ビットストリーム）に、オフセットデータＦＰＯＳ＿ＥおよびサイズデータＦＳＺ＿Ｅが含められ、多重化部２６に供給される。なお、上位レイヤ符号化部２３においては、上位レイヤ画像データの符号化は、解像度変換部２４から供給される拡大画像をも参照画像として用いて行われる。

多重化部２６では、上位レイヤ符号化部２３および下位レイヤ符号化部２５の出力が多重化されて出力される。

なお、下位レイヤ符号化部２５から上位レイヤ符号化部２３に対しては、下位レイヤのサイズデータＦＳＺ＿Ｂ、オフセットデータＦＰＯＳ＿Ｂ、動きベクトルＭＶ、フラグＣＯＤなどが供給されており、上位レイヤ符号化部２３では、これらのデータを必要に応じて参照しながら、処理を行うようになされているが、この詳細については、後述する。

次に、図１１は、図５の下位レイヤ符号化部２５の詳細構成例を示している。なお、図中、図１における場合と対応する部分については、同一の符号を付してある。即ち、下位レイヤ符号化部２５は、基本的には、図１のエンコーダと同様に構成されている。

画像階層化部２１（図５）からの画像データ、即ち、下位レイヤのＶＯＰは、図１における場合と同様に、フレームメモリ３１に供給されて記憶され、動きベクトル検出器３２において、マクロブロック単位で動きベクトルの検出が行われる。

下位レイヤ符号化部２５では、下位レイヤの画像データが符号化され、その結果得られる符号化データ（ビットストリーム）に、オフセットデータＦＰＯＳ＿Ｂおよびサイズデータ ＦＳＺ＿Ｂが含められ、多重化部２６に供給される。

また、下位レイヤ符号化部２５は、符号化データを局所復号し、その結果局所復号結果である下位レイヤの画像データを、解像度変換部２４に出力する。解像度変換部２４は、下位レイヤ符号化部２５からの下位レイヤの画像データを、倍率ＦＲにしたがって拡大（または縮小）することにより、元の大きさに戻し、これにより得られる拡大画像を、上位レイヤ符号化部２３に出力する。

一方、上位レイヤ符号化部２３では、上位レイヤの画像データが符号化され、その結果得られる符号化データ（ビットストリーム）に、オフセットデータＦＰＯＳ＿ＥおよびサイズデータＦＳＺ＿Ｅが含められ、多重化部２６に供給される。なお、上位レイヤ符号化部２３においては、上位レイヤ画像データの符号化は、解像度変換部２４から供給される拡大画像をも参照画像として用いて行われる。

多重化部２６では、上位レイヤ符号化部２３および下位レイヤ符号化部２５の出力が多重化されて出力される。

なお、下位レイヤ符号化部２５から上位レイヤ符号化部２３に対しては、下位レイヤのサイズデータＦＳＺ＿Ｂ、オフセットデータＦＰＯＳ＿Ｂ、動きベクトルＭＶ、フラグＣＯＤなどが供給されており、上位レイヤ符号化部２３では、これらのデータを必要に応じて参照しながら、処理を行うようになされているが、この詳細については、後述する。

次に、図１１は、図５の下位レイヤ符号化部２５の詳細構成例を示している。なお、図中、図１における場合と対応する部分については、同一の符号を付してある。即ち、下位レイヤ符号化部２５は、基本的には、図１のエンコーダと同様に構成されている。

画像階層化部２１（図５）からの画像データ、即ち、下位レイヤのＶＯＰは、図１における場合と同様に、フレームメモリ３１に供給されて記憶され、動きベクトル検出器３２において、マクロブロック単位で動きベクトルの検出が行われる。

但し、下位レイヤ符号化部２５の動きベクトル検出器３２には、下位レイヤのＶＯＰのサイズデータＦＳＺ＿ＢおよびオフセットデータＦＰＯＳ＿Ｂが供給されるようになされており、そこでは、このサイズデータＦＳＺ＿ＢおよびオフセットデータＦＰＯＳ＿Ｂに基づいて、マクロブロックの動きベクトルが検出される。

即ち、上述したように、ＶＯＰは、時刻（フレーム）によって、大きさや位置が変化するため、その動きベクトルの検出にあたっては、その検出のための基準となる座標系を設定し、その座標系における動きを検出する必要がある。そこで、ここでは、動きベクトル検出器３２は、上述の絶対座標系を基準となる座標系とし、その絶対座標系に、サイズデータＦＳＺ＿ＢおよびオフセットデータＦＰＯＳ＿Ｂにしたがって、符号化対象のＶＯＰおよび参照画像とするＶＯＰを配置して、動きベクトルを検出するようになされている。

なお、検出された動きベクトル（ＭＶ）は、予測モードとともに、ＶＬＣ器３６および動き補償器４２に供給される他、上位レイヤ符号化部２３（図５）にも供給される。

また、動き補償を行う場合においても、やはり、上述したように、基準となる座標系における動きを検出する必要があるため、動き補償器４２には、サイズデータＦＳＺ＿ＢおよびオフセットデータＦＰＯＳ＿Ｂが供給されるようになされている。

動きベクトルの検出されたＶＯＰは、図１における場合と同様に量子化係数とされてＶＬＣ器３６に供給される。ＶＬＣ器３６には、やはり図１における場合と同様に、量子化係数、量子化ステップ、動きベクトル、および予測モードが供給される他、画像階層化部２１からのサイズデータＦＳＺ＿ＢおよびオフセットデータＦＰＯＳ＿Ｂも供給されており、そこでは、これらのデータすべてが可変長符号化される。

動きベクトルの検出されたＶＯＰは、上述したように符号化される他、やはり図１における場合と同様に局所復号され、フレームメモリ４１に記憶される。この復号画像は、前述したように参照画像として用いられる他、解像度変換部２４（図５）に出力される。

なお、ＭＰＥＧ４においては、ＭＰＥＧ１および２と異なり、Ｂピクチャ（Ｂ－ＶＯＰ）も参照画像として用いられるため、Ｂピクチャも、局所復号され、フレームメモリ４１に記憶されるようになされている（但し、現時点においては、Ｂピクチャが参照画像として用いられるのは上位レイヤについてだけである）。

一方、ＶＬＣ器３６は、図１で説明したように、Ｉ，Ｐ，Ｂピクチャ（Ｉ－ＶＯＰ，Ｐ－ＶＯＰ，Ｂ－ＶＯＰ）のマクロブロックについて、スキップマクロブロックとするかどうかを決定し、その決定結果を示すフラグＣＯＤ，ＭＯＤＢを設定する。このフラグＣＯＤ，ＭＯＤＢは、やはり可変長符号化されて伝送される。さらに、フラグＣＯＤは、上位レイヤ符号化部２３にも供給される。

次に、図１２は、図５の上位レイヤ符号化部２３の構成例を示している。なお、図中、図１１または図１における場合と対応する部分については、同一の符号を付してある。即ち、上位レイヤ符号化部２３は、フレームメモリ５２が新たに設けられていることを除けば、基本的には、図１１の下位レイヤ符号化部２５または図１のエンコーダと同様に構成されている。

画像階層化部２１（図５）からの画像データ、即ち、上位レイヤのＶＯＰは、図１における場合と同様に、フレームメモリ３１に供給されて記憶され、動きベクトル検出器３２において、マクロブロック単位で動きベクトルの検出が行われる。なお、この場合も、動きベクトル検出器３２には、図１１における場合と同様に、上位レイヤのＶＯＰの他、そのサイズデータＦＳＺ＿ＥおよびオフセットデータＦＰＯＳ＿Ｅが供給されるようになされており、動きベクトル検出器３２では、上述の場合と同様に、このサイズデータＦＳＺ＿ＥおよびオフセットデータＦＰＯＳ＿Ｅに基づいて、絶対座標系における上位レイヤのＶＯＰの配置位置が認識され、マクロブロックの動きベクトルが検出される。

ここで、上位レイヤ符号化部２３および下位レイヤ符号化部２５における動きベクトル検出器３２では、図１で説明したように、予め設定されている所定のシーケンスにしたがって、ＶＯＰが処理されていくが、そのシーケンスは、ここでは、例えば、次のように設定されている。

即ち、空間スケーラビリティの場合においては、図１３（Ａ）または図１３（Ｂ）に示すように、上位レイヤまたは下位レイヤのＶＯＰは、例えば、Ｐ，Ｂ，Ｂ，Ｂ，・・・またはＩ，Ｐ，Ｐ，Ｐ，・・・の順でそれぞれ処理されていく。

そして、この場合、上位レイヤの最初のＶＯＰであるＰピクチャ（Ｐ－ＶＯＰ）は、例えば、同時刻における下位レイヤのＶＯＰ（ここでは、Ｉピクチャ（Ｉ－ＶＯＰ））を参照画像として用いて符号化される。また、上位レイヤの２番目以降のＶＯＰであるＢピクチャ（Ｂ－ＶＯＰ）は、例えば、その直前の上位レイヤのＶＯＰおよびそれと同時刻の下位レイヤのＶＯＰを参照画像として用いて符号化される。即ち、ここでは、上位レイヤのＢピクチャは、下位レイヤのＰピクチャと同様に他のＶＯＰを符号化する場合の参照画像として用いられる。

なお、下位レイヤについては、例えば、ＭＰＥＧ１や２、あるいはＨ．２６３における場合と同様に符号化が行われていく。

ＳＮＲスケーラビリティは、空間スケーラビリティにおける倍率ＦＲが１のときと考えられるから、上述の空間スケーラビリティの場合と同様に処理される。

テンポラルスケーラビリティの場合、即ち、例えば、上述したように、ＶＯが、ＶＯＰ０，ＶＯＰ１，ＶＯＰ２，ＶＯＰ３，・・・で構成され、ＶＯＰ１，ＶＯＰ３，ＶＯＰ５，ＶＯＰ７，・・・が上位レイヤとされ（図１４（Ａ））、ＶＯＰ０，ＶＯＰ２，ＶＯＰ４，ＶＯＰ６，・・・が下位レイヤとされた場合においては（図１４（Ｂ））、図１４に示すように、上位レイヤまたは下位レイヤのＶＯＰは、例えば、Ｂ，Ｂ，Ｂ，・・・またはＩ，Ｐ，Ｐ，Ｐ，・・・の順でそれぞれ処理されていく。

そして、この場合、上位レイヤの最初のＶＯＰ１（Ｂピクチャ）は、例えば、下位レイヤのＶＯＰ０（Ｉピクチャ）およびＶＯＰ２（Ｐピクチャ）を参照画像として用いて符号化される。また、上位レイヤの２番目のＶＯＰ３（Ｂピクチャ）は、例えば、その直前にＢピクチャとして符号化された上位レイヤのＶＯＰ１、およびＶＯＰ３の次の時刻（フレーム）における画像である下位レイヤのＶＯＰ４（Ｐピクチャ）を参照画像として用いて符号化される。上位レイヤの３番目のＶＯＰ５

（Ｂピクチャ）も、ＶＯＰ３と同様に、例えば、その直前にＢピクチャとして符号化された上位レイヤのＶＯＰ３、およびＶＯＰ５の次の時刻（フレーム）における画像である下位レイヤのＶＯＰ６（Ｐピクチャ）を参照画像として用いて符号化される。

以上のように、あるレイヤのＶＯＰ（ここでは、上位レイヤ）については、ＰおよびＢピクチャを符号化するための参照画像として、他のレイヤ（スケーラブルレイヤ）（ここでは、下位レイヤ）のＶＯＰを用いることができる。このように、あるレイヤのＶＯＰを符号化するのに、他のレイヤのＶＯＰを参照画像として用いる場合、即ち、ここでは、上位レイヤのＶＯＰを予測符号化するのに、下位レイヤのＶＯＰを参照画像として用いる場合、上位レイヤ符号化部２３（図１２）の動きベクトル検出器３２は、その旨を示すフラグｒｅｆ＿ｌａｙｅｒ＿ｉｄ（階層数が３以上存在する場合、フラグｒｅｆ＿ｌａｙｅｒ＿ｉｄは、参照画像として用いるＶＯＰが属するレイヤを表す）を設定して出力するようになされている。

さらに、上位レイヤ符号化部２３の動きベクトル検出器３２は、ＶＯＰについてのフラグｒｅｆ＿ｌａｙｅｒ＿ｉｄにしたがい、前方予測符号化または後方予測符号化を、それぞれ、どのレイヤのＶＯＰを参照画像として行うかを示すフラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅ（参照画像情報）を設定して出力するようにもなされている。

即ち、例えば、上位レイヤ（Enhancement Layer）のＰピクチャが、その直前に復号（局所復号）される、それと同一のレイヤに属するＶＯＰを参照画像として用いて符号化される場合、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは「００」とされる。また、Ｐピクチャが、その直前に表示される、それと異なるレイヤ（ここでは、下位レイヤ）（Reference Layer）に属するＶＯＰを参照画像として用いて符号化される場合、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは「０１」とされる。さらに、Ｐピクチャが、その直後に表示される、それと異なるレイヤに属するＶＯＰを参照画像として用いて符号化される場合、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは「１０」とされる。また、Ｐピクチャが、それと同時刻における、異なるレイヤの

ＶＯＰを参照画像として用いて符号化される場合、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは「１１」とされる。

一方、例えば、上位レイヤのＢピクチャが、それと同時刻における、異なるレイヤのＶＯＰを前方予測のための参照画像として用い、かつ、その直前に復号される、それと同一のレイヤに属するＶＯＰを後方予測のための参照画像として用いて符号化される場合、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは「００」とされる。また、上位レイヤのＢピクチャが、それと同一のレイヤに属するＶＯＰを前方予測のための参照画像として用い、かつ、その直前に表示される、それと異なるレイヤに属するＶＯＰを後方予測のための参照画像として用いて符号化される場合、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは「０１」とされる。さらに、上位レイヤのＢピクチャが、その直前に復号される、それと同一のレイヤに属するＶＯＰを前方予測のための参照画像として用い、かつその直後に表示される、それと異なるレイヤに属するＶＯＰを後方予測のための参照画像として用いて符号化される場合、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは「１０」とされる。また、上位レイヤのＢピクチャが、その直前に表示される、それと異なるレイヤに属するＶＯＰを前方予測のための参照画像として用い、かつその直後に表示される、それと異なるレイヤに属するＶＯＰを後方予測のための参照画像として用いて符号化される場合、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは「１１」とされる。

ここで、図１３および図１４で説明した予測符号化の方法は、１つの例であり、前方予測符号化、後方予測符号化、または両方向予測符号化における参照画像として、どのレイヤの、どのＶＯＰを用いるかは、例えば、上述した範囲で、自由に設定することが可能である。

なお、上述の場合においては、便宜的に、「空間スケーラビリティ」、「時間スケーラビリティ」、「ＳＮＲスケーラビリティ」という語を用いたが、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅによって、予測符号化に用いる参照画像を設定する場合、空間スケーラビリティや、テンポラルスケーラビリティ、ＳＮＲスケーラビリティを明確に区別することは困難となる。即ち、逆にいえば、フラグｒｅｆ＿ｓｅ

ｌｅｃｔ＿ｃｏｄｅを用いることによって、上述のようなスケーラビリティの区別をせずに済むようになる。

ここで、上述のスケーラビリティとフラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅとを対応付けるとすれば、例えば、次のようになる。即ち、Ｐピクチャについては、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅが「１１」の場合が、フラグｒｅｆ＿ｌａｙｅｒ＿ｉｄが示すレイヤの同時刻におけるＶＯＰを参照画像（前方予測のための参照画像）として用いる場合であるから、これは、空間スケーラビリティまたはＳＮＲスケーラビリティに対応する。そして、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅが「１１」の場合以外は、テンポラルスケーラビリティに対応する。

また、Ｂピクチャについては、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅが「００」の場合が、やはり、フラグｒｅｆ＿ｌａｙｅｒ＿ｉｄが示すレイヤの同時刻におけるＶＯＰを前方予測のための参照画像として用いる場合であるから、これが、空間スケーラビリティまたはＳＮＲスケーラビリティに対応する。そして、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅが「００」の場合以外は、テンポラルスケーラビリティに対応する。

なお、上位レイヤのＶＯＰの予測符号化のために、それと異なるレイヤ（ここでは、下位レイヤ）の、同時刻におけるＶＯＰを参照画像として用いる場合、両者の間に動きはないので、動きベクトルは、常に０（（０，０））とされる。

図１２に戻り、上位レイヤ符号化部２３の動きベクトル検出器３２では、以上のようなフラグｒｅｆ＿ｌａｙｅｒ＿ｉｄおよびｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅが設定され、動き補償器４２およびＶＬＣ器３６に供給される。

また、動きベクトル検出器３２では、フラグｒｅｆ＿ｌａｙｅｒ＿ｉｄおよびｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅにしたがって、フレームメモリ３１を参照するだけでなく、必要に応じて、フレームメモリ５２をも参照して、動きベクトルが検出される。

ここで、フレームメモリ５２には、解像度変換部２４（図５）から、局所復号された下位レイヤの拡大画像が供給されるようになされている。即ち、解像度変換部

２４では、局所復号された下位レイヤのＶＯＰが、例えば、いわゆる補間フィルタなどによって拡大され、これにより、そのＶＯＰを、ＦＲ倍だけした拡大画像、つまり、その下位レイヤのＶＯＰに対応する上位レイヤのＶＯＰと同一の大きさとした拡大画像が生成され、上位レイヤ符号化部２３に供給される。フレームメモリ５２では、このようにして解像度変換部２４から供給される拡大画像が記憶される。

従って、倍率ＦＲが１の場合は、解像度変換部２４は、下位レイヤ符号化部２５からの局所復号されたＶＯＰに対して、特に処理を施すことなく、そのまま、上位レイヤ符号化部２３に供給する。

動きベクトル検出器３２には、下位レイヤ符号化部２５からサイズデータＦＳＺ＿ＢおよびオフセットデータＦＰＯＳ＿Ｂが供給されるとともに、遅延回路２２（図５）からの倍率ＦＲが供給されるようになされており、動きベクトル検出器３２は、フレームメモリ５２に記憶された拡大画像を参照画像として用いる場合、即ち、上位レイヤのＶＯＰの予測符号化に、そのＶＯＰと同時刻における下位レイヤのＶＯＰを参照画像として用いる場合（この場合、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは、Ｐピクチャについては「１１」に、Ｂピクチャについては「００」にされる）、その拡大画像に対応するサイズデータＦＳＺ＿ＢおよびオフセットデータＦＰＯＳ＿Ｂに、倍率ＦＲを乗算する。そして、その乗算結果に基づいて、絶対座標系における拡大画像の位置を認識し、動きベクトルの検出を行う。

なお、動きベクトル検出器３２には、下位レイヤの動きベクトルと予測モードが供給されるようになされており、これは、次のような場合に使用される。即ち、動きベクトル検出部３２は、例えば、上位レイヤのＢピクチャについてのフラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅが「００」である場合において、倍率ＦＲが１であるとき、即ち、ＳＮＲスケーラビリティのとき（但し、この場合、上位レイヤの予測符号化に、上位レイヤのＶＯＰが用いられるので、この点で、ここでいうＳＮＲスケーラビリティは、ＭＰＥＧ２に規定されているものと異なる）、上位レイヤと下位レイヤは同一の画像であるから、上位レイヤのＢピクチャの予測符号化には、下位レイヤの同時刻における画像の動きベクトルと予測モードをそのまま用いること

ができる。そこで、この場合、動きベクトル検出部３２は、上位レイヤのＢピクチャについては、特に処理を行わず、下位レイヤの動きベクトルと予測モードをそのまま採用する。

なお、この場合、上位レイヤ符号化部２３では、動きベクトル検出器３２からＶＬＣ器３６には、動きベクトルおよび予測モードは出力されない（従って、伝送されない）。これは、受信側において、上位レイヤの動きベクトルおよび予測モードを、下位レイヤの復号結果から認識することができるからである。

以上のように、動きベクトル検出器３２は、上位レイヤのＶＯＰの他、拡大画像をも参照画像として用いて、動きベクトルを検出し、さらに、図１で説明したように、予測誤差（あるいは分散）を最小にする予測モードを設定する。また、動きベクトル検出器３２は、例えば、フラグｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅやｒｅｆ＿ｌａｙｅｒ＿ｉｄその他の必要な情報を設定して出力する。

なお、図１２では、下位レイヤ符号化部２５から、下位レイヤにおけるＩまたはＰピクチャを構成するマクロブロックがスキップマクロブロックであるかどうかを示すフラグＣＯＤが、動きベクトル検出器３２、ＶＬＣ器３６、および動き補償器４２に供給されるようになされている。

動きベクトルの検出されたマクロブロックは、上述した場合と同様に符号化され、これにより、ＶＬＣ器３６からは、その符号化結果としての可変長符号が出力される。

なお、上位レイヤ符号化部２３のＶＬＣ器３６は、下位レイヤ符号化部２５における場合と同様に、フラグＣＯＤ，ＭＯＤＢを設定して出力するようになされている。ここで、フラグＣＯＤは、上述したように、ＩまたはＰピクチャのマクロブロックがスキップマクロブロックであるかどうかを示すものであるが、フラグＭＯＤＢは、Ｂピクチャのマクロブロックがスキップマクロブロックであるかどうかを示すものである。

また、ＶＬＣ器３６には、量子化係数、量子化ステップ、動きベクトル、および予測モードの他、倍率ＦＲ、フラグｒｅｆ＿ｓｅｒｅｃｔ＿ｃｏｄｅ，ｒｅｆ＿ｌ

ａｙｅｒ＿ｉｄ、サイズデータＦＳＺ＿Ｅ、オフセットデータＦＰＯＳ＿Ｅ、も供給されるようになされており、ＶＬＣ器３６では、これらのデータがすべて可変長符号化されて出力される。

一方、動きベクトルの検出されたマクロブロックは符号化された後、やはり上述したように局所復号され、フレームメモリ４１に記憶される。そして、動き補償器４２において、動きベクトル検出器３２における場合と同様にして、フレームメモリ４１に記憶された、局所復号された上位レイヤのＶＯＰだけでなく、フレームメモリ５２に記憶された、局所復号されて拡大された下位レイヤのＶＯＰをも参照画像として用いて動き補償が行われ、予測画像が生成される。

即ち、動き補償器４２には、動きベクトルおよび予測モードの他、フラグｒｅｆ＿ｓｅｒｅｃｔ＿ｃｏｄｅ，ｒｅｆ＿ｌａｙｅｒ＿ｉｄ、倍率ＦＲ、サイズデータＦＳＺ＿Ｂ，ＦＳＺ＿Ｅ、オフセットデータＦＰＯＳ＿Ｂ，ＦＰＯＳ＿Ｅが供給されるようになされており、動き補償器４２は、フラグｒｅｆ＿ｓｅｒｅｃｔ＿ｃｏｄｅ，ｒｅｆ＿ｌａｙｅｒ＿ｉｄに基づいて、動き補償すべき参照画像を認識し、さらに、参照画像として、局所復号された上位レイヤのＶＯＰ、または拡大画像を用いる場合には、その絶対座標系における位置と大きさを、サイズデータＦＳＺ＿ＥおよびオフセットデータＦＰＯＳ＿Ｅ、またはサイズデータＦＳＺ＿ＢおよびオフセットデータＦＰＯＳ＿Ｂに基づいて認識し、必要に応じて、倍率ＦＲを用いて予測画像を生成する。

次に、図１５は、図３のエンコーダから出力されるビットストリームを復号するデコーダの一実施の形態の構成例を示している。

このデコーダには、図３のエンコーダから伝送路５または記録媒体６を介して提供されるビットストリームが供給される。即ち、図３のエンコーダから出力され、伝送路５を介して伝送されてくるビットストリームは、図示せぬ受信装置で受信され、あるいは、記録媒体６に記録されたビットストリームは、図示せぬ再生装置で再生され、逆多重化部７１に供給される。

逆多重化部７１では、そこに入力されたビットストリーム（後述するＶＳ

（Video Stream））が受信される。さらに、逆多重化部７１では、入力されたビットストリームが、ＶＯごとのビットストリームＶＯ＃１，ＶＯ＃２，・・・に分離され、それぞれ、対応するＶＯＰ復号部７２ｎに供給される。ＶＯＰ復号部７２ｎでは、逆多重化部７１からのビットストリームから、ＶＯを構成するＶＯＰ（画像データ）、サイズデータ（VOP size）、およびオフセットデータ（VOP offset）が復号され、画像再構成部７３に供給される。

画像再構成部７３では、ＶＯＰ復号部72_1～72_Nそれぞれからの出力に基づいて、元の画像が再構成される。この再構成された画像は、例えば、モニタ７４に供給されて表示される。

次に、図１６は、スケーラビリティを実現する、図１５のＶＯＰ復号部72_Nの構成例を示している。

逆多重化部７１（図１５）から供給されるビットストリームは、逆多重化部９１に入力され、そこで、上位レイヤのＶＯＰのビットストリームと、下位レイヤのＶＯＰのビットストリームとに分離される。上位レイヤのＶＯＰのビットストリームは、遅延回路９２において、下位レイヤ復号部９５における処理の時間だけ遅延された後、上位レイヤ復号部９３に供給され、また、下位レイヤのＶＯＰのビットストリームは、下位レイヤ復号部９５に供給される。

下位レイヤ復号部９５では、下位レイヤのビットストリームが復号され、その結果得られる下位レイヤの復号画像が解像度変換部９４に供給される。また、下位レイヤ復号部９５は、下位レイヤのビットストリームを復号することにより得られるサイズデータＦＳＺ＿Ｂ、オフセットデータＦＰＯＳ＿Ｂ、動きベクトル（ＭＶ）、予測モード、フラグＣＯＤなどの、上位レイヤのＶＯＰを復号するのに必要な情報を、上位レイヤ復号部９３に供給する。

上位レイヤ復号部９３では、遅延回路９２を介して供給される上位レイヤのビットストリームが、下位レイヤ復号部９５および解像度変換部９４の出力を必要に応じて参照することにより復号され、その結果得られる上位レイヤの復号画像、サイズデータＦＳＺ＿Ｅ、およびオフセットデータＦＰＯＳ＿Ｅが出力される。さらに、

上位レイヤ復号部９３は、上位レイヤのビットストリームを復号することにより得られる倍率ＦＲを、解像度変換部９４に出力する。解像度変換部９４では、上位レイヤ復号部９３からの倍率ＦＲを用いて、図５における解像度変換部２４における場合と同様にして、下位レイヤの復号画像が変換される。この変換により得られる拡大画像は、上位レイヤ復号部９３に供給され、上述したように、上位レイヤのビットストリームの復号に用いられる。

次に、図１７は、図１６の下位レイヤ復号部９５の構成例を示している。なお、図中、図２のデコーダにおける場合と対応する部分については、同一の符号を付してある。即ち、下位レイヤ復号部９５は、基本的に、図２のデコーダと同様に構成されている。

逆多重化部９１からの下位レイヤのビットストリームは、バッファ１０１に供給され、そこで受信されて一時記憶される。ＩＶＬＣ器１０２は、その後段のブロックの処理状態に対応して、バッファ１０１からビットストリームを適宜読み出し、そのビットストリームを可変長復号することで、量子化係数、動きベクトル、予測モード、量子化ステップ、サイズデータＦＳＺ＿Ｂ、オフセットデータＦＰＯＳ＿Ｂ、およびフラグＣＯＤなどを分離する。量子化係数および量子化ステップは、逆量子化器１０３に供給され、動きベクトルおよび予測モードは、動き補償器１０７と上位レイヤ復号部９３（図１６）に供給される。また、サイズデータＦＳＺ＿Ｂおよびオフセットデータ ＦＰＯＳ＿Ｂは、動き補償器１０７、画像再構成部７３（図１５）、および上位レイヤ復号部９３に供給され、フラグＣＯＤは、上位レイヤ復号部９３に供給される。

逆量子化器１０３、ＩＤＣＴ器１０４、演算器１０５、フレームメモリ１０６、または動き補償器１０７では、図１１の下位レイヤ符号化部２５の逆量子化器３８、ＩＤＣＴ器３９、演算器４０、フレームメモリ４１、または動き補償器４２における場合とそれぞれ同様の処理が行われることで、下位レイヤのＶＯＰが復号され、画像再構成部７３、上位レイヤ復号部９３、および解像度変換部９４（図１６）に供給される。

次に、図１８は、図１６の上位レイヤ復号部９３の構成例を示している。なお、図中、図２における場合と対応する部分については、同一の符号を付してある。即ち、上位レイヤ復号部９３は、フレームメモリ１１２が新たに設けられていることを除けば、基本的に、図２のエンコーダと同様に構成されている。

逆多重化部９１からの上位レイヤのビットストリームは、バッファ１０１を介してＩＶＬＣ器１０２に供給される。ＩＶＬＣ器１０２は、上位レイヤのビットストリームを可変長復号することで、量子化係数、動きベクトル、予測モード、量子化ステップ、サイズデータＦＳＺ＿Ｅ、オフセットデータＦＰＯＳ＿Ｅ、倍率ＦＲ、フラグｒｅｆ＿ｌａｙｅｒ＿ｉｄ，ｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅ，ＣＯＤ，ＭＯＤＢなどを分離する。量子化係数および量子化ステップは、図１７における場合と同様に、逆量子化器１０３に供給され、動きベクトルおよび予測モードは、動き補償器１０７に供給される。また、サイズデータＦＳＺ＿Ｅおよびオフセットデータ ＦＰＯＳ＿Ｅは、動き補償器１０７および画像再構成部７３（図１５）に供給され、フラグＣＯＤ，ＭＯＤＢ，ｒｅｆ＿ｌａｙｅｒ＿ｉｄ、およびｒｅｆ＿ｓｅｌｅｃｔ＿ｃｏｄｅは、動き補償器１０７に供給される。さらに、倍率ＦＲは、動き補償器１０７および解像度変換部９４（図１６）に供給される。

なお、動き補償器１０７には、上述したデータの他、下位レイヤ復号部９５（図１６）から、下位レイヤの動きベクトル、フラグＣＯＤ、サイズデータＦＳＺ＿Ｂ、およびオフセットデータＦＰＯＳ＿Ｂが供給されるようになされている。また、フレームメモリ１１２には、解像度変換部９４から拡大画像が供給される。

逆量子化器１０３、ＩＤＣＴ器１０４、演算器１０５、フレームメモリ１０６、動き補償器１０７、またはフレームメモリ１１２では、図１２の上位レイヤ符号化部２３の逆量子化器３８、ＩＤＣＴ器３９、演算器４０、フレームメモリ４１、動き補償器４２、またはフレームメモリ５２における場合とそれぞれ同様の処理が行われることで、上位レイヤのＶＯＰが復号され、画像再構成部７３に供給される。

ここで、以上のように構成される上位レイヤ復号部９３および下位レイヤ復号部９５を有するＶＯＰ復号部７２ｎにおいては、上位レイヤについての復号画像、サ

イズデータＦＳＺ＿Ｅ、およびオフセットデータＦＰＯＳ＿Ｅ（以下、適宜、これらをすべて含めて、上位レイヤデータという）と、下位レイヤについての上位レイヤについての復号画像、サイズデータＦＳＺ＿Ｂ、およびオフセットデータＦＰＯＳ＿Ｂ（以下、適宜、これらをすべて含めて、下位レイヤデータという）が得られるが、画像再構成部７３では、この上位レイヤデータまたは下位レイヤデータから、例えば、次のようにして画像が再構成されるようになされている。

即ち、例えば、第１の空間スケーラビリティ（図６）が行われた場合（入力されたＶＯＰ全体が上位レイヤとされるとともに、そのＶＯＰ全体を縮小したものが下位レイヤされた場合）において、下位レイヤデータおよび上位レイヤデータの両方のデータが復号されたときには、画像再構成部７３は、上位レイヤデータのみに基づき、サイズデータＦＳＺ＿Ｅに対応する大きさの上位レイヤの復号画像（ＶＯＰ）を、オフセットデータＦＰＯＳ＿Ｅによって示される位置に配置する。また、例えば、上位レイヤのビットストリームにエラーが生じたり、また、モニタ７４が、低解像度の画像にしか対応していないため、下位レイヤデータのみの復号が行われたときには、画像再構成部７３は、その下位レイヤデータのみに基づき、サイズデータＦＳＺ＿Ｂに対応する大きさの上位レイヤの復号画像（ＶＯＰ）を、オフセットデータＦＰＯＳ＿Ｂによって示される位置に配置する。

また、例えば、第２の空間スケーラビリティ（図７）が行われた場合（入力されたＶＯＰの一部が上位レイヤとされるとともに、そのＶＯＰ全体を縮小したものが下位レイヤとされた場合）において、下位レイヤデータおよび上位レイヤデータの両方のデータが復号されたときには、画像再構成部７３は、サイズデータＦＳＺ＿Ｂに対応する大きさの下位レイヤの復号画像を、倍率ＦＲにしたがって拡大し、その拡大画像を生成する。さらに、画像再構成部７３は、オフセットデータＦＰＯＳ＿ＢをＦＲ倍し、その結果得られる値に対応する位置に、拡大画像を配置する。そして、画像再構成部７３は、サイズデータＦＳＺ＿Ｅに対応する大きさの上位レイヤの復号画像を、オフセットデータＦＰＯＳ＿Ｅによって示される位置に配置する。

この場合、上位レイヤの復号画像の部分が、それ以外の部分に比較して高い解像

度で表示されることになる。

なお、上位レイヤの復号画像を配置する場合においては、その復号画像と、拡大画像とは合成される。

また、図１６（図１５）には図示しなかったが、上位レイヤ復号部９３（ＶＯＰ復号部７２ｎ）から画像再構成部７３に対しては、上述したデータの他、倍率ＦＲも供給されるようになされており、画像再構成部７３は、これを用いて、拡大画像を生成するようになされている。

一方、第２の空間スケーラビリティが行われた場合において、下位レイヤデータのみが復号されたときには、上述の第１の空間スケーラビリティが行われた場合と同様にして、画像が再構成される。

さらに、第３の空間スケーラビリティ（図８、図９）が行われた場合（入力されたＶＯＰを構成する物体ごとに、その物体（オブジェクト）全体を上位レイヤとするとともに、その物体全体を間引いたものを下位レイヤとした場合）においては、上述の第２の空間スケーラビリティが行われた場合と同様にして、画像が再構成される。

上述したように、オフセットデータＦＰＯＳ＿ＢおよびＦＰＯＳ＿Ｅは、下位レイヤの拡大画像および上位レイヤの画像を構成する、対応する画素どうしが、絶対座標系において同一の位置に配置されるようになっているため、以上のように画像を再構成することで、正確な（位置ずれのない）画像を得ることができる。

次に、図３のエンコーダが出力する符号化ビットストリームのシンタクスについて、例えば、MPEG4 規格の Video Verification Model(Version6.0)（以下、適宜、VM6.0 と記述する）を例に説明する。

図１９は、VM6.0 における符号化ビットストリームの構成を示している。

符号化ビットストリームは、ＶＳ（Video Session Class）を単位として構成され、各ＶＳは、１以上のＶＯ（Video Object Class）から構成される。そして、ＶＯは、１以上のＶＯＬ（Video Object Layer Class）から構成され（画像を階層化しないときは１のＶＯＬで構成され、画像を階層化する場合には、その階層数だけ

のＶＯＬで構成される）、ＶＯＬは、ＶＯＰ（Video Object Plane Class）から構成される。

なお、ＶＳは、画像シーケンスであり、例えば、一本の番組や映画などに相当する。

図２０または図２１は、ＶＳまたはＶＯのシンタクスをそれぞれ示している。ＶＯは、画像全体または画像の一部（物体）のシーケンスに対応するビットストリームであり、従って、ＶＳは、そのようなシーケンスの集合で構成される（よって、ＶＳは、例えば、一本の番組などに相当する）。

図２２は、ＶＯＬのシンタクスを示している。

ＶＯＬは、上述したようなスケーラビリティのためのクラスであり、video_object_layer_id で示される番号によって識別される。即ち、例えば、下位レイヤのＶＯＬについての video_object_layer_id は０とされ、また、例えば、上位レイヤのＶＯＬについての video_object_layer_id は１とされる。なお、上述したように、スケーラブルのレイヤの数は２に限られることなく、１や３以上を含む任意の数とすることができる。

また、各ＶＯＬについて、それが画像全体であるのか、画像の一部であるのかは、video_object_layer_shape で識別される。この video_object_layer_shape は、ＶＯＬの形状を示すフラグで、例えば、以下のように設定される。

即ち、ＶＯＬの形状が長方形状であるとき、video_object_layer_shape は、例えば「００」とされる。また、ＶＯＬが、ハードキー（０または１のうちのいずれか一方の値をとる２値（Binary）の信号）によって抜き出される領域の形状をしているとき、video_object_layer_shape は、例えば「０１」とされる。さらに、ＶＯＬが、ソフトキー（０～１の範囲の連続した値（Gray-Scale）をとることが可能な信号）によって抜き出される領域の形状をしているとき（ソフトキーを用いて合成されるものであるとき）、video_object_layer_shape は、例えば「１０」とされる。

ここで、video_object_layer_shape が「００」とされるのは、ＶＯＬの形状が長方形状であり、かつ、そのＶＯＬの絶対座標形における位置および大きさが、時間

とともに変化しない、即ち、一定の場合である。なお、この場合、その大きさ（横の長さと縦の長さ）は、video_object_layer_width と video_object_layer_height によって示される。video_object_layer_width および video_object_layer_height は、いずれも１０ビットの固定長のフラグで、video_object_layer_shape が「００」の場合には、最初に、一度だけ伝送される（これは、video_object_layer_shape が「００」の場合、上述したように、ＶＯＬの絶対座標系における大きさが一定であるからである）。

また、ＶＯＬが、下位レイヤまたは上位レイヤのうちのいずれであるかは、１ビットのフラグである scalability によって示される。ＶＯＬが下位レイヤの場合、scalability は、例えば１とされ、それ以外の場合、scalability は、例えば０とされる。

さらに、ＶＯＬが、自身以外のＶＯＬにおける画像を参照画像として用いる場合、その参照画像が属するＶＯＬは、上述したように、ref_layer_id で表される。なお、ref_layer_id は、上位レイヤについてのみ伝送される。

また、図２２において、hor_sampling_factor_n と hor_sampling_factor_m は、下位レイヤのＶＯＰの水平方向の長さに対応する値と、上位レイヤのＶＯＰの水平方向の長さに対応する値をそれぞれ示す。従って、下位レイヤに対する上位レイヤの水平方向の長さ（水平方向の解像度の倍率）は、式 hor_sampling_factor_n/hor_sampling_factor_m で与えられる。

さらに、図２２において、ver_sampling_factor_n と ver_sampling_factor_m は、下位レイヤのＶＯＰの垂直方向の長さに対応する値と、上位レイヤのＶＯＰの垂直方向の長さに対応する値をそれぞれ示す。従って、下位レイヤに対する上位レイヤの垂直方向の長さ（垂直方向の解像度の倍率）は、式 ver_sampling_factor_n/ver_sampling_factor_m で与えられる。

次に、図２３は、ＶＯＰ（Video Object Plane Class）のシンタクスを示している。

ＶＯＰの大きさ（横と縦の長さ）は、例えば、１０ビット固定長の VOP_width と

VOP_height で表される。また、ＶＯＰの絶対座標系における位置は、例えば、１０ビット固定長の VOP_horizontal_spatial_mc_ref と VOP_vertical_mc_ref で表される。なお、VOP_width または VOP_height は、ＶＯＰの水平方向または垂直方向の長さをそれぞれ表し、これらは、上述のサイズデータＦＳＺ＿ＢやＦＳＺ＿Ｅに相当する。また、VOP_horizontal_spatial_mc_ref または VOP_vertical_mc_ref は、ＶＯＰの水平方向または垂直方向の座標（ｘまたはｙ座標）をそれぞれ表し、これらは、上述のオフセットデータＦＰＯＳ＿ＢやＦＰＯＳ＿Ｅに相当する。

VOP_width, VOP_height, VOP_horizontal_spatial_mc_ref、および VOP_vertical_mc_ref は、video_object_layer_shape が「００」以外の場合にのみ伝送される。即ち、video_object_layer_shape が「００」の場合、上述したように、ＶＯＰの大きさおよび位置はいずれも一定であるから、VOP_width, VOP_height, VOP_horizontal_spatial_mc_ref、および VOP_vertical_mc_ref は伝送する必要がない。この場合、受信側では、ＶＯＰは、その左上の頂点が、例えば、絶対座標系の原点に一致するように配置され、また、その大きさは、図２２で説明した video_object_layer_width および video_object_layer_height から認識される。

図２３において、ref_select_code は、図１９で説明したように、参照画像として用いる画像を表すもので、ＶＯＰのシンタクスにおいて規定されている。

ところで、VM6.0 では、各 VOP(Video Object Plane:従来の Frame に相当する)の表示時刻は、modulo_time_base と、VOP_time_increment（図２３）によって、次のように定められる。

即ち、modulo_time_base は、エンコーダのローカルな時間軸上における時刻を、１秒（1000ms（ミリ秒））の精度で表す。modulo_time_base は、VOP ヘッダの中で伝送されるマーカ（marker）で表現され、必要な数の「1」と、１の「0」とで構成される。modulo_time_base を構成する「1」の数が、最後に（現在から遡って、最も最近に）（直前に）符号化／復号された modulo_time_base によって示された同期点（１秒精度の時刻）からの累積時間を表す。即ち、modulo_time_base が、例えば、「０」の場合は、直前に符号化／復号された modulo_time_base によって示さ

れた同期点からの累積時間が０秒であることを表す。また、modulo_time_base が、例えば、「１０」の場合は、直前に符号化／復号された modulo_time_base によって示された同期点からの累積時間が１秒であることを表す。さらに、modulo_time_base が、例えば、「１１０」の場合は、直前に符号化／復号された modulo_time_base によって示された同期点からの累積時間が２秒であることを表す。以上のように、modulo_time_base の「１」の数が、直前に符号化／復号された modulo_time_base によって示された同期点からの秒数になっている。

なお、VM6.0 では、modulo_time_base について、「This value represents the local time base at the one second resolution unit (1000 milliseconds). It is represented as a marker transmitted in the VOP header. The number of consecutive "1" followed by a "0" indicates the number of seconds has elapsed since the synchronization point marked by the last encoded/decoded modulo_time_base.」と記載されている。

VOP_time_increment は、エンコーダのローカルな時間軸上における時刻を、1ms の精度で表す。VM6.0 では、VOP_time_increment は、I-VOP および P-VOP については、直前に符号化／復号された modulo_time_base によって示された同期点からの時間を表し、B-VOP については、直前に符号化／復号された I-VOP または P-VOP からの相対時間を表す。

なお、VM6.0 では、VOP_time_increment について、「This value represents the local time base in the units of milliseconds. For I and P-VOP's this value is the absolute VOP_time_increment from the synchronization point marked by the last modulo_time_base. For the B-VOP's this value is the relative VOP_time_increment from the last encoded/decoded I- or P-VOP.」と記載されている。

そして、VM6.0 では、「At the encoder, the following formula are used to determine the absolute and relative VOP_time_increments for I/P-VOP's and B-VOP's, respectively.」と記載されている。

即ち、エンコーダにおいて、以下の式を使って、I-VOPおよびP-VOPと、B-VOPとについて、それぞれの表示時刻を符号化する旨が規定されている。

$$tGTB(n) = n \times 1000ms + tEST$$

$$tAVTI = tETB(I/P) - tGTB(n)$$

$$tRVTI = tETB(B) - tETB(I/P) \quad \cdots (1)$$

但し、式（1）において、tGTB（n）は、n番目に符号化されたmodulo_time_baseによって示された同期点の時刻（上述したように、秒精度）を表し、tESTは、エンコーダにおけるVOの符号化開始時刻（VOの符号化が開始された絶対時刻）を表す。また、tAVTIは、I-VOPまたはP-VOPについてのVOP_time_incrementを表し、tETB（I/P）は、エンコーダにおけるI-VOPまたはP-VOPの符号化開始時刻（VOPの符号化が開始された絶対時刻）を表す。さらに、tRVTIは、B-VOPについてのVOP_time_incrementを表し、tETB（B）は、エンコーダにおけるB-VOPの符号化開始時刻を表す。

なお、VM6.0では、式（1）におけるtGTB（n），tEST，tAVTI，tETB（I/P），tRVTI，tETB（B）について、「tGTB（n） is the encoder time base marked by the nth encoded modulo_time_base, tEST is the encoder time base start time, tAVTI is the absolute VOP_time_increment for the I or P-VOP, tETB（I/P） is the encoder time base at the start of the encoding of the I or P-VOP, tRVTI is the relative VOP_time_increment for the B-VOP, and tETB（B） is the encoder time base at the start of the encoding of the B-VOP.」と記載されている。

また、VM6.0では、「At the decoder, the following formula are used to determine the recovered time base of the I/P-VOP's and B-VOP's, respectively:」と記載されている。

即ち、デコーダ側では、以下の式を使って、I-VOPおよびP-VOPと、B-VOPについ

て、それぞれの表示時刻を復号する旨が規定されている。

$$tGTB(n) = n \times 1000ms + tDST$$
$$tDTB(I/P) = tAVTI + tGTB(n)$$
$$tDTB(B) = tRVTI + tDTB(I/P) \quad \cdots (2)$$

但し、式（2）において、tGTB（n）は、n番目に復号されたmodulo_time_baseによって示された同期点の時刻を表し、tDSTは、デコーダにおけるVOの復号開始時刻（VOの復号が開始された絶対時刻）を表す。また、tDTB（I／P）は、デコーダにおけるI-VOPまたはP-VOPの復号開始時刻を表し、tAVTIは、I-VOPまたはP-VOPについてのVOP_time_incrementを表す。さらに、tDTB（B）は、デコーダにおけるB-VOPの復号開始時刻（VOPの復号が開始された絶対時刻）を表し、tRVTIは、B-VOPについてのVOP_time_incrementを表す。

なお、VM6.0では、式（2）におけるtGTB（n），tDST，tDTB（I／P），tAVTI，tDTB（B），tRVTIについて、「tGTB（n） is the encoding time base marked by the nth decoded modulo_time_base, tDST is the decoding time base start time, tDTB（I／P） is the decoding time base at the start of the decoding of the I or P-VOP, tAVTI is the decoding absolute VOP_time_increment for the I or P-VOP, tDTB（B） is the decoding time base at the start of the decoding of the B-VOP, and tRVTI is the decoded relative VOP_time_increment for the B-VOP.」と記載されている。

図２４は、以上の定義に基づいて、modulo_time_baseとVOP_time_incrementとの関係を示した図である。

図２４において、VOは、Ｉ１（Ｉ－ＶＯＰ），Ｂ２（Ｂ－ＶＯＰ），Ｂ３，Ｐ４（Ｐ－ＶＯＰ），Ｂ５，Ｐ６，・・・というＶＯＰのシーケンスで構成されている。いま、VOの符号化／復号開始時刻（絶対時刻）をｔ０とすると、

modulo_time_baseは、時刻ｔ０からの経過時間を、１秒精度で表すから、ｔ０＋１秒、ｔ０＋２秒，・・・という時刻（同期点）を表す。なお、図２４において、表示順は、Ｉ１，Ｂ２，Ｂ３，Ｐ４，Ｂ５，Ｐ６，・・・であるが、符号化／復号順は、Ｉ１，Ｐ４，Ｂ２，Ｂ３，Ｐ６，・・・である。

図２４では（後述する図２８～図３１、および図３６においても同様）、各ＶＯＰについてのVOP_time_incrementを、四角形で囲んだ数字（単位はms）で示してあり、modulo_time_baseによって示される同期点の切り替わりを、▼印で示してある。従って、図２４では、Ｉ１，Ｂ２，Ｂ３，Ｐ４，Ｂ５，Ｐ６についてのVOP_time_incrementが、３５０ｍｓ，４００ｍｓ，８００ｍｓ，５５０ｍｓ，４００ｍｓ，３５０ｍｓとそれぞれされており、Ｐ４およびＰ６において、同期点が切り替わっている。

いま、図２４において、Ｉ１のVOP_time_incrementは、３５０ msであるから、Ｉ１の符号化／復号時刻は、直前に符号化／復号されたmodulo_time_baseによって示された同期点から３５０ｍｓ後の時刻となる。なお、符号化／復号の開始直後は、その開始時刻（符号化／復号開始時刻）ｔ０が同期点となるので、Ｉ１の符号化／復号時刻は、符号化／復号開始時刻ｔ０から３５０ｍｓ後の時刻ｔ０＋３５０ｍｓということになる。

そして、Ｂ２またはＢ３の符号化／復号時刻は、直前に符号化／復号されたI-VOPまたはP-VOPから、VOP_time_incrementだけ経過した時刻であるから、いまの場合、最後の符号化／復号されたＩ１の符号化／復号時刻ｔ０＋３５０ｍｓから、４００ｍｓまたは８００ｍｓ後の時刻ｔ０＋７５０ｍｓまたはｔ０＋１２００ｍｓということに、それぞれなる。

次に、Ｐ４についてであるが、Ｐ４では、modulo_time_baseによって示される同期点が切り替わっており、従って、同期点は時刻ｔ０＋１秒となる。その結果、Ｐ４の符号化／復号時刻は、時刻ｔ０＋１秒から５５０ｍｓ後の時刻（ｔ０＋１）秒＋５５０ｍｓということになる。

Ｂ５の符号化／復号時刻は、直前に符号化／復号されたI-VOPまたはP-VOPから、

VOP_time_increment だけ経過した時刻であるから、いまの場合、最後の符号化／復号されたＰ４の符号化／復号時刻（ｔ０＋１）秒＋５５０ｍｓから、４００ｍｓ後の時刻（ｔ０＋１）秒＋９５０ｍｓということになる。

次に、Ｐ６についてであるが、Ｐ６では、modulo_time_base によって示される同期点が切り替わっており、従って、同期点は時刻ｔ０＋２秒となる。その結果、Ｐ６の符号化／復号時刻は、時刻ｔ０＋２秒から３５０ｍｓ後の時刻（ｔ０＋２）秒＋３５０ｍｓということになる。

なお、VM6.0では、modulo_time_base によって示される同期点の切り替わりは、Ｉ－ＶＯＰとＰ－ＶＯＰとに対してだけ許されており、Ｂ－ＶＯＰに対しては許されていない。

また、VM6.0において、VOP_time_increment が、I－VOP と P－VOP については、直前に符号化／復号された modulo_time_base によって示された同期点からの時間を表すのに対し、B-VOP についてだけは、直前に符号化／復号された I-VOP または P-VOP からの相対時間を表すこととされているのは、主として、次のような理由による。即ち、B-VOP は、表示順で、その B-VOP を挟むＩ－VOP またはＰ－VOP を参照画像として予測符号化されるので、その予測符号化時に参照画像として用いるＩ－VOP またはＰ－VOP に対する重みを、B-VOP から、それを挟むＩ－VOP またはＰ－VOP までの時間的距離に基づいて決めるために、その時間的距離を、B-VOP についての VOP_time_increment としたことによる。

ところで、上述した VM6.0 の VOP_time_increment の定義では、不都合が生じる。即ち、図２４では、B-VOP についての VOP_time_increment が、その B-VOP の直前に符号化／復号される I-VOP または P-VOP からの相対時間ではなく、直前に表示される I-VOP または P-VOP からの相対時間を表すものとしてある。これは、次のような理由による。即ち、例えば、Ｂ２やＢ３に注目した場合、その直前に符号化／復号される I-VOP または P-VOP は、上述した符号化／復号順からいって、Ｐ４である。従って、B-VOP について、VOP_time_increment が、その B-VOP の直前に符号化／復号された I-VOP または P-VOP からの相対時間を表すとした場合、Ｂ２やＢ３につい

てのVOP_time_incrementは、Ｐ４の符号化／復号時刻からの相対時間を表すこととなり、負の値になる。

一方、MPEG4規格では、VOP_time_incrementは、１０ビットとされており、０以上の値のみをとるものとすれば、０～１０２３の範囲の値を表現することができるから、隣接する同期点の間の位置を、時間的に前（図２４において左方向）に位置する同期点を基準として、1ms単位で表すことができる。

しかしながら、VOP_time_incrementが、０以上の値だけでなく、負の値もとることを許すと、例えば、隣接する同期点の間の位置が、時間的に前に位置する同期点を基準として表されたり、また、時間的に後に位置する同期点を基準として表されたりすることになるため、ＶＯＰの符号化時刻や復号時刻を求める処理が煩雑になる。

従って、VM6.0では、上述したように、VOP_time_incrementが、「This value represents the local time base in the units of milliseconds. For I and P-VOP's this value is the absolute VOP_time_increment from the synchronization point marked by the last modulo_time_base. For the B-VOP's this value is the relative VOP_time_increment from the last encoded/decoded I- or P-VOP.」と定義されているが、最後の文の“For the B-VOP's this value is the relative VOP_time_increment from the last encoded/decoded I- or P-VOP”は、“For the B-VOP's this value is the relative VOP_time_increment from the last displayed I- or P-VOP”と変更するべきであり、これにより、VOP_time_incrementが、直前に符号化／復号されたI-VOPまたはP-VOPからの相対時間ではなく、直前に表示されるI-VOPまたはP-VOPからの相対時間を表すものと定義すべきである。

VOP_time_incrementを、このような定義にすることにより、B-VOPについての符号化／復号時刻の計算の基準が、B-VOPよりも過去の表示時刻を持つI/P-VOP（I-VOPまたはP-VOP）の表示時刻になるので、B-VOPについてのVOP_time_incrementは、それが参照するI-VOPが、そのB-VOPよりも先に表示されない限り、常に、正

の値をとることになり、従って、I/P-VOPのVOP_time_incrementも、常に正の値をとることになる。

また、図２４では、さらにVM6.0の定義を変更して、modulo_time_baseおよびVOP_time_incrementによって表される時刻が、符号化／復号時刻ではなく、VOPの表示時刻であるとしてある。即ち、図２４では、VOPのシーケンス上の絶対時刻を考えた場合に、式（１）におけるｔＥＳＴ（Ｉ／Ｐ）および式（２）におけるｔＤＴＢ（Ｉ／Ｐ）は、ＩまたはP-VOPが位置するシーケンス上の絶対時刻を、式（１）におけるｔＥＳＴ（Ｂ）および式（２）におけるｔＤＴＢ（Ｂ）は、B-VOPが位置するシーケンス上の絶対時刻を、それぞれ表すものとしてある。

次に、VM6.0では、式（１）における符号化開始時刻ｔＥＳＴ（the encoder time base start time)は符号化されず、その符号化開始時刻ｔＥＳＴと、各VOPの表示時刻（VOPのシーケンス上の各VOPの位置を表す絶対時刻）との差分情報としてのmodulo_time_baseおよびVOP_time_incrementが符号化される。このため、デコーダ側では、modulo_time_baseおよびVOP_time_incrementを用いて、各VOPの間の相対的な時間関係は定めることができるが、各VOPの絶対的な表示時刻、即ち、各VOPが、VOPのシーケンスの中のどの位置にあるものなのかを定めることはできない。従って、modulo_time_baseおよびVOP_time_incrementだけでは、ビットストリームの途中にアクセスすること、つまり、ランダムアクセスを行うことはできない。

一方、単に符号化開始時刻ｔＥＳＴを符号化すると、デコーダでは、それを用いて、各VOPの絶対時刻を復号することはできるが、常に、符号化ビットストリームの先頭から、符号化開始時刻ｔＥＳＴと、各VOPの相対的な時間情報であるmodulo_time_baseおよびVOP_time_incrementを復号しながら、それを累積して、絶対時刻を管理する必要があり、これは面倒であり、効率的なランダムアクセスができない。

そこで、本実施の形態では、容易に、効率的なランダムアクセスを行うことができるように、VM6.0の符号化ビットストリームの構成（階層）の中に、VOPのシー

ケンス上の絶対時刻を符号化する階層（この階層は、スケーラビリティを実現する階層（上述の下位レイヤや上位レイヤ）ではなく、符号化ビットストリームの階層である）を導入する。この階層は、符号化ビットストリームの先頭だけでなく、適当な位置に挿入できるような符号化ビットストリームの階層とする。

ここでは、この階層として、例えば、MPEG1/2で用いられているGOP(Group of Picture)層と同様に規定されるものを導入する。これにより、MPEG4に独自な符号化ストリームの階層を用いる場合に比べて、MPEG4と、MPEG1/2とのコンパチビリティ（Compatibility）を高めることができる。この新規に導入する階層を、ここでは、ＧＯＶ（またはＧＶＯＰ）（Group Of Video Object Plane）と呼ぶ。

図２５は、VOPのシーケンス上の絶対時刻を符号化するGOV層を導入した符号化ビットストリームの構成例を示している。

GOV層は、ビットストリームの先頭だけでなく、符号化ビットストリームの任意の位置に挿入することができるように、VOL層とVOP層との間に規定されている。

これにより、あるVOL#0が、VOP#0, VOP#1, ・・・, VOP#n, VOP#(n+1), ・・・, VOP#mといったVOPのシーケンスで構成される場合において、GOV層は、その先頭のVOP#0の直前だけでなく、VOP#(n+1)の直前にも挿入することができる。従って、エンコーダにおいて、GOV層は、例えば、符号化ストリームの中の、ランダムアクセスさせたい位置に挿入することができ、従って、GOV層を挿入することで、あるVOLを構成するVOPの一連のシーケンスは、GOV層によって、複数のグループ（以下、適宜、GOVという）に分けられて符号化されることになる。

GOV層のシンタクス（Syntax）は、例えば、図２６に示すように定義される。

同図に示すように、GOV層は、グループスタートコード（group_start_code）、タイムコード（time_code）、クローズドGOP（closed_gop）、ブロークンリンク（broken_link）、ネクストスタートコード（next_start_code()）が順次配置されて構成される。

次に、GOV層のセマンティクス（Semantics）について説明する。なお、GOV層のセマンティクスは、基本的には、MPEG2のGOP層と同様であり、従って、特に記述

しない部分については、MPEG2Video 規格(ISO/IEC13818-2)を参照されたい。

group_start_code は、000001B8 （１６進数）で、GOV の開始位置を示す。

time_code は、図２７に示すような、１ビットの drop_frame_flag、５ビットの time_code_hours、６ビットの time_code_minutes、１ビットの marker_bit、６ビットの time_code_seconds、および６ビットの time_code_pictures の合計２５ビットで構成される。

time_code は、IEC standard publication 461 で規定されている「time and control codes for video tape recorders」に相当する。ここで、MPEG4 では、ビデオのフレームレート（Video の Frame Rate）の概念がないので（従って、VOP は、任意の時刻に表示することができる）、ここでは、time_code がドロップフレームモード（drop_frame_mode）で記述されているか否かを示す drop_frame_flag を利用せず、その値は、例えば、0 に固定する。同様の理由で、time_code_pictures も利用せず、その値は、例えば、0 に固定する。従って、ここでは、time_code は、時刻の時間の単位を表す time_code_hours、時刻の分の単位を表す time_code_minutes、および時刻の秒の単位を表す time_code_seconds によって、GOV の先頭の時刻を表す。その結果、GOV 層の time_code（符号化開始秒精度絶対時刻）は、秒精度で、その先頭の時刻、即ち、その GOV 層の符号化が開始された、VOP のシーケンス上の絶対時刻を表現することとなる。このため、本実施の形態では、秒より細かい精度の時刻（時間）（ここでは、ミリ秒）は、VOP 毎に設定する。

なお、time_code の marker_bit は、符号化ビットストリームにおいて、０が２３個以上連続しないように１とされる。

closed_gop は、MPEG2Video 規格(ISO/IEC 13818-2)における close_gop の定義の記載の中のＩ，Ｐ、またはＢピクチャを、I-VOP, P-VOP、または B-VOP にそれぞれ置き換えたものを意味し、従って、ある GOV の中の B-VOP が、その GOV を構成する VOP だけでなく、他の GOV を構成する VOP を参照画像として符号化されているかどうかを表す。ここで、以下に、MPEG2Video 規格(ISO/IEC 13818-2)における

close_gop の定義について、上述のような置き換えを行った文を示す。

This is a one-bit flag which indicates the nature of the predictions used in the first consecutive B-VOPs (if any) immediately following the first coded I-VOP following the group of plane header. The closed_gop is set to 1 to indicate that these B-VOPs have been encoded using only backward prediction or intra coding. This bit is provided for use during any editing which occurs after encoding. If the previous pictures have been removed by editing, broken_link may be set to 1 so that a decoder may avoid displaying these B-VOPs following the first I-VOP following the group of plane header. However if the closed_gop bit is set to 1, then the editor may choose not to set the broken_link bit as these B-VOPs can be correctly decoded.

broken_link も、MPEG2Video 規格(ISO/IEC 13818-2)における broken_link の記載について、closed_gop における場合と同様の置き換を行ったものを意味し、従って、GOV の先頭の B-VOP が正確に再生することができるかどうかを表す。ここで、以下に、MPEG2Video 規格(ISO/IEC 13818-2)における broken_link の定義について、上述のような置き換えを行った文を示す。

This is a one-bit flag which shall be set to 0 during encoding. It is set to 1 to indicate that the first consecutive B-VOPs (if any) immediately following the first coded I-VOP following the group of plane header may not be correctly decoded because the reference frame which is used for prediction is not available (because of the action of editing). A decoder may use this flag to avoid displaying frames that cannot be correctly decoded.

next_start_code()は、次の GOV の先頭の位置を与える。

以上のような GOV 層を導入し、GOV の符号化を開始する、GOV のシーケンス上の絶対時刻（以下、適宜、符号化開始絶対時刻という）を、GOV のタイムコード

time_code に設定する。さらに、上述のように、GOV 層の time_code は秒精度なので、ここでは、各 VOP の、VOP のシーケンス上の絶対時刻の、さらに細かい精度の部分を、VOP 毎に設定する。

即ち、図２８は、図２６の GOV 層を導入した場合の time_code と、modulo_time_base および VOP_time_increment との関係を示している。

図２８において、GOV は、その先頭から、表示順で、Ｉ１，Ｂ２，Ｂ３，Ｐ４，Ｂ５，Ｐ６が配置されて構成されている。

いま、例えば、GOV の符号化開始絶対時刻を、0h:12m:35sec:350msec（０時１２分３５秒３５０ミリ秒）とすると、GOV の time_code は、上述したように、秒精度（秒単位）なので、0h:12m:35sec とされる（time_code を構成する time_code_hours, time_code_minutes、または time_code_seconds が、それぞれ０，１２、または３５とされる）。一方、Ｉ１の、VOP のシーケンス上の絶対時刻（図２８の GOV を含む VS の符号化前（または復号後）の VOP のシーケンスの絶対時刻）（これは、VOP のシーケンスが表示されるときの、Ｉ１が表示される時刻に相当するので、以下、適宜、表示時刻という）が、例えば、0h:12m:35sec:350msec である場合には、その表示時刻の、秒精度より細かい精度である 350ms は、Ｉ１についての I-VOP の VOP_time_increment に設定されて符号化されるように（Ｉ１についての VOP_time_increment＝350 とされて符号化されるように）、VOP_time_increment のセマンティクスを変更する。

即ち、図２８において、GOV の、表示順で先頭の I-VOP（Ｉ１）の VOP_time_increment は、GOV の time_code と、I-VOP の表示時刻の差分値とする。従って、秒精度による time_code で表された時刻が、GOV の最初の同期点（ここでは、秒精度の時刻を表す点）となる。

なお、図２８において、GOV の２番目以降に配置された VOP であるＢ２，Ｂ３，Ｐ４，Ｂ５，Ｐ６についての VOP_time_increment のセマンティクスは、図２４で説明したように、VM6.0 の定義を変更したものと同様である。

従って、図２８において、Ｂ２またはＢ３の表示時刻は、直前に表示される I-

VOP またはP-VOP の表示時刻から、VOP_time_increment だけ経過した時刻であるから、いまの場合、直前に表示されるＩ１の表示時刻 0h:12m:35s+350ms から、400ms または 800ms 後の時刻 0h:12m:35s:750ms または 0h:12m:36s:200ms ということに、それぞれなる。

次に、Ｐ４についてであるが、Ｐ４では、modulo_time_base によって示される同期点が切り替わっており、従って、同期点は時刻 0h:12m:35s から１秒経過した 0h:12m:36s となる。その結果、Ｐ４の表示時刻は、時刻 0h:12m:36s から 550ms 後の時刻 0h:12m:36:550ms ということになる。

Ｂ５の表示時刻は、直前に表示される I-VOP またはP-VOP から、VOP_time_increment だけ経過した時刻であるから、いまの場合、直前に表示されるＰ４の表示時刻 0h:12m:36:550ms から、400ms 後の時刻 0h:12m:36s:950ms ということになる。

そして、Ｐ６についてであるが、Ｐ６では、modulo_time_base によって示される同期点が切り替わっており、従って、同期点は時刻 0h:12m:35s+2 秒、即ち、0h:12m:37s となる。その結果、Ｐ６の表示時刻は、時刻 0h:12m:37s から 350ms 後の時刻 0h:12m:37s:350ms ということになる。

次に、図２９は、表示順で、先頭の VOP が B-VOP になっている場合の、GOV についての time_code と、modulo_time_base および VOP_time_increment との関係を示している。

図２９において、GOV は、その先頭から、表示順で、Ｂ０，Ｉ１，Ｂ２，Ｂ３，Ｐ４，Ｂ５，Ｐ６が配置されて構成されている。即ち、図２９では、図２８において、Ｉ１の前にＢ０が追加されて、GOV が構成されている。

この場合、GOV の先頭のＢ０についての VOP_time_increment を、その GOV を構成する I/P-VOP の表示時刻を基準として定めることとすると、即ち、例えば、Ｉ１の表示時刻を基準として定めることとすると、その値は負になり、上述したように、都合が悪い。

そこで、GOV の中の、I-VOP よりも先に表示される B-VOP（GOV の中で、最初に表

示されるI-VOPよりも先行して表示されるB-VOP）のVOP_time_incrementについては、そのセマンティクスを、以下のように変更する。

すなわち、そのようなB-VOPのVOP_time_incrementは、GOVのtime_codeの時刻と、B-VOPの表示時刻との差分値とする。この場合、図２９に示すように、Ｂ０の表示時刻が、例えば、0h:12m:35s:200msであり、GOVのtime_codeが、例えば、0h:12m:35sであるときには、Ｂ０のVOP_time_incrementは、350ms（＝0h:12m:35s:200ms－0h:12m:35s）になる。このようにすることで、VOP_time_incrementは、常に正の値になる。

以上のような、VOP_time_incrementについてのセマンティクスの２つの変更により、GOVのtime_codeと、VOPのmodulo_time_baseおよびVOP_time_incrementとを関係付けることができ、さらに、これにより、各VOPが表示される絶対時刻（表示時刻）を特定することができる。

次に、図３０は、I-VOPの表示時刻と、それから予測されるB-VOPの表示時刻との間隔が１秒（正確には、１．０２３秒）より大きい場合の、GOVについてのtime_codeと、modulo_time_baseおよびVOP_time_incrementとの関係を示している。

図３０において、GOVは、表示順で、Ｉ１，Ｂ２，Ｂ３，Ｂ４，Ｐ６が順次配置されて構成されており、Ｂ４が、直前に表示されるI-VOPであるＩ１の表示時刻よりも、１秒より後の時刻において表示されるようになされている。

この場合、上述のようにセマンティクスを変更したVOP_time_incrementによって、Ｂ４の表示時刻を符号化しようとしても、VOP_time_incrementは、上述のように１０ビットであるため、1023までしか表現できず、1.023秒より長い時間を表現することはできない。そこで、VOP_time_incrementのセマンティクスをさらに変更するとともに、modulo_time_baseのセマンティクスをも変更し、このような場合であっても対応できるようにする。

ここでは、例えば、次のような第１または第２の方法のいずれかによって対応する。

即ち、第１の方法では、I/P-VOPの表示時刻と、それから予測されるB-VOPの表

示時刻との間の時間を、ミリ秒精度で求め、その時間を、秒の単位までは、modulo_time_base で表現し、残りのミリ秒の精度を、VOP_time_increment で表現する。

図３０に示した場合において、第１の方法にしたがって、modulo_time_base および VOP_time_increment を符号化した場合の、GOV についての time_code と、modulo_time_base および VOP_time_increment との関係を、図３１に示す。

即ち、第１の方法では、modulo_time_base の付加を、I-VOP および P-VOP だけでなく、B-VOP に対しても許可する。そして、B-VOP に付加されている modulo_time_base は、同期点の切り替わりではなく、直前に表示される I/P-VOP の表示時刻からの秒単位の繰り上がりを表すものとする。

さらに、第１の方法では、B-VOP に付加される modulo_time_base によって示される、直前に表示される I/P-VOP の表示時刻からの秒単位の繰り上がり後の時刻を、その B-VOP の表示時刻から減算した値を、その VOP_time_increment として設定する。

従って、第１の方法によれば、図３０において、例えば、Ｉ１の表示時刻を、0h:12m:35s:350ms とするとともに、Ｂ４の表示時刻を、0h:12m:36s:550ms とすると、Ｉ１とＢ４との表示時刻の差は、１秒以上の 1200msec であるから、Ｂ４には、図３１に示すように、直前に表示されるＩ１の表示時刻からの秒単位の繰り上がりを示す modulo_time_base（図３１において、▼印で示す）が付加される。具体的には、Ｂ４に付加される modulo_time_base は、1200ms の１秒の位の値である１秒の繰り上がりを表す「１０」とされる。そして、Ｂ４の VOP_time_increment は、図３１に示すように、Ｉ１とＢ４との表示時刻の差の、１秒未満の値（Ｂ４の表示時刻から、その modulo_time_base によって示される、直前に表示される I/P-VOP であるＩ１の表示時刻からの秒単位の繰り上がり後の時刻を減算した値）であるである２００とされる。

以上のような、第１の方法による modulo_time_base と VOP_time_increment についての処理は、エンコーダ側では、例えば、図１１および図１２に示したＶＬＣ器３６において、デコーダ側では、例えば、図１７および図１８に示したＩＶＬＣ器

１０２において、それぞれ行われる。

そこで、まず、図３２のフローチャートを参照して、ＶＬＣ器３６が行うI/P-VOPのmodulo_time_baseおよびVOP_time_incrementに関する処理について説明する。

ＶＬＣ器３６は、VOPのシーケンスを、GOVごとに分けて処理を行うようになされている。なお、GOVは、少なくとも１のイントラ符号化されるVOPを含むように構成される。

ＶＬＣ器３６は、GOVを受信すると、例えば、その受信時刻を、そのGOVの符号化開始絶対時刻とし、その符号化開始絶対時刻の秒精度まで（秒の桁までの符号化開始絶対時刻）を、time_codeとして符号化して、符号化ビットストリームの中に含める。その後、ＶＬＣ器３６は、GOVを構成するI/P-VOPを受信するごとに、そのI/P-VOPを注目I/P-VOPとして、図３２のフローチャートにしたがい、注目I/P-VOPのmodulo_time_baseおよびVOP_time_incrementを求めて、符号化する。

即ち、ＶＬＣ器３６では、まず最初に、ステップＳ１において、modulo_time_baseに０Ｂ（Ｂは２進数を表す）がセットされるとともに、VOP_time_incrementに０がセットされることにより、modulo_time_baseおよびVOP_time_incrementがリセットされる。

そして、ステップＳ２に進み、注目I/P-VOPが、処理対象としているGOV（処理対象GOV）の中で、最初に表示されるI-VOP（First I-VOP）であるかどうかが判定される。ステップＳ２において、注目I/P-VOPが、処理対象GOVの中で、最初に表示されるI-VOPであると判定された場合、ステップＳ４に進み、処理対象GOVのtime_codeと、注目I/P-VOP（ここでは、処理対象GOVの中で、最初に表示されるI-VOP）の表示時刻の秒精度との差分、即ち、time_codeと、注目I/P-VOPの表示時刻の秒の桁までとの差分が求められ、変数Ｄにセットされて、ステップＳ５に進む。

また、ステップＳ２において、注目I/P-VOPが、処理対象GOVの中で、最初に表示されるI-VOPでないと判定された場合、ステップＳ３に進み、注目I/P-VOPの表示時刻の秒の桁までと、その直前に表示されるI/P-VOP（処理対象GOVを構成する

VOPのうちの、注目I/P-VOPの直前に表示されるI/P-VOP）（Last display I/P-VOP）の表示時刻の秒の桁までとの差分値が求められ、その差分値が、変数Ｄにセットされて、ステップＳ５に進む。

ステップＳ５では、変数Ｄが０に等しいかどうか、即ち、time_codeと、注目I/P-VOPの表示時刻の秒の桁までとの差分、または注目I/P-VOPの表示時刻の秒の桁までと、その直前に表示されるI/P-VOPの表示時刻の秒の桁までとの差分値が０秒であるかどうかが判定される。ステップＳ５において、変数Ｄが０に等しくないと判定された場合、即ち、変数Ｄが１以上である場合、ステップＳ６に進み、modulo_time_baseのＭＳＢ（Most Significant Bit）として、１が付加される。即ち、この場合、modulo_time_baseが、例えば、リセット直後の０Ｂであるときには、１０Ｂとされ、また、modulo_time_baseが、例えば、１０Ｂであるときには、１１０Ｂとされる。

そして、ステップＳ７に進み、変数Ｄが１だけデクリメントされ、ステップＳ５に戻る。その後、ステップＳ５において、変数Ｄが０に等しいと判定されるまで、ステップＳ５～Ｓ７の処理が繰り返される。即ち、これにより、modulo_time_baseは、time_codeと、注目I/P-VOPの表示時刻の秒の桁までとの差分、または注目I/P-VOPの表示時刻の秒の桁までと、その直前に表示されるI/P-VOPの表示時刻の秒の桁までとの差分値に対応する秒数と同一の数だけ「１」が連続し、その最後に０が付加された値とされる。

そして、ステップＳ５において、変数Ｄが０に等しいと判定された場合、ステップＳ８に進み、VOP_time_incrementに、注目I/P-VOPの表示時刻の秒精度より細かい精度の時刻、即ち、ミリ秒単位の時刻がセットされ、処理を終了する。

以上のようにして求められた注目I/P-VOPのmodulo_time_baseおよびVOP_time_incrementは、ＶＬＣ回路３６において、注目I/P-VOPに付加され、これにより、符号化ビットストリームの中に含められる。

なお、modulo_time_baseおよびVOP_time_increment、並びにtime_codeは、ＶＬＣ回路３６において可変長符号化される。

次に、ＶＬＣ器３６は、処理対象GOVを構成するB-VOPを受信するごとに、そのB-VOPを注目B-VOPとして、図３３のフローチャートにしたがい、注目B-VOPのmodulo_time_baseおよびVOP_time_incrementを求めて、符号化する。

即ち、ＶＬＣ器３６では、まず最初に、ステップＳ１１において、図３２のステップＳ１における場合と同様に、modulo_time_baseおよびVOP_time_incrementがリセットされる。

そして、ステップＳ１２に進み、注目B-VOPが、処理対象GOVの中で、最初に表示されるI-VOP（First I-VOP）よりも先に表示されるものであるかどうかが判定される。ステップＳ１２において、注目B-VOPが、処理対象GOVの中で、最初に表示されるI-VOPよりも先に表示されるものであると判定された場合、ステップＳ１４に進み、処理対象GOVのtime_codeと、注目B-VOP（ここでは、処理対象GOVの中で、最初に表示されるI-VOPよりも先行して表示されるB-VOP）の表示時刻との差分が求められ、変数Ｄにセットされて、ステップＳ１５に進む。従って、ここでは、変数Ｄには、ミリ秒精度の時間（ミリ秒の桁までの時間）がセットされる（これに対して、図３２における変数Ｄには、上述したように、秒精度の時間がセットされる）。

また、ステップＳ１２において、注目B-VOPが、処理対象GOVの中で、最初に表示されるI-VOPよりも後に表示されるものであると判定された場合、ステップＳ１４に進み、注目B-VOPの表示時刻と、その直前に表示されるI/P-VOP（処理対象GOVを構成するVOPのうちの、注目B-VOPの直前に表示されるI/P-VOP）（Last display I/P-VOP）の表示時刻との差分値が求められ、その差分値が、変数Ｄにセットされて、ステップＳ１５に進む。

ステップＳ１５では、変数Ｄが１より大であるかどうか、即ち、time_codeと、注目B-VOPの表示時刻との差分値、または注目B-VOPの表示時刻と、その直前に表示されるI/P-VOPの表示時刻との差分値が、１秒より大であるかどうかが判定される。ステップＳ５において、変数Ｄが１より大であると判定された場合、即ち、modulo_time_baseのＭＳＢとして、１が付加され、ステップＳ１７に進む。ステッ

プＳ１７では、変数Ｄが１だけデクリメントされ、ステップＳ１５に戻る。そして、ステップＳ１５において、変数Ｄが１より大でないと判定されるまで、ステップＳ１５～Ｓ１７の処理が繰り返される。即ち、これにより、modulo_time_base は、time_code と、注目 B-VOP の表示時刻との差分値、または注目 B-VOP の表示時刻と、その直前に表示される I/P-VOP の表示時刻との差分値に対応する時間の秒数と同一の数だけ「１」が連続し、その最後に０が付加された値とされる。

そして、ステップＳ１５において、変数Ｄが１より大でないと判定された場合、ステップＳ８に進み、そのときの変数Ｄの値、即ち、time_code と、注目 B-VOP の表示時刻との差分値、または注目 B-VOP の表示時刻と、その直前に表示される I/P-VOP の表示時刻との差分値の秒の位より下の位（ミリ秒単位の時間）が、VOP_time_increment にセットされ、処理を終了する。

以上のようにして求められた注目 B-VOP の modulo_time_base および VOP_time_increment は、ＶＬＣ回路３６において、注目 B-VOP に付加され、これにより、符号化ビットストリームの中に含められる。

次に、ＩＶＬＣ器１０２は、上述のようにして、ＶＬＣ器３６が、VOP のシーケンスを、GOV ごとに分けて処理を行うことにより出力する符号化ストリームの中の VOP の表示時刻を、各 VOP についての符号化データを受信するごとに、その VOP を注目 VOP として処理を行うことにより認識し、その表示時刻に VOP が表示されるように、可変長復号を行うようになされている。即ち、ＩＶＬＣ器１０２は、GOV を受信すると、その GOV の time_code を認識し、その GOV を構成する I/P-VOP を受信するごとに、その I/P-VOP を注目 I/P-VOP として、図３４のフローチャートにしたがい、注目 I/P-VOP の modulo_time_base および VOP_time_increment に基づき、その表示時刻を求める。

即ち、ＩＶＬＣ器１０２では、まず最初に、ステップＳ２１において、注目 I/P-VOP が、処理対象 GOV の中で、最初に表示される I-VOP (First I-VOP) であるかどうかが判定される。ステップＳ２１において、注目 I/P-VOP が、処理対象 GOV の中で、最初に表示される I-VOP であると判定された場合、ステップＳ２３に進み、変

数Tに、処理対象GOVのtime_codeがセットされ、ステップＳ２４に進む。

また、ステップＳ２１において、注目I/P-VOPが、処理対象GOVの中で、最初に表示されるI-VOPでないと判定された場合、ステップＳ２２に進み、注目I/P-VOPの直前に表示されるI/P-VOP（処理対象GOVを構成するVOPのうちの、注目I/P-VOPの直前に表示されるI/P-VOP）（Last display I/P-VOP）の表示時刻の秒の桁までが、変数Tにセットされて、ステップＳ２４に進む。

ステップＳ２４では、注目I/P-VOPに付加されているmodulo_time_baseが０Ｂに等しいかどうかが判定される。ステップＳ２４において、注目I/P-VOPに付加されているmodulo_time_baseが０Ｂに等しくないと判定された場合、即ち、注目I/P-VOPに付加されているmodulo_time_baseに１が含まれる場合、ステップＳ２５に進み、そのmodulo_time_baseのＭＳＢの１が削除され、ステップＳ２６に進む。ステップＳ２６では、変数Tが１秒だけインクリメントされ、ステップＳ２４に戻り、以下、ステップＳ２４において、注目I/P-VOPに付加されているmodulo_time_baseが０Ｂに等しいと判定されるまで、ステップＳ２４～Ｓ２６の処理を繰り返す。これにより、変数Tは、注目I/P-VOPに、最初に付加されていたmodulo_time_baseの１の数に対応する秒数だけインクリメントされる。

そして、ステップＳ２４において、注目I/P-VOPに付加されているmodulo_time_baseが０Ｂに等しいと判定された場合、ステップＳ２７に進み、変数Tに、VOP_time_incrementが表すミリ秒精度の時刻が加算され、その加算値が、注目I/P-VOPの表示時刻として認識されて、処理を終了する。

次に、ＩＶＬＣ器１０２では、GOVを構成するB-VOPを受信した場合には、図３５のフローチャートにしたがい、注目B-VOPのmodulo_time_baseおよびVOP_time_incrementに基づき、その表示時刻が求められる。

即ち、ＩＶＬＣ器１０２では、まず最初に、ステップＳ３１において、注目B-VOPが、処理対象GOVの中で、最初に表示されるI-VOP（First I-VOP）よりも先に表示されるものであるかどうかが判定される。ステップＳ３１において、注目B-VOPが、処理対象GOVの中で、最初に表示されるI-VOPよりも先に表示されるもの

であると判定された場合、ステップＳ３３に進み、以下、ステップＳ３３～Ｓ３７において、図３４のステップＳ２３～Ｓ２７における場合とそれぞれ同様の処理が行われることにより、注目B-VOPの表示時刻が求められる。

一方、ステップＳ３１において、注目B-VOPが、処理対象GOVの中で、最初に表示されるI-VOPよりも後に表示されるものであると判定された場合、ステップＳ３２に進み、以下、ステップＳ３２，Ｓ３４～Ｓ３７において、図３４のステップＳ２２，Ｓ２４～Ｓ２７における場合とそれぞれ同様の処理が行われることにより、注目B-VOPの表示時刻が求められる。

次に、第２の方法では、I-VOPの表示時刻と、それから予測されるB-VOPの表示時刻との間の時間を、秒の桁まで求め、その値を、modulo_time_baseで表現し、B-VOPの表示時刻のミリ秒精度を、VOP_time_incrementで表現する。即ち、VM6.0では、上述したように、B-VOPの予測符号化時に参照画像として用いるI－VOPまたはP－VOPに対する重みを、B-VOPから、それを挟むI－VOPまたはP－VOPまでの時間的距離に基づいて決めるのに、その時間的距離を、B-VOPについてのVOP_time_incrementとしており、このため、I－VOPおよびP－VOPについてのVOP_time_incrementが、直前に符号化／復号されたmodulo_time_baseによって示された同期点からの時間を表すのと異なっているが、B-VOPの表示時刻と、それを挟むI－VOPまたはP－VOPの表示時刻が分かれば、それらの間の時間的距離は、差分をとるだけで求めることができ、従って、B-VOPについてのVOP_time_incrementだけを、I－VOPおよびP－VOPについてのVOP_time_incrementと異なる取り扱いとする必要性は小さい。むしろ、処理効率の観点からは、Ｉ，Ｂ，ＰのすべてのVOPのVOP_time_incrment（詳細時間情報）、さらには、modulo_time_base（秒精度時刻情報）は、同一の取り扱いとするのが望ましい。

そこで、第２の方法では、B-VOPについてのmodulo_time_baseおよびVOP_time_incrementを、I/P－VOPについてのものと同一の取り扱いとしている。

例えば、図３０に示した場合において、第２の方法にしたがって、modulo_time_baseおよびVOP_time_incrementを符号化した場合の、GOVについての

time_code と、modulo_time_base および VOP_time_increment との関係を、図３６に示す。

即ち、第２の方法でも、modulo_time_base の付加を、I-VOP および P-VOP だけでなく、B-VOP に対しても許可する。そして、B-VOP に付加される modulo_time_base も、I/P-VOP に付加される modulo_time_base と同様に、同期点の切り替わりを表すものとする。

さらに、第２の方法では、B-VOP に付加されている modulo_time_base によって示される同期点の時刻を、その B-VOP の表示時刻から減算した値が、その VOP_time_increment として設定される。

従って、第２の方法によれば、図３０において、GOV の time_code が表す時刻である、GOV の最初の同期点から、時刻 time_code＋１秒で示される同期点までの間に表示されるＩ１またはＢ２の modulo_time_base は、いずれも０Ｂとされるとともに、それぞれの VOP_time_increment は、Ｉ１またはＢ２の表示時刻の秒の位より下の位のミリ秒単位の値がセットされる。また、時刻 time_code＋１秒で示される同期点から、時刻 time_code＋２秒で示される同期点までの間に表示されるＢ３またはＢ４の modulo_time_base は、いずれも１０Ｂとされるとともに、それぞれの VOP_time_increment は、Ｂ３またはＢ４の表示時刻の秒の位より下の位のミリ秒単位の値がセットされる。さらに、時刻 time_code＋２秒で示される同期点から、時刻 time_code＋３秒で示される同期点までの間に表示されるＰ５の modulo_time_base は１１０Ｂとされるとともに、その VOP_time_increment は、Ｐ５の表示時刻の秒の位より下の位のミリ秒単位の値がセットされる。

図３０において、例えば、上述したように、Ｉ１の表示時刻を、0h:12m:35s:350ms とするとともに、Ｂ４の表示時刻を、0h:12m:36s:550ms とすると、Ｉ１またはＢ４の modulo_time_base は、上述したように、それぞれ０Ｂまたは１０Ｂとされる。また、Ｉ１またはＢ４の VOP_time_increment は、それぞれ、表示時刻のミリ秒単位である 350ms または 550ms とされる。

以上のような、第２の方法による modulo_time_base と VOP_time_increment につ

いての処理も、例えば、第１の方法による場合と同様に、図１１および図１２に示したＶＬＣ器３６、並びに図１７および図１８に示したＩＶＬＣ器１０２において行われる。

即ち、ＶＬＣ器３６では、I/P-VOPについては、図３２における場合と同様にして、modulo_time_baseおよびVOP_time_incrementが求められる。

また、B-VOPについては、ＶＬＣ器３６において、GOVを構成するB-VOPを受信するごとに、そのB-VOPを注目B-VOPとして、図３７のフローチャートにしたがい、注目B-VOPのmodulo_time_baseおよびVOP_time_incrementが求められる。

即ち、ＶＬＣ器３６では、まず最初に、ステップＳ４１において、図３２のステップＳ１における場合と同様にして、modulo_time_baseおよびVOP_time_incrementがリセットされる。

そして、ステップＳ４２に進み、注目B-VOPが、処理対象としているGOV（処理対象GOV）の中で、最初に表示されるI-VOP（First I-VOP）よりも先行して表示されるものであるかどうかが判定される。ステップＳ１２において、注目B-VOPが、処理対象GOVの中で、最初に表示されるI-VOPよりも先行して表示されるものであると判定された場合、ステップＳ４４に進み、処理対象GOVのtime_codeと、注目B-VOPの表示時刻の秒精度との差分、即ち、time_codeと、注目B-VOPの表示時刻の秒の桁までとの差分が求められ、変数Ｄにセットされて、ステップＳ４５に進む。

また、ステップＳ４２において、注目B-VOPが、処理対象GOVの中で、最初に表示されるI-VOPよりも後に表示されるものであると判定された場合、ステップＳ４３に進み、注目B-VOPの表示時刻の秒の桁までと、その直前に表示されるI/P-VOP（処理対象GOVを構成するVOPのうちの、注目B-VOPの直前に表示されるI/P-VOP）（Last display I/P-VOP）の表示時刻の秒の桁までとの差分値が求められ、その差分値が、変数Ｄにセットされて、ステップＳ４５に進む。

ステップＳ４５では、変数Ｄが０に等しいかどうか、即ち、time_codeと、注目B-VOPの表示時刻の秒の桁までとの差分、または注目B-VOPの表示時刻の秒の桁までと、その直前に表示されるI/P-VOPの表示時刻の秒の桁までとの差分値が０秒で

あるかどうかが判定される。ステップＳ４５において、変数Ｄが０に等しくないと判定された場合、即ち、変数Ｄが１以上である場合、ステップＳ４６に進み、modulo_time_base のＭＳＢとして、１が付加される。

そして、ステップＳ４７に進み、変数Ｄが１だけデクリメントされ、ステップＳ４５に戻る。その後、ステップＳ４５において、変数Ｄが０に等しいと判定されるまで、ステップＳ４５～Ｓ４７の処理が繰り返される。即ち、これにより、modulo_time_base は、time_code と、注目 B-VOP の表示時刻の秒の桁までとの差分、または注目 B-VOP の表示時刻の秒の桁までと、その直前に表示される I/P-VOP の表示時刻の秒の桁までとの差分値に対応する秒数と同一の数だけ「１」が連続し、その最後に０が付加された値とされる。

そして、ステップＳ４５において、変数Ｄが０に等しいと判定された場合、ステップＳ４８に進み、VOP_time_increment に、注目 B-VOP の表示時刻の秒精度より細かい精度の時刻、即ち、ミリ秒単位の時刻がセットされ、処理を終了する。

一方、ＩＶＬＣ器１０２では、I/P-VOP については、上述の図３４における場合と同様にして、modulo_time_base および VOP_time_increment に基づき、その表示時刻が求められる。

また、Ｂ－VOP については、ＩＶＬＣ器１０２において、GOV を構成する B-VOP を受信するごとに、その B-VOP を注目 B-VOP として、図３８のフローチャートにしたがい、注目 B-VOP の modulo_time_base および VOP_time_increment に基づき、その表示時刻が求められる。

即ち、ＩＶＬＣ器１０２では、まず最初に、ステップＳ５１において、注目 B-VOP が、処理対象 GOV の中で、最初に表示される I-VOP (First I-VOP) よりも先行して表示されるものであるかどうかが判定される。ステップＳ５１において、注目 B-VOP が、処理対象 GOV の中で、最初に表示される I-VOP よりも先行して表示されるものであると判定された場合、ステップＳ５２に進み、変数Ｔに、処理対象 GOV の time_code がセットされ、ステップＳ５４に進む。

また、ステップＳ５１において、注目 B-VOP が、処理対象 GOV の中で、最初に表

示されるI-VOPよりも後に表示されるものであると判定された場合、ステップＳ５３に進み、注目B-VOPの直前に表示されるI/P-VOP（処理対象GOVを構成するVOPのうちの、注目B-VOPの直前に表示されるI/P-VOP）（Last display I/P-VOP）の表示時刻の秒の桁までが、変数Ｔにセットされて、ステップＳ５４に進む。

ステップＳ５４では、注目B-VOPに付加されているmodulo_time_baseがＯＢに等しいかどうかが判定される。ステップＳ５４において、注目B-VOPに付加されているmodulo_time_baseがＯＢに等しくないと判定された場合、即ち、注目B-VOPに付加されているmodulo_time_baseに１が含まれる場合、ステップＳ５５に進み、そのmodulo_time_baseのＭＳＢの１が削除され、ステップＳ５６に進む。ステップＳ５６では、変数Ｔが１秒だけインクリメントされ、ステップＳ５４に戻り、以下、ステップＳ５４において、注目B-VOPに付加されているmodulo_time_baseがＯＢに等しいと判定されるまで、ステップＳ５４～Ｓ５６の処理を繰り返す。これにより、変数Ｔは、注目B-VOPに、最初に付加されていたmodulo_time_baseの１の数に対応する秒数だけインクリメントされる。

そして、ステップＳ５４において、注目B-VOPに付加されているmodulo_time_baseがＯＢに等しいと判定された場合、ステップＳ５７に進み、変数Ｔに、VOP_time_incrementが表すミリ秒精度の時刻が加算され、その加算値が、注目B-VOPの表示時刻として認識されて、処理を終了する。

以上のように、符号化ビットストリームの構成（階層）の中に、符号化開始絶対時刻を符号化するGOV層を導入し、このGOV層を、ビットストリームの先頭だけでなく、適当な位置に挿入できるようなものとするとともに、VM6.0で規定されていたmodulo_time_baseとVOP_time_incrementの定義を、上述のように変更したので、VOPのピクチャタイプの並びや、隣接するVOPの時間間隔などによらず、すべての場合において、各VOPの表示時刻（絶対時刻）を求めることが可能となる。

従って、エンコーダにおいて、符号化開始絶対時刻を、GOV単位で符号化するとともに、各VOPのmodulo_time_baseとVOP_time_incrementを符号化し、符号化ビットストリームに含めることで、デコーダでは、符号化開始絶対時刻を、GOV単位

で復号するとともに、VOP の modulo_time_base および VOP_time_increment を復号し、それらから、各 VOP の表示時刻を復号することができるので、ランダムアクセスを、GOV 単位で、効率的に行うことが可能となる。

なお、modulo_time_base に付加する１の数を、同期点の切り替わりにしたがって、単純に増加していくと、time_code が示す時刻から、例えば、１時間（３６００秒）経過後は（但し、GOV が、それだけの時間に相当する VOP で構成されるとする）、modulo_time_base は、３６００ビットの１と、１ビットの０とで構成されるから、３６０１ビットという莫大なビット数になることになる。

そこで、MPEG4 では、modulo_time_base は、同期点の切り替わり後に最初に現れる I/P-VOP においてリセットされるように規定されている。

従って、例えば、図３９に示すように、GOV が、その time_code が表す時刻である、GOV の最初の同期点から、時刻 time_code＋１秒で示される同期点までの間に表示されるＩ１およびＢ２、時刻 time_code＋１秒で示される同期点から、時刻 time_code＋２秒で示される同期点までの間に表示されるＢ３およびＢ４、時刻 time_code＋２秒で示される同期点から、時刻 time_code＋３秒で示される同期点までの間に表示されるＰ５およびＢ６、時刻 time_code＋３秒で示される同期点から、時刻 time_code＋４秒で示される同期点までの間に表示されるＢ７、並びに時刻 time_code＋４秒で示される同期点から、時刻 time_code＋５秒で示される同期点までの間に表示されるＢ８で構成される場合には、GOV の最初の同期点から、時刻 time_code＋１秒で示される同期点までの間に表示されるＩ１およびＢ２の modulo_time_base は、０Ｂとされる。

また、時刻 time_code＋１秒で示される同期点から、時刻 time_code＋２秒で示される同期点までの間に表示されるＢ３およびＢ４の modulo_time_base は、１０Ｂとされる。さらに、時刻 time_code＋２秒で示される同期点から、時刻 time_code＋３秒で示される同期点までの間に表示されるＰ５の modulo_time_base は、１１０Ｂとされる。

そして、Ｐ５は、GOV の最初の同期点から、時刻 time_code＋１秒で示される同

期点に切り替わった後に、最初に表示されるP-VOPであるから、modulo_time_baseは、０Ｂにリセットされ、その後に表示されるＢ６のmodulo_time_baseは、Ｐ５の表示時刻を求めるときに参照される同期点、即ち、いまの場合、時刻time_code＋２秒で示される同期点を、GOVの最初の同期点とみなして設定される。従って、Ｂ６のmodulo_time_baseは、０Ｂとされる。

その後、時刻time_code＋３秒で示される同期点から、時刻time_code＋４秒で示される同期点までの間に表示されるＢ７のmodulo_time_baseは、１０Ｂとされ、時刻time_code＋４秒で示される同期点から、時刻time_code＋５秒で示される同期点までの間に表示されるＢ８のmodulo_time_baseは、１１０Ｂとされる。

図３２、図３３、および図３７で説明したエンコーダ側（ＶＬＣ器３６）の処理は、上述のようにして、modulo_time_baseを設定するようになっている。

また、この場合、デコーダ側（ＩＶＬＣ器１０２）では、同期点の切り替わり後に最初に表示されるI/P-VOPを検出した場合は、それに付加されているmodulo_time_baseによって示される秒数を、time_codeに累積加算して、表示時刻を求める必要があるが、即ち、例えば、図３９に示した場合においては、Ｉ１～Ｐ５の表示時刻は、time_codeに、各VOPに付加されているmodulo_time_baseに対応する秒数と、VOP_time_incrementとを加算して求めれば良いが、同期点の切り替わり後に最初に表示されるＰ５の後に表示されるＢ６～Ｂ８の表示時刻は、time_codeに、各VOPに付加されているmodulo_time_baseに対応する秒数と、VOP_time_incrementとを加算する他に、さらに、Ｐ５のmodulo_time_baseに対応する秒数である２秒を加算して求める必要があるが、図３４、図３５、および図３８で説明した処理は、そのようにして、表示時刻を求めるようになされている。

次に、以上説明したエンコーダおよびデコーダは、それ専用のハードウェアによって実現することもできるし、コンピュータに、上述したような処理を行わせるためのプログラムを実行させることによっても実現することができる。

図４０は、図３のエンコーダまたは図１５のデコーダとして機能するコンピュータの一実施の形態の構成例を示している。

ＲＯＭ（Read Only Memory）２０１は、例えば、ブートプログラムなどを記憶している。ＣＰＵ（Central Processing Unit）２０２は、例えば、ＨＤ（Hard Disk）２０６に記憶されたプログラムを、ＲＡＭ（Read Only Memory）２０３上に展開して実行することで、各種の処理を行うようになされている。ＲＡＭ２０３は、ＣＰＵ２０２が実行するプログラムや、ＣＰＵ２０２の処理上必要なデータを一時記憶するようになされている。入力部２０４は、例えば、キーボードやマウスなどでなり、必要なコマンドやデータを入力するときなどに操作される。出力部２０５は、例えば、ディスプレイなどでなり、ＣＰＵ２０２の制御にしたがったデータを表示する。ＨＤ２０６は、ＣＰＵ２０２が実行すべきプログラム、さらには、エンコード対象の画像データや、エンコード後のデータ（符号化ビットストリーム）、デコード後の画像データなどを記憶するようになされている。通信Ｉ／Ｆ（Interface）２０７は、外部との通信を制御することにより、例えば、エンコード対象の画像データを、外部から受信したり、また、エンコード後の符号化ビットストリームを外部に送信したりするようになされている。また、通信Ｉ／Ｆ２０７は、外部でエンコードされた符号化ビットストリームを受信したり、また、デコード後の画像データを、外部に送信するようにもなされている。

以上のように構成されるコンピュータのＣＰＵ２０２に、上述したような処理を行うためのプログラムを実行させることにより、このコンピュータは、図３に示したエンコーダや、図１５に示したデコーダとして機能することになる。

なお、本実施の形態では、VOP_time_increment は、VOP の表示時刻を、1ms 単位で表すものとしたが、VOP_time_increment は、その他、例えば、次のようなものとすることも可能である。即ち、１の同期点から、その次の同期点までの間を、Ｎ個に分割し、VOP の表示時刻に対応する分割点が、１の同期点から何番目の分割点かを表す値を、VOP_time_increment とすることが可能である。このように VOP_time_increment を定義した場合、Ｎ＝１０００とすると、VOP_time_increment は、VOP の表示時刻を、1ms 単位で表すものとなる。なお、この場合、デコーダでは、１の同期点から、その次の同期点までの間を幾つに分割したかという情報が必

要となるが、同期点の間の分割数は、あらかじめ定めておいても良いし、あるいは、GOV層よりも上位の階層に含めて、デコーダに提供するようにしても良い。

請求の範囲第１項に記載の画像符号化装置および請求の範囲第６項に記載の画像符号化方法によれば、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものが、複数のグループに分けて符号化される。従って、その符号化結果に対して、グループ単位でのランダムアクセスを行うことが可能となる。

請求の範囲第１１項に記載の画像復号装置および請求の範囲第１６項に記載の画像復号方法によれば、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られた符号化ビットストリームが復号される。従って、符号化ビットストリームに対して、グループ単位でのランダムアクセスを行い、復号することが可能となる。

請求の範囲第２１項に記載の提供媒体によれば、画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られる符号化ビットストリームが提供される。従って、符号化ビットストリームに対して、グループ単位でのランダムアクセスを行うことが可能となる。

請求の範囲第２６項に記載の画像符号化装置および請求の範囲第２９項に記載の画像符号化方法によれば、秒精度の時刻を表す秒精度時刻情報が生成され、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報が生成される。従って、秒精度時刻情報および詳細時間情報に基づいて、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を認識し、その認識結果に基づいて、ランダムアクセスを行うことが可能となる。

請求の範囲第３２項に記載の画像復号装置および請求の範囲第３５項に記載の画像復号方法によれば、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻が、秒精度時刻情報および詳細時間情報に基づいて求められる。従って、その表示時刻に基づいて、ランダムアクセスを行うことが可能となる。

請求の範囲第３８項に記載の提供媒体によれば、秒精度の時刻を表す秒精度時刻

情報を生成し、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成し、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時刻を表す情報として、秒精度時刻情報および詳細時間情報を、対応するＩ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加することにより得られる符号化ストリームが提供される。従って、秒精度時刻情報および詳細時間情報に基づいて、Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を認識し、その認識結果に基づいて、ランダムアクセスを行うことが可能となる。

産業上の利用可能性

本発明は、動画像データを、光磁気デイスクや磁気テープなどの記録媒体に記録し、これを再生してデイスプレイなどに表示する画像情報記録再生装置や、動画像データを伝送路を介して送信側から受信側に伝送し、受信側において、受信された動画像データを表示したり、編集して記録したりするテレビ会議システム、テレビ電話システム、放送用機器、マルチメデイアデータベース検索システムなどに利用できる。

* C L M S *

請求の範囲

1． 画像を符号化し、その結果得られる符号化ビットストリームを出力する画像符号化装置であって、

上記画像を受信する受信手段と、

上記画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化する符号化手段と

を備えることを特徴とする画像符号化装置。

2． 上記オブジェクトがイントラ符号化またはインター符号化される場合において、

上記符号化手段は、上記オブジェクトのシーケンスを、少なくとも１のイントラ符号化されるオブジェクトを含むグループに分ける

ことを特徴とする請求の範囲第１項に記載の画像符号化装置。

3． 上記符号化手段は、上記オブジェクトのグループに、その符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻を含める

ことを特徴とする請求の範囲第１項に記載の画像符号化装置。

4． イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、

上記符号化開始秒精度絶対時刻を基準とする秒精度の時刻を表す秒精度時刻情報を生成する秒精度時刻情報生成手段と、

上記オブジェクトのグループに含まれる上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時

刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成する詳細時間情報生成手段と、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時刻を表す情報として、上記秒精度時刻情報および詳細時間情報を、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加する付加手段と

をさらに備えることを特徴とする請求の範囲第３項に記載の画像符号化装置。

５．　上記秒精度時刻情報生成手段は、所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したものを生成する

ことを特徴とする請求の範囲第４項に記載の画像符号化装置。

６．　画像を符号化し、その結果得られる符号化ビットストリームを出力する画像符号化方法であって、

上記画像を受信する受信ステップと、

上記画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化する符号化ステップと

を備えることを特徴とする画像符号化方法。

７．　上記オブジェクトがイントラ符号化またはインター符号化される場合において、

上記符号化ステップにおいて、上記オブジェクトのシーケンスを、少なくとも１のイントラ符号化されるオブジェクトを含むグループに分ける

ことを特徴とする請求の範囲第６項に記載の画像符号化方法。

８．　上記符号化ステップにおいて、上記オブジェクトのグループに、その符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻を含める

ことを特徴とする請求の範囲第６項に記載の画像符号化方法。

９．　イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、

上記符号化開始秒精度絶対時刻を基準とする秒精度の時刻を表す秒精度時刻情報を生成する秒精度時刻情報生成ステップと、

上記オブジェクトのグループに含まれる上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成する詳細時間情報生成ステップと、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時刻を表す情報として、上記秒精度時刻情報および詳細時間情報を、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加する付加ステップと

をさらに備えることを特徴とする請求の範囲第８項に記載の画像符号化方法。

１０．　上記秒精度時刻情報生成ステップにおいて、所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したものを生成する

ことを特徴とする請求の範囲第９項に記載の画像符号化方法。

１１．　画像を復号する画像復号装置であって、

上記画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られた符号化ビットストリームを受信する受信手段と、

上記符号化ビットストリームを復号する復号手段と

* C L M S *

を備えることを特徴とする画像復号装置。

１２．　上記オブジェクトがイントラ符号化またはインター符号化されている場合において、

上記オブジェクトのシーケンスが、少なくとも１のイントラ符号化されるオブジェクトを含むグループに分けられている

ことを特徴とする請求の範囲第１１項に記載の画像復号装置。

１３．　上記オブジェクトのグループには、その符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻が含められており、

上記オブジェクトのグループに含まれる上記オブジェクトの表示時刻を、上記符号化開始秒精度絶対時刻に基づいて求める表示時刻算出手段をさらに備える

ことを特徴とする請求の範囲第１１項に記載の画像復号装置。

１４．　イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、

上記符号化開始秒精度絶対時刻を基準とする秒精度の時刻を表す秒精度時刻情報と、上記オブジェクトのグループに含まれる上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報とが、上記表示時刻を表す情報として、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加されており、

上記表示時刻算出手段は、上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を、上記符号化開始秒精度絶対時刻、並びに上記秒精度時刻情報および詳細時間情報に基づいて求める

* C L M S *

ことを特徴とする請求の範囲第１３項に記載の画像復号装置。

１５．　所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したものが用いられている

ことを特徴とする請求の範囲第１４項に記載の画像復号装置。

１６．　画像を復号する画像復号方法であって、

上記画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化することにより得られた符号化ビットストリームを受信する受信ステップと、

上記符号化ビットストリームを復号する復号ステップと

を備えることを特徴とする画像復号方法。

１７．　上記オブジェクトがイントラ符号化またはインター符号化されている場合において、

上記オブジェクトのシーケンスが、少なくとも１のイントラ符号化されるオブジェクトを含むグループに分けられている

ことを特徴とする請求の範囲第１６項に記載の画像復号方法。

１８．　上記オブジェクトのグループには、その符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻が含められており、

上記オブジェクトのグループに含まれる上記オブジェクトの表示時刻を、上記符号化開始秒精度絶対時刻に基づいて求める表示時刻算出ステップをさらに備える

ことを特徴とする請求の範囲第１６項に記載の画像復号方法。

１９．　イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符

* C L M S *

号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、

上記符号化開始秒精度絶対時刻を基準とする秒精度の時刻を表す秒精度時刻情報と、上記オブジェクトのグループに含まれる上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報とが、上記表示時刻を表す情報として、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加されており、

上記表示時刻算出ステップにおいて、上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を、上記符号化開始秒精度絶対時刻、並びに上記秒精度時刻情報および詳細時間情報に基づいて求める

ことを特徴とする請求の範囲第１８項に記載の画像復号方法。

２０．　所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したものが用いられている

ことを特徴とする請求の範囲第１９項に記載の画像復号方法。

２１．　画像を符号化して得られる符号化ビットストリームを提供する提供媒体であって、

上記画像を受信し、

上記画像を構成するオブジェクトのシーケンスを１以上の階層に分けたものを、複数のグループに分けて符号化する

ことにより得られる上記符号化ビットストリームを提供する

ことを特徴とする提供媒体。

２２．　上記オブジェクトがイントラ符号化またはインター符号化される場合にお

* C L M S *

いて、

上記オブジェクトのシーケンスが、少なくとも１のイントラ符号化されるオブジェクトを含むグループに分けられている

ことを特徴とする請求の範囲第２１項に記載の提供媒体。

２３．　上記オブジェクトのグループには、その符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻が含められている

ことを特徴とする請求の範囲第２１項に記載の提供媒体。

２４．　イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、

上記符号化開始秒精度絶対時刻を基準とする秒精度の時刻を表す秒精度時刻情報と、上記オブジェクトのグループに含まれる上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報とが、上記表示時刻を表す情報として、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加されている

ことを特徴とする請求の範囲第２３項に記載の提供媒体。

２５．　所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したものが用いられている

ことを特徴とする請求の範囲第２４項に記載の提供媒体。

* C L M S *

２６．　画像を、その画像を構成するオブジェクトごとに符号化し、その結果得られる符号化ビットストリームを出力する画像符号化装置であって、

イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、

秒精度の時刻を表す秒精度時刻情報を生成する秒精度時刻情報生成手段と、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成する詳細時間情報生成手段と、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時刻を表す情報として、上記秒精度時刻情報および詳細時間情報を、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加する付加手段と

を備えることを特徴とする画像符号化装置。

２７．　上記秒精度時刻情報生成手段は、上記オブジェクトのシーケンスを、１以上のグループに分けて符号化する場合の、そのグループのオブジェクトの符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻を基準とした秒精度の時刻を、上記秒精度時刻情報として生成する

ことを特徴とする請求の範囲第２６項に記載の画像符号化装置。

２８．　上記秒精度時刻情報生成手段は、所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したものを生成する

ことを特徴とする請求の範囲第２７項に記載の画像符号化装置。

２９．　画像を、その画像を構成するオブジェクトごとに符号化し、その結果得られる符号化ビットストリームを出力する画像符号化方法であって、

イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、

秒精度の時刻を表す秒精度時刻情報を生成する秒精度時刻情報生成ステップと、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成する詳細時間情報生成ステップと、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時刻を表す情報として、上記秒精度時刻情報および詳細時間情報を、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加する付加ステップと

を備えることを特徴とする画像符号化方法。

３０．　上記秒精度時刻情報生成ステップにおいて、上記オブジェクトのシーケンスを、１以上のグループに分けて符号化する場合の、そのグループのオブジェクトの符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻を基準とした秒精度の時刻を、上記秒精度時刻情報として生成する

ことを特徴とする請求の範囲第２９項に記載の画像符号化方法。

３１．　上記秒精度時刻情報生成ステップにおいて、所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したものを生成する

ことを特徴とする請求の範囲第３０項に記載の画像符号化方法。

* C L M S *

３２．　画像を、その画像を構成するオブジェクトごとに符号化して得られる符号化ビットストリームを復号する画像復号装置であって、

イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとともに、

秒精度の時刻を表す秒精度時刻情報と、上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報とが、上記表示時刻を表す情報として、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加されているとき、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を、上記秒精度時刻情報および詳細時間情報に基づいて求める表示時刻算出手段と、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰを、対応する表示時刻にしたがって復号する復号手段と

を備えることを特徴とする画像復号装置。

３３．　上記秒精度時刻情報は、上記オブジェクトのシーケンスが、１以上のグループに分けて符号化されている場合の、そのグループのオブジェクトの符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻を基準とした秒精度の時刻を表す

ことを特徴とする請求の範囲第３２項に記載の画像復号装置。

３４．　所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒

* C L M S *

精度で表したものが用いられている

ことを特徴とする請求の範囲第３３項に記載の画像復号装置。

３５．　画像を、その画像を構成するオブジェクトごとに符号化して得られる符号化ビットストリームを復号する画像復号方法であって、

イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとともに、

秒精度の時刻を表す秒精度時刻情報と、上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報とが、上記表示時刻を表す情報として、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加されているとき、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻を、上記秒精度時刻情報および詳細時間情報に基づいて求める表示時刻算出ステップと、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰを、対応する表示時刻にしたがって復号する復号ステップと

を備えることを特徴とする画像復号方法。

３６．　上記秒精度時刻情報は、上記オブジェクトのシーケンスが、１以上のグループに分けて符号化されている場合の、そのグループのオブジェクトの符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻を基準とした秒精度の時刻を表す

ことを特徴とする請求の範囲第３５項に記載の画像復号方法。

３７．　所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表

したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したものが用いられている

ことを特徴とする請求の範囲第３６項に記載の画像復号方法。

３８．　画像を、その画像を構成するオブジェクトごとに符号化して得られる符号化ビットストリームを提供する提供媒体であって、

イントラ符号化されるオブジェクトをＩ－ＶＯＰ（Intra-Video Object Plane）と、イントラ符号化または前方予測符号化のうちのいずれかで符号化されるオブジェクトをＰ－ＶＯＰ（Predictive-VOP）と、イントラ符号化、前方予測符号化、後方予測符号化、または両方向予測符号化のうちのいずれかで符号化されるオブジェクトをＢ－ＶＯＰ（Biderectionally Predictive-VOP）と、それぞれするとき、

秒精度の時刻を表す秒精度時刻情報を生成し、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰそれぞれの表示時刻の直前の上記秒精度時刻情報から、それぞれの表示時刻までの時間を、秒精度より細かい精度で表す詳細時間情報を生成し、

上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰの表示時刻を表す情報として、上記秒精度時刻情報および詳細時間情報を、対応する上記Ｉ－ＶＯＰ，Ｐ－ＶＯＰ、またはＢ－ＶＯＰにそれぞれ付加する

ことにより得られる上記符号化ストリームを提供する

ことを特徴とする提供媒体。

３９．　上記オブジェクトのシーケンスを、１以上のグループに分けて符号化する場合の、そのグループのオブジェクトの符号化を開始した、上記画像のシーケンス上の絶対時刻であって、秒精度の時刻である符号化開始秒精度絶対時刻を基準とした秒精度の時刻を、上記秒精度時刻情報として生成する

ことを特徴とする請求の範囲第３８項に記載の提供媒体。

４０．　所定のオブジェクトについての秒精度時刻情報として、上記符号化開始秒精度絶対時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表

* C L M S *

したもの、または上記所定のオブジェクトの直前に表示されるＩ－ＶＯＰもしくはＰ－ＶＯＰの表示時刻から、上記所定のオブジェクトの表示時刻までの時間を、秒精度で表したものを生成する

ことを特徴とする請求の範囲第３９項に記載の提供媒体。

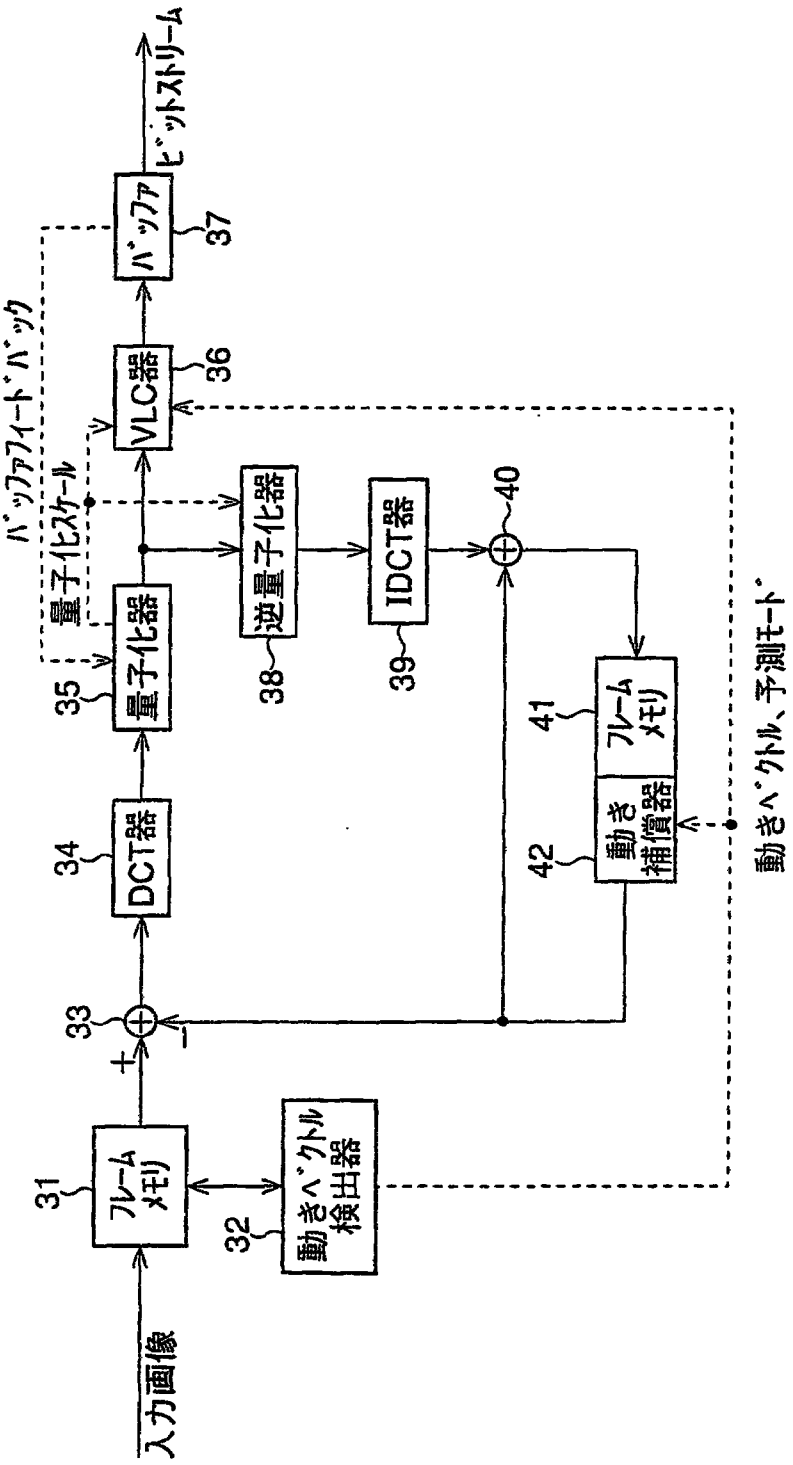

入力画像
フレームメモリ
31
動きベクトル検出器
32
+
−
33
DCT器
34
量子化器
35
量子化スケール
バッファフィードバック
逆量子化器
38
IDCT器
39
40
VLC器
36
バッファ
37
ビットストリーム
フレームメモリ
41
動き補償器
42
動きベクトル、予測モード

図1

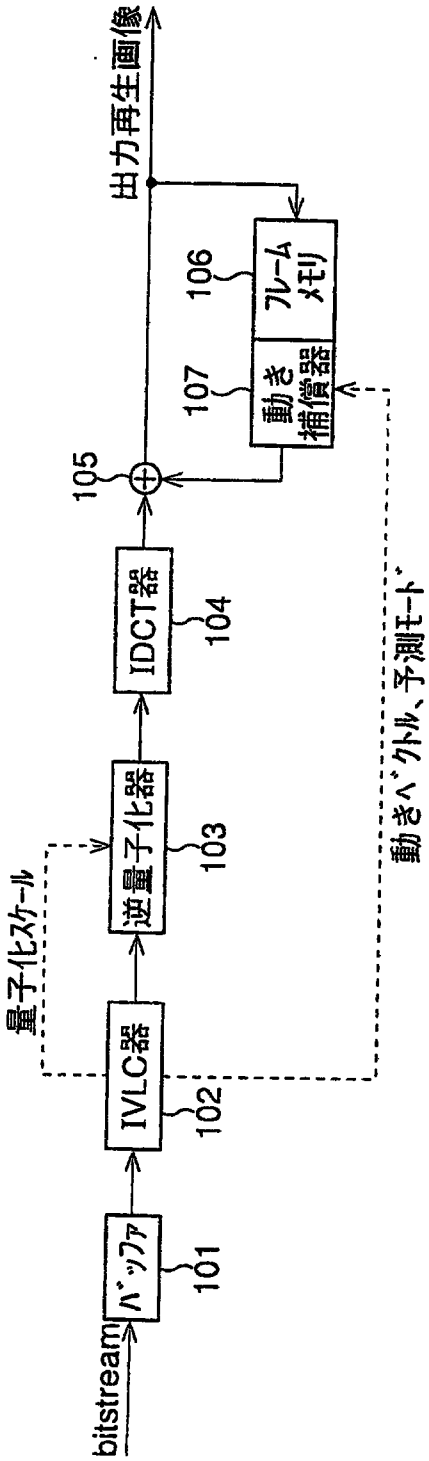

bitstream
ﾊﾞｯﾌｧ
101
IVLC器
102
逆量子化器
103
IDCT器
104
105
動き
補償器
107
ﾌﾚｰﾑ
ﾒﾓﾘ
106
出力再生画像
量子化ｽｹｰﾙ
動きﾍﾞｸﾄﾙ、予測ﾓｰﾄﾞ

図2

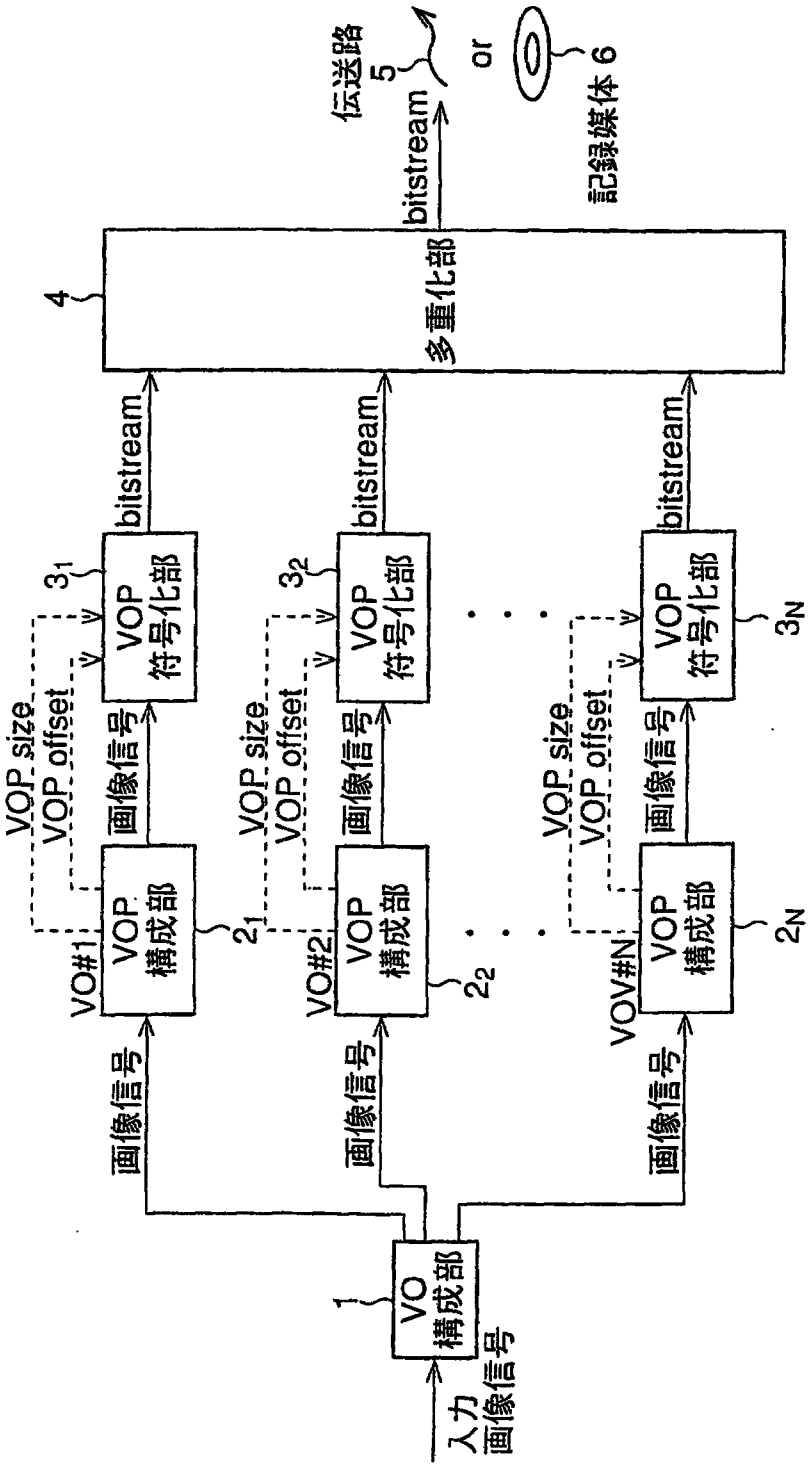

入力画像信号
VO構成部
1
画像信号
VOP構成部
2_1
2_2
2_N
VO#1
VO#2
VOV#N
VOP size
VOP offset
画像信号
VOP符号化部
3_1
3_2
3_N
bitstream
多重化部
4
伝送路
5
or
記録媒体 6

図3

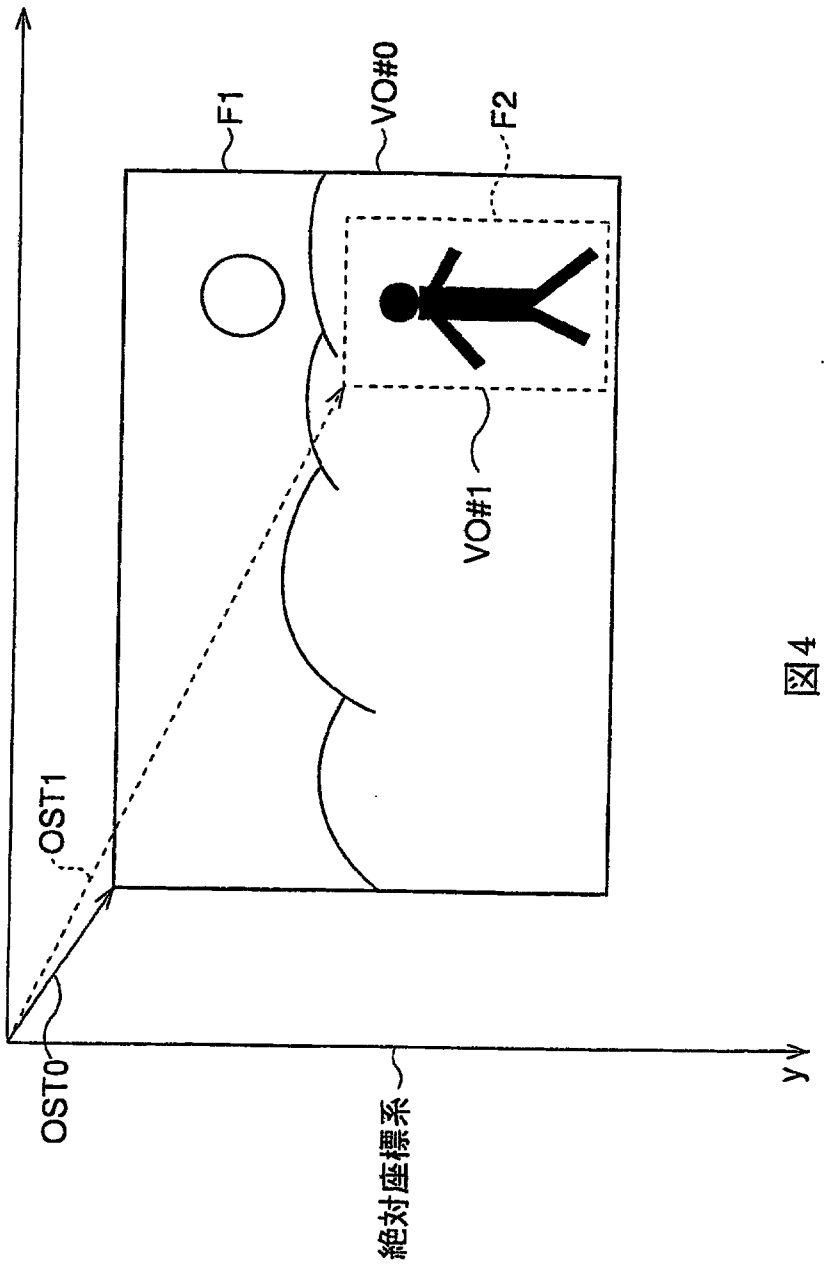
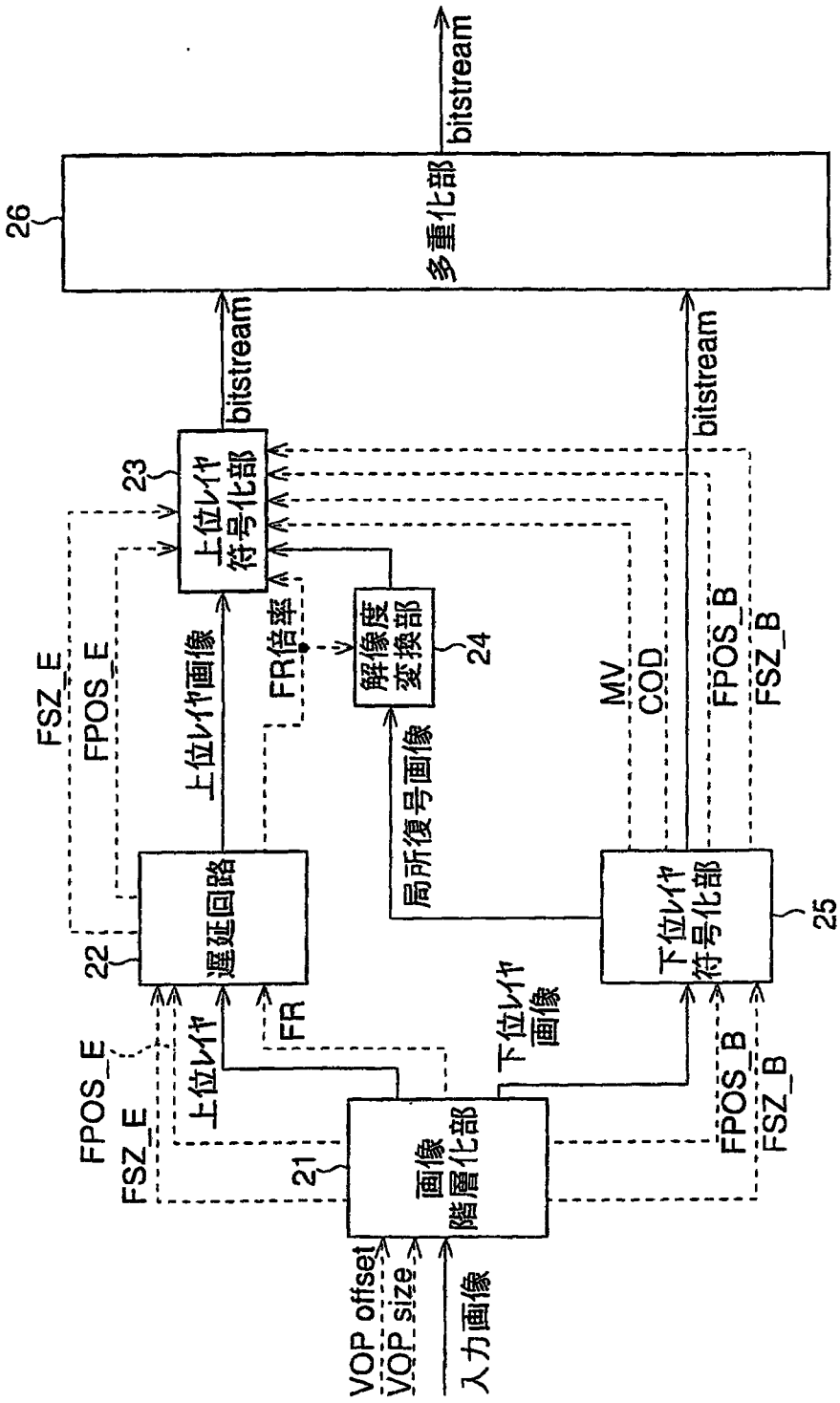


図4



VOP符号化部 3n

図5

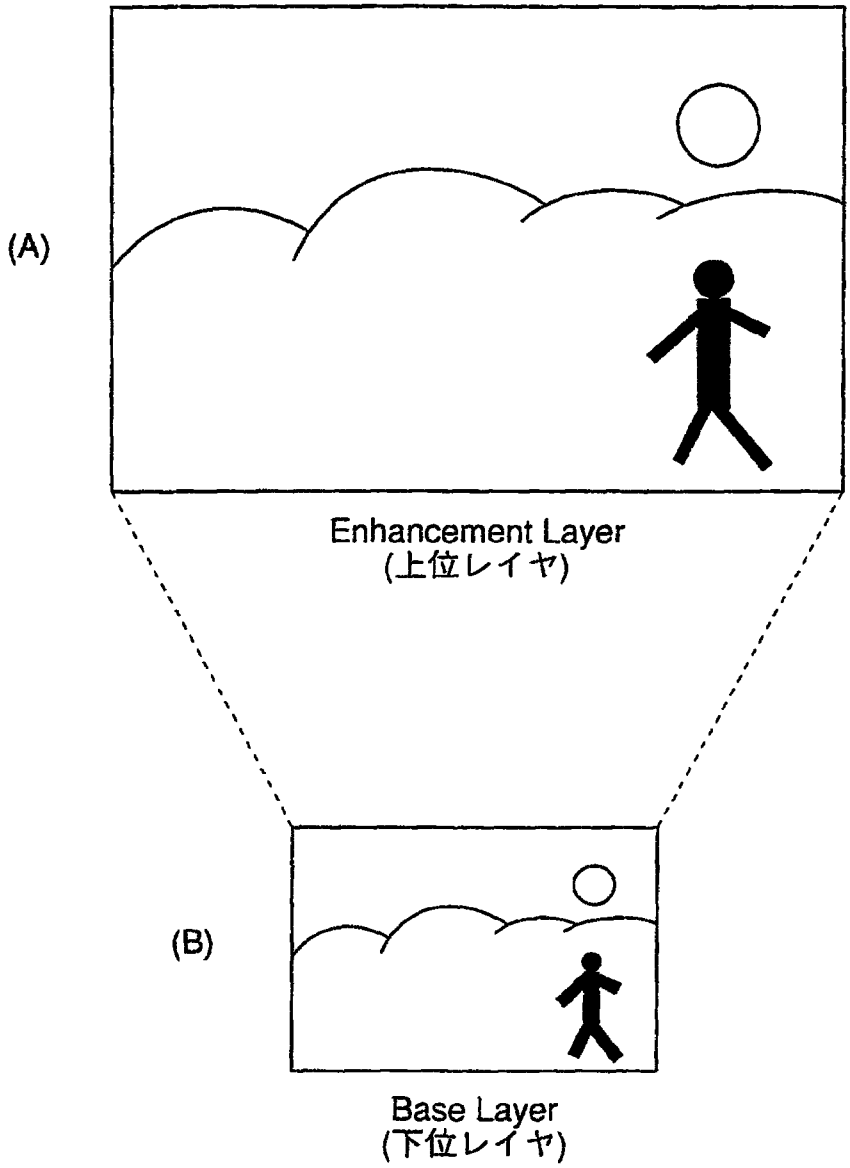


図6

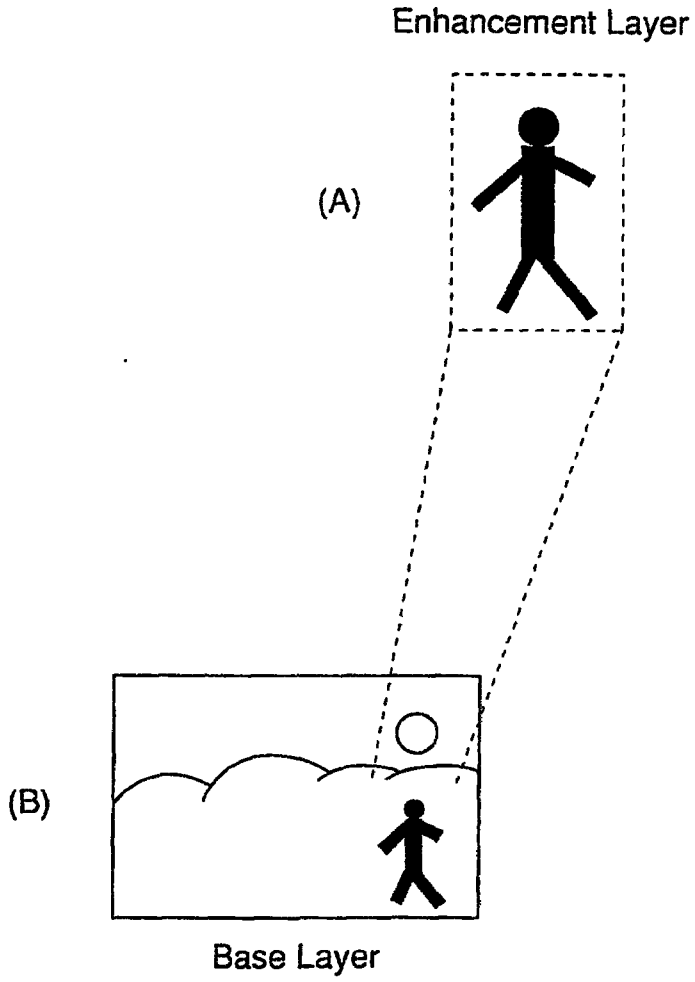


図 7

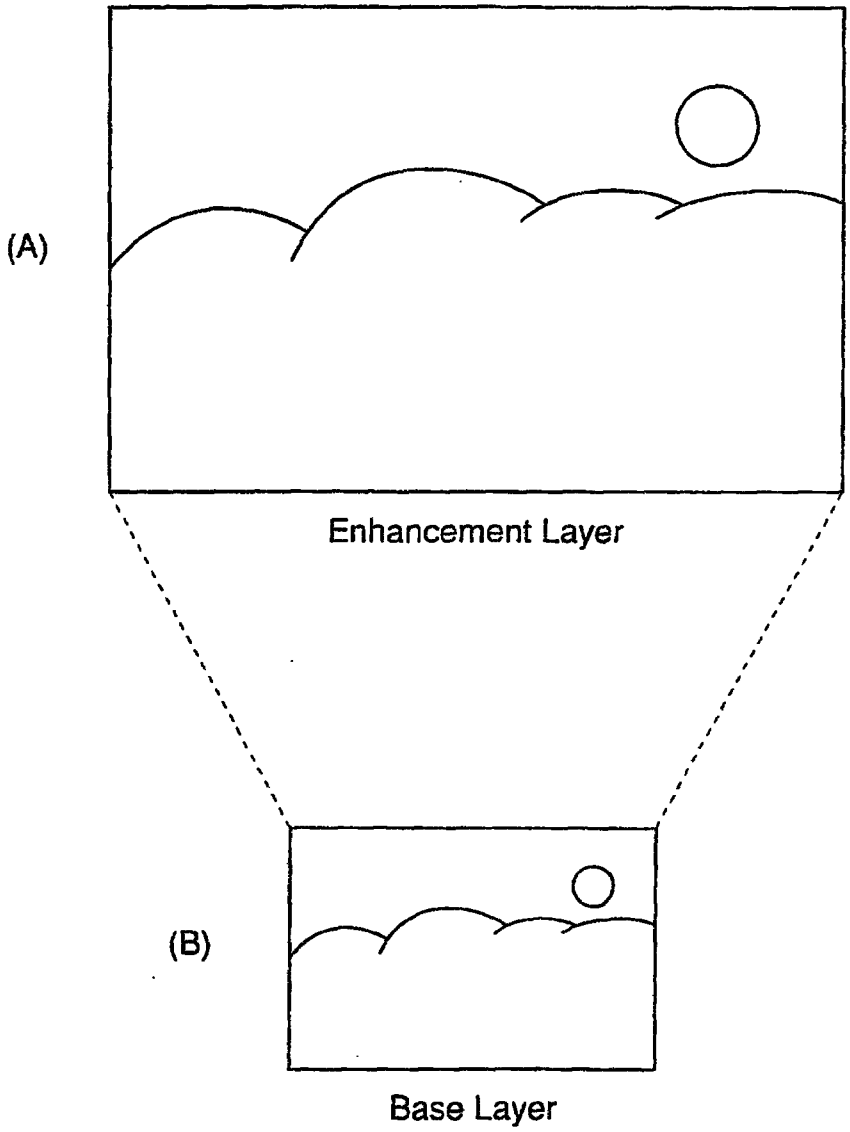


図8

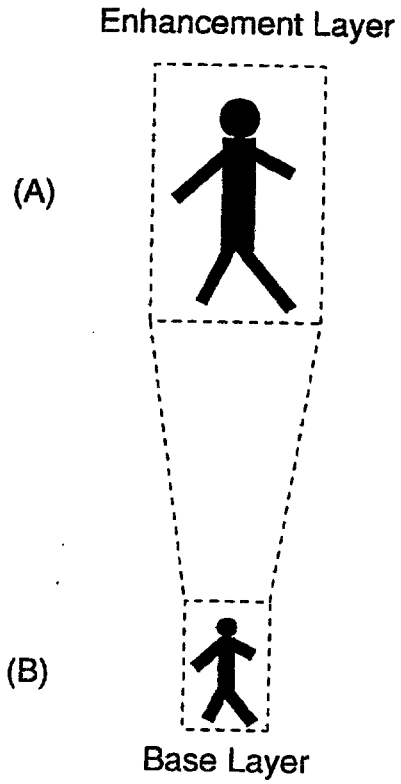
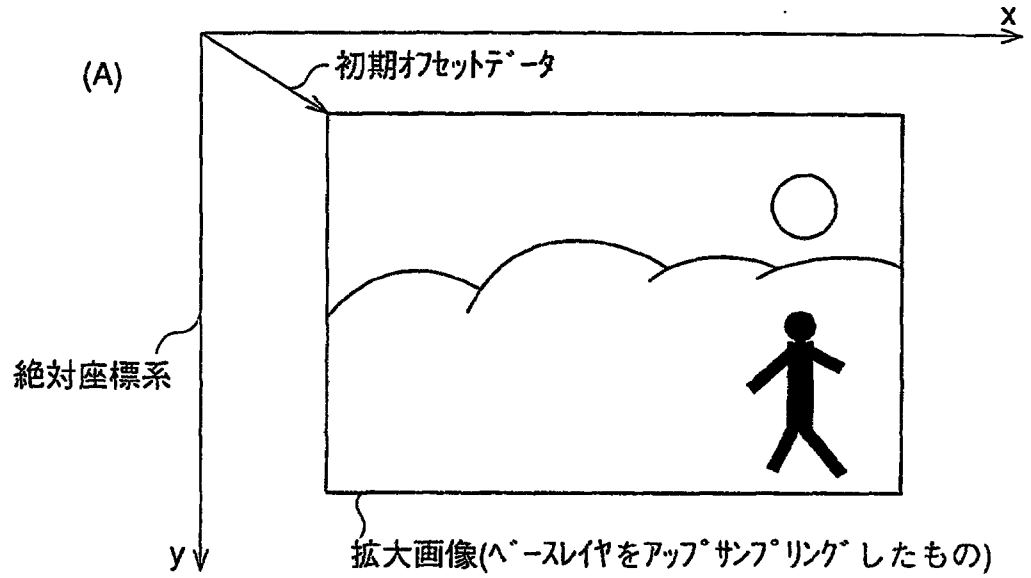


図9


x
(A)
初期オフセットデータ
絶対座標系
y
拡大画像(ベースレイヤをアップサンプリングしたもの)

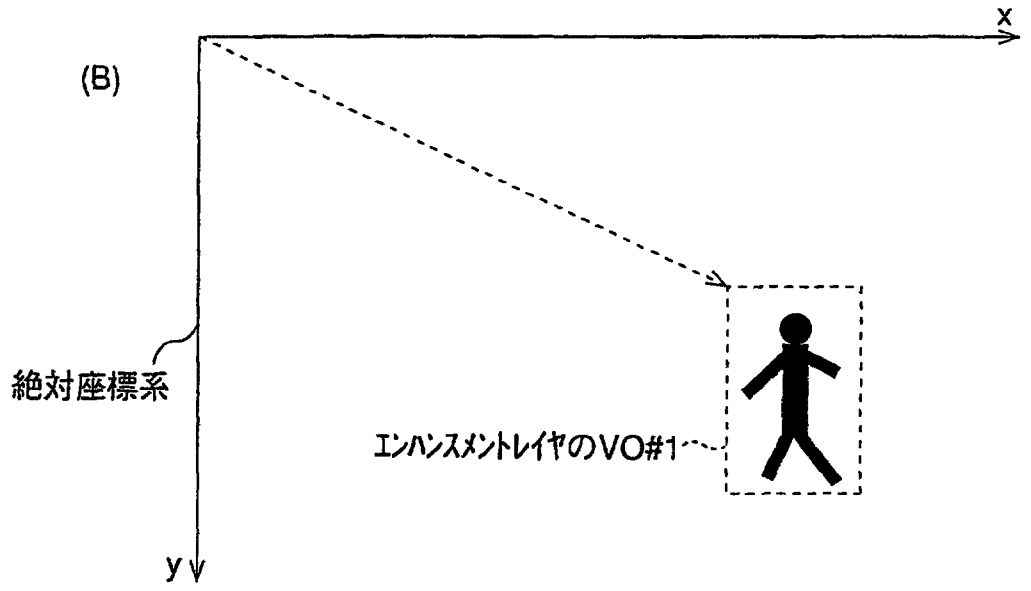
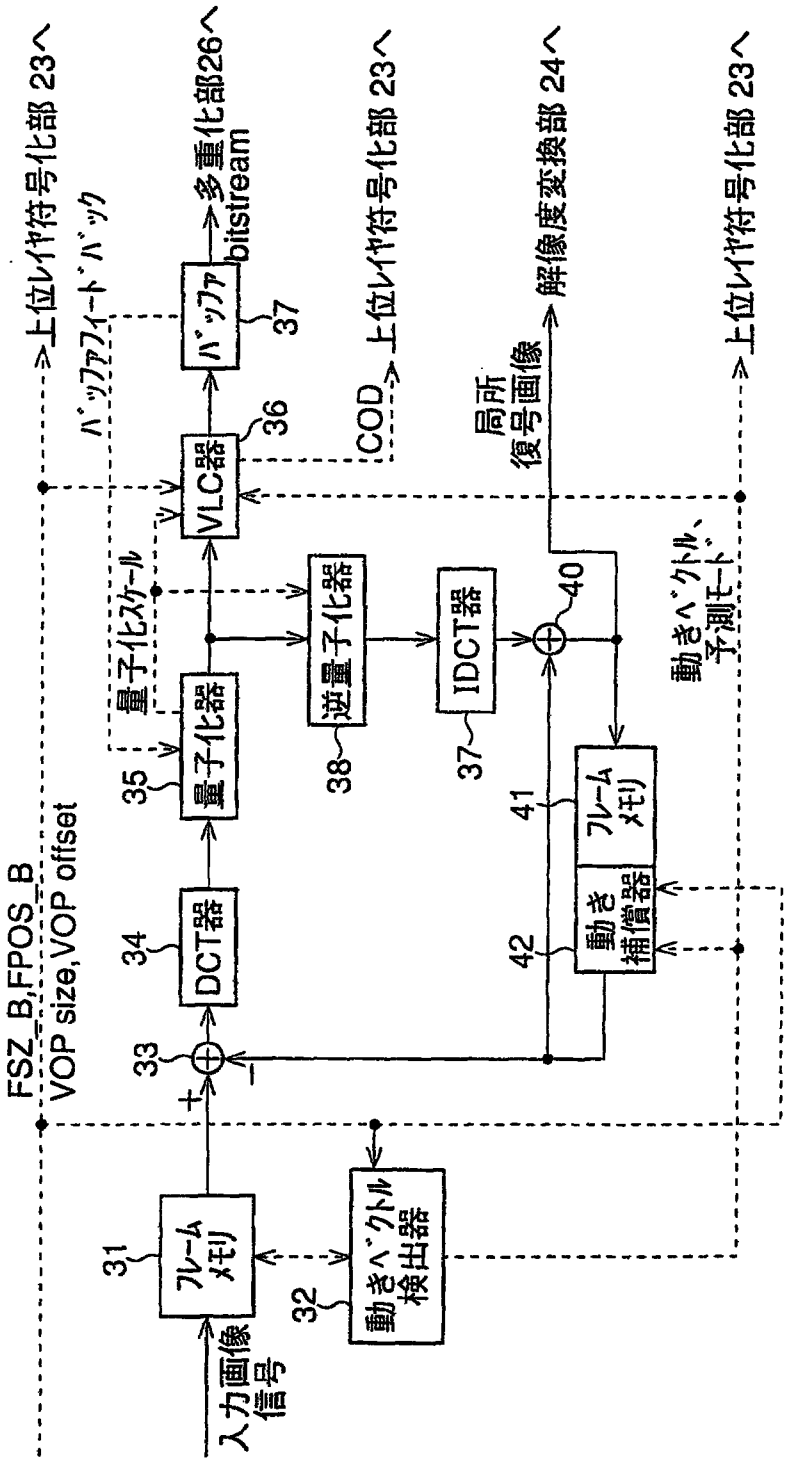

x
(B)
絶対座標系
エンハンスメントレイヤのVO#1
y

図10



下部レイヤ符号化部 25

図11

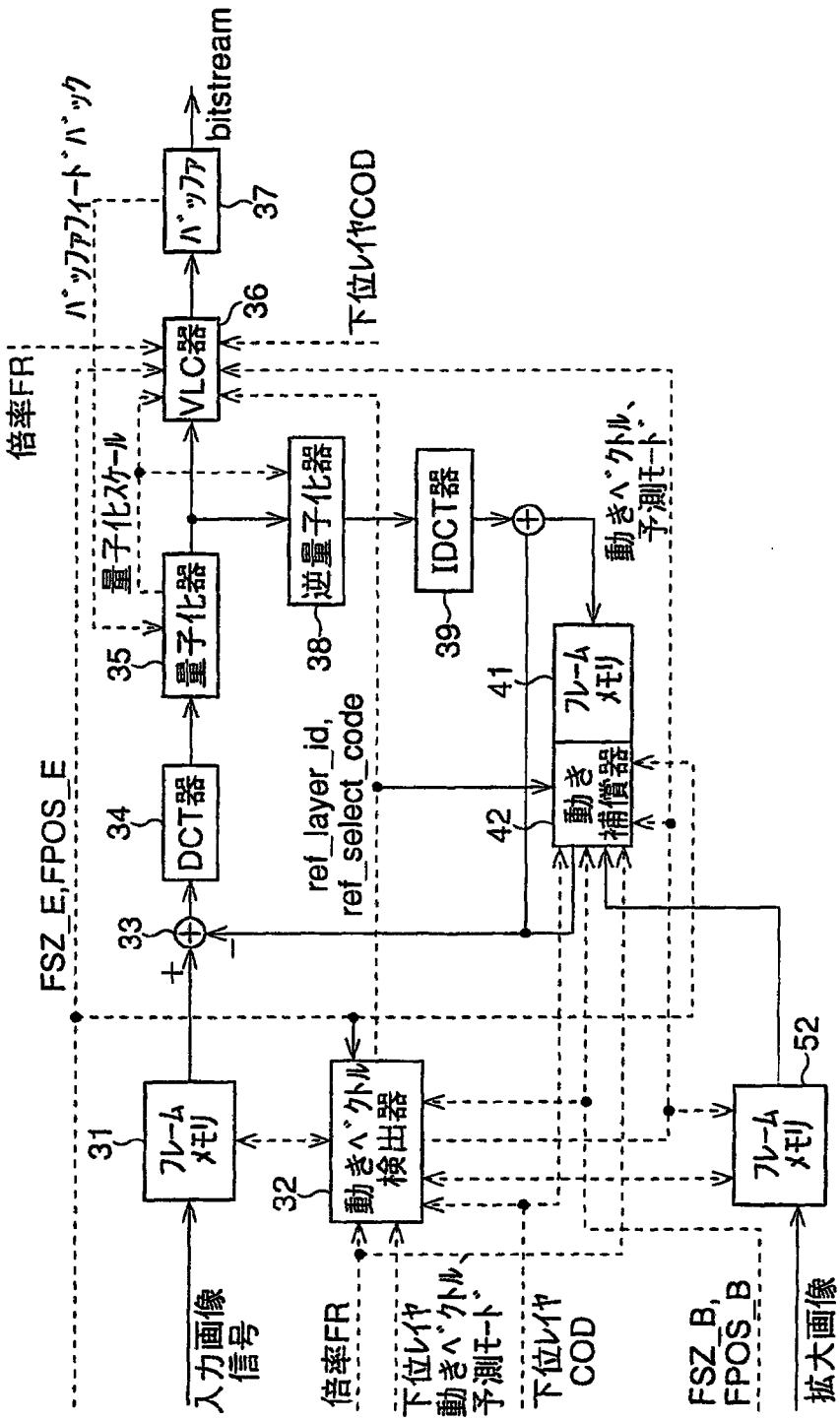

上部レイヤ符号化部 23
入力画像信号
フレームメモリ
31
動きベクトル検出器
32
倍率FR
下位レイヤ動きベクトル、予測モード
下位レイヤCOD
FSZ_B, FPOS_B
フレームメモリ
52
拡大画像
FSZ_E,FPOS_E
33
DCT器
34
量子化器
35
量子化スケール
VLC器
36
バッファ
37
bitstream
バッファフィードバック
倍率FR
下位レイヤCOD
逆量子化器
38
IDCT器
39
ref_layer_id,
ref_select_code
動き補償器
42
フレームメモリ
41
動きベクトル、予測モード

図12

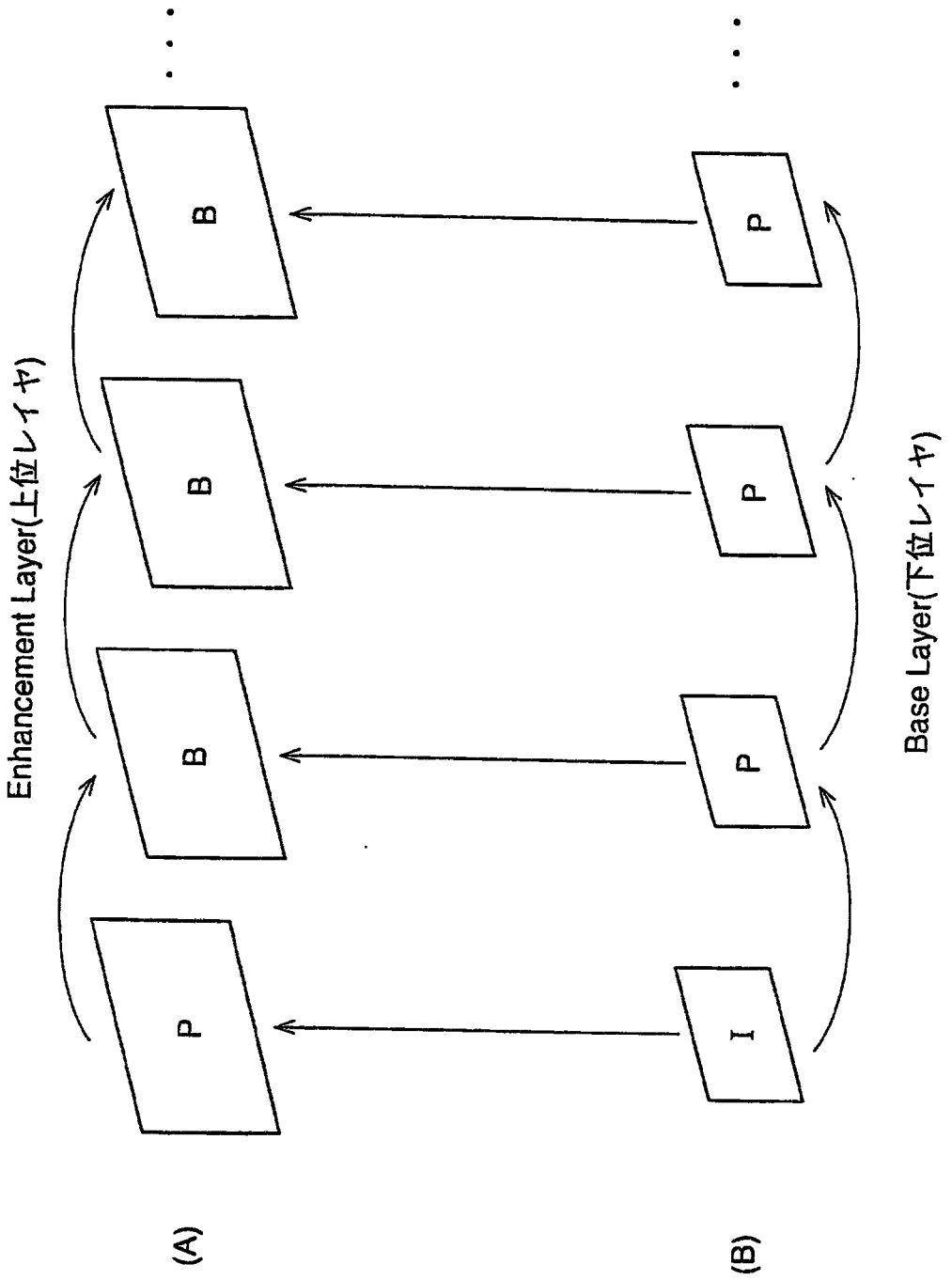


図13

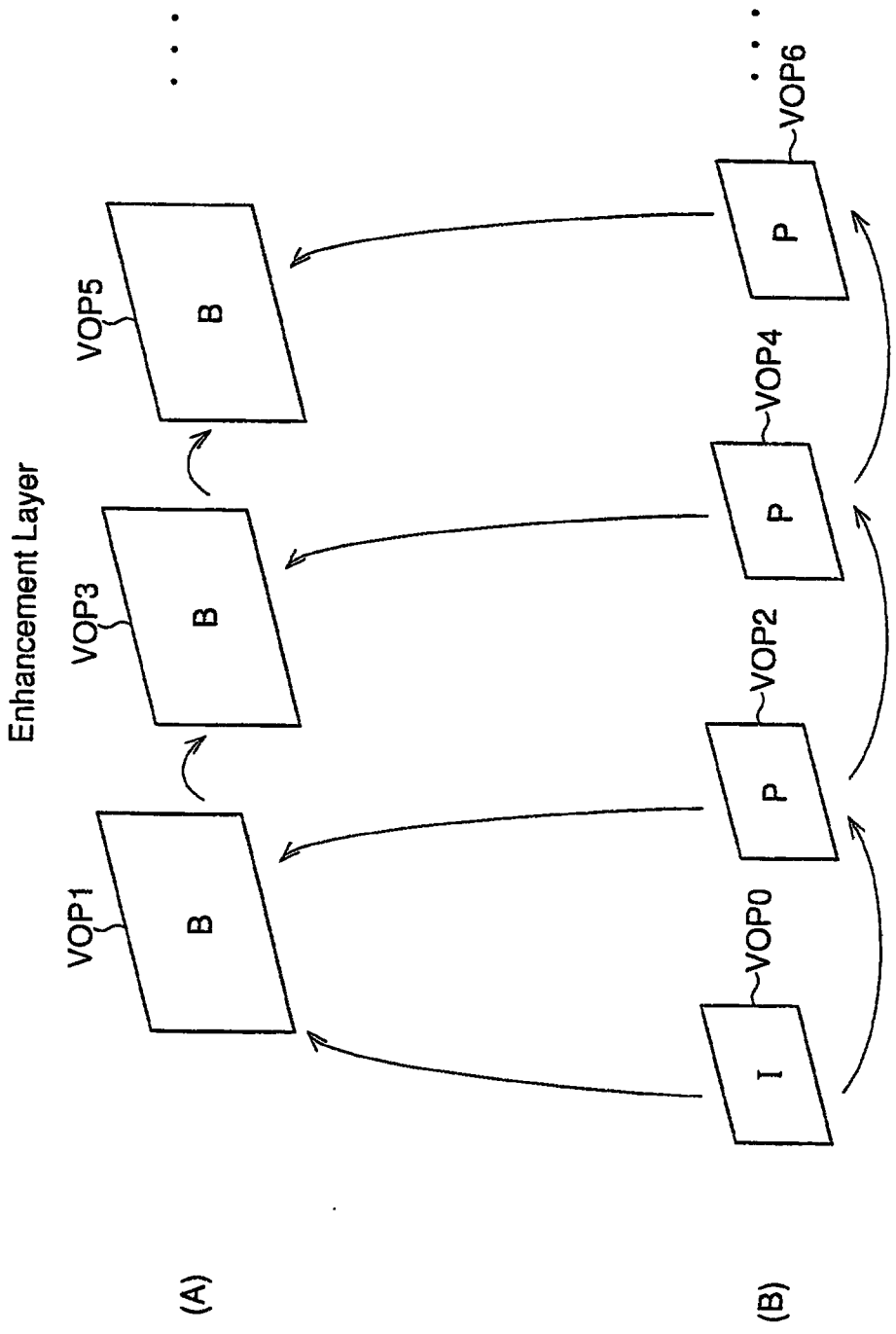

Enhancement Layer
Base Layer
VOP1
VOP3
VOP5
VOP0
VOP2
VOP4
VOP6
B
B
B
I
P
P
P
(A)
(B)

図14

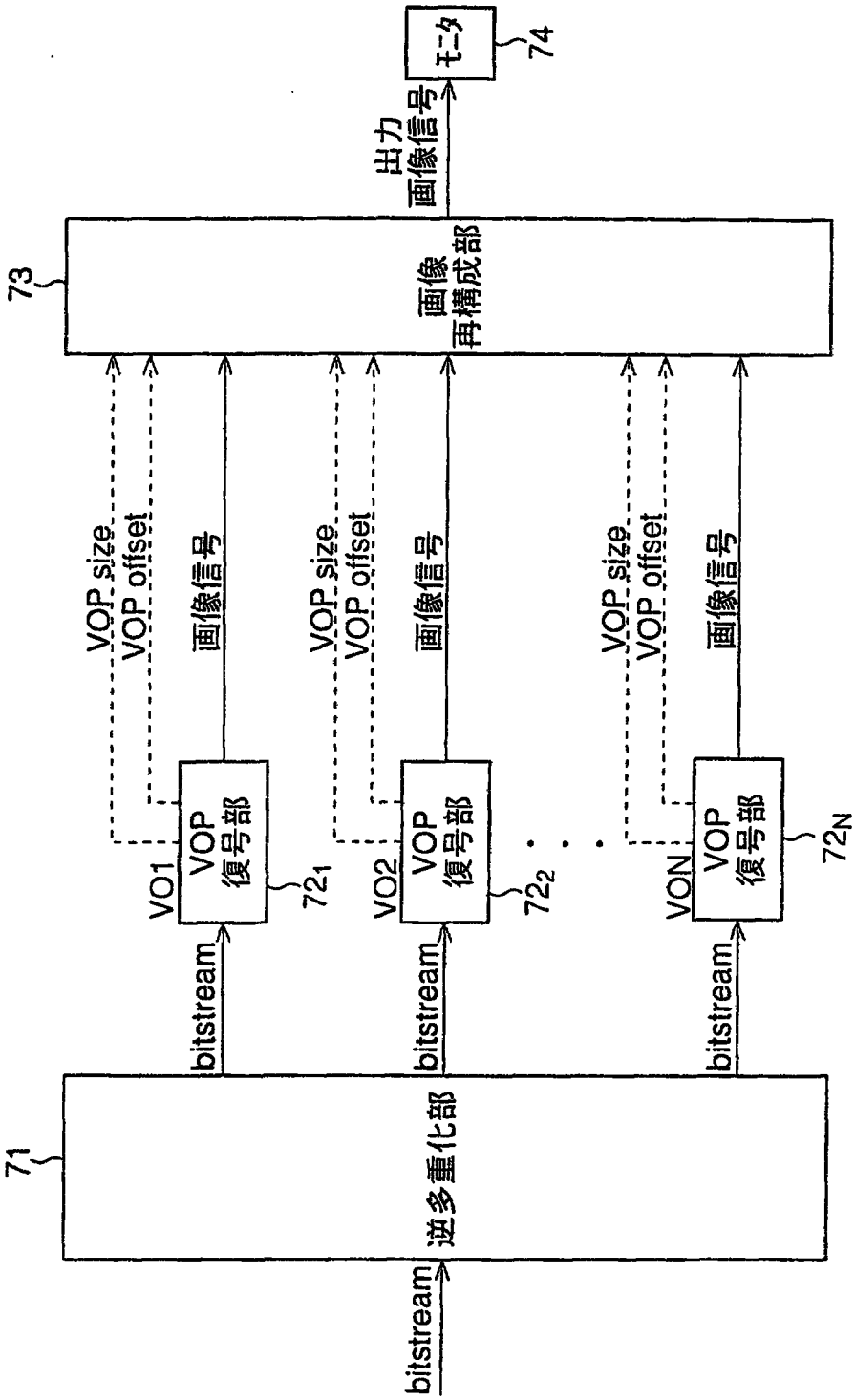
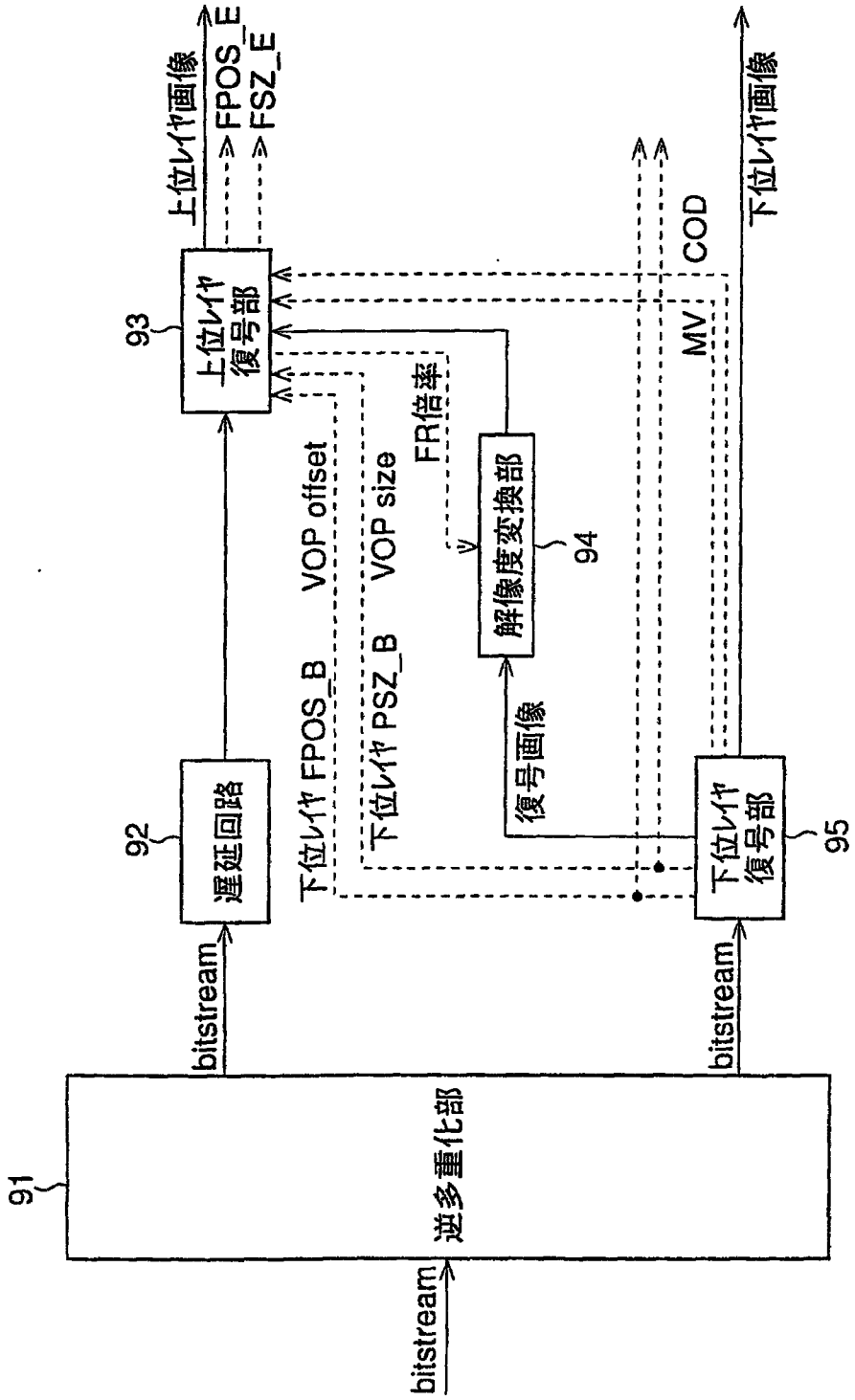
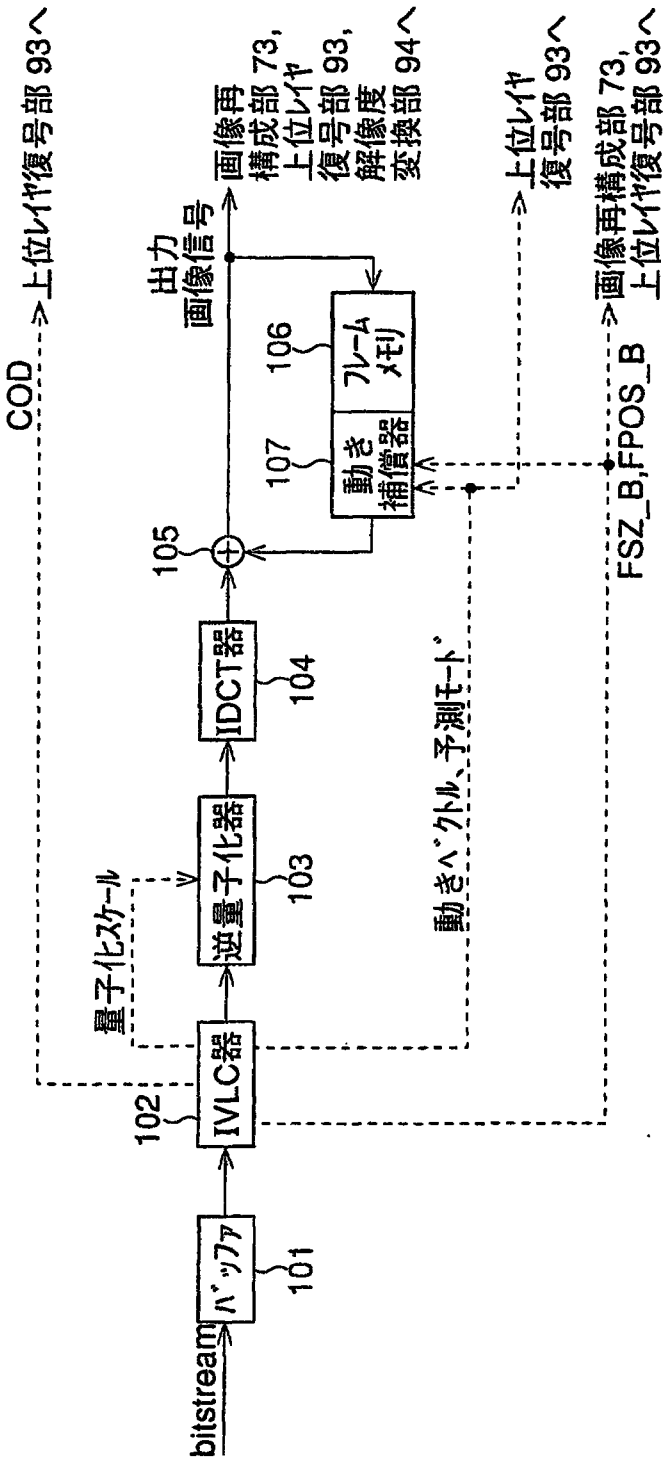

bitstream
逆多重化部
71
bitstream
bitstream
bitstream
VOP
復号部
72_1
72_2
72_N
VO1
VO2
VON
VOP size
VOP offset
画像信号
画像
再構成部
73
出力
画像信号
モニタ
74

図15



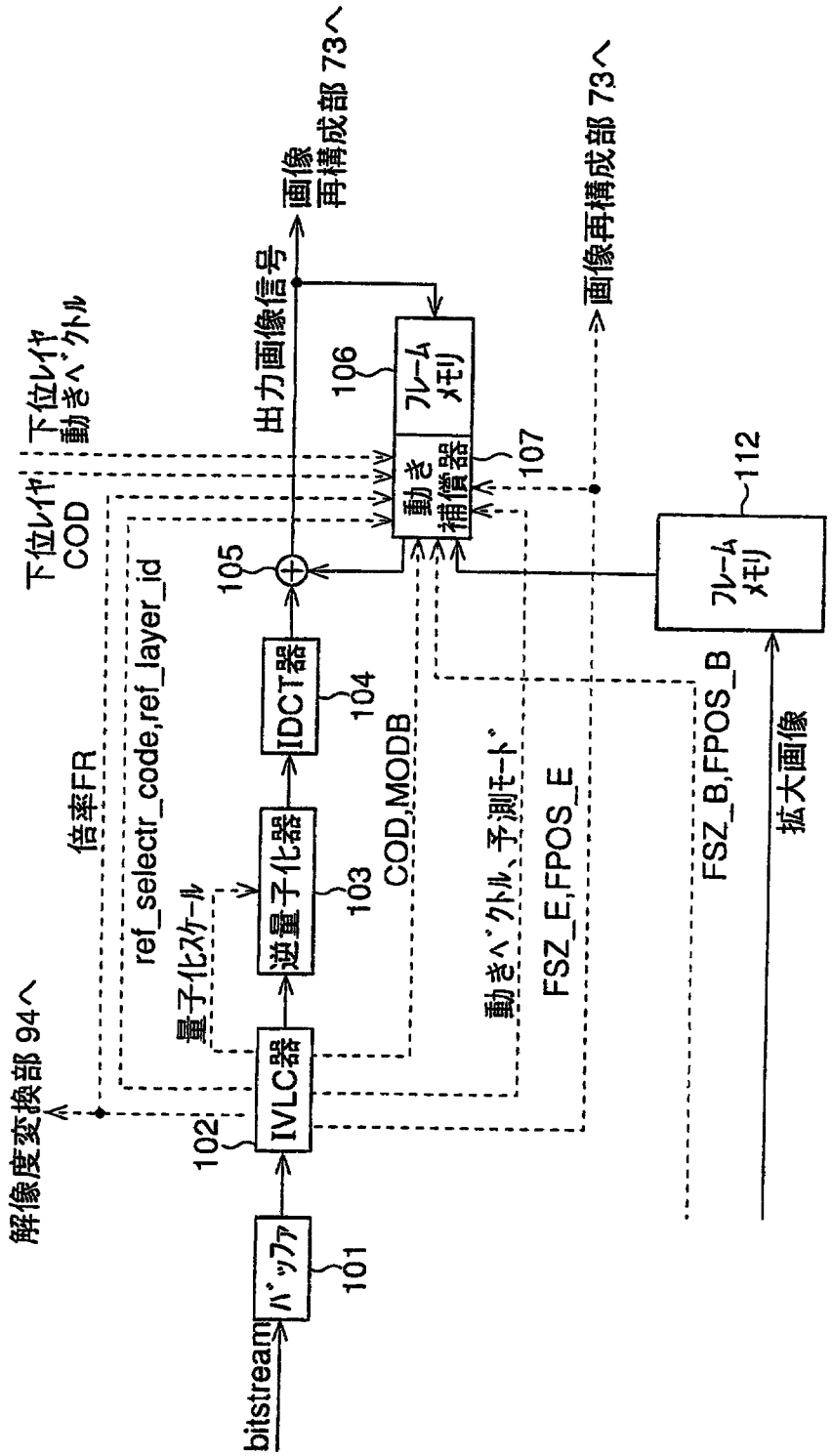
VOP復号部 72n

図16



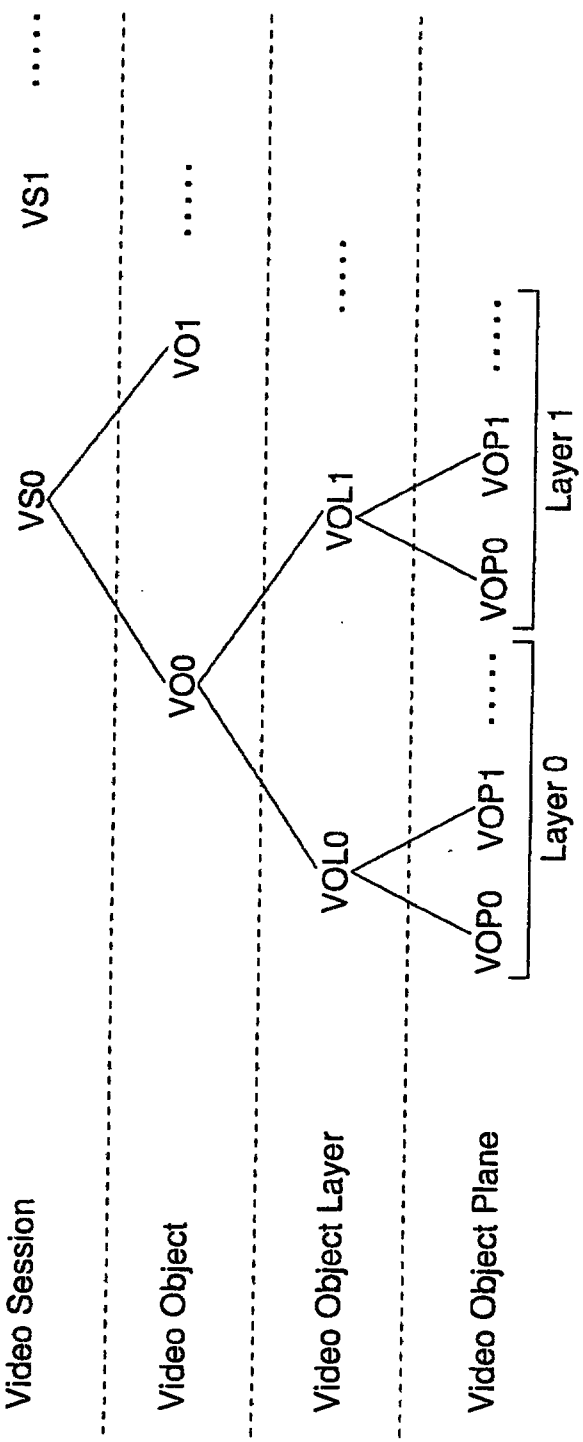
下位レイヤ復号部 95

図17



上位レイヤ復号部 93

図18



Hierarchy in the proposed video syntax

図19

Video Session Class

Syntax	No.of bits	Mnemonie
Video Seeion(){		
Video_session_start_code	sc+8=32	
do*{		
Video Object()		
}while(nextbits()==video_object_start_code)		
vedeo_session_end_code	sc+8=32	
}		

*concurrent loop solution to be provided by MSDL.

図20

Video Object

Syntax	No.of bits	Mnemonie
Video Object(){		
Video_object_start_code	sc+3=27	
Video_object_Id	5	
do{		
Video Object Layer()		
}while(nextbits()==		
video_object_layer_start_code)		
}		

図21

Video Object Layer

Syntax	No.of bits	Mnemonie
Video Object Layer(){		
video_object_layer_start_code	sc+4=28	
video_object_layer_id	4	
video_object_layer_shape	2	
if(video_object_layer_shape== "00"){		
video_object_layer_width	10	
video_object_layer_height	10	
}		
video_object_layer_quant_type	1	
if(video_object_layer_quant_type){		
load_intra_quant_mat	1	
If(load_intra_quant_mat)		
intra_quant_mat[64]	8*64	
load_nonintra_quant_mat	1	
if(load_nonintra_quant_mat)		
nonintra_quant_mat[64]	8*64	
}		
intra_dcpred_disable	1	
video_object_layer_fcode_forward	2	
video_object_layer_fcode_backward	2	
separate_motion_shape_texture	1	
scalability	1	
if(scalability){		
ref_layer_id	4	
ref_layer_sampling_direc	1	
hor_sampling_factor_n	5	
hor_sampling_factor_m	5	
vert_sampling_factor_n	5	
vert_sampling_factor_m	5	
enhancement_type	1	
}		
do{		
video object Plane()		
}while(nextbits()==		
video_object_plane_plane_start_code)		
}		

図22

Video Object Plane

Syntax	No.of bits	Mnemonie
Video Object Plane(){		
VOP_start_code	sc+8=32	
do{		
modulo_time_base	1	
}while(modulo_time_base_!= "0")		
VOP_time_increment	10	
VOP_prediction_type	2	
if(video_object_layer_shape!= "0") {		
VOP_width	10	
VOP_height	10	
VOP_horizontal_mc_spatial_ref	10	
marker_bit	1	
VOP_vertical_mc_spatial_ref	10	
if(scalability && enhancement_type)	1	
background_composition		
}		
if(VOP_prediction_type== "10")		
VOP_dbquant	2	
else		
VOP_quant	5	
if(!scalability){		
if(!separate_motion_shape_texture)		
combined_motion_shape_texture_coding()		
else{		
do{		
first_MMR_code	1-2	
}while(count of macroblocks! =total number of macroblocks)		
motion_coding()		
shade_coding()		
texture_coding()		
}		
}		
else{		
if(background_composition){		
load_backward_shape	1	
if(load_backward_shape){		
bacward_shape_coding()		
load_forward_shape	1	
if(load_forward_shape)		
forward_shape_coding()		
}		
}		
ref_select_code	2	
if(VOP_prediction_type== "01"		
|| VOP_prediction_type== "10") {		
forward_temporal_ref	10	
if(VOP_prediction_type_== "10"){		
marker_bit	1	
backward_temporal_ref	10	
}		
}		
combined_motion_shape_texture_coding()		
}		
}		

図23

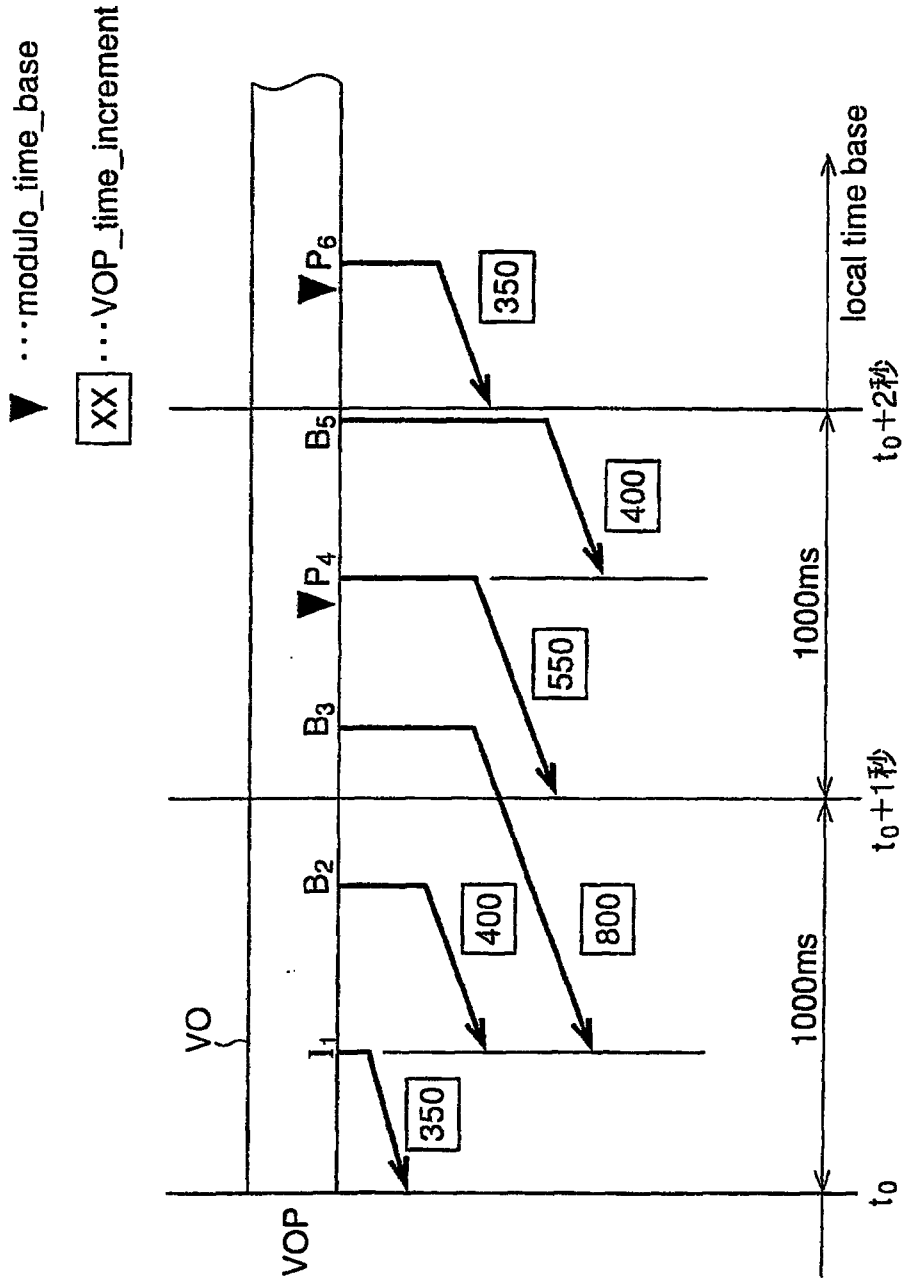

▼ ・・・modulo_time_base
XX ・・・VOP_time_increment
VO
VOP
I1
B2
B3
P4
B5
P6
350
400
800
550
400
350
1000ms
1000ms
local time base
t0
t0+1秒
t0+2秒

図24

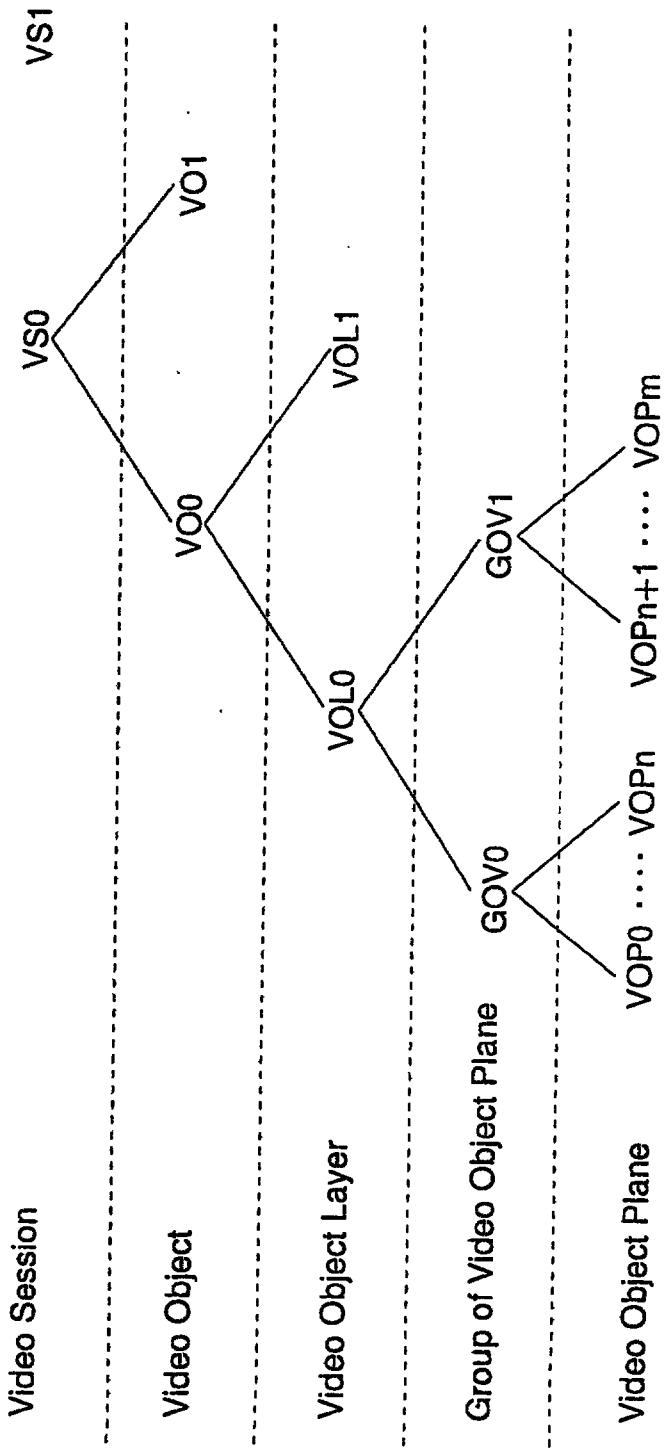

Video Session
Video Object
Video Object Layer
Group of Video Object Plane
Video Object Plane
VS0
VS1
VO0
VO1
VOL0
VOL1
GOV0
GOV1
VOP0 .... VOPn
VOPn+1 .... VOPm

図25

Group of Video Object Plane

Syntax	No.of bits	Mnemonic
group_of_plane(){		
group_start_code	32	bslbf
time_code	25	bslbf
closed_gop	1	uimsbf
broken_link	1	uimsbf
next_start_code()		
}		

図26

time_code

time_code	range of value	No. of bits	Mnemonic
drop_frame_flag		1	uimsbf
time_code_hours	0−23	5	uimsbf
time_code_minutes	0−59	6	uimsbf
marker_bit	1	1	bslbf
time_code_seconds	0−59	6	uimsbf
time_code_pictures	0−59	6	uimsbf

図27

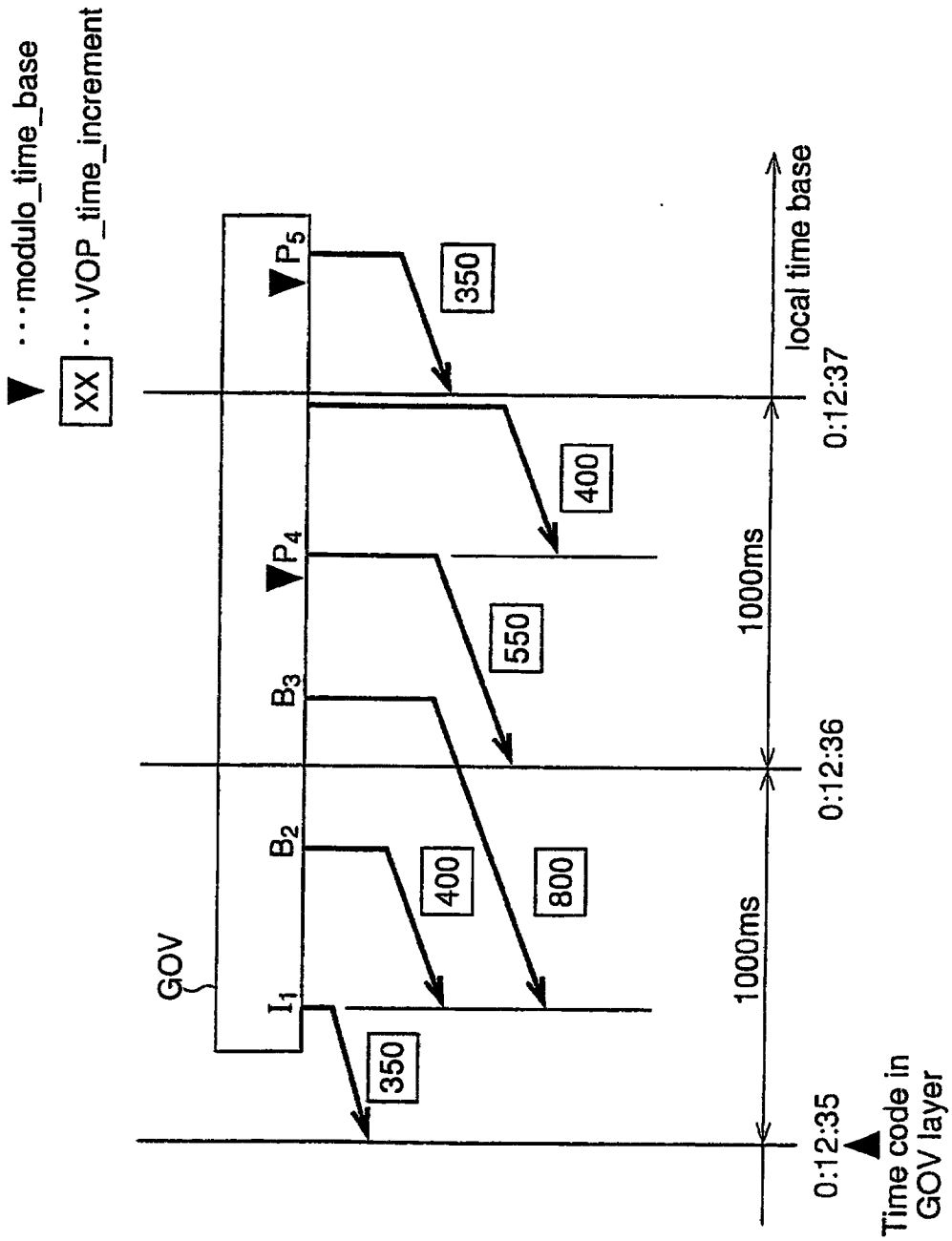

▼ ・・・modulo_time_base
XX ・・・VOP_time_increment
GOV
I1
B2
B3
P4
P5
350
400
800
550
400
350
1000ms
1000ms
local time base
0:12:37
0:12:36
0:12:35
Time code in GOV layer

図28

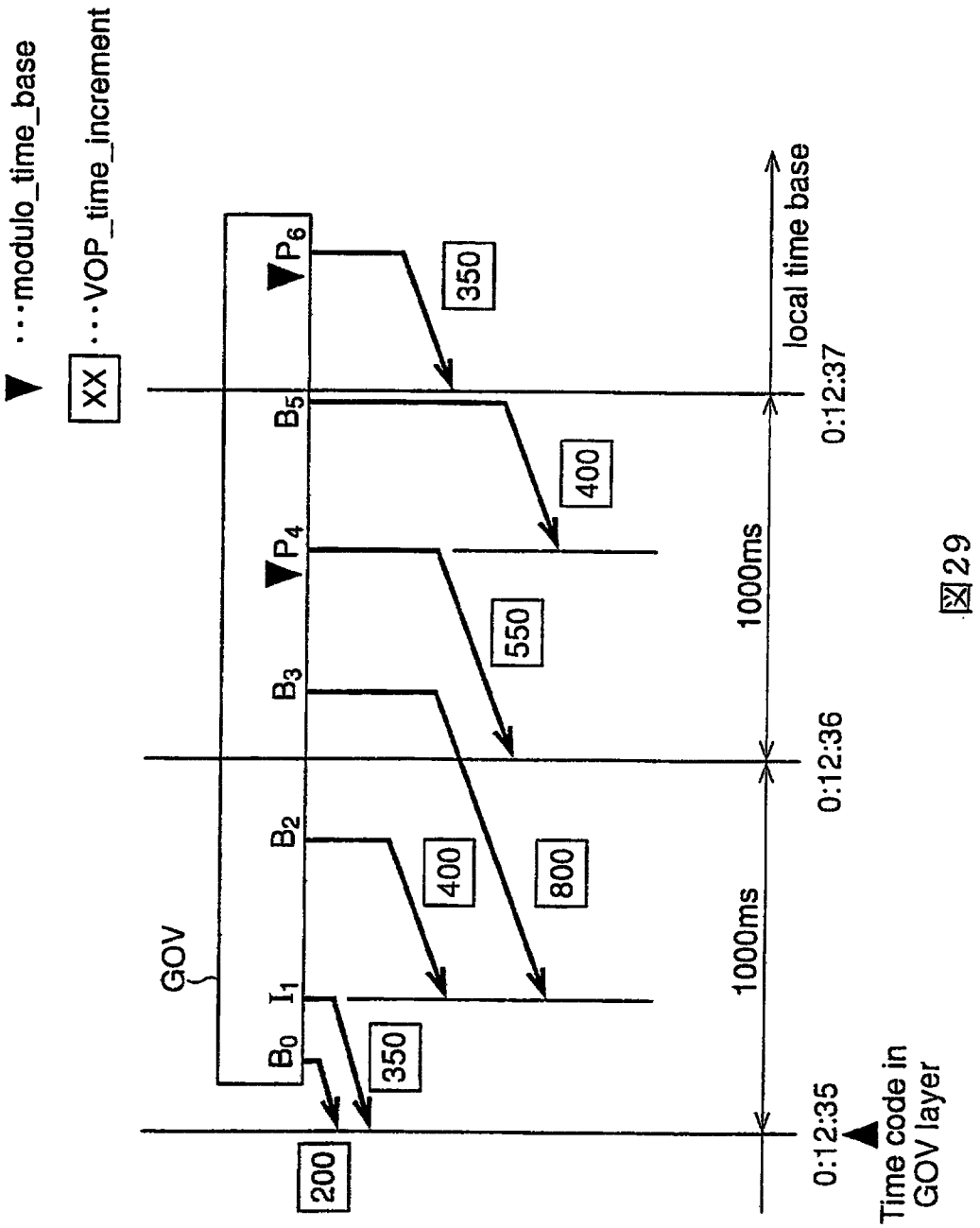

▼ ・・・modulo_time_base
XX ・・・VOP_time_increment
GOV
I_1
P_4
P_6
B_0
B_2
B_3
B_5
200
350
400
800
550
400
350
1000ms
1000ms
local time base
0:12:35
0:12:36
0:12:37
Time code in GOV layer

図29

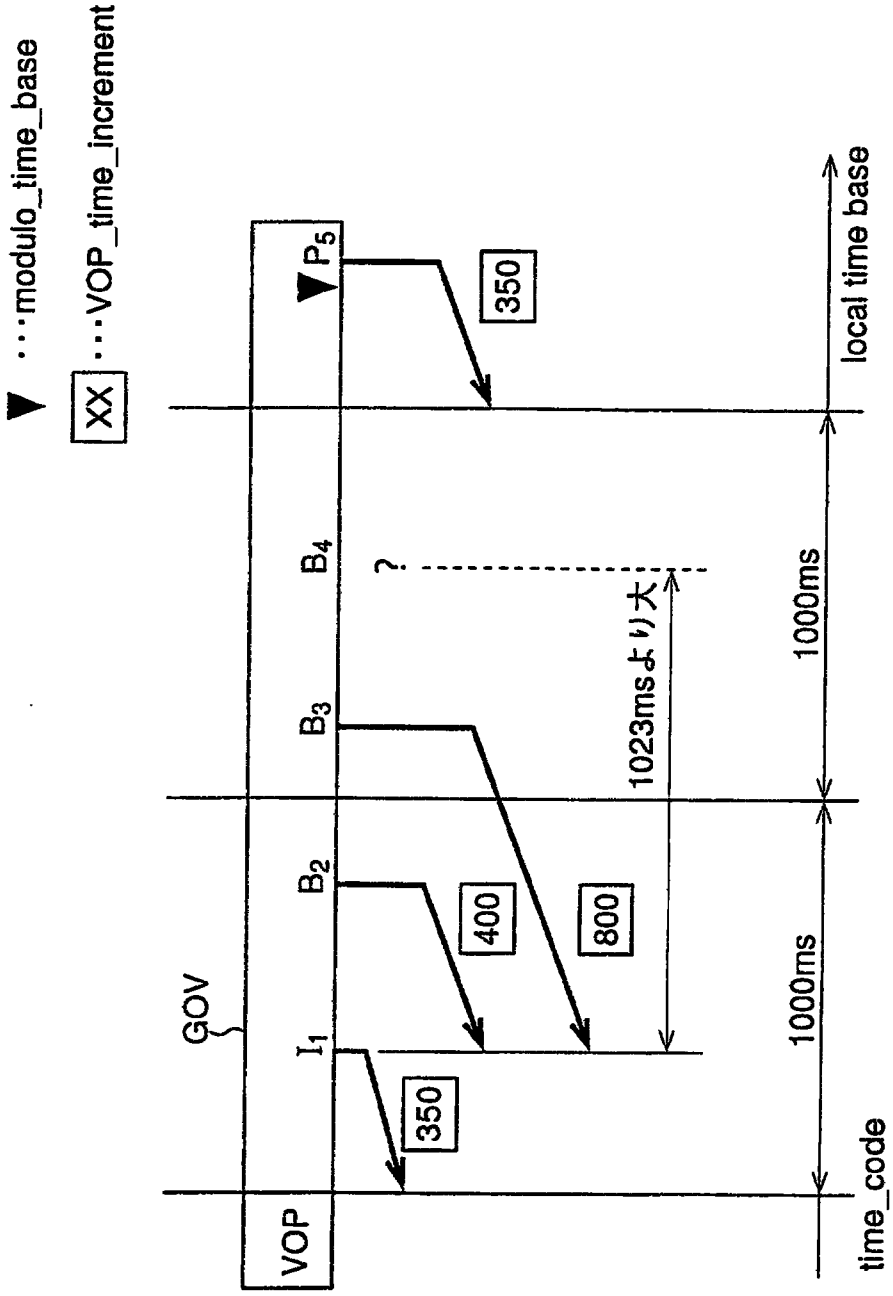

▼ ・・・modulo_time_base
XX ・・・VOP_time_increment
VOP
GOV
I1
B2
B3
B4
P5
350
400
800
350
?
1023msより大
1000ms
1000ms
local time base
time_code

図30

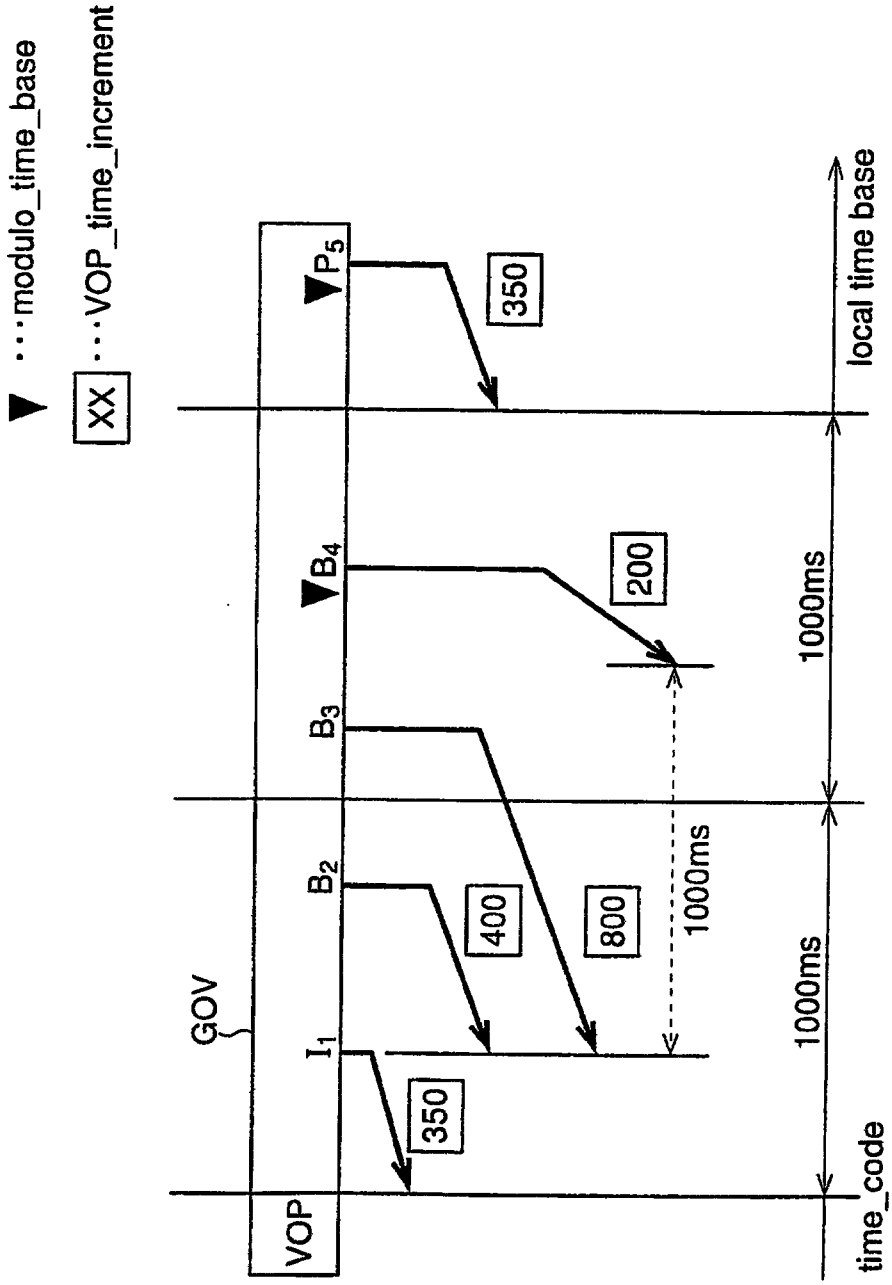

▼ ・・・modulo_time_base
XX ・・・VOP_time_increment
VOP
GOV
I_1
B_2
B_3
B_4
P_5
350
400
800
200
350
1000ms
1000ms
1000ms
local time base
time_code

図31

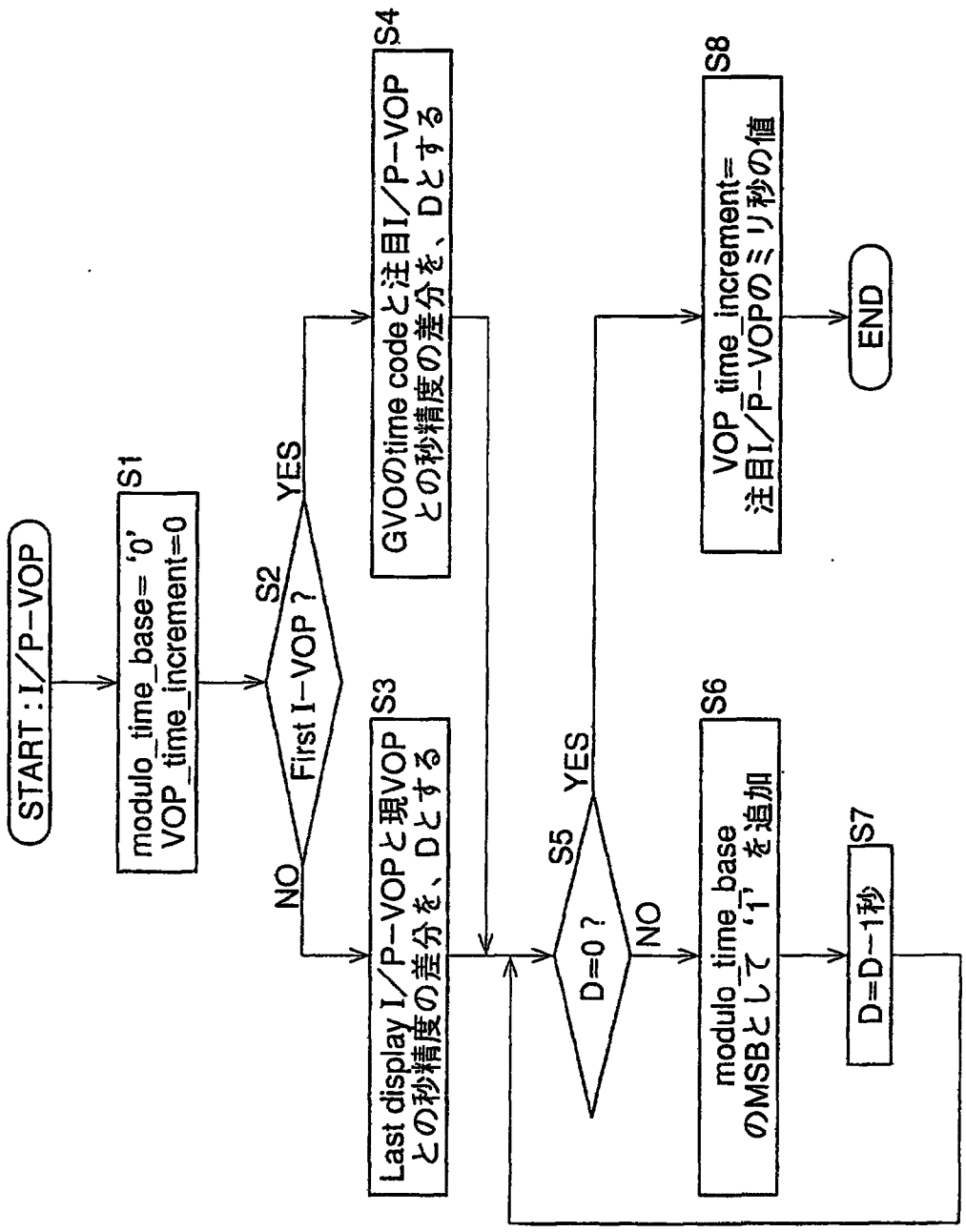

START : I/P-VOP
S1
modulo_time_base= '0'
VOP_time_increment=0
S2
First I-VOP ?
YES
NO
S3
Last display I/P-VOPと現VOP
との秒精度の差分を、Dとする
S4
GVOのtime codeと注目I/P-VOP
との秒精度の差分を、Dとする
S5
D=0 ?
YES
NO
S6
modulo_time_base
のMSBとして '1' を追加
S7
D=D-1秒
S8
VOP_time_increment=
注目I/P-VOPのミリ秒の値
END

図32

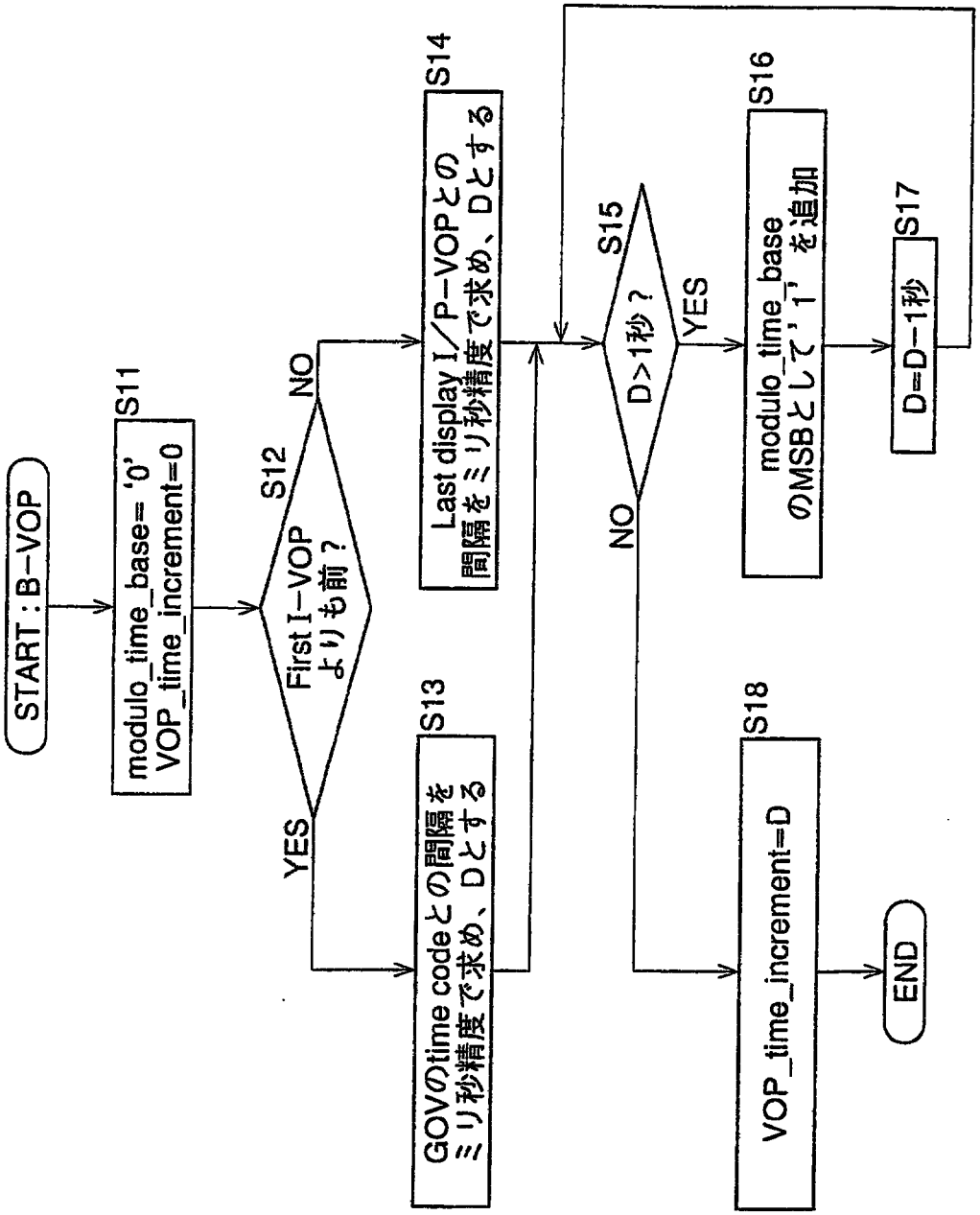

START : B-VOP
S11
modulo_time_base= '0'
VOP_time_increment=0
S12
First I-VOP
よりも前？
NO
YES
S13
GOVのtime codeとの間隔を
ミリ秒精度で求め、Dとする
S14
Last display I/P-VOPとの
間隔をミリ秒精度で求め、Dとする
S15
D>1秒？
YES
NO
S16
modulo_time_base
のMSBとして'1'を追加
S17
D=D-1秒
S18
VOP_time_increment=D
END

図33

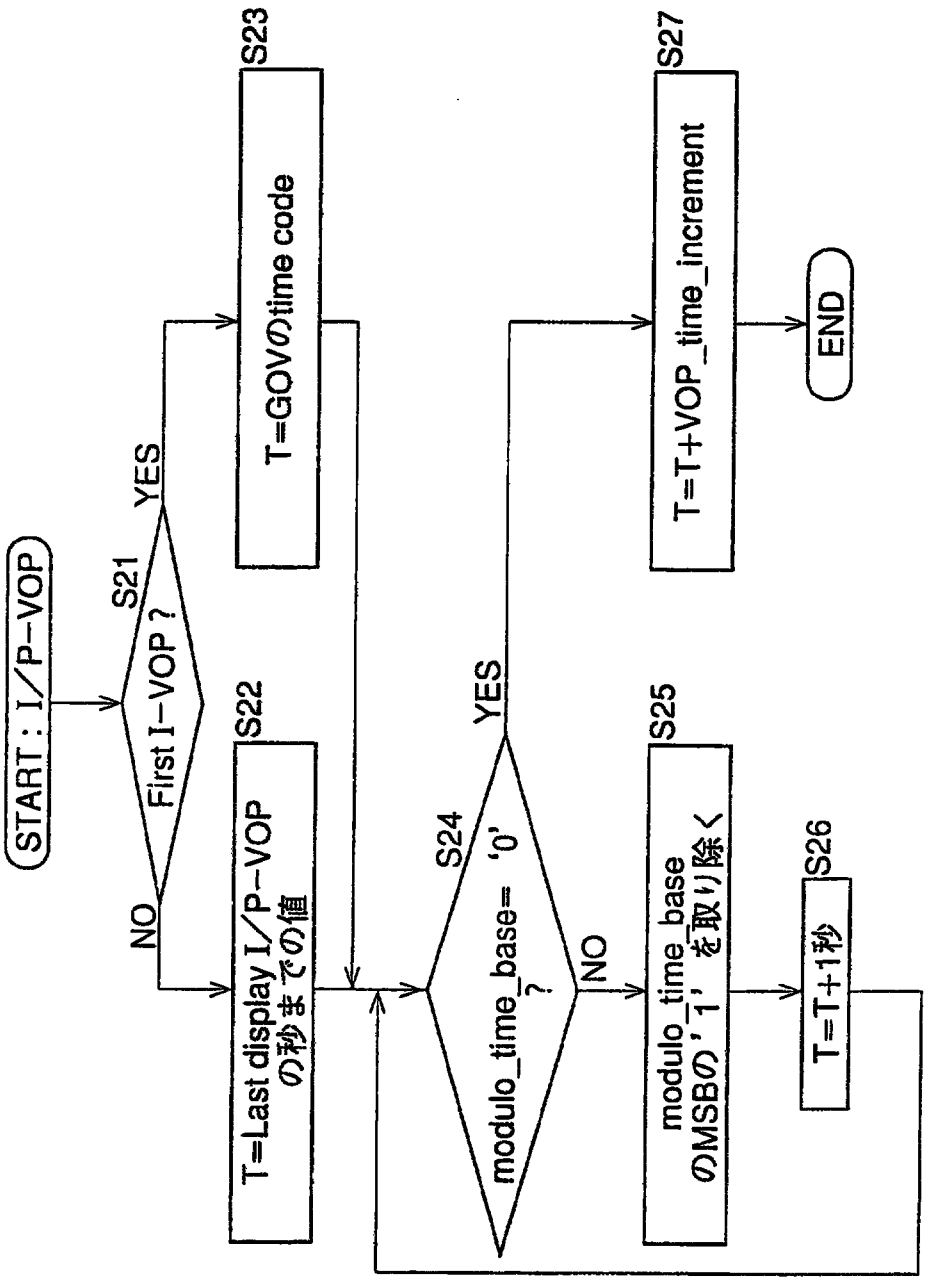

START : I/P-VOP
S21
First I-VOP ?
YES
NO
S23
T=GOVのtime code
S22
T=Last display I/P-VOP
の秒までの値
S24
modulo_time_base= '0'
?
YES
NO
S27
T=T+VOP_time_increment
END
S25
modulo_time_base
のMSBの'1'を取り除く
S26
T=T+1秒

図34

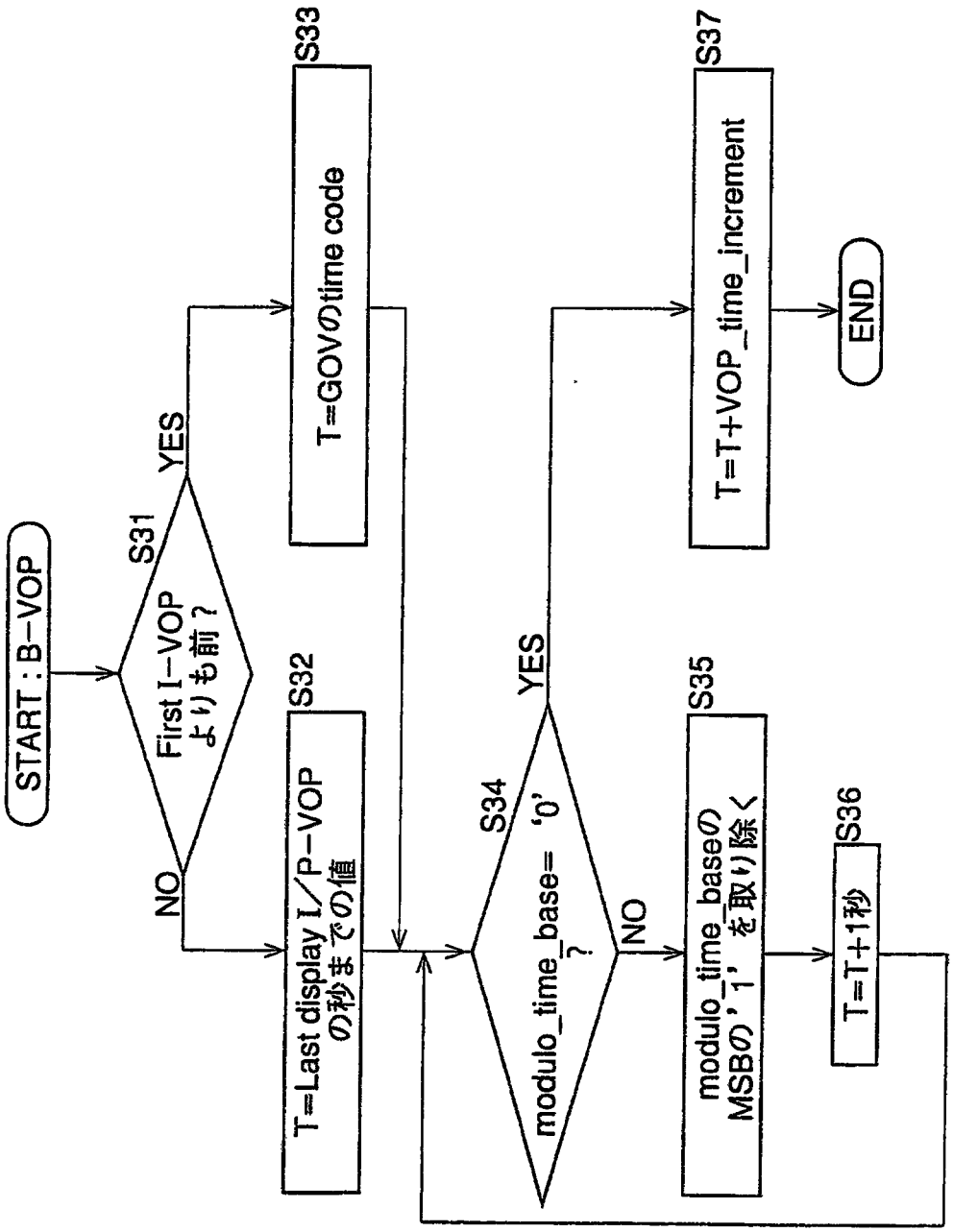

START : B-VOP
S31
First I-VOP よりも前？
YES
NO
S33
T=GOVのtime code
S32
T=Last display I/P-VOP の秒までの値
S34
modulo_time_base= '0' ？
YES
NO
S37
T=T+VOP_time_increment
END
S35
modulo_time_baseの MSBの'1'を取り除く
S36
T=T+1秒

図35

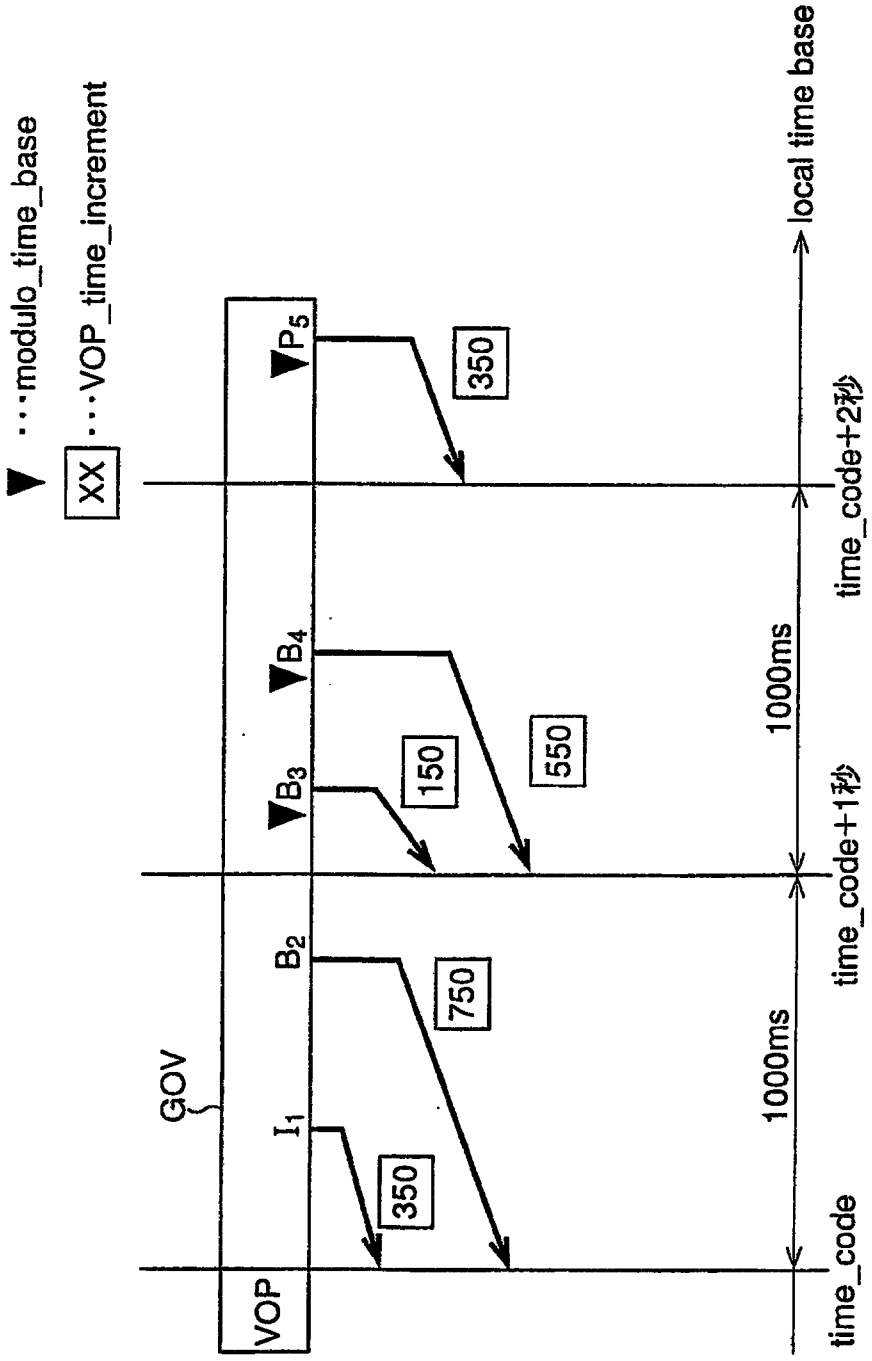

▼ ・・・modulo_time_base
XX ・・・VOP_time_increment
local time base
VOP
GOV
I_1
B_2
B_3
B_4
P_5
350
750
150
550
350
1000ms
1000ms
time_code
time_code+1秒
time_code+2秒

図36

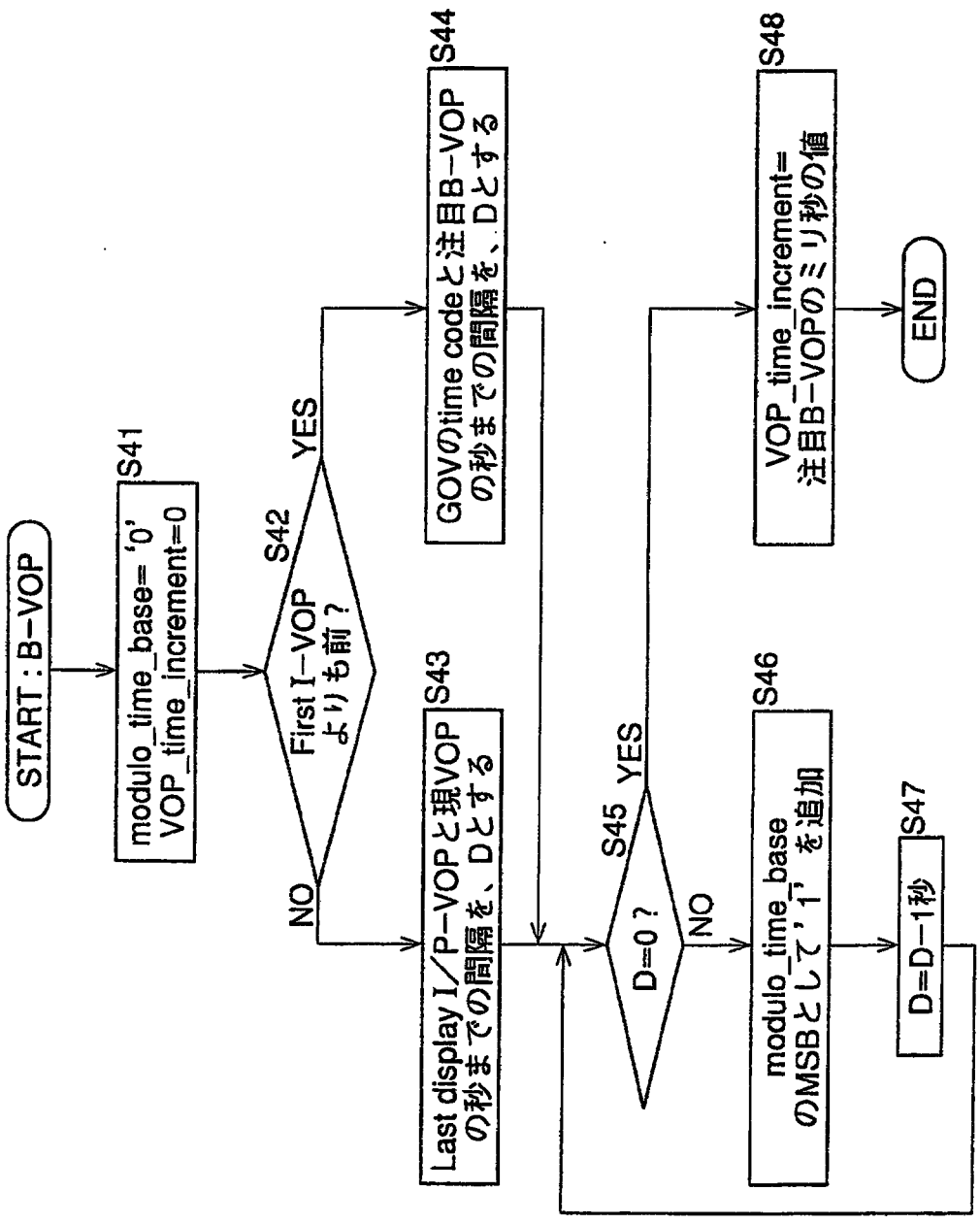

START : B-VOP
S41
modulo_time_base= '0'
VOP_time_increment=0
S42
First I-VOP
よりも前 ?
YES
NO
S44
GOVのtime codeと注目B-VOP
の秒までの間隔を、Dとする
S43
Last display I/P-VOPと現VOP
の秒までの間隔を、Dとする
S45
D=0 ?
YES
NO
S48
VOP_time_increment=
注目B-VOPのミリ秒の値
END
S46
modulo_time_base
のMSBとして'1'を追加
S47
D=D-1秒

図37

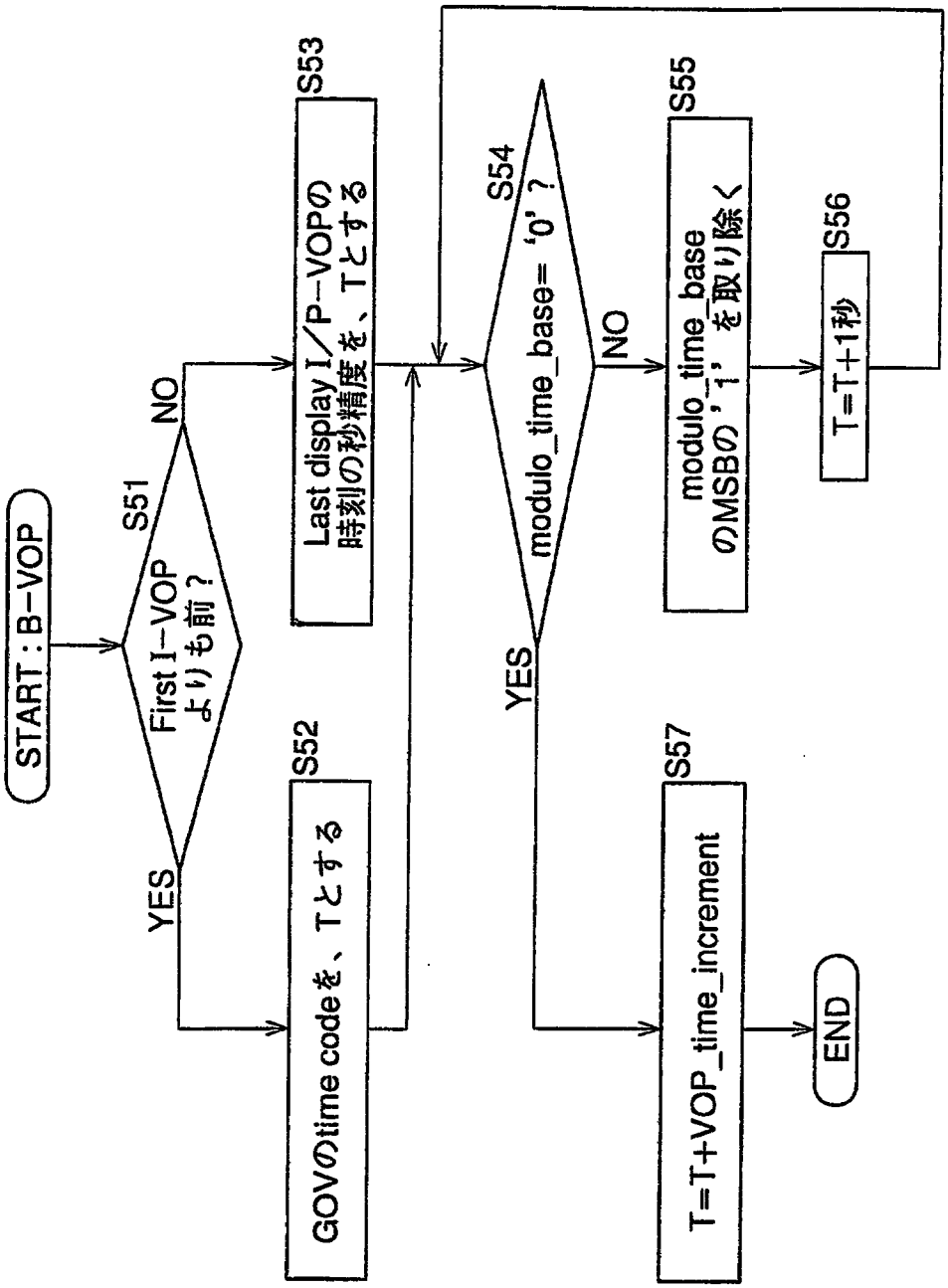

START : B-VOP
S51
First I-VOP よりも前 ?
NO
YES
S52
GOVのtime codeを、Tとする
S53
Last display I/P-VOPの時刻の秒精度を、Tとする
S54
modulo_time_base= '0' ?
NO
YES
S55
modulo_time_base のMSBの '1' を取り除く
S56
T=T+1秒
S57
T=T+VOP_time_increment
END

図38

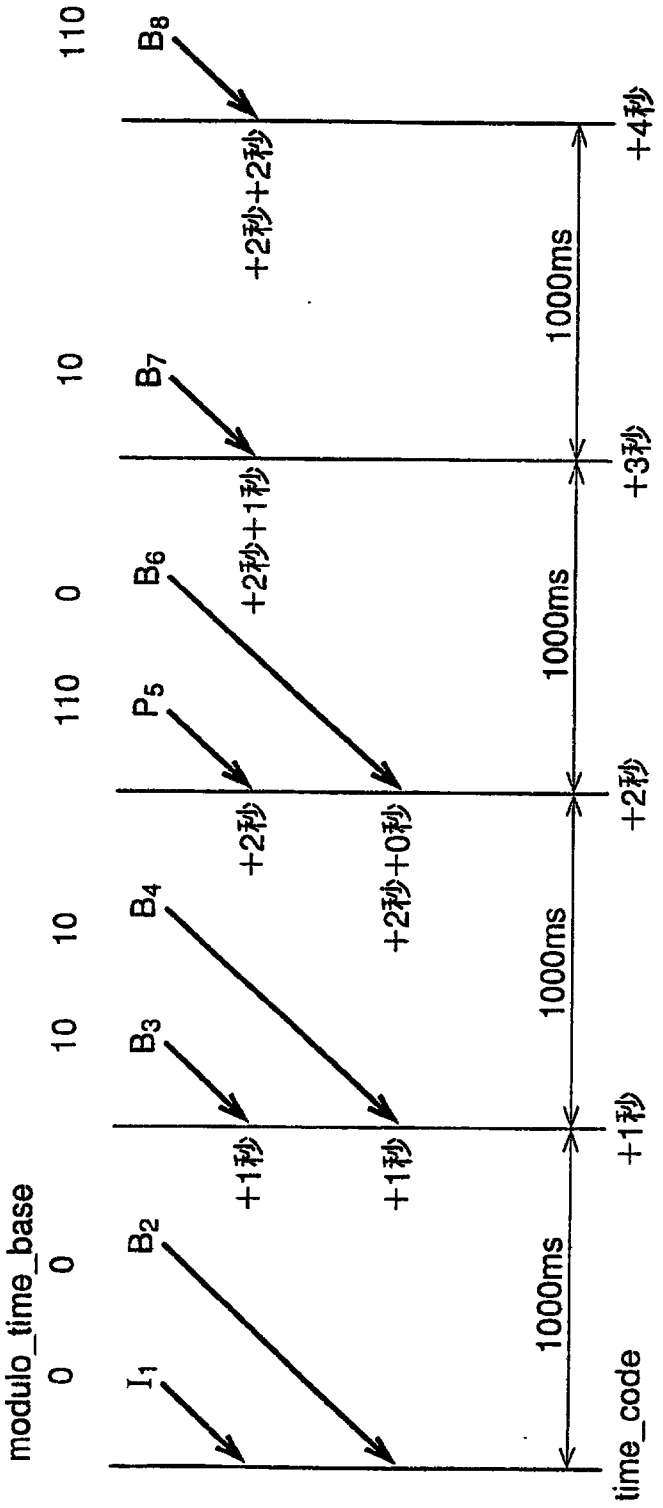

modulo_time_base
0
0
10
10
110
0
10
110
I1
B2
B3
B4
P5
B6
B7
B8
+1秒
+1秒
+2秒
+2秒+0秒
+2秒+1秒
+2秒+2秒
1000ms
1000ms
1000ms
1000ms
time_code
+1秒
+2秒
+3秒
+4秒

図39

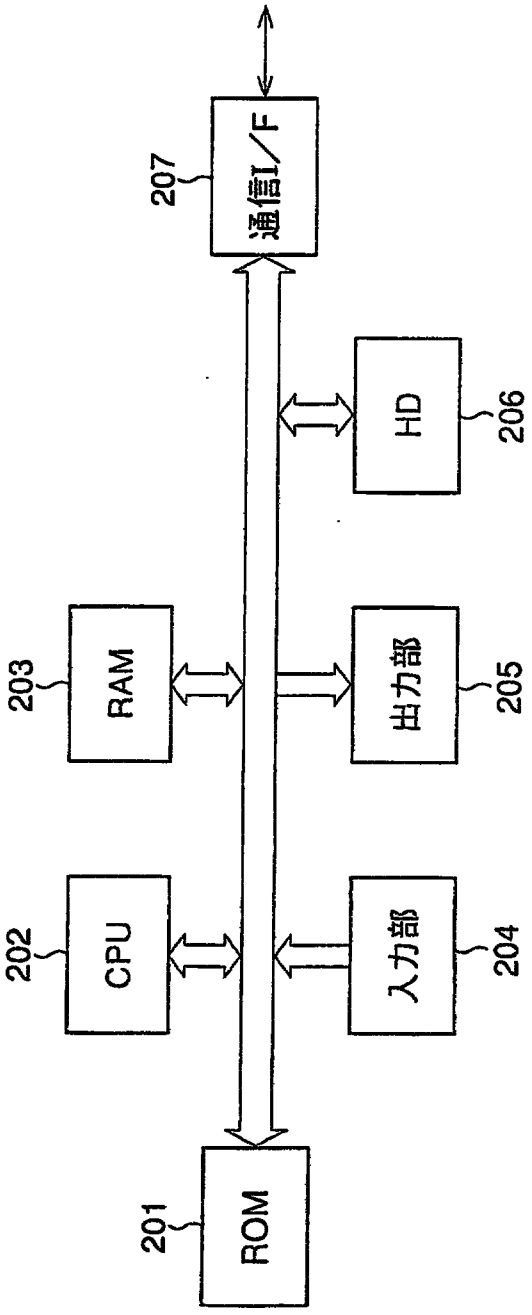


図40

符号の説明

1...VO構成部, 2_1～2_N...VOP構成部, 3_1～3_N...VOP符号化部, 4...多重化部, 21...画像階層化部, 23...上位レイヤ符号化部, 24...解像度変換部, 25...下位レイヤ符号化部, 26...多重化部, 31...フレームメモリ, 32...動きベクトル検出器, 33...演算器, 34...DCT器, 35...量子化器, 36...VLC器, 38...逆量子化器, 39...IDCT器, 40...演算器, 41...フレームメモリ, 42...動き補償器, 53...フレームメモリ, 71...逆多重化部, 72_1～72_N...VOP復号部, 73...画像再構成部, 91...逆多重化部, 93...上位レイヤ復号部, 94...解像度変換部, 95...下位レイヤ復号部, 102...IVLC器, 103...逆量子化器, 104...IDCT器, 105...演算器, 106...フレームメモリ, 107...動き補償器, 112...フレームメモリ, 201...ROM, 202...CPU, 203...RAM, 204...入力部, 205...出力部, 206...HD, 207...通信I／F

SPECIFICATION

Image Encoder, Image Encoding Method, Image Decoder, Image Decoding Method, and Distribution Media

Technical Field

The present invention relates to an image encoder, an image encoding method, an image decoder, an image decoding method, and distribution media. More particularly, the invention relates to an image encoder, an image encoding method, an image decoder, an image decoding method, and distribution media suitable for use, for example, in the case where dynamic image data is recorded on storage media, such as a magneto-optical disk, magnetic tape, etc., and also the recorded data is regenerated and displayed on a display, or in the case where dynamic image data is transmitted from a transmitter side to a receiver side through a transmission path and, on the receiver side, the received dynamic image data is displayed or it is edited and recorded, as in videoconference systems, videophone systems, broadcasting equipment, and multimedia data base retrieval systems.

Background Art

For instance, as in videoconference systems and videophone systems, in systems which transmit dynamic image data to a remote place, image data is compressed and encoded by taking advantage of the line correlation and interframe correlation in order to take efficient advantage of transmission paths.

As a representative high-efficient dynamic image encoding system, there is a dynamic image encoding system for storage media, based on Moving Picture Experts Group (MPEG) standard. This MPEG standard has been discussed by the International Organization for Standardization (ISO)-IEC/JTC1/SC2/WG11 and has been proposed as a proposal for standard. The MPEG standard has adopted a hybrid system using a combination of motion compensative predictive coding and discrete cosine transform (DCT) coding.

The MPEG standard defines some profiles and levels in order to support a wide range of applications and functions. The MPEG standard is primarily based on Main Profile at Main level (MP@ML).

Figure 1 illustrates the constitution example of an MP@ML encoder in the MPEG standard system.

Image data to be encoded is input to frame memory 31 and stored temporarily. A motion vector detector 32 reads out image data stored in the frame memory 31, for example, at a macroblock unit constituted by 16 (16 pixels, and

detects the motion vectors.

Here, the motion vector detector 32 processes the image data of each frame as any one of an intracoded picture (I-picture), a forward predictive-coded picture (P-picture), or a bidirectionally predictive-coded picture (B-picture). Note that how images of frames input in sequence are processed as I-, P-, and B-pictures has been predetermined (e.g., images are processed as I-picture, B-picture, P-picture, B-picture, P-picture, ..., B-picture, and P-picture in the recited order).

That is, in the motion vector detector 32, reference is made to a predetermined reference frame in the image data stored in the frame memory 31, and a small block of 16 pixels (16 lines (macroblock) in the current frame to be encoded is matched with a set of blocks of the same size in the reference frame. With block matching, the motion vector of the macroblock is detected.

Here, in the MPEG standard, predictive modes for an image include four kinds: intracoding, forward predictive coding, backward predictive coding, and bidirectionally predictive coding. An I-picture is encoded by intracoding. A P-picture is encoded by either intracoding or forward predictive coding. A B-picture is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive coding.

That is, the motion vector detector 32 sets the intracoding mode to an I-picture as a predictive mode. In this case, the motion vector detector 32 outputs the predictive mode (intracoding mode) to a variable word length coding (VLC) unit 36 and a motion compensator 42 without detecting the motion vector.

The motion vector detector 32 also performs forward prediction for a P-picture and detects the motion vector. Furthermore, in the motion vector detector 32, a prediction error caused by performing forward prediction is compared with dispersion, for example, of macroblocks to be encoded (macroblocks in the P-picture). As a result of the comparison, when the dispersion of the macroblocks is smaller than the prediction error, the motion vector detector 32 sets an intracoding mode as the predictive mode and outputs it to the VLC unit 36 and motion compensator 42. Also, if the prediction error caused by performing forward prediction is smaller, the motion vector detector 32 sets a forward predictive coding mode as the predictive mode. The forward predictive coding mode, along with the detected motion vector, is output to the VLC unit 36 and motion compensator 42.

The motion vector detector 32 further performs forward prediction, backward prediction, and bidirectional prediction for a B-picture and detects the respective motion

vectors. Then, the motion vector detector 32 detects the minimum error from among the prediction errors in the forward prediction, backward prediction, and bidirectional prediction (hereinafter referred to the minimum prediction error as needed), and compares the minimum prediction error with dispersion, for example, of macroblocks to be encoded (macroblocks in the B-picture). As a result of the comparison, when the dispersion of the macroblocks is smaller than the minimum prediction error, the motion vector detector 32 sets an intracoding mode as the predictive mode and outputs it to the VLC unit 36 and motion compensator 42. Also, if the minimum prediction error is smaller, the motion vector detector 32 sets as the predictive mode a predictive mode in which the minimum prediction error was obtained. The predictive mode, along with the corresponding motion vector, is output to the VLC unit 36 and motion compensator 42.

If the motion compensator 42 receives both the predictive mode and the motion vector from the motion vector detector 32, the motion compensator 42 will read out the coded and previously locally decoded image data stored in the frame memory 41 in accordance with the received predictive mode and motion vector. This read image data is supplied to arithmetic units 33 and 40 as predicted image data.

The arithmetic unit 33 reads from the frame memory 31 the same macroblock as the image data read out from the frame memory 31 by the motion vector detector 32, and computes the difference between the macroblock and the predicted image which was supplied from the motion compensator 42. This differential value is supplied to a DCT unit 34.

On the other hand, in the case where a predictive mode alone is received from the motion vector detector 32, i.e., the case where a predictive mode is an intracoding mode, the motion compensator 42 does not output a predicted image. In this case, the arithmetic unit 33 (the arithmetic unit 40 as well) outputs to the DCT unit 34 the macroblock read out from the frame memory 31 without processing it.

In the DCT unit 34, DCT is applied to the output data of the arithmetic unit 33, and the resultant DCT coefficients are supplied to a quantizer 35. In the quantizer 35, a quantization step (quantization scale) is set in correspondence to the data storage quantity of the buffer 37 (which is the quantity of the data stored in a buffer 37) (buffer feedback). In the quantization step, the DCT coefficients from the DCT unit 34 are quantized. The quantized DCT coefficients (hereinafter referred to as quantized coefficients as needed), along with the set quantization step, are supplied to the VLC unit 36.

In the VLC unit 36, the quantized coefficients supplied by the quantizer 35 are transformed to variable word length codes such as Huffman codes and output to the buffer 37. Furthermore, in the VLC unit 36, the quantization step from the quantizer 35 is encoded by variable word length coding, and likewise the predictive mode (indicating either intracoding (image predictive intracoding), forward predictive coding, backward predictive coding, or bidirectionally predictive coding) and motion vector from the motion vector detector 32 are encoded. The resultant coded data is output to the buffer 37.

The buffer 37 temporarily stores the coded data supplied from the VLC unit 36, thereby smoothing the stored quantity of data. For example, the smoothed data is output to a transmission path or recorded on a storage medium, as a coded bit stream.

The buffer 37 also outputs the stored quantity of data to the quantizer 35. The quantizer 35 sets a quantization step in correspondence to the stored quantity of data output by this buffer 37. That is, when there is a possibility that the capacity of the buffer 37 will overflow, the quantizer 35 increases the size of the quantization step, thereby reducing the data quantity of quantized coefficients. When there is a possibility that the capacity of the buffer 37 will be caused to be in a state of underflow, the

quantizer 35 reduces the size of the quantization step, thereby increasing the data quantity of quantized coefficients. In this manner, the overflow and underflow of the buffer 37 are prevented.

The quantized coefficients and quantization step, output by the quantizer 35, are not supplied only to the VLC unit 36 but also to an inverse quantizer 38. In the inverse quantizer 38, the quantized coefficients from the quantizer 35 are inversely quantized according to the quantization step supplied from the quantizer 35, whereby the quantized coefficients are transformed to DCT coefficients. The DCT coefficients are supplied to an inverse DCT unit (IDCT unit) 39. In the IDCT 39, an inverse DCT is applied to the DCT coefficients and the resultant data is supplied to the arithmetic unit 40.

In addition to the output data of the IDCT unit 39, the same data as the predicted image supplied to the arithmetic unit 33 is supplied from the motion compensator 42 to the arithmetic unit 40, as described above. The arithmetic unit 40 adds the output data (prediction residual (differential data)) of the IDCT unit 39 and the predicted image data of the motion compensator 42, thereby decoding the original image data locally. The locally decoded image data is output. (However, in the case where a predictive mode is an intracoding mode, the output data of the IDCT 39 is passed

through the arithmetic unit 40 and supplied to the frame memory 41 as locally decoded image data without being processed.) Note that this decoded image data is consistent with decoded image data that is obtained at the receiver side.

The decoded image data obtained in the arithmetic unit 40 (locally decoded image data) is supplied to the frame memory 41 and stored. Thereafter, the decoded image data is employed as reference image data (reference frame) with respect to an image to which intracoding (forward predictive coding, backward predictive coding, or bidirectionally predictive coding) is applied.

Next, Figure 2 illustrates the constitution example of an MP@ML decoder in the MPEG standard system which decodes the coded data output from the encoder of Figure 1.

The coded bit stream (coded data) transmitted through a transmission path is received by a receiver (not shown), or the coded bit stream (coded data) recorded in a storage medium is regenerated by a regenerator (not shown). The received or regenerated bit stream is supplied to a buffer 101 and stored.

An inverse VLC unit (IVLC unit (variable word length decoder) 102 reads out the coded data stored in the buffer 101 and performs variable length word decoding, thereby separating the coded data into the motion vector, predictive

mode, quantization step, and quantized coefficients at a macroblock unit. Among them, the motion vector and the predictive mode are supplied to a motion compensator 107, while the quantization step and the quantized macroblock coefficients are supplied to an inverse quantizer 103.

In the inverse quantizer 103, the quantized macroblock coefficients supplied from the IVLC unit 102 are inversely quantized according to the quantization step supplied from the same IVLC unit 102. The resultant DCT coefficients are supplied to an IDCT unit 104. In the IDCT 104, an inverse DCT is applied to the macroblock DCT coefficients supplied from the inverse quantizer 103, and the resultant data is supplied to an arithmetic unit 105.

In addition to the output data of the IDCT unit 104, the output data of the motion compensator 107 is also supplied to the arithmetic unit 105. That is, in the motion compensator 107, as in the case of the motion compensator 42 of Figure 1, the previously decoded image data stored in the frame memory 106 is read out according to the motion vector and predictive mode supplied from the IVLC unit 102 and is supplied to the arithmetic unit 105 as predicted image data. The arithmetic unit 105 adds the output data (prediction residual (differential value)) of the IDCT unit 104 and the predicted image data of the motion compensator 107, thereby decoding the original image data. This decoded image data

is supplied to the frame memory 106 and stored. Note that, in the case where the output data of the IDCT unit 104 is intracoded data, the output data is passed through the arithmetic unit 105 and supplied to the frame memory 106 as decoded image data without being processed.

The decoded image data stored in the frame memory 106 is employed as reference image data for the next image data to be decoded. Furthermore, the decoded image data is supplied, for example, to a display (not shown) and displayed as an output reproduced image.

Note that in MPEG-1 standard and MPEG-2 standard, a B-picture is not stored in the frame memory 41 in the encoder (Figure 1) and the frame memory 106 in the decoder (Figure 2), because it is not employed as reference image data.

The aforementioned encoder and decoder shown in Figures 1 and 2 are based on MPEG-1/2 standard. Currently a system for encoding video at a unit of the video object (VO) of an object sequence constituting an image is being standardized as MPEG-4 standard by the ISO-IEC/JTC1/SC29/WG11.

Incidentally, since the MPEG-4 standard is being standardized on the assumption that it is primarily used in the field of communication, it does not prescribe the group of pictures (GOP) prescribed in the MPEG-1/2 standard. Therefore, in the case where the MPEG-4 standard is utilized in storage media, efficient random access will be difficult.

Disclosure of Invention

The present invention has been made in view of such circumstances and therefore the object of the invention is to make efficient random access possible.

An image encoder as set forth in claim 1 is characterized in that it comprises encoding means for partitioning one or more layers of each sequence of objects constituting an image into a plurality of groups and encoding the groups.

An image encoding method as set forth in claim 6 is characterized in thàt it comprises an encoding step of partitioning one or more layers of each sequence of objects constituting an image into a plurality of groups and encoding the groups.

An image encoder as set forth in claim 11 is characterized in that it comprises decoding means for decoding a coded bit stream obtained by partitioning one or more layers of each sequence of objects constituting an image into a plurality of groups and also by encoding the groups.

An image decoding method as set forth in claim 16 is characterized in that it comprises an decoding step of decoding a coded bit stream obtained by partitioning one or more layers of each sequence of objects constituting an

image into a plurality of groups and also by encoding the groups.

A distribution medium as set forth in claim 21 is characterized in that it distributes the coded bit stream which is obtained by partitioning one or more layers of each sequence of objects constituting an image into a plurality of groups and encoding the groups.

An image encoder as set forth in claim 26 is characterized in that it comprises: second-accuracy time information generation means for generating second-accuracy time information which indicates time within accuracy of a second; and detailed time information generation means for generating detailed time information which indicates a time period between the second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second.

An image decoding method as set forth in claim 29 is characterized in that it comprises: a second-accuracy time information generation step of generating second-accuracy time information which indicates time within accuracy of a second; and a detailed time information generation step of generating detailed time information which indicates a time period between the second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second.

An image decoder as set forth in claim 32 is characterized in that it comprises display time computation means for computing display time of I-VOP, P-VOP, or B-VOP on the basis of the second-accuracy time information and detailed time information.

An image decoding method as set forth in claim 35 is characterized in that it comprises a display time computation step of computing display time of I-VOP, P-VOP, or B-VOP on the basis of the second-accuracy time information and detailed time information.

A distribution medium as set forth in claim 38 is characterized in that it distributes a coded bit stream which is obtained by generating second-accuracy time information which indicates time within accuracy of a second, also by generating detailed time information which indicates a time period between the second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second, and furthermore by adding the second-accuracy time information and detailed time information to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates display time of the I-VOP, P-VOP, or B-VOP.

In the image encoder as set forth in claim 1, the encoding means partitions one or more layers of each of objects constituting an image into a plurality of groups and

encodes the groups.

In the image encoding method as set forth in claim 6, one or more layers of each sequence of objects constituting an image is partitioned into a plurality of groups, and the groups are encoded.

In the image encoder as set forth in claim 11, the decoding means decodes a coded bit stream obtained by partitioning one or more layers of each sequence of objects constituting an image into a plurality of groups and also by encoding the groups.

In the image decoding method as set forth in claim 16, a coded bit stream, obtained by partitioning one or more layers of each of objects constituting an image into a plurality of groups and also by encoding the groups is decoded.

In the distribution medium as set forth in claim 21, a coded bit stream, which is obtained by partitioning one or more layers of each sequence of objects constituting an image into a plurality of groups and encoding the groups is distributed.

In the image encoder as set forth in claim 26, the second-accuracy time information generation means generates second-accuracy time information which indicates time within accuracy of a second, and the detailed time information generation means generates detailed time information which

indicates a time period between the second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second.

In the image encoding method as set forth in claim 29, second-accuracy time information which indicates time within accuracy of a second is generated, and detailed time information, which indicates a time period between the second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second is generated.

In the image decoder as set forth in claim 32, the display time computation means computes display time of I-VOP, P-VOP, or B-VOP on the basis of the second-accuracy time information and detailed time information.

In the image decoding method as set forth in claim 35, display time of I-VOP, P-VOP, or B-VOP is computed on the basis of the second-accuracy time information and detailed time information.

In the distribution medium as set forth in claim 38, the medium distributes a coded bit stream which is obtained by generating second-accuracy time information which indicates time within accuracy of a second, also by generating detailed time information which indicates a time period between the second-accuracy time information directly

before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second, and furthermore by adding the second-accuracy time information and detailed time information to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates display time of the I-VOP, P-VOP, or B-VOP.

Brief Description of the Drawings

Figure 1 is a block diagram showing the constitution example of a conventional encoder;

Figure 2 is a block diagram showing the constitution example of a conventional decoder;

Figure 3 is a block diagram showing the constitution example of an embodiment of an encoder to which the present invention is applied;

Figure 4 is a diagram for explaining that the position and size of a video object (VO) vary with time;

Figure 5 is a block diagram showing the constitution example of the VOP encoding sections 31 to 3N of Figure 3;

Figure 6 is a diagram for explaining spatial scalability;

Figure 7 is a diagram for explaining spatial scalability;

Figure 8 is a diagram for explaining spatial scalability;

Figure 9 is a diagram for explaining spatial scalability;

Figure 10 is a diagram for explaining a method of determining the size data and offset data of a video object plane (VOP);

Figure 11 is a block diagram showing the constitution example of the base layer encoding section 25 of Figure 5;

Figure 12 is a block diagram showing the constitution example of the enhancement layer encoding section 23 of Figure 5;

Figure 13 is a diagram for explaining spatial scalability;

Figure 14 is a diagram for explaining time scalability;

Figure 15 is a block diagram showing the constitution example of an embodiment of a decoder to which the present invention is applied;

Figure 16 is a block diagram showing another constitution example of the VOP decoding sections 72_1 to 72_N of Figure 15;

Figure 17 is a block diagram showing the constitution example of the base layer decoding section 95 of Figure 16;

Figure 18 is a block diagram showing the constitution example of the enhancement layer decoding section 93 of Figure 16;

Figure 19 is a diagram showing the syntax of a bit

stream obtained by scalable coding;

Figure 20 is a diagram showing the syntax of VS;

Figure 21 is a diagram showing the syntax of a VO;

Figure 22 is a diagram showing the syntax of a VOL;

Figure 23 is a diagram showing the syntax of a VOP;

Figure 24 is a diagram showing the relation between modulo_time_base and VOP_time_increment;

Figure 25 is a diagram showing the syntax of a bit stream according to the present invention;

Figure 26 is a diagram showing the syntax of a GOV;

Figure 27 is a diagram showing the constitution of time_code;

Figure 28 is a diagram showing a method of encoding the time_code of the GOV layer and the modulo_time_base and VOP_time_increment of the first I-VOP of the GOV;

Figure 29 is a diagram showing a method of encoding the time_code of the GOV layer and also the modulo_time_base and VOP_time_increment of the B-VOP located before the first I-VOP of the GOV;

Figure 30 is a diagram showing the relation between the modulo_time_base and the VOP_time_increment when the definitions thereof are not changed;

Figure 31 is a diagram showing a process of encoding the modulo_time_base and VOP_time_increment of the B-VOP, based on a first method;

Figure 32 is a flowchart showing a process of encoding the modulo_time_base and VOP_time_increment of I/P-VOP, based on a first method and a second method;

Figure 33 is a flowchart showing a process of encoding the modulo_time_base and VOP_time_increment of the B-VOP, based on a first method;

Figure 34 is a flowchart showing a process of decoding the modulo_time_base and VOP_time_increment of the I/P-VOP encoded by the first and second methods;

Figure 35 is a flowchart showing a process of decoding the modulo_time_base and VOP_time_increment of the B-VOP encoded by the first method;

Figure 36 is a diagram showing a process of encoding the modulo_time_base and VOP_time_increment of the B-VOP, based on a second method;

Figure 37 is a flowchart showing the process of encoding the modulo_time_base and VOP_time_increment of the B-VOP, based on the second method;

Figure 38 is a flowchart showing a process of decoding the modulo_time_base and VOP_time_increment of the B-VOP encoded by the second method;

Figure 39 is a diagram for explaining the modulo_time_base; and

Figure 40 is a block diagram showing the constitution example of another embodiment of an encoder and a decoder to

which the present invention is applied.

Best Mode for Carrying Out the Invention

Embodiments of the present invention will hereinafter be described in detail with reference to the drawings. Before that, in order to make clear the corresponding relation between each means of the present invention as set forth in claims and the following embodiments, the characteristics of the present invention will hereinafter be described in detail by adding a corresponding embodiment within a parenthesis after each means. The corresponding embodiment is merely an example.

That is, the image encoder as set forth in claim 1 is characterized in that it is an image encoder for encoding an image and outputting the resultant coded bit stream, the image encoder comprising: receiving means for receiving the image (e.g., frame memory 31 shown in Figure 11 or 12, etc.); and encoding means for partitioning one or more layers of each of objects constituting the image into a plurality of groups and encoding the groups (e.g., VLC unit 36 shown in Figure 11 or 12, etc.).

The image encoder as set forth in claim 4 is characterized in that, when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either

intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive coding is a bidirectionally predictive-VOP (B-VOP), the image encoder further comprises: second-accuracy time information generation means for generating second-accuracy time information which indicates time within accuracy of a second based on encoding start second-accuracy absolute time (e.g., processing steps S3 to S7 in the program shown in Figure 32, processing steps S43 to S47 in the program shown in Figure 37, etc.); detailed time information generation means for generating detailed time information which indicates a time period between the second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP included in the object group and the display time within accuracy finer than accuracy of a second (e.g., processing step S8 in the program shown in Figure 32, processing step S48 in the program shown in Figure 37, etc.); and addition means for adding the second-accuracy time information and detailed time information to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates display time of the I-VOP, P-VOP, or B-VOP (e.g., VLC unit 36 shown in Figure 11 or 12, etc.).

The image decoder as set forth in claim 11 is

characterized in that it is an image decoder for decoding an image, the image decoder comprising: receiving means for receiving a coded bit stream obtained by partitioning one or more layers of each of objects constituting the image into a plurality of groups and also by encoding the groups (e.g., buffer 101 shown in Figure 17 or 18, etc.); and decoding means for decoding the coded bit stream (e.g., IVLC unit 102 shown in Figure 17 or 18, etc.).

The image decoder as set forth in claim 13 is characterized in that a group of objects includes encoding start second-accuracy absolute time which is absolute time on an image sequence that started the encoding and also which is time within accuracy of a second and in that the image decoder further comprises display time computation means for computing the display time of an object included in the group of objects on the basis of the encoding start second-accuracy absolute time (e.g., processing steps S22 to S27 in the program shown in Figure 34, processing steps S52 to S57 in the program shown in Figure 38, etc.).

The image encoder as set forth in claim 26 is characterized in that, in an image encoder for encoding an image for each of objects constituting the image and outputting the resultant coded bit stream, when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded

by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-coding is a bidirectionally predictive-VOP (B-VOP), the image encoder comprises: second-accuracy time information generation means for generating second-accuracy time information which indicates time within accuracy of a second (e.g., processing steps S3 to S7 in the program shown in Figure 32, processing steps S43 to S47 in the program shown in Figure 37, etc.); detailed time information generation means for generating detailed time information which indicates a time period between the second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second (e.g., processing step S8 in the program shown in Figure 32, processing step S48 in the program shown in Figure 37, etc.); and addition means for adding the second-accuracy time information and detailed time information to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates the display time of the I-VOP, P-VOP, and B-VOP (e.g., VLC unit 36 shown in Figure 11 or 12, etc.).

The image decoder as set forth in claim 32 is characterized in that, in an image decoder for decoding a coded bit stream obtained by encoding an image for each of

objects constituting the image, when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive coding is a bidirectionally predictive-VOP (B-VOP) and when both second-accuracy time information indicating time within accuracy of a second and detailed time information indicating a time period between the second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second have been added to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates the display time, the image decoder comprises: display time computation means for computing the display time of the I-VOP, P-VOP, or B-VOP on the basis of the second-accuracy time information and detailed time information (e.g., processing steps S22 to S27 in the program shown in Figure 34, processing steps S52 to S57 in the program shown in Figure 38, etc.); and decoding means for decoding the I-VOP, P-VOP, or B-VOP in accordance with the corresponding display time (e.g., IVLC unit 102 shown in Figure 17 or 18, etc.).

Note that, of course, this description does not mean

that each means is limited to the aforementioned.

Figure 3 shows the constitution example of an embodiment of an encoder to which the present invention is applied.

Image (dynamic image) data to be encoded is input to a video object (VO) constitution section 1. In the VO constitution section 1, the image is constituted for each object by a sequence of VOs. The sequence of VOs are output to VOP constitution sections 21 to 2N. That is, in the VO constitution section 1, in the case where N video objects (VO#1 to VO#N) are produced, the VO#1 to VO#N are output to the VOP constitution sections 21 to 2N, respectively.

More specifically, for example, when image data to be encoded is constituted by a sequence of independent background F1 and foreground F2, the VO constitution section 1 outputs the foreground F2, for example, to the VOP constitution section 21 as VO#1 and also outputs the background F1 to the VOP constitution section 22 as VO#2.

Note that, in the case where image data to be encoded is, for example, an image previously synthesized by background F1 and foreground F2, the VO constitution section 1 partitions the image into the background F1 and foreground F2 in accordance with a predetermined algorithm. The background F1 and foreground F2 are output to corresponding VOP constitution sections 2n (where n = 1, 2, ..., and N).

The VOP constitution sections 2n produce VO planes (VOPs) from the outputs of the VO constitution section 1. That is, for example, an object is extracted from each frame. For example, the minimum rectangle surrounding the object (hereinafter referred to as the minimum rectangle as needed) is taken to be the VOP. Note that, at this time, the VOP constitution sections 2n produce the VOP so that the number of horizontal pixels and the number of vertical pixels are a multiple of 16. If the VO constitution sections 2n produce VOPs, the VOPs are output to VOP encoding sections 3n, respectively.

Furthermore, the VOP constitution sections 2n detect size data (VOP size) indicating the size of a VOP (e.g., horizontal and vertical lengths) and offset data (VOP offset) indicating the position of the VOP in a frame (e.g., coordinates as the left uppermost of a frame is the origin). The size data and offset data are also supplied to the VOP encoding sections 3n.

The VOP encoding sections 3n encode the outputs of the VOP constitution sections 2n, for example, by a method based on MPEG standard or H.263 standard. The resulting bit streams are output to a multiplexing section 4 which multiplexes the bit streams obtained from the VOP encoding sections 31 to 3N. The resulting multiplexed data is transmitted through a ground wave or through a transmission

path 5 such as a satellite line, a CATV network, etc. Alternatively, the multiplexed data is recorded on storage media 6 such as a magnetic disk, a magneto-optical disk, an optical disk, magnetic tape, etc.

Here, a description will be made of the video object (VO) and the video object plane (VOP).

In the case of a synthesized image, each of the images constituting the synthesized image is referred to as the VO, while the VOP means a VO at a certain time. That is, for example, in the case of a synthesized image F3 constituted by images F1 and F2, when the image F1 and F2 are arranged in a time series manner, they are VOs. The image F1 or F2 at a certain time is a VOP. Therefore, it may be said that the VO is a set of the VOPs of the same object at different times.

For instance, if it is assumed that image F1 is background and also image F2 is foreground, synthesized image F3 will be obtained by synthesizing the images F1 and F2 with a key signal for extracting the image F2. The VOP of the image F2 in this case is assumed to include the key signal in addition to image data (luminance signal and color difference signal) constituting the image F2.

An image frame does not vary in both size and position, but there are cases where the size or position of a VO changes. That is, even in the case a VOP constitutes the

same VO, there are cases where the size or position varies with time.

Specifically, Figure 4 illustrates a synthesized image constituted by image F1 (background) and image F2 (foreground).

For example, assume that the image F1 is an image obtained by photographing a certain natural scene and that the entire image is a single VO (e.g., VO#0). Also assume that the image F2 is an image obtained by photographing a person who is walking and that the minimum rectangular surrounding the person is a single VO (e.g., VO#1).

In this case, since the VO#0 is the image of a scene, basically both the position and the size do not change as in a normal image frame. On the other hand, since the VO#1 is the image of a person, the position or the size will change if the person moves right and left or moves toward this side or depth side in Figure 4. Therefore, although Figure 4 shows VO#0 and VO#1 at the same time, there are cases where the position or size of the VO varies with time.

Hence, the output bit stream of the VOP encoding sections 3n of Figure 3 includes information on the position (coordinates) and size of a VOP on a predetermined absolute coordinate system in addition to data indicating a coded VOP. Note in Figure 4 that a vector indicating the position of the VOP of VO#0 (image F1) at a certain time is represented

by OST0 and also a vector indicating the position of the VOP of VO#1 (image F2) at the certain time is represented by OST1.

Next, Figure 5 shows the constitution example of the VOP encoding sections 3n of Figure 3 which realize scalability. That is, the MPEG standard introduces a scalable encoding method which realizes scalability coping with different image sizes and frame rates. The VOP encoding sections 3n shown in Figure 5 are constructed so that such scalability can be realized.

The VOP (image data), the size data (VOP size), and offset data (VOP offset) from the VOP constitution sections 2n are all supplied to an image layering section 21.

The image layering section 21 generates one or more layers of image data from the VOP (layering of the VOP is performed). That is, for example, in the case of performing encoding of spatial scalability, the image data input to the image layering section 21, as it is, is output as an enhancement layer of image data. At the same time, the number of pixels constituting the image data is reduced (resolution is reduced) by thinning out the pixels, and the image data reduced in number of pixels is output as a base layer of image data.

Note that an input VOP can be employed as a base layer of data and also the VOP increased in pixel number

(resolution) by some other methods can be employed as an enhancement layer of data.

In addition, although the number of layers can be made 1, this case cannot realize scalability. In this case, the VOP encoding sections 3n are constituted, for example, by a base layer encoding section 25 alone.

Furthermore, the number of layers can be made 3 or more. But in this embodiment, the case of two layers will be described for simplicity.

For example, in the case of performing encoding of temporal scalability, the image layering section 21 outputs image data, for example, alternately base layer data or enhancement layer data in correspondence to time. That is, for example, when it is assumed that the VOPs constituting a certain VO are input in order of VOP0, VOP1, VOP2, VOP3, ..., the image layering section 21 outputs VOP0, VOP2, VOP4, VOP6, ... as base layer data and VOP1, VOP3, VOP5, VOP7, ... as enhancement layer data. Note that, in the case of temporal scalability, the VOPs thus thinned out are merely output as base layer data and enhancement layer data and the enlargement or reduction of image data (resolution conversion) is not performed (But it is possible to perform the enlargement or reduction).

Also, for example, in the case of performing the encoding of signal-to-noise ratio (SNR) scalability, the

image data input to the image layering section 21, as it is, is output as enhancement layer data or base layer data. That is, in this case, the base layer data and the enhancement layer data are consistent with each other.

Here, for the spatial scalability in the case of performing an encoding operation for each VOP, there are, for example, the following three kinds.

That is, for example, if it is now assumed that a synthesized image consisting of images F1 and F2 such as the one shown in Figure 4 is input as a VOP, in the first spatial scalability the input entire VOP (Figure 6(A)) is taken to be an enhancement layer, as shown in Figure 6, and the entire VOP reduced (Figure 6(B)) is taken to be a base layer.

Also, in the second spatial scalability, as shown in Figure 7, an object constituting part of an input VOP (Figure 7(A) (which corresponds to image F2)) is extracted. The extracted object is taken to be an enhancement layer, while the reduced entire VOP (Figure 7(B)) is taken to be a base layer. (Such extraction is performed, for example, in the same manner as the case of the VOP constitution sections 2n. Therefore, the extracted object is also a single VOP.)

Furthermore, in the third scalability, as shown in Figures 8 and 9, objects (VOP) constituting an input VOP are extracted, and an enhancement layer and a base layer are

generated for each object. Note that Figure 8 shows an enhancement layer and a base layer generated from the background (image F1) constituting the VOP shown in Figure 4, while Figure 9 shows an enhancement layer and a base layer generated from the foreground (image F2) constituting the VOP shown in Figure 4.

It has been predetermined which of the aforementioned scalabilities is employed. The image layering section 21 performs layering of a VOP so that encoding can be performed according to a predetermined scalability.

Furthermore, the image layering section 21 computes (or determines) the size data and offset data of generated base and enhancement layers from the size data and offset data of an input VOP (hereinafter respectively referred to as initial size data and initial offset data as needed). The offset data indicates the position of a base or enhancement layer in a predetermined absolute coordinate system of the VOP, while the size data indicates the size of the base or enhancement layer.

Here, a method of determining the offset data (position information) and size data of VOPs in base and enhancement layers will be described, for example, in the case where the above-mentioned second scalability (Figure 7) is performed.

In this case, for example, the offset data of a base layer, FPOS_B, as shown in Figure 10(A), is determined so

that, when the image data in the base layer is enlarged (upsampled) based on the difference between the resolution of the base layer and the resolution of the enhancement layer, i.e., when the image in the base layer is enlarged with a magnification ratio such that the size is consistent with that of the image in the enhancement layer (a reciprocal of the demagnification ratio as the image in the base layer is generated by reducing the image in the enhancement layer) (hereinafter referred to as magnification FR as needed), the offset data of the enlarged image in the absolute coordinate system is consistent with the initial offset data. The size data of the base layer, FSZ_B, is likewise determined so that the size data of an enlarged image, obtained when the image in the base layer is enlarged with magnification FR, is consistent with the initial size data. That is, the offset data FPOS_B is determined so that it is FR times itself or consistent with the initial offset data. Also, the size data FSZ_B is determined in the same manner.

On the other hand, for the offset data FPOS_E of an enhancement layer, the coordinates of the left upper corner of the minimum rectangle (VOP) surrounding an object extracted from an input VOP, for example, are computed based on the initial offset data, as shown in Figure 10(B), and this value is determined as offset data FPOS_E. Also, the

size data FPOS_E of the enhancement layer is determined to the horizontal and vertical lengths, for example, of the minimum rectangle surrounding an object extracted from an input VOP.

Therefore, in this case, the offset data FPOS_B and size data FPOS_B of the base layer are first transformed according to magnification FR. (The offset data FPOS_B and size data FPOS_B after transformation are referred to as transformed offset data FPOS_B and transformed size data FPOS_B, respectively.) Then, at a position corresponding to the transformed offset data FPOS_B in the absolute coordinate system, consider an image frame of the size corresponding to the transformed size data FSZ_B. If an enlarged image obtained by enlarging the image data in the base layer by FR times is arranged at the aforementioned corresponding position (Figure 10(A)) and also if the image in the enhancement layer is likewise arranged in the absolute coordinate system in accordance with the offset data FPOS_E and size data FPOS_E of the enhancement layer (Figure 10(B)), the pixels constituting the enlarged image and the pixels constituting the image in the enhancement layer will be arranged so that mutually corresponding pixels are located at the same position. That is, for example, in Figure 10, the person in the enhancement layer and the person in the enlarged image will be arranged at the same

position.

Even in the case of the first scalability and the third scalability, the offset data FPOS_B, offset data FPOS_E, size data FSZ_B, and size data FSZ_E are likewise determined so that mutually corresponding pixels constituting an enlarged image in a base layer and an image in an enhancement layer are located at the same position in the absolute coordinate system.

Returning to Figure 5, the image data, offset data FPOS_E, and size data FSZ_E in the enhancement layer, generated in the image layering section 21, are delayed by a delay circuit 22 by the processing period of a base layer encoding section 25 to be described later and are supplied to an enhancement layer encoding section 23. Also, the image data, offset data FPOS_B, and size data FSZ_B in the base layer are supplied to the base layer encoding section 25. In addition, magnification FR is supplied to the enhancement layer encoding section 23 and resolution transforming section 24 through the delay circuit 22.

In the base layer encoding section 25, the image data in the base layer is encoded. The resultant coded data (bit stream) includes the offset data FPOS_B and size data FSZ_B and is supplied to a multiplexing section 26.

Also, the base layer encoding section 25 decodes the coded data locally and outputs the locally decoded image

data in the base layer to the resolution transforming section 24. In the resolution transforming section 24, the image data in the base layer from the base layer encoding section 25 is returned to the original size by enlarging (or reducing) the image data in accordance with magnification FR. The resultant enlarged image is output to the enhancement layer encoding section 23.

On the other hand, in the enhancement layer encoding section 23, the image data in the enhancement layer is encoded. The resultant coded data (bit stream) includes the offset data FPOS_E and size data FSZ_E and is supplied to the multiplexing section 26. Note that in the enhancement layer encoding section 23, the encoding of the enhancement layer image data is performed by employing as a reference image the enlarged image supplied from the resolution transforming section 24.

The multiplexing section 26 multiplexes the outputs of the enhancement layer encoding section 23 and base layer encoding section 25 and outputs the multiplexed bit stream.

Note that the size data FSZ_B, offset data FPOS_B, motion vector (MV), flag COD, etc. of the base layer are supplied from the base layer encoding section 25 to the enhancement layer encoding section 23 and that the enhancement layer encoding section 23 is constructed so that it performs processing, making reference to the supplied

data as needed. The details will be described later.

Next, Figure 11 shows the detailed constitution example of the base layer encoding section 25 of Figure 5. In Figure 11, the same reference numerals are applied to parts corresponding to Figure 1. That is, basically the base layer encoding section 25 is constituted as in the encoder of Figure 1.

The image data from the image layering section 21 (Figure 5), i.e., the VOP in the base layer, as with Figure 1, is supplied to a frame memory 31 and stored. In a motion vector detector 32, the motion vector is detected at a macroblock unit.

But the size data FSZ_B and offset data FPOS_B of the VOP of a base layer are supplied to the motion vector detector 32 of the base layer encoding section 25, which in turn detects the motion vector of a macroblock, based on the supplied size data FSZ_B and offset data FPOS_B.

That is, as described above, the size and position of a VOP vary with time (frame). Therefore, in detecting the motion vector, there is a need to set a reference coordinate system for the detection and detect motion in the coordinate system. Hence, in the motion vector detector 32 here, the above-mentioned absolute coordinate system is employed as a reference coordinate system, and a VOP to be encoded and a reference VOP are arranged in the absolute coordinate system

in accordance with the size data FSZ_B and offset data FPOS_B, whereby the motion vector is detected.

Note that the detected motion vector (MV), along with the predictive mode, is supplied to a VLC unit 36 and a motion compensator 42 and is also supplied to the enhancement layer encoding section 23 (Figure 5).

Even in the case of performing motion compensation, there is also a need to detect motion in a reference coordinate system, as described above. Therefore, size data FSZ_B and offset data FPOS_B are supplied to the motion compensator 42.

A VOP whose motion vector was detected is quantized as in the case of Figure 1, and the quantized coefficients are supplied to the VLC unit 36. Also, as in the case of Figure 1, the size data FSZ_B and offset data FPOS_B from the image layering section 21 are supplied to the VLC unit 36 in addition to the quantized coefficients, quantization step, motion vector, and predictive mode. In the VLC unit 36, the supplied data is encoded by variable word length coding.

In addition to the above-mentioned encoding, the VOP whose motion vector was detected is locally decoded as in the case of Figure 1 and stored in frame memory 41. This decoded image is employed as a reference image, as previously described, and furthermore, it is output to the resolution transforming section 24 (Figure 5).

Note that, unlike the MPEG-1 standard and the MPEG-2 standard, in the MPEG-4 standard a B-picture (B-VOP) is also employed as a reference image. For this reason, a B-picture is also decoded locally and stored in the frame memory 41. (However, a B-picture is presently employed only in an enhancement layer as a reference image.)

On the other hand, as described in Figure 1, the VLC unit 36 determines whether the macroblock in an I-picture, a P-picture, or a B-picture (I-VOP, P-VOP, or B-VOP) is made a skip macroblock. The VLC unit 36 sets flags COD and MODB indicating the determination result. The flags COD and MODB are also encoded by variable word length coding and are transmitted. Furthermore, the flag COD is supplied to the enhancement layer encoding section 23.

Next, Figure 12 shows the constitution example of the enhancement layer encoding section 23 of Figure 5. In Figure 12, the same reference numerals are applied to parts corresponding to Figure 11 or 1. That is, basically the enhancement layer encoding section 23 is constituted as in the base layer encoding section 25 of Figure 11 or the encoder of Figure 1 except that frame memory 52 is newly provided.

The image data from the image layering section 21 (Figure 5), i.e., the VOP of the enhancement layer, as in the case of Figure 1, is supplied to the frame memory 31 and

stored. In the motion vector detector 32, the motion vector is detected at a macroblock unit. Even in this case, as in the case of Figure 11, the size data FSZ_E and offset data FPOS_E are supplied to the motion vector detector 32 in addition to the VOP of the enhancement layer, etc. In the motion vector detector 32, as in the above-mentioned case, the arranged position of the VOP of the enhancement layer in the absolute coordinate system is recognized based on the size data FSZ_E and offset data FPOS_E, and the motion vector of the macroblock is detected.

Here, in the motion vector detectors 32 of the enhancement layer encoding section 23 and base layer encoding section 25, VOPs are processed according to a predetermined sequence, as described in Figure 1. For example, the sequence is set as follows.

That is, in the case of spatial scalability, as shown in Figure 13(A) or 13(B), the VOPs in an enhancement layer or a base layer are processed, for example, in order of P, B, B, B, ... or I, P, P, P, ...

And in this case, the first P-picture (P-VOP) in the enhancement layer is encoded, for example, by employing as a reference image the VOP of the base layer present at the same time as the P-picture (here, I-picture (I-VOP)). Also, the second B-picture (B-VOP) in the enhancement layer is encoded, for example, by employing as reference images the

picture in the enhancement layer immediately before that and also the VOP in the base layer present at the same time as the B-picture. That is, in this example, the B-picture in the enhancement layer, as with the P-picture in base layer, is employed as a reference image in encoding another VOP.

For the base layer, encoding is performed, for example, as in the case of the MPEG-1 standard, MPEG-2 standard, or H.263 standard.

The SNR scalability is processed in the same manner as the above-mentioned spatial scalability, because it is the same as the spatial scalability when the magnification FR in the spatial scalability is 1.

In the case of the temporal scalability, i.e., for example, in the case where a VO is constituted by VOP0, VOP1, VOP2, VOP3, ..., and also VOP1, VOP3, VOP5, VOP7, ... are taken to be in an enhancement layer (Figure 14(A)) and VOP0, VOP2, VOP4, VOP6, ... to be in a base layer (Figure 14(B)), as described above, the VOPs in the enhancement and base layers are respectively processed in order of B, B, B, ... and in order of I, P, P, P, ..., as shown in Figure 14.

And in this case, the first VOP1 (B-picture) in the enhancement layer is encoded, for example, by employing the VOP0 (I-picture) and VOP2 (P-picture) in the base layer as reference images. The second VOP3 (B-picture) in the enhancement layer is encoded, for example, by employing as

reference images the first coded VOP1 (B-picture) in the enhancement layer immediately before that and the VOP4 (P-picture) in the base layer present at the time (frame) next to the VOP3. The third VOP5 (B-picture) in the enhancement layer, as with the encoding of the VOP3, is encoded, for example, by employing as reference images the second coded VOP3 (B-picture) in the enhancement layer immediately before that and the VOP6 (P-picture) in the base layer which is an image present at the time (frame) next to the VOP5.

As described above, for VOPs in one layer (here, enhancement layer), VOPs in another layer (scalable layer) (here, base layer) can be employed as reference images for encoding a P-picture and a B-picture. In the case where a VOP in one layer is thus encoded by employing a VOP in another layer as a reference image, i.e., like this embodiment, in the case where a VOP in the base layer is employed as a reference image in encoding a VOP in the enhancement layer predictively, the motion vector detector 32 of the enhancement layer encoding section 23 (Figure 12) is constructed so as to set and output flag ref_layer_id indicating that a VOP in the base layer is employed to encode a VOP in the enhancement layer predictively. (In the case of 3 or more layers, the flag ref_layer_id represents a layer to which a VOP, employed as a reference image, belongs.)

Furthermore, the motion vector detector 32 of the enhancement layer encoding section 23 is constructed so as to set and output flag ref_select_code (reference image information) in accordance with the flag ref_layer_id for a VOP. The flag ref_select_code (reference image information) indicates which layer and which VOP in the layer are employed as a reference image in performing forward predictive coding or backward predictive coding.

More specifically, for example, in the case where a P-picture in an enhancement layer is encoded by employing as a reference image a VOP which belongs to the same layer as a picture decoded (locally decoded) immediately before the P-picture, the flag ref_select_code is set to 00. Also, in the case where the P-picture is encoded by employing as a reference image a VOP which belongs to a layer (here, base layer (reference layer)) different from a picture displayed immediately before the P-picture, the flag ref_select_code is set to 01. In addition, in the case where the P-picture is encoded by employing as a reference image a VOP which belongs to a layer different from a picture to be displayed immediately after the P-picture, the flag ref_select_code is set to 10. Furthermore, in the case where the P-picture is encoded by employing as a reference image a VOP which belongs to a different layer present at the same time as the P-picture, the flag ref_select_code is set to 11.

On the other hand, for example, in the case where a B-picture in an enhancement layer is encoded by employing as a reference image for forward prediction a VOP which belongs to a different layer present at the same time as the B-picture and also by employing as a reference image for backward prediction a VOP which belongs to the same layer as a picture decoded immediately before the B-picture, the flag ref_select_code is set to 00. Also, in the case where the B-picture in the enhancement layer is encoded by employing as a reference image for forward prediction a VOP which belongs to the same layer as the B-picture and also by employing as a reference image for backward prediction a VOP which belongs to a layer different from a picture displayed immediately before the B-picture, the flag ref_select_code is set to 01. In addition, in the case where the B-picture in the enhancement layer is encoded by employing as a reference image for forward prediction a VOP which belongs to the same layer as a picture decoded immediately before the B-picture and also by employing as a reference image for backward prediction a VOP which belongs to a layer different from a picture to be displayed immediately after the B-picture, the flag ref_select_code is set to 10. Furthermore, in the case where the B-picture in the enhancement layer is encoded by employing as a reference image for forward prediction a VOP which belongs to a layer different from a

picture displayed immediately before the B-picture and also by employing as a reference image for backward prediction a VOP which belongs to a layer different from a picture to be displayed immediately after the B-picture, the flag ref_select_code is set to 11.

Here, the predictive coding shown in Figures 13 and 14 is merely a single example. Therefore, it is possible within the above-mentioned range to set freely which layer and which VOP in the layer are employed as a reference image for forward predictive coding, backward predictive coding, or bidirectionally predictive coding.

In the above-mentioned case, while the terms spatial scalability, temporal scalability, and SNR scalability have been employed for the convenience of explanation, it becomes difficult to discriminate the spatial scalability, temporal scalability, and SNR scalability from each other in the case where a reference image for predictive coding is set by the flag ref_select_code. That is, conversely speaking, the employment of the flag ref_select_code renders the above-mentioned discrimination between scalabilites unnecessary.

Here, if the above-mentioned scalability and flag ref_select_code are correlated with each other, the correlation will be, for example, as follows. That is, with respect to a P-picture, since the case of the flag ref_select_code being 11 is a case where a VOP at the same

time in the layer indicated by the flag ref_layer_id is employed as a reference image (for forward prediction), this case corresponds to spatial scalability or SNR scalability. And the cases other than the case of the flag ref_select_code being 11 correspond to temporal scalability.

Also, with respect to a B-picture, the case of the flag ref_select_code being 00 is also the case where a VOP at the same time in the layer indicated by the flag ref_layer_id is employed as a reference image for forward prediction, so this case corresponds to spatial scalability or SNR scalability. And the cases other than the case of the flag ref_select_code being 00 correspond to temporal scalability.

Note that, in the case where in order to encode a VOP in an enhancement layer predictively, a VOP at the same time in a layer (here, base layer) different from the enhancement layer is employed as a reference image, there is no motion therebetween, so the motion vector is always made 0 ((0,0)).

Returning to Figure 12, the aforementioned flag ref_layer_id and flag ref_select_code are set to the motion vector detector 32 of the enhancement layer encoding section 23 and supplied to the motion compensator 42 and VLC unit 36.

Also, the motion vector detector 32 detects a motion vector by not making reference only to the frame memory 31 in accordance with the flag ref_layer_id and flag ref_select_code but also making reference to the frame

memory 52 as needed.

Here, a locally decoded enlarged image in the base layer is supplied from the resolution transforming section 24 (Figure 5) to the frame memory 52. That is, in the resolution transforming section 24, the locally decoded VOP in the base layer is enlarged, for example, by a so-called interpolation filter, etc. With this, an enlarged image which is FR times the size of the VOP, i.e., an enlarged image of the same size as the VOP in the enhancement layer corresponding to the VOP in the base layer is generated. The generated image is supplied to the enhancement layer encoding section 23. The frame memory 52 stores the enlarged image supplied from the resolution transforming section 24 in this manner.

Therefore, when magnification FR is 1, the resolution transforming section 24 does not process the locally decoded VOP supplied from the base layer encoding section 25. The locally decoded VOP from the base layer encoding section 25, as it is, is supplied to the enhancement layer encoding section 23.

The size data FSZ_B and offset data FPOS_B are supplied from the base layer encoding section 25 to the motion vector detector 32, and the magnification FR from the delay circuit 22 (Figure 5) is also supplied to the motion vector detector 32. In the case where the enlarged image stored in the

frame memory 52 is employed as a reference image, i.e., in the case where in order to encode a VOP in an enhancement layer predictively, a VOP in a base layer at the same time as the enhancement layer VOP is employed as a reference image (in this case, the flag ref_select_code is made 11 for a P-picture and 00 for a B-picture), the motion vector detector 32 multiplies the size data FSZ_B and offset data FPOS_B corresponding to the enlarged image by magnification FR. And based on the multiplication result, the motion vector detector 32 recognizes the position of the enlarged image in the absolute coordinate system, thereby detecting the motion vector.

Note that the motion vector and predictive mode in a base layer are supplied to the motion vector detector 32. This data is used in the following case. That is, in the case where the flag ref_select_code for a B-picture in an enhancement layer is 00, when magnification FR is 1, i.e., in the case of SNR scalability (in this case, since a VOP in an enhancement layer is employed in encoding the enhancement layer predictively, the SNR scalability used herein differs in this respect from that prescribed in the MPEG-2 standard), images in the enhancement layer and base layer are the same. Therefore, when the predictive coding of a B-picture in an enhancement layer is performed, the motion vector detector 32 can employ the motion vector and predictive mode in a

base layer present at the same time as the B-picture, as they are. Hence, in this case the motion vector detector 32 does not process the B-picture of the enhancement layer, but it adopts the motion vector and predictive mode of the base layer as they are.

In this case, in the enhancement layer encoding section 23, a motion vector and a predictive mode are not output from the motion vector detector 32 to the VLC unit 36. (Therefore, they are not transmitted.) This is because a receiver side can recognize the motion vector and predictive mode of an enhancement layer from the result of the decoding of a base layer.

As previously described, the motion vector detector 32 detects a motion vector by employing both a VOP in an enhancement layer and an enlarged image as reference images. Furthermore, as shown in Figure 1, the motion vector detector 32 sets a predictive mode which makes a prediction error (or dispersion) minimum. Also, the motion vector detector 32 sets and outputs necessary information, such as flag ref_select_code, flag ref_layer_id, etc.

In Figure 12, flag COD indicates whether a macroblock constituting an I-picture or a P-picture in a base layer is a skip macroblock, and the flag COD is supplied from the base layer encoding section 25 to the motion vector detector 32, VLC unit 36, and motion compensator 42.

The macroblock whose motion vector was detected is encoded in the same manner as the above-mentioned case. As a result of the encoding, variable-length codes are output from the VLC unit 36.

The VLC unit 36 of the enhancement layer encoding section 23, as in the case of the base layer encoding section 25, is constructed so as to set and output flags COD and MODB. Here, the flag COD, as described above, indicates whether a macroblock in an I- or P-picture is a skip macroblock, while the flag MODB indicates whether a macroblock in a B-picture is a skip macroblock.

The quantized coefficients, quantization step, motion vector, predictive mode, magnification FR, flag ref_select_code, flag ref_layer_id, size data FSZ_E, and offset data FPOS_E are also supplied to the VLC unit 36. In the VLC unit 36, these are encoded by variable word length coding and are output.

On the other hand, after a macroblock whose motion vector was detected has been encoded, it is also decoded locally as described above and is stored in the frame memory 41. And in the motion compensator 42, as in the case of the motion vector detector 32, motion compensation is performed by employing as reference images both a locally decoded VOP in an enhancement layer, stored in the frame memory 41, and a locally decoded and enlarged VOP in a base layer, stored

in the frame memory 52. With this compensation, a predicted image is generated.

That is, in addition to the motion vector and predictive mode, the flag ref_select_code, flag ref_layer_id, magnification FR, size data FSZ_B, size data FSZ_E, offset data FPOS_B, and offset data FPOS_E are supplied to the motion compensator 42. The motion compensator 42 recognizes a reference image to be motion-compensated, based on the flags ref_select_code and ref_layer_id. Furthermore, in the case where a locally decoded VOP in an enhancement layer or an enlarged image is employed as a reference image, the motion compensator 42 recognizes the position and size of the reference image in the absolute coordinate system, based on the size data FSZ_E and offset data FPOS_E, or the size data FSZ_B and offset data FPOS_B. The motion compensator 42 generates a predicted image by employing magnification FR, as needed.

Next, Figure 15 shows the constitution example of an embodiment of a decoder which decodes the bit stream output from the encoder of Figure 3.

This decoder receives the bit stream supplied by the encoder of Figure 3 through the transmission path 5 or storage medium 6. That is, the bit stream, output from the encoder of Figure 3 and transmitted through the transmission path 5, is received by a receiver (not shown).

Alternatively, the bit stream recorded on the storage medium 6 is regenerated by a regenerator (not shown). The received or regenerated bit stream is supplied to an inverse multiplexing section 71.

The inverse multiplexing section 71 receives the bit stream (video stream (VS) described later) input thereto. Furthermore, in the inverse multiplexing section 71, the input bit stream is separated into bit streams VO#1, VO#2, ... The bit streams are supplied to corresponding VOP decoding sections 72n, respectively. In the VOP decoding sections 72n, the VOP (image data) constituting a VO, the size data (VOP size), and the offset data (VOP offset) are decoded from the bit stream supplied from the inverse multiplexing section 71. The decoded data is supplied to an image reconstituting section 73.

The image reconstituting section 73 reconstitutes the original image, based on the respective outputs of the VOP decoding sections 72_1 to 72_N. This reconstituted image is supplied, for example, to a monitor 74 and displayed.

Next, Figure 16 shows the constitution example of the VOP decoding section 72_N of Figure 15 which realizes scalability.

The bit stream supplied from the inverse multiplexing section 71 (Figure 15) is input to an inverse multiplexing section 91, in which the input bit stream is separated into

a bit stream of a VOP in an enhancement layer and a bit stream of a VOP in a base layer. The bit stream of a VOP in an enhancement layer is delayed by a delay circuit 92 by the processing period in the base layer decoding section 95 and supplied to the enhancement layer decoding section 93. Also, the bit stream of a VOP in a base layer is supplied to the base layer decoding section 95.

In the base layer decoding section 95, the bit stream in a base layer is decoded, and the resulting decoded image in a base layer is supplied to a resolution transforming section 94. Also, in the base layer decoding section 95, information necessary for decoding a VOP in an enhancement layer, obtained by decoding the bit stream of a base layer, is supplied to the enhancement layer decoding section 93. The necessary information includes size data FSZ_B, offset data FPOS_B, motion vector (MV), predictive mode, flag COD, etc.

In the enhancement layer decoding section 93, the bit stream in an enhancement layer supplied through the delay circuit 92 is decoded by making reference to the outputs of the base layer decoding section 95 and resolution transforming section 94 as needed. The resultant decoded image in an enhancement layer, size data FSZ_E, and offset data FPOS_E are output. Furthermore, in the enhancement layer decoding section 93, the magnification FR, obtained by

decoding the bit stream in an enhancement layer, is output to the resolution transforming section 94. In the resolution transforming section 94, as in the case of the resolution transforming section 24 in Figure 5, the decoded image in a base layer is transformed by employing the magnification FR supplied from the enhancement layer decoding section 93. An enlarged image obtained with this transformation is supplied to the enhancement layer decoding section 93. As described above, the enlarged image is employed in decoding the bit stream of an enhancement layer.

Next, Figure 17 shows the constitution example of the base layer decoding section 95 of Figure 16. In Figure 17, the same reference numerals are applied to parts corresponding to the case of the decoder in Figure 2. That is, basically the base layer decoding section 95 is constituted in the same manner as the decoder of Figure 2.

The bit stream of a base layer from the inverse multiplexing section 91 is supplied to a buffer 101 and stored temporarily. An IVLC unit 102 reads out the bit stream from the buffer 101 in correspondence to a block processing state of the following stage, as needed, and the bit stream is decoded by variable word length decoding and is separated into quantized coefficients, a motion vector, a predictive mode, a quantization step, size data FSZ_B, offset data FPOS_B, and flag COD. The quantized

coefficients and quantization step are supplied to an inverse quantizer 103. The motion vector and predictive mode are supplied to a motion compensator 107 and enhancement layer decoding section 93 (Figure 16). Also, the size data FSZ_B and offset data FPOS_B are supplied to the motion compensator 107, image reconstituting section 73 (Figure 15), and enhancement layer decoding section 93, while the flag COD is supplied to the enhancement layer decoding section 93.

The inverse quantizer 103, IDCT unit 104, arithmetic unit 105, frame memory 106, and motion compensator 107 perform similar processes corresponding to the inverse quantizer 38, IDCT unit 39, arithmetic unit 40, frame memory 41, and motion compensator 42 of the base layer encoding section 25 of Figure 11, respectively. With this, the VOP of a base layer is decoded. The decoded VOP is supplied to the image reconstituting section 73, enhancement layer decoding section 93, and resolution transforming section 94 (Figure 16).

Next, Figure 18 shows the constitution example of the enhancement layer decoding section 93 of Figure 16. In Figure 18, the same reference numerals are applied to parts corresponding to the case in Figure 2. That is, basically the enhancement layer decoding section 93 is constituted in the same manner as the decoder of Figure 2 except that frame

memory 112 is newly provided.

The bit stream of an enhancement layer from the inverse multiplexing section 91 is supplied to an IVLC 102 through a buffer 101. The IVLC unit 102 decodes the bit stream of an enhancement layer by variable word length decoding, thereby separating the bit stream into quantized coefficients, a motion vector, a predictive mode, a quantization step, size data FSZ_E, offset data FPOS_E, magnification FR, flag ref_layer_id, flag ref_select_code, flag COD, and flag MODB. The quantized coefficients and quantization step, as in the case of Figure 17, are supplied to an inverse quantizer 103. The motion vector and predictive mode are supplied to a motion compensator 107. Also, the size data FSZ_E and offset data FPOS_E are supplied to the motion compensator 107 and image reconstituting section 73 (Figure 15). The flag COD, flag MODB, flag ref_layer_id, and flag ref_select_code are supplied to the motion compensator 107. Furthermore, the magnification FR is supplied to the motion compensator 107 and resolution transforming section 94 (Figure 16).

Note that the motion vector, flag COD, size data FSZ_B, and offset data FPOS_B of a base layer are supplied from the base layer decoding section 95 (Figure 16) to the motion compensator 107 in addition to the above-mentioned data. Also, an enlarged image is supplied from the resolution

transforming section 94 to frame memory 112.

The inverse quantizer 103, IDCT unit 104, arithmetic unit 105, frame memory 106, motion compensator 107, and frame memory 112 perform similar processes corresponding to the inverse quantizer 38, IDCT unit 39, arithmetic unit 40, frame memory 41, motion compensator 42, and frame memory 52 of the enhancement layer encoding section 23 of Figure 12, respectively. With this, the VOP of an enhancement layer is decoded. The decoded VOP is supplied to the image reconstituting section 73.

Here, in the VOP decoding sections 72n having both the enhancement layer decoding section 93 and base layer decoding section 95 constituted as described above, both the decoded image, size data FSZ_E, and offset data FPOS_E in an enhancement layer (hereinafter referred to as enhancement layer data as needed) and the decoded image, size data FSZ_B, and offset data FPOS_B in a base layer (hereinafter referred to as base layer data as needed) are obtained. In the image reconstituting section 73, an image is reconstituted from the enhancement layer data or base layer data, for example, in the following manner.

That is, for instance, in the case where the first spatial scalability (Figure 6) is performed (i.e., in the case where the entire input VOP is made an enhancement layer and the entire VOP reduced is made a base layer), when both

the base layer data and the enhancement layer data are decoded, the image reconstituting section 73 arranges the decoded image (VOP) of the enhancement layer of the size corresponding to size data FSZ_E at the position indicated by offset data FPOS_E, based on enhancement layer data alone. Also, for example, when an error occurs in the bit stream of an enhancement layer, or when the monitor 74 processes only an image of low resolution and therefore only base layer data is decoded, the image reconstituting section 73 arranges the decoded image (VOP) of an enhancement layer of the size corresponding to size data FSZ_B at the position indicated by offset data FPOS_B, based on the base layer data alone.

Also, for instance, in the case where the second spatial scalability (Figure 7) is performed (i.e., in the case where part of an input VOP is made an enhancement layer and the entire VOP reduced is made a base layer), when both the base layer data and the enhancement layer data are decoded, the image reconstituting section 73 enlarges the decoded image of the base layer of the size corresponding to size data FSZ_B in accordance with magnification FR and generates the enlarged image. Furthermore, the image reconstituting section 73 enlarges offset data FPOS_B by FR times and arranges the enlarged image at the position corresponding to the resulting value. And the image

reconstituting section 73 arranges the decoded image of the enhancement layer of the size corresponding to size data FSZ_E at the position indicated by offset data FPOS_E.

In this case, the portion of the decoded image of an enhancement layer is displayed with higher resolution than the remaining portion.

Note that in the case where the decoded image of an enhancement layer is arranged, the decoded image and an enlarged image are synthesized with each other.

Also, although not shown in Figure 16 (Figure 15), magnification FR is supplied from the enhancement layer decoding section 93 (VOP decoding sections 72n) to the image reconstituting section 73 in addition to the above-mentioned data. The image reconstituting section 73 generates an enlarged image by employing the supplied magnification FR.

On the other hand, in the case where the second spatial scalability is performed, when base layer data alone is decoded, an image is reconstituted in the same manner as the above-mentioned case where the first spatial scalability is performed.

Furthermore, in the case where the third spatial scalability (Figures 8 and 9) is performed (i.e., in the case where each of the objects constituting an input VOP is made an enhancement layer and the VOP excluding the objects is made a base layer), an image is reconstituted in the same

manner as the above-mentioned case where the second spatial scalability is performed.

As described above, the offset data FPOS_B and offset data FPOS_E are constructed so that mutually corresponding pixels, constituting the enlarged image of a base layer and an image of an enhancement layer, are arranged at the same position in the absolute coordinate system. Therefore, by reconstituting an image in the aforementioned manner, an accurate image (with no positional offset) can be obtained.

Next, the syntax of the coded bit stream output by the encoder of Figure 3 will be described, for example, with the video verification model (version 6.0) of the MPEG-4 standard (hereinafter referred to as VM-6.0 as needed) as an example.

Figure 19 shows the syntax of a coded bit stream in VM-6.0.

The coded bit stream is constituted by video session classes (VSs). Each VS is constituted by one or more video object classes (VOs). Each VO is constituted by one or more video object layer classes (VOLs). (When an image is not layered, it is constituted by a single VOL. In the case where an image is layered, it is constituted by VOLs corresponding to the number of layers.) Each VOL is constituted by video object plane classes (VOP).

Note that VSs are a sequence of images and equivalent,

for example, to a single program or movie.

Figures 20 and 21 show the syntax of a VS and the syntax of a VO. The VO is a bit stream corresponding to an entire image or a sequence of objects constituting an image. Therefore, VSs are constituted by a set of such sequences. (Therefore, VSs are equivalent, for example, to a single program.)

Figure 22 shows the syntax of a VOL.

The VOL is a class for the above-mentioned scalability and is identified by a number indicated with video_object_layer_id. For example, the video_object_layer_id for a VOL in a base layer is made a 0, while the video_object_layer_id for a VOL in an enhancement layer is made a 1. Note that, as described above, the number of scalable layers is not limited to 2, but it may be an arbitrary number including 1, 3, or more.

Also, whether a VOL is an entire image or part of an image is identified by video_object_layer_shape. This video_object_layer_shape is a flag for indicating the shape of a VOL and is set as follows.

When the shape of a VOL is rectangular, the video_object_layer_shape is made, for example, 00. Also, when a VOL is in the shape of an area cut out by a hard key (a binary signal which takes either a 0 or a 1), the video_object_layer_shape is made, for example, 01.

Furthermore, when a VOL is in the shape of an area cut out by a soft key (a signal which can take a continuous value (gray-scale) in a range of 0 to 1) (when synthesized by a soft key), the video_object_layer_shape is made, for example, 10.

Here, when video_object_layer_shape is made 00, the shape of a VOP is rectangular and also the position and size of a VOL in the absolute coordinate system do not vary with time, i.e., are constant. In this case, the sizes (horizontal length and vertical length) are indicated by video_object_layer_width and video_object_layer_height. The video_object_layer_width and video_object_layer_height are both 10-bit fixed-length flags. In the case where video_object_layer_shape is 00, it is first transmitted only once. (This is because, in the case where video_object_layer_shape is 00, as described above, the size of a VOL in the absolute coordinate system is constant.)

Also, whether a VOL is a base layer or an enhancement layer is indicated by scalability which is a 1-bit flag. When a VOL is a base layer, the scalability is made, for example, a 1. In the case other than that, the scalability is made, for example, a 0.

Furthermore, in the case where a VOL employs an image in a VOL other than itself as a reference image, the VOL to which the reference image belongs is represented by

ref_layer_id, as described above. Note that the ref_layer_id is transmitted only when a VOL is an enhancement layer.

In Figure 22 the hor_sampling_factor_n and the hor_sampling_factor_m indicate a value corresponding to the horizontal length of a VOP in a base layer and a value corresponding to the horizontal length of a VOP in an enhancement layer, respectively. The horizontal length of an enhancement layer to a base layer (magnification of horizontal resolution) is given by the following equation: hor_sampling_factor_n / hor_sampling_factor_m.

In Figure 22 the ver_sampling_factor_n and the ver_sampling_factor_m indicate a value corresponding to the vertical length of a VOP in a base layer and a value corresponding to the vertical length of a VOP in an enhancement layer, respectively. The vertical length of an enhancement layer to a base layer (magnification of vertical resolution) is given by the following equation: ver_sampling_factor_n / ver_sampling_factor_m.

Next, Figure 23 shows the syntax of a VOP.

The sizes (horizontal length and vertical length) of a VOP are indicated, for example, by VOP_width and VOP_height having a 10-bit fixed-length. Also, the positions of a VOP in the absolute coordinate system are indicated, for example, by 10-bit fixed-length VOP_horizontal_spatial_mc_ref and

VOP_vertical_mc_ref. The VOP_width and VOP_height represent the horizontal length and vertical length of a VOP, respectively. These are equivalent to size data FSZ_B and size data FSZ_E described above. The VOP_horizontal_spatial_mc_ref and VOP_vertical_mc_ref represent the horizontal and vertical coordinates (x and y coordinates) of a VOP, respectively. These are equivalent to offset data FPOS_B and offset data FPOS_E described above.

The VOP_width, VOP_height, VOP_horizontal_mc_ref, and VOP_vertical_mc_ref are transmitted only when video_object_layer_shape is not 00. That is, when video_object_layer_shape is 00, as described above, the size and position of a VOP are both constant, so there is no need to transmit the VOP_width, VOP_height, VOP_horizontal_spatial_mc_ref, and VOP_vertical_mc_ref. In this case, on a receiver side a VOP is arranged so that the left upper corner is consistent, for example, with the origin of the absolute coordinate system. Also, the sizes are recognized from the video_object_layer_width and video_object_layer_height described in Figure 22.

In Figure 23 the ref_select_code, as described in Figure 19, represents an image which is employed as a reference image, and is prescribed by the syntax of a VOP.

Incidentally, in VM-6.0 the display time of each VOP (equivalent to a conventional frame) is determined by

modulo_time_base and VOP_time_increment (Figure 23) as follows:

That is, the modulo_time_base represents the encoder time on the local time base within accuracy of one second (1000 milliseconds). The modulo_time_base is represented as a marker transmitted in the VOP header and is constituted by a necessary number of 1's and a 0. The number of consecutive "1" constituting the modulo_time_base followed by a "0" is the cumulative period from the synchronization point (time within accuracy of a second) marked by the last encoded/decoded modulo_time_base. For example, when the modulo_time_base indicates a 0, the cumulative period from the synchronization point marked by the last encoded/decoded modulo_time_base is 0 second. Also, when the modulo_time_base indicates 10, the cumulative period from the synchronization point marked by the last encoded/decoded modulo_time_base is 1 second. Furthermore, when the modulo_time_base indicates 110, the cumulative period from the synchronization point marked by the last encoded/decoded modulo_time_base is 2 seconds. Thus, the number of 1's in the modulo_time_base is the number of seconds from the synchronization point marked by the last encoded/decoded modulo_time_base.

Note that, for the modulo_time_base, the VM-6.0 states that:

This value represents the local time base at the one second resolution unit (1000 milliseconds). It is represented as a marker transmitted in the VOP header. The number of consecutive "1" followed by a "0" indicates the number of seconds has elapsed since the synchronization point marked by the last encoded/decoded modulo_time_base.

The VOP_time_increment represents the encoder time on the local time base within accuracy of 1 ms. In VM-6.0, for I-VOPs and P-VOPs the VOP_time_increment is the time from the synchronization point marked by the last encoded/decoded modulo_time_base. For the B-VOPs the VOP_time_increment is the relative time from the last encoded/decoded I- or P-VOP.

Note that, for the VOP_time_increment, the VM-6.0 states that:

This value represents the local time base in the units of milliseconds. For I- and P-VOPs this value is the absolute VOP_time_increment from the synchronization point marked by the last modulo_time_base. For the B-VOPs this value is the relative VOP_time_increment from the last encoded/decoded I- or P-VOP.

And the VM-6.0 states that:

At the encoder, the following formula are used to determine the absolute and relative VOP_time_increments for I/P-VOPs and B-VOPs, respectively.

That is, VM-6.0 prescribes that at the encoder, the

display times for I/P-VOPs and B-VOPs are respectively encoded by the following formula:

$$tGTB(n) = n \times 1000ms + tEST$$

$$tAVTI = tETB(I/P) - tGTB(n)$$

$$tRVTI = tETB(B) - tETB(I/P) \quad \ldots \quad (1)$$

where tGTB(n) represents the time of the synchronization point (as described above, accuracy of a second) marked by the nth encoded modulo_time_base, tEST represents the encoder time at the start of the encoding of the VO (the absolute time at which the encoding of the VO was started), tAVTI represents the VOP_time_increment for the I or P-VOP, tETB(I/P) represents the encoder time at the start of the encoding of the I or P-VOP (the absolute time at which encoding of the VOP was started), tRVTI represents the VOP_time_increment for the B-VOP, and tETB(B) represents the encoder time at the start of the encoding of the B-VOP.

Note that, for the tGTB(n), tEST, tAVTI, tETB(I/P), tRVTI, and tETB(B) in Formula (1), the VM-6.0 states that: tGTB(n) is the encoder time base marked by the nth encoded modulo_time_base, tEST is the encoder time base start time, tAVTI is the absolute VOP_time_increment for the I or P-VOP, tETB(I/P) is the encoder time base at the start of the encoding of the I or P-VOP, tRVTI is the relative VOP_time_increment for the B-VOP, and tETB(B) is the encoder time base at the start of the encoding of the B-VOP.

Also, the VM-6.0 states that:

At the decoder, the following formula are used to determine the recovered time base of the I/P-VOPs and B-VOPs, respectively.

That is, VM-6.0 prescribes that at the decoder side, the display times for I/P-VOPs and B-VOPs are respectively decoded by the following formula:

$$tGTB(n) = n \times 1000ms + tDST$$

$$tDTB(I/P) = tAVTI + tGTB(n)$$

$$tDTB(B) = tRVTI + tDTB(I/P) \quad \ldots (2)$$

where tGTB(n) represents the time of the synchronization point marked by the nth decoded modulo_time_base, tDST represents the decoder time at the start of the decoding of the VO (the absolute time at which the decoding of the VO was started), tDTB(I/P) represents the decoder time at the start of the decoding of the I-VOP or P-VOP, tAVTI represents the VOP_time_increment for the I-VOP or P-VOP, tDTB(B) represents the decoder time at the start of the decoding of the B-VOP (the absolute time at which the decoding of the VOP was started), tRVTI represents the VOP_time_increment for the B-VOP.

Note that, for the tGTB(n), tDST, tDTB(I/P), tAVTI, tDTB(B), and tRVTI in Formula (2), the VM-6.0 states that:

tGTB(n) is the encoding time base marked by the nth decoded modulo_time_base, tDST is the decoding time base

start time, tDTB(I/P) is the decoding time base at the start of the decoding of the I or P-VOP, tAVTI is the decoding absolute VOP_time_increment for the I- or P-VOP, tDTB(B) is the decoding time base at the start of the decoding of the B-VOP, and tRVTI is the decoded relative VOP_time_increment for the B-VOP.

Figure 24 shows the relation between modulo_time_base and VOP_time_increment based on the above definition.

In the figure, a VO is constituted by a sequence of VOPs, such as I1 (I-VOP), B2 (B-VOP), B3, P4 (P-VOP), B5, P6, etc. Now, assuming the encoding/decoding start time (absolute time) of the VO is t0, the modulo_time_base will represent time (synchronization point), such as t0 + 1 sec, t0 + 2 sec, etc., because the elapsed time from the start time t0 is represented within accuracy of one second. In Figure 24, although the display order is I1, B2, B3, P4, B5, P6, etc., the encoding/decoding order is I1, P4, B2, B3, P6, etc.

In Figure 24 (as are Figures 28 to 31 and Figure 36 to be described later), the VOP_time_increment for each VOP is indicated by a numeral (in the units of milliseconds) enclosed within a square. The switch of synchronization points indicated by modulo_time_base is indicated by a mark of ▼. In Figure 24, therefore, the VOP_time_increments for the I1, B2, B3, P4, B5, and P6 are 350 ms, 400 ms, 800 ms,

550 ms, 400 ms, and 350 ms, and at P4 and P6, the synchronization point is switched.

Now, in Figure 24 the VOP_time_increment for the I1 is 350 ms. The encoding/decoding time of the I1, therefore, is the time after 350 ms from the synchronization point marked by the last encoded/decoded modulo_time_base. Note that, immediately after the start of the encoding/decoding of the I1, the start time (encoding/decoding start time) t0 becomes a synchronization point. The encoding/decoding time of the I1, therefore, will be the time t0 + 350 ms after 350 ms from the start time (encoding/decoding start time) t0.

And the encoding/decoding time of the B2 or B3 is the time of the VOP_time_increment which has elapsed since the last encoded/decoded I-VOP or P-VOP. In this case, since the encoding/decoding time of the last encoded/decoded I1 is t0 + 350 ms, the encoding/decoding time of the B2 or B3 is the time t0 + 750 ms or t0 + 1200 ms after 400 ms or 800 ms.

Next, for the P4, at the P4 the synchronization point indicated by modulo_time_base is switched. Therefore, the synchronization point is time t0 + 1 sec. As a result, the encoding/decoding time of the P4 is the time (t0 + 1) sec + 550 ms after 550 ms from the time t0 + 1 sec.

The encoding/decoding time of the B5 is the time of the VOP_time_increment which has elapsed since the last encoded/decoded I-VOP or P-VOP. In this case, since the

encoding/decoding time of the last encoded/decoded P4 is (t0 + 1) sec + 550 ms, the encoding/decoding time of the B5 is the time (t0 + 1) sec + 950 ms after 400 ms.

Next, for the P6, at the P6 the synchronization point indicated by modulo_time_base is switched. Therefore, the synchronization point is time t0 + 2 sec. As a result, the encoding/decoding time of the P6 is the time (t0 + 2) sec + 350 ms after 350 ms from the time t0 + 2 sec.

Note that in VM-6.0, the switch of the synchronization points indicated by modulo_time_base is allowed only for I-VOPs and P-VOPs and is not allowed for B-VOPs.

Also the VM-6.0 states that for I-VOPs and P-VOPs the VOP_time_increment is the time from the synchronization point marked by the last encoded/decoded modulo_time_base, while for B-VOPs the VOP_time_increment is the relative time from the synchronization point marked by the last encoded/decoded I-VOP or P-VOP. This is mainly for the following reason. That is, a B-VOP is predictively encoded by employing as a reference image the I-VOP or P-VOP arranged across the B-VOP in display order. Therefore, the temporal distance to the I-VOP or P-VOP is set to the VOP_time_increment for the B-VOP so that the weight, relative to the I-VOP or P-VOP which is employed as a reference image in performing the predictive coding, is determined from the B-VOP on the basis of the temporal

distance to the I-VOP or P-VOP arranged across the B-VOP. This is the main reason.

Incidentally, the definition of the VOP_time_increment of the above-mentioned VM-6.0 has a disadvantage. That is, in Figure 24 the VOP_time_increment for a B-VOP is not the relative time from the I-VOP or P-VOP encoded/decoded immediately before the B-VOP but it is the relative time from the last displayed I-VOP or P-VOP. This is for the following reason. For example, consider B2 or B3. The I-VOP or P-VOP which is encoded/decoded immediately before the B2 or B3 is the P4 from the standpoint of the above-mentioned encoding/decoding order. Therefore, when it is assumed that the VOP_time_increment for a B-VOP is the relative time from the I-VOP or P-VOP encoded/decoded immediately before the B-VOP, the VOP_time_increment for the B2 or B3 is the relative time from the encoding/decoding time of the P4 and becomes a negative value.

On the other hand, in the MPEG-4 standard the VOP_time_increment is 10 bits. If the VOP_time_increment has only a value equal to or greater than 0, it can express a value in a range of 0 to 1023. Therefore, the position between adjacent synchronization points can be represented in the units of milliseconds with the previous temporal synchronization point (in the left direction in Figure 24) as reference.

However, if the VOP_time_increment is allowed to have not only a value equal to or greater than 0 but also a negative value, the position between adjacent synchronization points will be represented with the previous temporal synchronization point as reference, or it will be represented with the next temporal synchronization point as reference. For this reason, the process of computing the encoding time or decoding time of a VOP becomes complicated.

Therefore, as described above, for the VOP_time_increment the VM-6.0 states that:
This value represents the local time base in the units of milliseconds. For I- and P-VOPs this value is the absolute VOP_time_increment from the synchronization point marked by the last modulo_time_base. For the B-VOPs this value is the relative VOP_time_increment from the last encoded/decoded I- or P-VOP.

However, the last sentence "For the B-VOPs this value is the relative VOP_time_increment from the last encoded/decoded I- or P-VOP" should be changed to "For the B-VOPs this value is the relative VOP_time_increment from the last displayed I- or P-VOP". With this, the VOP_time_increment should not be defined as the relative time from the last encoded/decoded I-VOP or P-VOP, but it should be defined as the relative time from the last displayed I- or P-VOP.

By defining the VOP_time_increment in this manner, the computation base of the encoding/decoding time for a B-VOP is the display time of the I/P-VOP (I-VOP or P-VOP) having display time prior to the B-VOP. Therefore, the VOP_time_increment for a B-VOP always has a positive value, so long as a reference image I-VOP for the B-VOP is not displayed prior to the B-VOP. Therefore, the VOP_time_increments for I/P-VOPs also have a positive value at all times.

Also, in Figure 24 the definition of the VM-6.0 is further changed so that the time represented by the modulo_time_base and VOP_time_increment is not the encoding/decoding time of a VOP but is the display time of a VOP. That is, in Figure 24, when the absolute time on a sequence of VOPs is considered, the tEST(I/P) in Formula (1) and the tDTB(I/P) in Formula (2) represent absolute times present on a sequence of I-VOPs or P-VOPs, respectively, and the tEST(B) in Formula (1) and the tDTB(B) in Formula (2) represent absolute times present on a sequence of B-VOPs, respectively.

Next, in the VM-6.0 the encoder time base start time tEST in Formula (1) is not encoded, but the modulo_time_base and VOP_time_increment are encoded as the differential information between the encoder time base start time tEST and the display time of each VOP (absolute time representing

the position of a VOP present on a sequence of VOPs). For this reason, at the decoder side, the relative time between VOPs can be determined by employing the modulo_time_base and VOP_time_increment, but the absolute display time of each VOP, i.e., the position of each VOP in a sequence of VOPs cannot be determined. Therefore, only the modulo_time_base and VOP_time_increment cannot perform access to a bit stream, i.e., random access.

On the other hand, if the encoder time base start time tEST is merely encoded, the decoder can decode the absolute time of each VOP by employing the encoded tEST. However, by decoding from the head of the coded bit stream the encoder time base start time tEST and also the modulo_time_base and VOP_time_increment which are the relative time information of each VOP, there is a need to control the cumulative absolute time. This is troublesome, so efficient random access cannot be carried out.

Hence, in the embodiment of the present invention, a layer for encoding the absolute time present on a VOP sequence is introduced into the hierarchical constitution of the encoded bit stream of the VM-6.0 so as to easily perform an effective random access. (This layer is not a layer which realizes scalability (above-mentioned base layer or enhancement layer) but is a layer of encoded bit stream.) This layer is an encoded bit stream layer which can be

inserted at an appropriate position as well as at the head of the encoded bit stream.

As this layer, this embodiment introduces, for example, a layer prescribed in the same manner as a GOP (group of picture) layer employed in the MPEG-1/2 standard. With this, the compatibility between the MPEG-4 standard and the MPEG-1/2 standard can be enhanced as compared with the case where an original encoded bit stream layer is employed in the MPEG-4 standard. This newly introduced layer is referred to as a GOV (or a group of video object plane (GVOP)).

Figure 25 shows a constitution of the encoded bit stream into which a GOV layer is introduced for encoding the absolute times present on a sequence of VOPs.

The GOV layer is prescribed between a VOL layer and a VOP layer so that it can be inserted at the arbitrary position of an encoded bit stream as well as at the head of the encoded bit stream.

With this, in the case where a certain VOL#0 is constituted by a VOP sequence such as VOP#0, VOP#1, ..., VOP#n, VOP#(n+1), ..., and VOP#m, the GOV layer can be inserted, for example, directly before the VOP#(n+1) as well as directly before the head VOP#0. Therefore, at the encoder, the GOV layer can be inserted, for example, at the position of an encoded bit stream where random access is performed. Therefore, by inserting the GOV layer, a VOP

sequence constituting a certain VOL is separated into a plurality of groups (hereinafter referred to as a GOV as needed) and is encoded.

The syntax of the GOV layer is defined, for example, as shown in Figure 26.

As shown in the figure, the GOV layer is constituted by a group_start_code, a time_code, a closed_gop, a broken_link, and a next_start_code(), arranged in sequence.

Next, a description will be made of the semantics of the GOV layer. The semantics of the GOV layer is basically the same as the GOP layer in the MPEG-2 standard. Therefore, for the parts not described here, see the MPEG-2 video standard (ISO/IEC-13818-2).

The group_start_code is 000001B8 (hexadecimal) and indicates the start position of a GOV.

The time_code, as shown in Figure 27, consists of a 1-bit drop_frame_flag, a 5-bit time_code_hours, a 6-bit time_code_minutes, a 1-bit marker_bit, a 6-bit time_code_seconds, and a 6-bit time_code_pictures. Thus, the time_code is constituted by 25 bits in total.

The time_code is equivalent to the "time and control codes for video tape recorders" prescribed in IEC standard publication 461. Here, the MPEG-4 standard does not have the concept of the frame rate of video. (Therefore, a VOP can be represented at an arbitrary time.) Therefore, this

embodiment does not take advantage of the drop_frame_flag indicating whether or not the time_code is described in drop_frame_mode, and the value is fixed, for example, to 0. Also, this embodiment does not take advantage of the time_code_pictures for the same reason, and the value is fixed, for example, to 0. Therefore, the time_code used herein represents the time of the head of a GOV by the time_code_hours representing the hour unit of time representing the hour unit of time, time_code_minutes representing the minute unit of time, and time_code_seconds representing the second unit of time. As a result, the time_code (encoding start second-accuracy absolute time) in a GOV layer expresses the time of the head of the GOV layer, i.e., the absolute time on a VOP sequence when the encoding of the GOV layer is started, within accuracy of a second. For this reason, this embodiment of the present invention sets time within accuracy finer than a second (here, milliseconds) for each VOP.

Note that the marker_bit in the time_code is made 1 so that 23 or more 0's do not continue in a coded bit stream.

The closed_gop means one in which the I-, P- and B-pictures in the definition of the close_gop in the MPEG-2 video standard (ISO/IEC 13818-2) have been replaced with an I-VOP, a P-VOP, and a B-VOP, respectively. Therefore, the B-VOP in one VOP represents not only a VOP constituting the

GOV but whether the VOP has been encoded with a VOP in another GOV as a reference image. Here, for the definition of the close_gop in the MPEG-2 video standard (ISO/IEC 13818-29) the sentences performing the above-mentioned replacement are shown as follows:

This is a one-bit flag which indicates the nature of the predictions used in the first consecutive B-VOPs (if any) immediately following the first coded I-VOP following the group of plane header. The closed_gop is set to 1 to indicate that these B-VOPs have been encoded using only backward prediction or intra coding. This bit is provided for use during any editing which occurs after encoding. If the previous pictures have been removed by editing, broken_link may be set to 1 so that a decoder may avoid displaying these B-VOPs following the first I-VOP following the group of plane header. However if the closed_gop bit is set to 1, then the editor may choose not to set the broken_link bit as these B-VOPs can be correctly decoded.

The broken_link also means one in which the same replacement as in the case of the closed_gop has been performed on the definition of the broken_link in the MPEG-2 video standard (ISO/IEC 13818-29). The broken_link, therefore, represents whether the head B-VOP of a GOV can be correctly regenerated. Here, for the definition of the broken_link in the MPEG-2 video standard (ISO/IEC 13818-2)

the sentences performing the above-mentioned replacement are shown as follows:

This is a one-bit flag which shall be set to 0 during encoding. It is set to 1 to indicate that the first consecutive B-VOPs (if any) immediately following the first coded I-VOP following the group of plane header may not be correctly decoded because the reference frame which is used for prediction is not available (because of the action of editing). A decoder may use this flag to avoid displaying frames that cannot be correctly decoded.

The next_start_code() gives the position of the head of the next GOV.

The above-mentioned absolute time in a GOV sequence which introduces the GOV layer and also starts the encoding of the GOV layer (hereinafter referred to as encoding start absolute time as needed) is set to the time_code of the GOV. Furthermore, as described above, since the time_code in the GOV layer has accuracy within a second, this embodiment sets a finer accuracy portion to the absolute time of each VOP present in a VOP sequence for each VOP.

Figure 28 shows the relation between the time_code, modulo_time_base, and VOP_time_increment in the case where the GOV layer of Figure 26 has been introduced.

In the figure, the GOV is constituted by I1, B2, B3, P4, B5, and P6 arranged in display order from the head.

Now, for example, assuming the encoding start absolute time of the GOV is 0h:12m:35sec:350msec (0 hour 12 minutes 35 second 350 milliseconds), the time_code of the GOV will be set to 0h:12m:35sec because it has accuracy within a second, as described above. (The time_code_hours, time_code_minutes, and time_code_seconds which constitute the time_code will be set to 0, 12, and 35, respectively.) On the other hand, in the case where the absolute time of the I1 in a VOP sequence (absolute time of a VOP sequence before the encoding (or after the decoding) of a VS including the GOV of Figure 28) (since this is equivalent to the display time of the I1 when a VOP sequence is displayed, it will hereinafter be referred to display time as needed) is, for example, 0h:12m:35sec:350msec, the semantics of VOP_time_increment is changed so that 350 ms which is accuracy finer than accuracy of a second is set to the VOP_time_increment of the I-VOP of the I1 and encoded (i.e., so that encoding is performed with the VOP_time_increment of the I1 = 350).

That is, in Figure 28, the VOP_time_increment of the head I-VOP (I1) of a GOV in display order has a differential value between the time_code of the GOV and the display time of the I-VOP. Therefore, the time within accuracy of a second represented by the time_code is the first synchronization point of the GOV (here, a point representing

time within accuracy of a second).

Note that, in Figure 28, the semantics of the VOP_time_increments for the B2, B3, P4, B5, and P6 of the GOV which is VOP arranged as the second or later is the same as the one in which the definition of the VM-6.0 has been changed, as described in Figure 24.

Therefore, in Figure 28 the display time of the B2 or B3 is the time when VOP_time_increment has elapsed since the last displayed I-VOP or P-VOP. In this case, since the display time of the last displayed I1 is 0h:12m:35s:350ms, the display time of the B2 or B3 is 0h:12m:35s:750ms or 0h:12m:36s:200ms after 400 ms or 800 ms.

Next, for the P4, at the P4 the synchronization point indicated by modulo_time_base is switched. Therefore, the time of the synchronization point is 0h:12m:36s after 1 second from 0h:12m:35s. As a result, the display time of the P4 is 0h:12m:36s:550ms after 550 ms from 0h:12m:36s.

The display time of the B5 is the time when VOP_time_increment has elapsed since the last displayed I-VOP or P-VOP. In this case, the display time of the B5 is 0h:12m:36s:950ms after 400 ms from the display time 0h:12m:36s:550ms of the last displayed P4.

Next, for the P6, at the P6 the synchronization point indicated by modulo_time_base is switched. Therefore, the time of the synchronization point is 0h:12m:35s + 2 sec,

i.e., 0h:12m:37s. As a result, the display time of the P6 is 0h:12m:37s:350ms after 350 ms from 0h:12m:37s.

Next, Figure 29 shows the relation between the time_code, modulo_time_base, and VOP_time_increment in the case where the head VOP of a GOV is a B-VOP in display order.

In the figure, the GOV is constituted by B0, I1, B2, B3, P4, B5, and P6 arranged in display order from the head. That is, in Figure 29 the GOV is constituted with the B0 added before the I1 in Figure 28.

In this case, if it is assumed that the VOP_time_increment for the head B0 of the GOV is determined with the display time of the I/P-VOP of the GOV as standard, i.e., for example, if it is assumed that it is determined with the display time of the I1 as standard, the value will be a negative value, which is disadvantageous as described above.

Hence, the semantics of the VOP_time_increment for the B-VOP which is displayed prior to the I-VOP in the GOV (the B-VOP which is displayed prior to the I-VOP in the GOV which is first displayed) is changed as follows.

That is, the VOP_time_increment for such a B-VOP has a differential value between the time_code of the GOV and the display time of the B-VOP. In this case, when the display time of the B0 is, for example, 0h:12m:35s:200ms and when the time_code of the GOV is, for example, 0h:12m:35s, as

shown in Figure 29, the VOP_time_increment for the B0 is 350 ms (= 0h:12m:35s:200ms - 0h:12m:35s). If done in this manner, VOP_time_increment will always have a positive value.

With the aforementioned two changes in the semantics of the VOP_time_increment, the time_code of a GOV and the modulo_time_base and VOP_time_increment of a VOP can be correlated with each other. Furthermore, with this, the absolute time (display time) of each VOP can be specified.

Next, Figure 30 shows the relation between the time_code of a GOV and the modulo_time_base and VOP_time_increment of a VOP in the case where the interval between the display time of the I-VOP and the display time of the B-VOP predicted from the I-VOP is equal to or greater than 1 sec (exactly speaking, 1.023 sec).

In Figure 30, the GOV is constituted by I1, B2, B3, B4, and P6 arranged in display order. The B4 is displayed at the time after 1 sec from the display time of the last displayed I1 (I-VOP).

In this case, when the display time of the B4 is encoded by the above-mentioned VOP_time_increment whose semantics has been changed, the VOP_time_increment is 10 bits as described above and can express only time up to 1023. For this reason, it cannot express time longer than 1.023 sec. Hence, the semantics of the VOP_time_increment is further changed and also the semantics of modulo_time_base

is changed in order to cope with such a case.

In this embodiment, such changes are performed, for example, by either the following first method or second method.

That is, in the first method, the time between the display time of an I/P-VOP and the display time of a B-VOP predicted from the I/P-VOP is detected within accuracy of a second. For the time, the unit of a second is expressed with modulo_time_base, while the unit of a millisecond is expressed with VOP_time_increment.

Figure 31 shows the relation between the time_code for a GOV and the modulo_time_base and VOP_time_increment for a VOP in the case where the modulo_time_base and VOP_time_increment have been encoded in the case shown in Figure 30 in accordance with the first method.

That is, in the first method, the addition of modulo_time_base is allowed not only for an I-VOP and a P-VOP but also for a B-VOP. And the modulo_time_base added to a B-VOP does not represent the switch of synchronization points but represents the carry of a second unit obtained from the display time of the last displayed I/P-VOP.

Furthermore, in the first method, the time after the carry of a second unit from the display time of the last displayed I/P-VOP, indicated by the modulo_time_base added to a B-VOP, is subtracted from the display time of the B-VOP,

and the resultant valve is set as the VOP_time_increment.

Therefore, according to the first method, in Figure 30, if it is assumed that the display time of the I1 is 0h:12m:35s:350ms and also the display time of the B4 is 0h:12m:36s:550ms, then the difference between the display times of the I1 and B4 is 1200 ms more than 1 sec, and therefore the modulo_time_base (shown by a "' mark in Figure 31) indicating the carry of a second unit from the display time of the last displayed I1 is added to the B4 as shown in Figure 31. More specifically, the modulo_time_base which is added to the B4 is 10 representing the carry of 1 sec which is the value of the 1-second digit of 1200 ms. And the VOP_time_increment for the B4 is 200 which is the value less than 1 sec, obtained from the difference between the display times between the I1 and B4 (the value is obtained by subtracting from the display time of the B4 the time after the carry of a second unit obtained from the display time of the last displayed I/P-VOP indicated by the modulo_time_base for the B4).

The aforementioned process for the modulo_time_base and VOP_time_increment according to the first method is performed at the encoder by the VLC unit 36 shown in Figures 11 and 12 and at the decoder by the IVLS unit 102 shown in Figures 17 and 18.

Hence, first, the process for the modulo_time_base and

VOP_time_increment which is performed by the VLC unit 36 will be described in reference to a flowchart of Figure 32.

The VLC UNIT 36 divides a VOP sequence into GOVs and performs processing for each GOV. Note that the GOV is constituted so as to include at least one VOP which is encoded by intra coding.

If a GOV is received, the VLC unit 36 will set the received time to the encoding start absolute time of the GOV, and the GOV will be encoded up to the second accuracy of the encoding start absolute time as the time_code (the encoding start absolute time up to the digit of a second is encoded). The encoded time_code is included in a coded bit stream. Each time an I/P-VOP constituting the GOV is received, the VLC unit 36 sets the I/P-VOP to an attention I/P-VOP, computes the modulo_time_base and VOP_time_increment of the attention I/P-VOP in accordance with the flowchart of Figure 32, and performs encoding.

That is, at the VLC unit 36, first, in step S1, 0B (where B represents a binary number) is set to modulo_time_base and also 0 is set to VOP_time_increment, whereby the modulo_time_base and VOP_time_increment are reset.

And in step S2 it is judged whether the attention I/P-VOP is the first I-VOP of a GOV to be processed (hereinafter referred to as a processing object GOV). In step S2, in the

case where the attention I/P-VOP is judged to be the first I-VOP of the processing object GOV, step S2 advances to step S4. In step S4, the difference between the time_code of the processing object GOV and the second-accuracy of the attention I/P-VOP (here, the first I-VOP in the processing object GOV), i.e., the difference between the time_code and the digit of the second of the display time of the attention I/P-VOP is computed and set to a variable D. Then, step S4 advances to step S5.

Also, in step S2, in the case where it is judged that the attention I/P-VOP is not the first I-VOP of the processing object GOV, step S2 advances to step S3. In step S3, the differential value between the digit of the second of the display time of the attention I/P-VOP and the digit of the second of the display time of the last displayed I/P-VOP (which is displayed immediately before the attention. IP/VOP of the VOP constituting the processing object GOV) is computed and the differential value is set to the variable D. Then, step S3 advances to step S5.

In step S5 it is judged whether the variable D is equal to 0. That is, it is judged whether the difference between the time_code and the digit of the second of the display time of the attention I/P-VOP is equal to 0, or it is judged whether the differential value between the digit of the second of the display time of the attention I/P-VOP and the

digit of the second of the display time of the last displayed I/P-VOP is equal to 0. In step S5, in the case where it is judged that the variable D is not equal to 0, i.e., in the case where the variable D is equal to or greater than 1, step S5 advances to step S6, in which 1 is added as the most significant bit (MSB) of the modulo_time_base. That is, in this case, when the modulo_time_base is, for example, 0B immediately after resetting, it is set to 10B. Also, when the modulo_time_base is, for example, 10B, it is set to 110B.

And step S6 advances to step S7, in which the variable D is incremented by 1. Then, step S7 returns to step S5. Thereafter, steps S5 through S7 are repeated until in step S5 it is judged that the variable D is equal to 0. That is, the number of consecutive 1's in the modulo_time_base is the same as the number of seconds corresponding to the difference between the time_code and the digit of the second of the display time of the attention I/P-VOP or the differential value between the digit of the second of the display time of the attention I/P-VOP and the digit of the second of the display time of the last displayed I/P-VOP. And the modulo_time_base has 0 at the least significant digit (LSD) thereof.

And in step S5, in the case where it is judged that the variable D is equal to 0, step S5 advances to step S8, in

which time finer than the accuracy of the second of the display time of the attention I/P-VOP, i.e., time in the units of milliseconds is set to VOP_time_increment, and the process ends.

At the VLC circuit 36, the modulo_time_base and VOP_time_increment of an attention I/P-VOP computed in the aforementioned manner are added to the attention I/P-VOP. With this, it is included in a coded bit stream.

Note that modulo_time_base, VOP_time_increment, and time_code are encoded at the VLC circuit 36 by variable word length coding.

Each time a B-VOP constituting a processing object GOV is received, the VLC unit 36 sets the B-VOP to an attention B-VOP, computes the modulo_time_base and VOP_time_increment of the attention B-VOP in accordance with a flowchart of Figure 33, and performs encoding.

That is, at the VLC unit 36, in step S11, as in the case of step S1 in Figure 32, the modulo_time_base and VOP_time_increment are first reset.

And step S11 advances to step S12, in which it is judged whether the attention B-VOP is displayed prior to the first I-VOP of the processing object GOV. In step S12, in the case where it is judged that the attention B-VOP is one which is displayed prior to the first I-VOP of the processing object GOV, step S12 advances to step S14. In

step S14, the difference between the time_code of the processing object GOV and the display time of the attention B-VOP (here, B-VOP which is displayed prior to the first I-VOP of the processing object GOV) is computed and set to a variable D. Then, step S13 advances to step S15. Therefore, in Figure 33, time within accuracy of a millisecond (the time up to the digit of the millisecond)is set to the variable D (on the other hand, time within accuracy of a second is set to the variable in Figure 32, as described above).

Also, in step S12, in the case where it is judged that the attention B-VOP is one which is displayed after the first I-VOP of the processing object GOV, step S12 advances to step S14. In step S14, the differential value between the display time of the attention B-VOP and the display time of the last displayed I/P-VOP (which is displayed immediately before the attention B-VOP of the VOP constituting the processing object GOV) is computed and the differential value is set to the variable D. Then, step S13 advances to step S15.

In step S15 it is judged whether the variable D is greater than 1. That is, it is judged whether the difference value between the time_code and the display time of the attention B-VOP is greater than 1, or it is judged whether the differential value between the display time of

the attention B-VOP and the display time of the last displayed I/P-VOP is greater than 1. In step S15, in the case where it is judged that the variable D is greater than 1, step S15 advances to step S17, in which 1 is added as the most significant bit (MSB) of the modulo_time_base. In step S17 the variable D is decremented by 1. Then, step S17 returns to step S15. And until in step S15 it is judged that the variable D is not greater than 1, steps S15 through S17 are repeated. That is, with this, the number of consecutive 1's in the modulo_time_base is the same as the number of seconds corresponding to the difference between the time_code and the display time of the attention B-VOP or the differential value between the display time of the attention B-VOP and the display time of the last displayed I/P-VOP. And the modulo_time_base has 0 at the least significant digit (LSD) thereof.

And in step S15, in the case where it is judged that the variable D is not greater than 1, step S15 advances to step S18, in which the value of the current variable D, i.e., the differential value between the time_code and the display time of the attention B-VOP, or the milliseconds digit to the right of the seconds digit of the differential between the display time of the attention B-VOP and the display time of the last displayed I/P-VOP, is set to VOP_time_increment, and the process ends.

At the VLC circuit 36, the modulo_time_base and VOP_time_increment of an attention B-VOP computed in the aforementioned manner are added to the attention B-VOP. With this, it is included in a coded bit stream.

Next, each time the coded data for each VOP is received, the IVLC unit 102 processes the VOP as an attention VOP. With this process, the IVLC unit 102 recognizes the display time of a VOP included in a coded stream which the VLC unit 36 outputs by dividing a VOP sequence into GOVs and also processing each GOV in the above-mentioned manner. Then, the IVLC unit 102 performs variable word length coding so that the VOP is displayed at the recognized display time. That is, if a GOV is received, the IVLC unit 102 will recognize the time_code of the GOV. Each time an I/P-VOP constituting the GOV is received, the IVLC unit 102 sets the I/P-VOP to an attention I/P-VOP and computes the display time of the attention I/P-VOP, based on the modulo_time_base and VOP_time_increment of the attention I/P-VOP in accordance with a flowchart of Figure 34.

That is, at the IVLC unit 102, first, in step S21 it is judged whether the attention I/P-VOP is the first I-VOP of the processing object GOV. In step S21, in the case where the attention I/P-VOP is judged to be the first I-VOP of the processing object GOV, step S21 advances to step S23. In step S23 the time_code of the processing object GOV is set

to a variable T, and step S23 advances to step S24.

Also, in step S21, in the case where it is judged that the attention I/P-VOP is not the first I-VOP of the processing object GOV, step S21 advances to step S22. In step S22, a value up to the seconds digit of the display time of the last displayed I/P-VOP (which is one of the VOPs constituting the processing object GOV) displayed immediately before the attention I/P-VOP is set to the variable T. Then, step S22 advances to step S24.

In step S24 it is judged whether the modulo_time_base added to the attention I/P-VOP is equal to 0B. In step S24, in the case where it is judged that the modulo_time_base added to the attention I/P-VOP is not equal to 0B, i.e., in the case where the modulo_time_base added to the attention I/P-VOP includes 1, step S24 advances to step S25, in which 1 in the MSB of the modulo_time_base is deleted. Step S25 advances to step S26, in which the variable T is incremented by 1. Then, step S26 returns to step S24. Thereafter, until in step S24 it is judged that the modulo_time_base added to the attention I/P-VOP is equal to 0B, steps S24 through S26 are repeated. With this, the variable T is incremented by the number of seconds which corresponds to the number of 1's in the first modulo_time_base added to the attention I/P-VOP.

And in step S24, in the case where the modulo_time_base

added to the attention I/P-VOP is equal to 0B, step S24 advances to step S27, in which time within accuracy of a millisecond, indicated by VOP_time_increment, is added to the variable T. The added value is recognized as the display time of the attention I/P-VOP, and the process ends.

Next, when a B-VOP constituting the processing object GOV is received, the IVLC unit 102 sets the B-VOP to an attention B-VOP and computes the display time of the attention B-VOP, based on the modulo_time_base and VOP_time_increment of the attention B-VOP in accordance with a flowchart of Figure 35.

That is, at the IVLC unit 102, first, in step S31 it is judged whether the attention B-VOP is one which is displayed prior to the first I-VOP of the processing object GOV. In step S31, in the case where the attention B-VOP is judged to be one which is displayed prior to the first I-VOP of the processing object GOV, step S31 advances to step S33. Thereafter, in steps S33 to S37, as in the case of steps S23 to S27 in Figure 34, a similar process is performed, whereby the display time of the attention B-VOP is computed.

On the other hand, in step S31, in the case where it is judged that the attention B-VOP is one which is displayed after the first I-VOP of the processing object GOV, step S31 advances to step S32. Thereafter, in steps s32 and S34 to S37, as in the case of steps S22 and S24 to S27 in Figure 34,

a similar process is performed, whereby the display time of the attention B-VOP is computed.

Next, in the second method, the time between the display time of an I-VOP and the display time of a B-VOP predicted from the I-VOP is computed up to the seconds digit. The value is expressed with modulo_time_base, while the millisecond accuracy of the display time of B-VOP is expressed with VOP_time_increment. That is, the VM-6.0, as described above, the temporal distance to an I-VOP or P-VOP is set to the VOP_time_increment for a B-VOP so that the weight, relative to the I-VOP or P-VOP which is employed as a reference image in performing the predictive coding of the B-VOP, is determined from the B-VOP on the basis of the temporal distance to the I-VOP or P-VOP arranged across the B-VOP. For this reason, the VOP_time_increment for the I-VOP or P-VOP is different from the time from the synchronization point marked by the last encoded/decoded modulo_time_base. However, if the display time of a B-VOP and also the I-VOP or P-VOP arranged across the B-VOP are computed, the temporal distance therebetween can be computed by the difference therebetween. Therefore, there is little necessity to handle only the VOP_time_increment for the B-VOP separately from the VOP_time_increments for the I-VOP and P-VOP. On the contrary, from the viewpoint of processing efficiency it is preferable that all

VOP_time_increments (detailed time information) for I-, B-, and P-VOPs and, furthermore, the modulo_time_bases (second-accuracy time information) be handled in the same manner.

Hence, in the second method, the modulo_time_base and VOP_time_increment for the B-VOP are handled in the same manner as those for the I/P-VOP.

Figure 36 shows the relation between the time_code for a GOV and the modulo_time_base and VOP_time_increment in the case where the modulo_time_base and VOP_time_increment have been encoded according to the second method, for example, in the case shown in Figure 30.

That is, even in the second method, the addition of modulo_time_base is allowed not only for an I-VOP and a P-VOP but also for a B-VOP. And the modulo_time_base added to a B-VOP, as with the modulo_time_base added to an I/P-VOP, represents the switch of synchronization points.

Furthermore, in the second method, the time of the synchronization point marked by the modulo_time_base added to a B-VOP is subtracted from the display time of the B-VOP, and the resultant valve is set as the VOP_time_increment.

Therefore, according to the second method, in Figure 30, the modulo_time_bases for I1 and B2, displayed between the first synchronization point of a GOV (which is time represented by the time_code of the GOV) and the synchronization point marked by the time_code + 1 sec, are

both 0B. And the values of the milliseconds unit lower than the seconds unit of the display times of the I1 and B2 are set to the VOP_time_increments for the I1 and B2, respectively. Also, the modulo_time_bases for B3 and B4, displayed between the synchronization point marked by the time_code + 1 sec and the synchronization point marked by the time_code + 2 sec, are both 10B. And the values of the milliseconds unit lower than the seconds unit of the display times of the B3 and B4 are set to the VOP_time_increments for the B3 and B4, respectively. Furthermore, the modulo_time_base for P5, displayed between the synchronization point marked by the time_code + 2 sec and the synchronization point marked by the time_code + 3 sec, is 110B. And the value of the milliseconds unit lower than the seconds unit of the display time of the P5 is set to the VOP_time_increment for the P5.

For example, in Figure 30 if it is assumed that the display time of the I1 is 0h:12m:35s:350ms and also the display time of the B4 is 0h:12m:36s:550ms, as described above, the modulo_time_bases for I1 and B4 are 0B and 10B, respectively. Also, the VOP_time_increments for I1 and B4 are 0B are 350 ms and 550 ms (which are the milliseconds unit of the display time), respectively.

The aforementioned process for the modulo_time_base and VOP_time_increment according to the second method, as in the

case of the first method, is performed by the VLC unit 36 shown in Figures 11 and 12 and also by the IVLC unit 102 shown in Figures 17 and 18.

That is, the VLC unit 36 computes the modulo_time_base and VOP_time_for an I/P-VOP in the same manner as the case in Figure 32.

Also, for a B-VOP, each time the B-VOP constituting a GOV is received, the VLC unit 36 sets the B-VOP to an attention B-VOP and computes the modulo_time_base and VOP_time_increment of the attention B-VOP in accordance with a flowchart of Figure 37.

That is, at the VLC unit 36, first, in step S41 the modulo_time_base and VOP_time_increment are reset in the same manner as the case in step S1 of Figure 32.

And step S41 advances to step S42, in which it is judged whether the attention B-VOP is one which is displayed prior to the first I-VOP of a GOV to be processed (a processing object GOV). In step S42, in the case where it is judged whether the attention B-VOP is one which is displayed prior to the first I-VOP of the processing object GOV, step S42 advances to step S44. In step S44, the difference between the time_code of the processing object GOV and the second-accuracy of the attention B-VOP, i.e., the difference between the time_code and the seconds digit of the display time of the attention B-VOP is computed and

set to a variable D. Then, step S44 advances to step S45.

Also, in step S42, in the case where it is judged that the attention B-VOP is one which is displayed after the first I-VOP of the processing object GOV, step S42 advances to step S43. In step S43, the differential value between the seconds digit of the display time of the attention B-VOP and the seconds digit of the display time of the last displayed I/P-VOP (which is one of the VOPs constituting the processing object GOV, displayed immediately before the attention B-VOP) is computed and the differential value is set to the variable D. Then, step S43 advances to step S45.

In step S45 it is judged whether the variable D is equal to 0. That is, it is judged whether the difference between the time_code and the seconds digit of the display time of the attention B-VOP is equal to 0, or it is judged whether the differential value between the seconds digit of the display time of the attention B-VOP and the seconds digit of the display time of the last displayed I/P-VOP is equal to 0 sec. In step S45, in the case where it is judged that the variable D is not equal to 0, i.e., in the case where the variable D is equal to or greater than 1, step S45 advances to step S46, in which 1 is added as the MSB of the modulo_time_base.

And step S46 advances to step S47, in which the variable D is incremented by 1. Then, step S47 returns to

step S45. Thereafter, until in step S45 it is judged that the variable D is equal to 0, steps S45 through S47 are repeated. That is, with this, the number of consecutive 1's in the modulo_time_base is the same as the number of seconds corresponding to the difference between the time_code and the seconds digit of the display time of the attention B-VOP or the differential value between the seconds digit of the display time of the attention B-VOP and the seconds digit of the display time of the last displayed I/P-VOP. And the modulo_time_base has 0 at the LSD thereof.

And in step S45, in the case where it is judged that the variable D is equal to 0, step S45 advances to step S48, in which time finer than the seconds accuracy of the display time of the attention B-VOP, i.e., time in the millisecond unit is set to the VOP_time_increment, and the process ends.

On the other hand, for an I/P-VOP the IVLC unit 102 computes the display time of the I/P-VOP, based on the modulo_time_base and VOP_time_increment in the same manner as the above-mentioned case in Figure 34.

Also, for a B-VOP, each time the B-VOP constituting a GOV is received, the IVLC unit 102 sets the B-VOP to an attention B-VOP and computes the display time of the attention B-VOP, based on the modulo_time_base and VOP_time_increment of the attention B-VOP in accordance with a flowchart of Figure 38.

That is, at the IVLC unit 102, first, in step S51 it is judged whether the attention B-VOP is one which is displayed prior to the first I-VOP of the processing object GOV. In step S51, in the case where it is judged that the attention B-VOP is one which is displayed prior to the first I-VOP of the processing object GOV, step S51 advances to step S52. In step S52 the time_code of the processing object GOV is set to a variable T, and step S52 advances to step S54.

Also, in step S51, in the case where it is judged that the attention B-VOP is one which is displayed after the first I-VOP of the processing object GOV, step S51 advances to step S53. In step S53, a value up to the seconds digit of the display time of the last displayed I/P-VOP (which is one of the VOPs constituting the processing object GOV, displayed immediately before the attention B-VOP) is set to the variable T. Then, step S53 advances to step S54.

In step S54 it is judged whether the modulo_time_base added to the attention B-VOP is equal to 0B. In step S54, in the case where it is judged that the modulo_time_base added to the attention B-VOP is not equal to 0B, i.e., in the case where the modulo_time_base added to the attention B-VOP includes 1, step S54 advances to step S55, in which the 1 in the MSB of the modulo_time_base is deleted. Step S55 advances to step S56, in which the variable T is incremented by 1. Then, step S56 returns to step S54.

Thereafter, until in step S54 it is judged that the modulo_time_base added to the attention B-VOP is equal to 0B, steps S54 through S56 are repeated. With this, the variable T is incremented by the number of seconds which corresponds to the number of 1's in the first modulo_time_base added to the attention B-VOP.

And in step S54, in the case where the modulo_time_base added to the attention B-VOP is equal to 0B, step S54 advances to step S57, in which time within accuracy of a millisecond, indicated by the VOP_time_increment, is added to the variable T. The added value is recognized as the display time of the attention B-VOP, and the process ends.

Thus, in the embodiment of the present invention, the GOV layer for encoding the encoding start absolute time is introduced into the hierarchical constitution of an encoded bit stream. This GOV layer can be inserted at an appropriate position of the encoded bit stream as well as at the head of the encoded bit stream. In addition, the definitions of the modulo_time_base and VOP_time_increment prescribed in the VM-6.0 have been changed as described above. Therefore, it becomes possible in all cases to compute the display time (absolute time) of each VOP regardless of the arrangement of picture types of VOPs and the time interval between adjacent VOPs.

Therefore, at the encoder, the encoding start absolute

time is encoded at a GOV unit and also the modulo_time_base and VOP_time_increment of each VOP are encoded. The coded data is included in a coded bit stream. With this, at the decoder, the encoding start absolute time can be decoded at a GOV unit and also the modulo_time_base and VOP_time_increment of each VOP can be decoded. And the display time of each VOP can be decoded, so it becomes possible to perform random access efficiently at a GOV unit.

Note if the number of 1's which are added to modulo_time_base is merely increased as a synchronization point is switched, it will reach the huge number of bits. For example, if 1 hr (3600 sec) has elapsed since the time marked by time_code (in the case where a GOV is constituted by VOPs equivalent to that time), the modulo_time_base will reach 3601 bits, because it is constituted by a 1 of 3600 bits and a 0 of 1 bit.

Hence, in the MPEG-4 the modulo_time_base is prescribed so that it is reset at an I/P-VOP which is first displayed after a synchronization point has been switched.

Therefore, for example, as shown in Figure 39, in the case where a GOV is constituted by I1 and B2 displayed between the first synchronization point of the GOV (which is time represented by the time_code of the GOV) and the synchronization point marked by time_code + 1 sec, B3 and B4 displayed between the synchronization point marked by the

time_code + 1 sec and the synchronization point marked by the time_code + 2 sec, P5 and B6 displayed between the synchronization point marked by the time_code + 2 sec and the synchronization point marked by the time_code + 3 sec, B7 displayed between the synchronization point marked by the time_code + 3 sec and the synchronization point marked by the time_code + 4 sec, and B8 displayed between the synchronization point marked by the time_code + 4 sec and the synchronization point marked by the time_code + 5 sec, the modulo_time_bases for the I1 and B2, displayed between the first synchronization point of the GOV and the synchronization point marked by the time_code + 1 sec, are set to 0B.

Also, the modulo_time_bases for the B3 and B4, displayed between the synchronization point marked by the time_code + 1 sec and the synchronization point marked by the time_code + 2 sec, are set to 10B. Furthermore, the modulo_time_base for the P5, displayed between the synchronization point marked by the time_code + 2 sec and the synchronization point marked by the time_code + 3 sec, is set to 110B.

Since the P5 is a P-VOP which is first displayed after the first synchronization point of a GOV has been switched to the synchronization point marked by the time_code + 1 sec, the modulo_time_base for the P5 is set to 0B. The

modulo_time_base for the B6, which is displayed after the B5, is set on the assumption that a reference synchronization point used in computing the display time of the P5, i.e., the synchronization point marked by the time_code + 2 sec in this case is the first synchronization point of the GOV. Therefore, the modulo_time_base for the B6 is set to 0B.

Thereafter, the modulo_time_base for the B7, displayed between the synchronization point marked by the time_code + 3 sec and the synchronization point marked by the time_code + 4 sec, is set to 10B. The modulo_time_base for the B8, displayed between the synchronization point marked by the time_code + 4 sec and the synchronization point marked by the time_code + 5 sec, is set to 110B.

The process at the encoder (VLC unit 36) described in Figures 32, 33, and 37 is performed so as to set the modulo_time_base in the above-mentioned manner.

Also, in this case, when the first displayed I/P-VOP after the switch of synchronization points is detected, at the decoder (IVLC unit 102) there is a need to add the number of seconds indicated by the modulo_time_base for the I/P-VOP to the time_code and compute the display time. For instance, in the case shown in Figure 39, the display times of I1 to P5 can be computed by adding both the number of seconds corresponding to the modulo_time_base for each VOP and the VOP_time_increment to the time_code. However, the

display times of B6 to B8, displayed after P5 which is first displayed after a switch of synchronization points, need to be computed by adding both the number of seconds corresponding to the modulo_time_base for each VOP and the VOP_time_increment to the time_code and, furthermore, by adding 2 seconds which is the number of seconds corresponding to the modulo_time_base for P5. For this reason, the process described in Figures 34, 35, and 38 is performed so as to compute display time in the aforementioned manner.

Next, the aforementioned encoder and decoder can also be realized by dedicated hardware or by causing a computer to execute a program which performs the above-mentioned process.

Figure 40 shows the constitution example of an embodiment of a computer which functions as the encoder of Figure 3 or the decoder of Figure 15.

A read only memory (ROM) 201 stores a boot program, etc. A central processing unit 202 performs various processes by executing a program stored on a hard disk (HD) 206 at a random access memory (RAM) 203. The RAM 203 temporarily stores programs which are executed by the CPU 202 or data necessary for the CPU 202 to process. An input section 204 is constituted by a keyboard or a mouse. The input section 204 is operated when a necessary command or data is input.

An output section 205 is constituted, for example, by a display and displays data in accordance with control of the CPU 202. The HD 206 stores programs to be executed by the CPU 202, image data to be encoded, coded data (coded bit stream), decoded image data, etc. A communication interface (I/F) 207 receives the image data of an encoding object from external equipment or transmits a coded bit stream to external equipment, by controlling communication between it and external equipment. Also, the communication I/F 207 receives a coded bit stream from an external unit or transmits decoded image data to an external unit.

By causing the CPU 202 of the thus-constituted computer to execute a program which performs the aforementioned process, this computer functions as the encoder of Figure 3 or the decoder of Figure 15.

In the embodiment of the present invention, although VOP_time_increment represents the display time of a VOP in the unit of a millisecond, the VOP_time_increment can also be made as follows. That is, the time between one synchronization point and the next synchronization point is divided into N points, and the VOP_time_increment can be set to a value which represents the nth position of the divided point corresponding to the display time of a VOP. In the case where the VOP_time_increment is thus defined, if N = 1000, it will represent the display time of a VOP in the

unit of a millisecond. In this case, although information on the number of divided points between two adjacent synchronization points is required, the number of divided points may be predetermined or the number of divided points included in an upper layer than a GOV layer may be transmitted to a decoder.

According to the image encoder as set forth in claim 1 and the image encoding method as set forth in claim 6, one or more layers of each sequence of objects constituting an image are partitioned into a plurality of groups, and the groups are encoded. Therefore, it becomes possible to have random access to the encoded result at a group unit.

According to the image decoder as set forth in claim 11 and the image decoding method as set forth in claim 16, a coded bit stream, obtained by partitioning one or more layers of each sequence of objects constituting the image into a plurality of groups and also by encoding the groups, is decoded. Therefore, it becomes possible to have random access to the coded bit stream at a group unit and decode the bit stream.

According to the distribution medium as set forth in claim 21, a coded bit stream, obtained by partitioning one or more layers of each sequence of objects constituting the image into a plurality of groups and also by encoding the groups, is distributed. Therefore, it becomes possible to

have random access to the coded bit stream at a group unit.

According to the image encoder as set forth in claim 26 and the image encoding method as set forth in claim 29, second-accuracy time information which indicates time within accuracy of a second is generated and detailed time information, which indicates a time period between the second-accuracy time information directly before display time of I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second, is generated. Therefore, it becomes possible to recognize the display times of the I-VOP, P-VOP, and B-VOP on the basis of the second-accuracy time information and detailed time information and perform random access on the basis of the recognition result.

According to the image decoder as set forth in claim 32 and the image decoding method as set forth in claim 35, the display times of the I-VOP, P-VOP, and B-VOP are computed based on the second-accuracy time information and detailed time information. Therefore, it becomes possible to perform random access, based on the display time.

According to the distribution medium as set forth in claim 38, there is distributed a coded bit stream which is obtained by generating second-accuracy time information which indicates time within accuracy of a second, also by generating detailed time information which indicates a time

period between the second-accuracy time information directly before display time of I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second, and furthermore by adding the second-accuracy time information and detailed time information to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates display time of said I-VOP, P-VOP, or B-VOP. Therefore, it becomes possible to recognize the display times of the I-VOP, P-VOP, and B-VOP on the basis of the second-accuracy time information and detailed time information and perform random access on the basis of the recognition result.

Industrial Applicability

The present invention can be utilized in image information recording-regenerating units in which dynamic image data is recorded on storage media, such as a magneto-optical disk, magnetic tape, etc., and also the recorded data is regenerated and displayed on a display. The invention can also be utilized in videoconference systems, videophone systems, broadcasting equipment, and multimedia data base retrieval systems, in which dynamic image data is transmitted from a transmitter side to a receiver side through a transmission path and, on the receiver side, the received dynamic data is displayed or it is edited and recorded.



CLAIMS

1. An image encoder for encoding an image and outputting the resultant coded bit stream, the image encoder comprising:

receiving means for receiving said image; and

encoding means for partitioning one or more layers of each sequence of objects constituting said image into a plurality of groups and encoding said groups.

2. The image encoder as set forth in claim 1, wherein, when said object is intra encoded or inter encoded, said encoding means partitions a sequence of said objects into groups which includes at least one object that is intra encoded.

3. The image encoder as set forth in claim 1, wherein said encoding means includes, in said object group, encoding start second-accuracy absolute time which is an absolute time in said image sequence at which the encoding is started and also which is time within accuracy of a second.

4. The image encoder as set forth in claim 1, wherein, when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an

* C L M S *

object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-VOP is a bidirectionally predictive-VOP (B-VOP), said encoder further comprises:

second-accuracy time information generation means for generating second-accuracy time information which indicates time within accuracy of a second based on said encoding start second-accuracy absolute time;

detailed time information generation means for generating detailed time information which indicates a time period between said second-accuracy time information directly before display time of said I-VOP, P-VOP, or B-VOP included in said object group and the display time within accuracy finer than accuracy of a second; and

addition means for adding said second-accuracy time information and detailed time information to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates display time of said I-VOP, P-VOP, or B-VOP.

5. The image encoder as set forth in claim 4, wherein said second-accuracy time information generation means generates, as second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute time and display time of said predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed

* C L M S *

immediately before said predetermined object and display time of said predetermined object which is represented within accuracy of a second.

6. An image encoding method of encoding an image and outputting the resultant coded bit stream, the image encoding method comprising:

a receiving step of receiving said image; and

an encoding step of partitioning one or more layers of each sequence of objects constituting said image into a plurality of groups and encoding said groups.

7. The image encoding method as set forth in claim 6, wherein, when a sequence of said object is intra encoded or inter encoded, in said encoding step a sequence of said objects is partitioned into groups which includes at least one object that is intra encoded.

8. The image encoding method as set forth in claim 6, wherein in said encoding step, encoding start second-accuracy absolute time, which is an absolute time in said image sequence at which the encoding is started and also which is time within accuracy of a second, is included in said object group.

9. The image encoding method as set forth in claim 8, wherein, when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward

predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-VOP is a bidirectionally predictive-VOP (B-VOP), said image encoding method further comprises:

a second-accuracy time information generation step of generating second-accuracy time information which indicates time within accuracy of a second based on said encoding start second-accuracy absolute time;

a detailed time information generation step of generating detailed time information which indicates a time period between said second-accuracy time information directly before display time of said I-VOP, P-VOP, or B-VOP included in said object group and the display time within accuracy finer than accuracy of a second; and

an addition step of adding said second-accuracy time information and detailed time information to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates display time of said I-VOP, P-VOP, or B-VOP.

10. The image encoding method as set forth in claim 9, wherein said second-accuracy time information generation step generates, as second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute time and display time of said

predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed immediately before said predetermined object and display time of said predetermined object which is represented within accuracy of a second.

11. An image decoder for decoding an image, comprising:

receiving means for receiving a coded bit stream obtained by partitioning one or more layers of each sequence of objects constituting said image into a plurality of groups and also by encoding the groups; and

decoding means for decoding said coded bit stream.

12. The image decoder as set forth in claim 11, wherein, when said object is intra encoded or inter encoded, a sequence of said objects is partitioned into groups which includes at least one object that is intra encoded.

13. The image decoder as set forth in claim 11, wherein said group of objects includes encoding start second-accuracy absolute time which is an absolute time in said image sequence at which the encoding is started and also which is time within accuracy of a second; and which further comprises display time computation means for computing the display time of said object included in said group of objects on the basis of said encoding start second accuracy absolute time.

14. The image decoder as set forth in claim 13, wherein,

when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-VOP is a bidirectionally predictive-VOP (B-VOP),

second-accuracy time information indicating time within accuracy of a second with said encoding start second-accuracy absolute time as standard, and detailed time information indicating a time period between said second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP included in said object group and the display time within accuracy finer than accuracy of a second, have been added to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates said display time, and

said display time computation means computes the display time of said I-VOP, P-VOP, or B-VOP on the basis of said encoding start second accuracy absolute time, said second-accuracy time information, and said detailed time information.

15. The image decoder as set forth in claim 14, wherein, as second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute

* C L M S *

time and display time of said predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed immediately before said predetermined object and display time of said predetermined object which is represented within accuracy of a second, is employed.

16. An image decoding method of decoding an image, comprising:

a receiving step of receiving a coded bit stream obtained by partitioning one or more layers of each sequence of objects constituting said image into a plurality of groups and also by encoding the groups; and

a decoding step of decoding said coded bit stream.

17. The image decoding method as set forth in claim 16, wherein, when said object is intra encoded or inter encoded, a sequence of said objects is partitioned into groups which includes at least one object that is intra encoded.

18. The image decoding method as set forth in claim 16, wherein said group of objects includes encoding start second-accuracy absolute time which is an absolute time in said image sequence at which the encoding is started and also which is time within accuracy of a second; and which further comprises display time computation means for computing the display time of said object included in said group of objects on the basis of said encoding start second

accuracy absolute time.

19. The image decoding method as set forth in claim 18, wherein, when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-VOP is a bidirectionally predictive-VOP (B-VOP),

second-accuracy time information indicating time within accuracy of a second with said encoding start second-accuracy absolute time as standard, and detailed time information indicating a time period between said second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP included in said object group and the display time within accuracy finer than accuracy of a second, have been added to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates said display time, and

said display time computation step computes the display time of said I-VOP, P-VOP, or B-VOP on the basis of said encoding start second accuracy absolute time, said second-accuracy time information, and said detailed time information.

20. The image decoding method as set forth in claim 19,



wherein, as second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute time and display time of said predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed immediately before said predetermined object and display time of said predetermined object which is represented within accuracy of a second, is employed.

21. In a distribution medium for distributing a coded bit stream obtained by encoding an image, the distribution medium which:

receives said image; and

distributes said coded bit stream which is obtained by partitioning one or more layers of each sequence of objects constituting said image into a plurality of groups and encoding said groups.

22. The distribution medium as set forth in claim 21, wherein, when said object is intra encoded or inter encoded, a sequence of said objects is partitioned into groups which includes at least one object that is intra encoded.

23. The distribution medium as set forth in claim 21, wherein said object group includes encoding start second-accuracy absolute time which is an absolute time in said image sequence at which the encoding is started and also which is time within accuracy of a second.

* C L M S *

24. The distribution medium as set forth in claim 23, wherein, when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-VOP is a bidirectionally predictive-VOP (B-VOP),

second-accuracy time information indicating time within accuracy of a second with said encoding start second-accuracy absolute time as standard, and detailed time information indicating a time period between said second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP included in said object group and the display time within accuracy finer than accuracy of a second, have been added to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates said display time.

25. The distribution medium as set forth in claim 24, wherein, as second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute time and display time of said predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed immediately before said predetermined object

and display time of said predetermined object which is represented within accuracy of a second, is employed.

26. In an image encoder for encoding an image for each of objects constituting the image and outputting the resultant coded bit stream,

when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-VOP is a bidirectionally predictive-VOP (B-VOP),

the image encoder comprising:

second-accuracy time information generation means for generating second-accuracy time information which indicates time within accuracy of a second;

detailed time information generation means for generating detailed time information which indicates a time period between said second-accuracy time information directly before display time of said I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second; and

addition means for adding said second-accuracy time information and detailed time information to a corresponding

I-VOP, P-VOP, or B-VOP as information which indicates the display time of said I-VOP, P-VOP, and B-VOP.

27. The image encoder as set forth in claim 26, wherein, when said object is partitioned into one or more groups and encoded, said detailed time information generation means generates time within accuracy of a second as said second-accuracy time information, the time within accuracy of a second being absolute time in said image sequence at which encoding of the object of the group is started and also being based on said encoding start second-accuracy absolute time.

28. The image encoder as set forth in claim 27, wherein said second-accuracy time information generation means generates, as second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute time and display time of said predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed immediately before said predetermined object and display time of said predetermined object which is represented within accuracy of a second.

29. In an image encoding method of encoding an image for each of objects constituting the image and outputting the resultant coded bit stream,

when it is assumed that an object which is encoded by

* C L M S *

intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-VOP is a bidirectionally predictive-VOP (B-VOP),

the image encoding method comprising:

a second-accuracy time information generation step of generating second-accuracy time information which indicates time within accuracy of a second;

a detailed time information generation step of generating detailed time information which indicates a time period between said second-accuracy time information directly before display time of said I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second; and

an addition step of adding said second-accuracy time information and detailed time information to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates the display time of said I-VOP, P-VOP, and B-VOP.

30. The image encoding method as set forth in claim 29, wherein, when a sequence of said object is partitioned into one or more groups and encoded, said second-accuracy time information generation steps generates time within accuracy

of a second as said second-accuracy time information, the time within accuracy of a second being an absolute time in said image sequence at which encoding of the object of the group is started and also being based on said encoding start second-accuracy absolute time.

31. The image encoder as set forth in claim 30, wherein said second-accuracy time information generation step generates, as second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute time and display time of said predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed immediately before said predetermined object and display time of said predetermined object which is represented within accuracy of a second.

32. In an image decoder for decoding a coded bit stream obtained by encoding an image for each of objects constituting the image,

when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-VOP is a bidirectionally

predictive-VOP (B-VOP), and

when both second-accuracy time information indicating time within accuracy of a second and detailed time information indicating a time period between said second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second have been added to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates said display time,

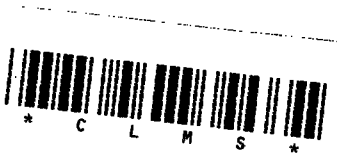
the image decoder comprising:

display time computation means for computing the display time of said I-VOP, P-VOP, or B-VOP on the basis of said second-accuracy time information and detailed time information; and

decoding means for decoding said I-VOP, P-VOP, or B-VOP in accordance with the corresponding display time.

33. The image decoder as set forth in claim 32, wherein, when a sequence of said object is partitioned into one or more groups and encoded, said second-accuracy time information is time within accuracy of a second which is an absolute time in said image sequence at which encoding of the object of the group is started, and represents time within accuracy of a second with encoding start second-accuracy absolute time as standard.

34. The image decoder as set forth in claim 33, wherein, as



second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute time and display time of said predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed immediately before said predetermined object and display time of said predetermined object which is represented within accuracy of a second, is employed.

35. In an image decoding of decoding a coded bit stream obtained by encoding an image for each of objects constituting the image,

when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, backward predictive coding, or bidirectionally predictive-VOP is a bidirectionally predictive-VOP (B-VOP), and

when both second-accuracy time information indicating time within accuracy of a second and detailed time information indicating a time period between said second-accuracy time information directly before display time of the I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second have been added to

a corresponding I-VOP, P-VOP, or B-VOP as information which indicates said display time,

the image decoding method comprising:

a display time computation step of computing the display time of said I-VOP, P-VOP, or B-VOP on the basis of said second-accuracy time information and detailed time information; and

a decoding step of decoding said I-VOP, P-VOP, or B-VOP in accordance with the corresponding display time.

36. The image decoding method as set forth in claim 35, wherein, when a sequence of said object is partitioned into one or more groups and encoded, said second-accuracy time information is time within accuracy of a second which is an absolute time in said image sequence at which encoding of the object of the group is started, and represents time within accuracy of a second with encoding start second-accuracy absolute time as standard.

37. The image decoding method as set forth in claim 36, wherein, as second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute time and display time of said predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed immediately before said predetermined object and display time of said predetermined object which is

represented within accuracy of a second, is employed.

38. In a distribution medium for distributing a coded bit stream obtained by encoding an image for each of objects constituting the image,

when it is assumed that an object which is encoded by intracoding is an intra-video object plane (I-VOP), an object which is encoded by either intracoding or forward predictive coding is a predictive-VOP (P-VOP), and that an object which is encoded by either intracoding, forward predictive coding, or bidirectionally predictive-VOP is a bidirectionally predictive-VOP (B-VOP),

the distribution medium which:

generates second-accuracy time information which indicates time within accuracy of a second;

generates detailed time information which indicates a time period between said second-accuracy time information directly before display time of said I-VOP, P-VOP, or B-VOP and the display time within accuracy finer than accuracy of a second; and

adds said second-accuracy time information and detailed time information to a corresponding I-VOP, P-VOP, or B-VOP as information which indicates display time of said I-VOP, P-VOP, or B-VOP.

39. The distribution medium as set forth in claim 38, wherein, when a sequence of said object is partitioned into

* C L M S *

one or more groups and encoded, time within accuracy of a second is generated as said second-accuracy time information, the time within accuracy of a second being an absolute time in said image sequence at which encoding of the object of the group is started and also being based on said encoding start second-accuracy absolute time.

40. The distribution medium as set forth in claim 39, wherein, as second-accuracy time information for a predetermined object, a time between said encoding start second-accuracy absolute time and display time of said predetermined object which is represented within accuracy of a second, or a time between display time of an I-VOP or P-VOP displayed immediately before said predetermined object and display time of said predetermined object which is represented within accuracy of a second, is generated.

THIS SHEET HAS BEEN REPLACED

ABSTRACT

Efficient random access is made possible. A group of video plane (GOV) layer in which encoding start absolute time within accuracy of a second is arranged is provided as the layer of a coded bit stream. The GOV layer can be inserted not only at the head of the coded bit stream but at an arbitrary position in the coded bit stream. The display time of each video object plane (VOP) constituting the GOV layer is represented by modulo_time_base which represents absolute time in the unit of a second with the encoding start absolute time as standard and VOP_time_increment which represents in the unit of a millisecond the time that has elapsed since time (synchronization point) indicated by the modulo_time_base.

(fig. 25)

IThis Sheet Has Been Replaced

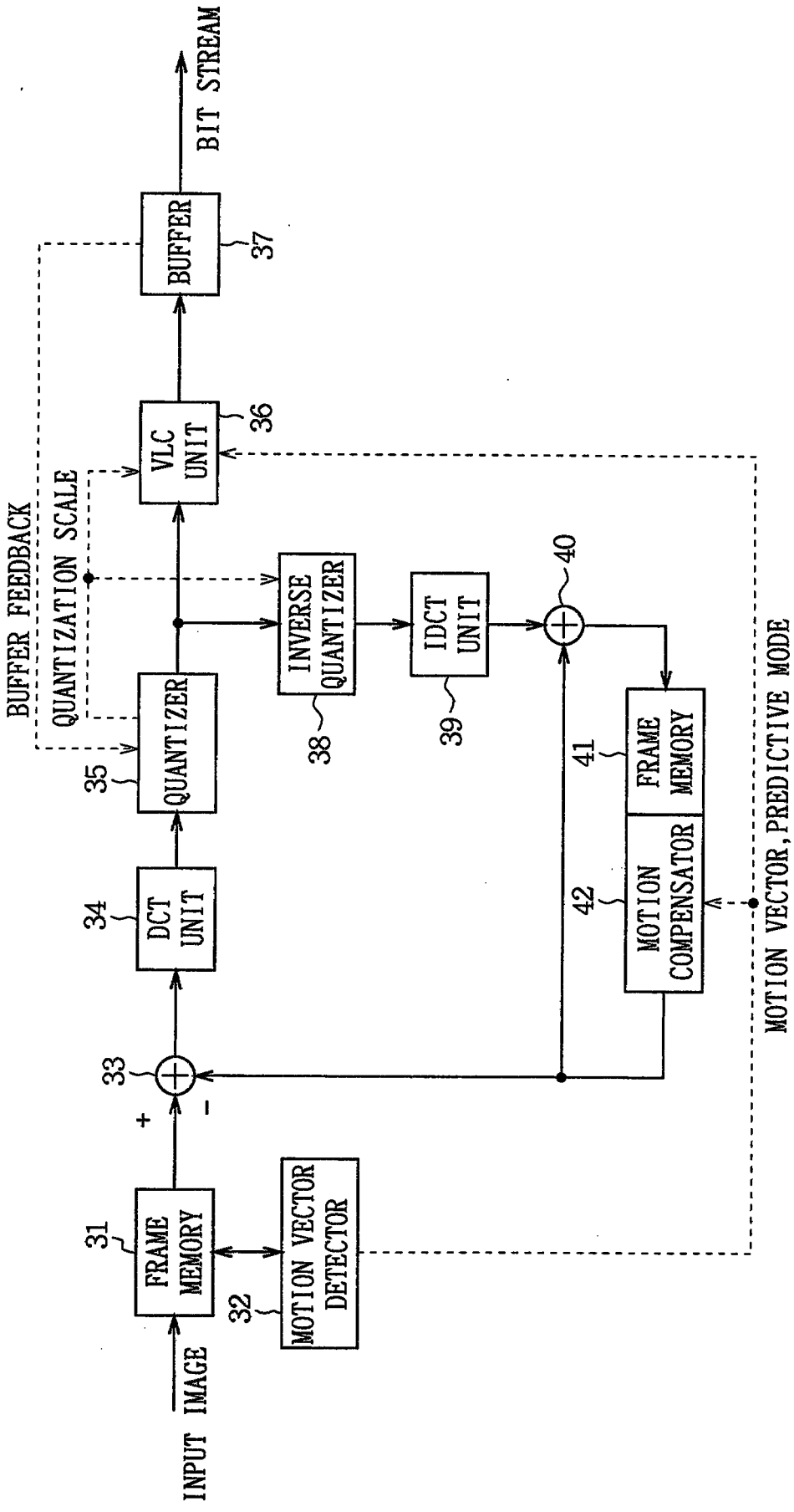

INPUT IMAGE
FRAME MEMORY
31
MOTION VECTOR DETECTOR
32
+
-
33
DCT UNIT
34
QUANTIZER
35
VLC UNIT
36
BUFFER
37
BIT STREAM
INVERSE QUANTIZER
38
IDCT UNIT
39
40
FRAME MEMORY
41
MOTION COMPENSATOR
42
BUFFER FEEDBACK
QUANTIZATION SCALE
MOTION VECTOR, PREDICTIVE MODE

FIG. 1

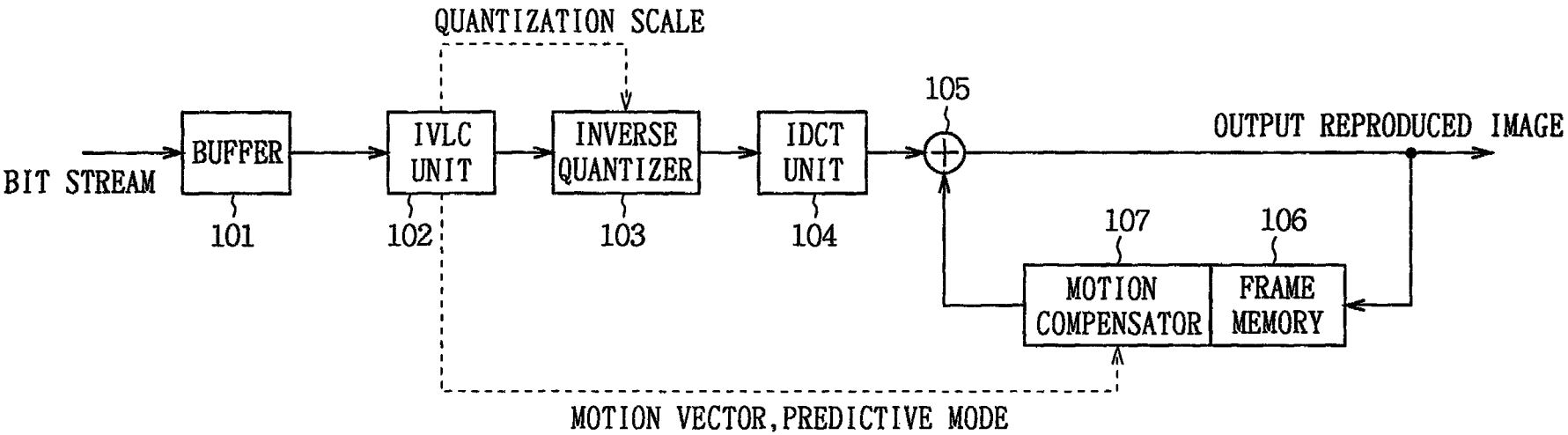

QUANTIZATION SCALE
BIT STREAM
BUFFER
101
IVLC UNIT
102
INVERSE QUANTIZER
103
IDCT UNIT
104
105
OUTPUT REPRODUCED IMAGE
107
106
MOTION COMPENSATOR
FRAME MEMORY
MOTION VECTOR, PREDICTIVE MODE

FIG. 2

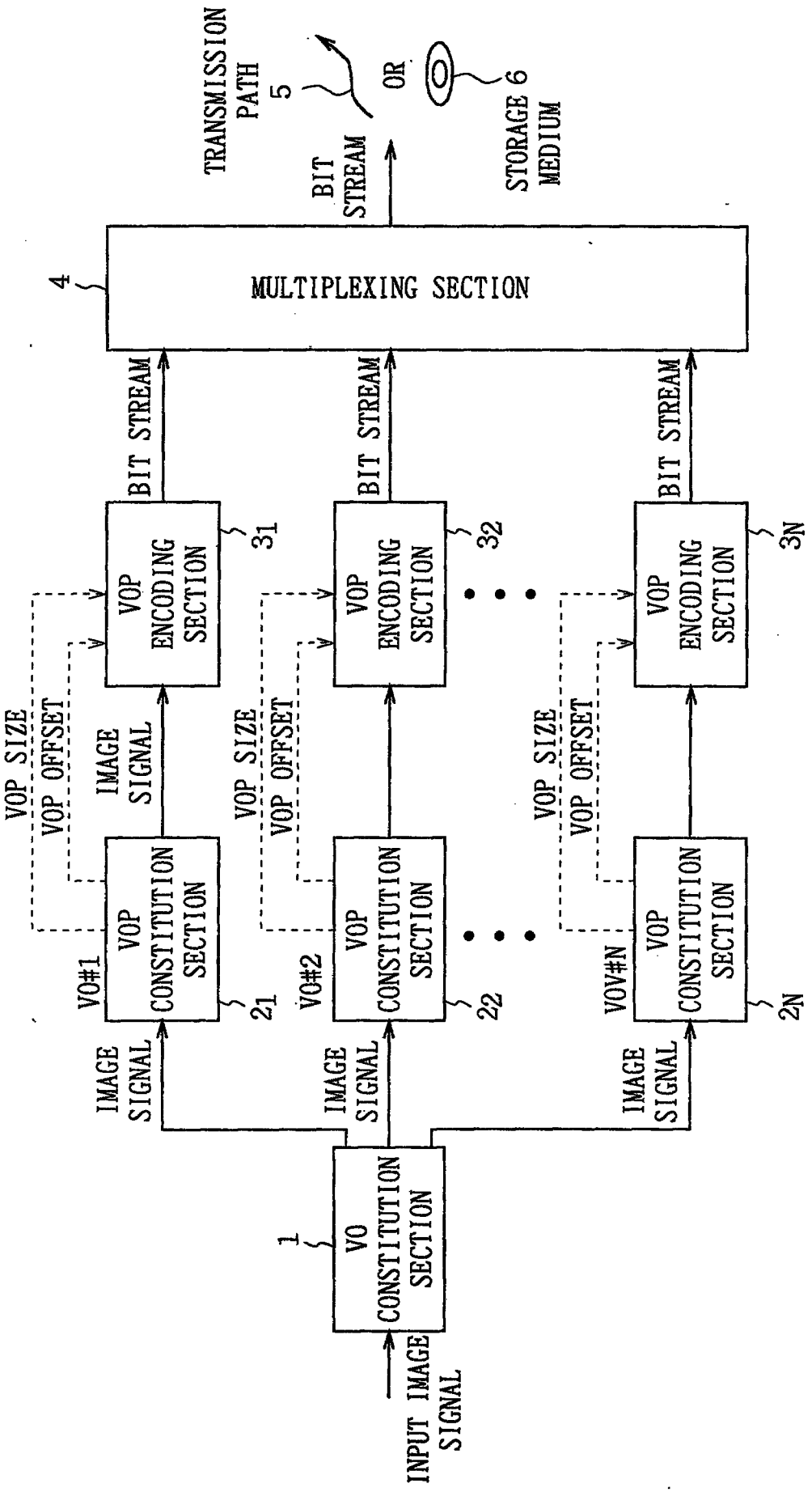

INPUT IMAGE SIGNAL
1
VO CONSTITUTION SECTION
IMAGE SIGNAL
VO#1
VOP CONSTITUTION SECTION
21
VOP SIZE
VOP OFFSET
IMAGE SIGNAL
VOP ENCODING SECTION
31
BIT STREAM
VO#2
VOP CONSTITUTION SECTION
22
VOP SIZE
VOP OFFSET
VOP ENCODING SECTION
32
BIT STREAM
VOV#N
VOP CONSTITUTION SECTION
2N
VOP SIZE
VOP OFFSET
VOP ENCODING SECTION
3N
BIT STREAM
4
MULTIPLEXING SECTION
BIT STREAM
TRANSMISSION PATH
5
OR
STORAGE MEDIUM
6

FIG. 3

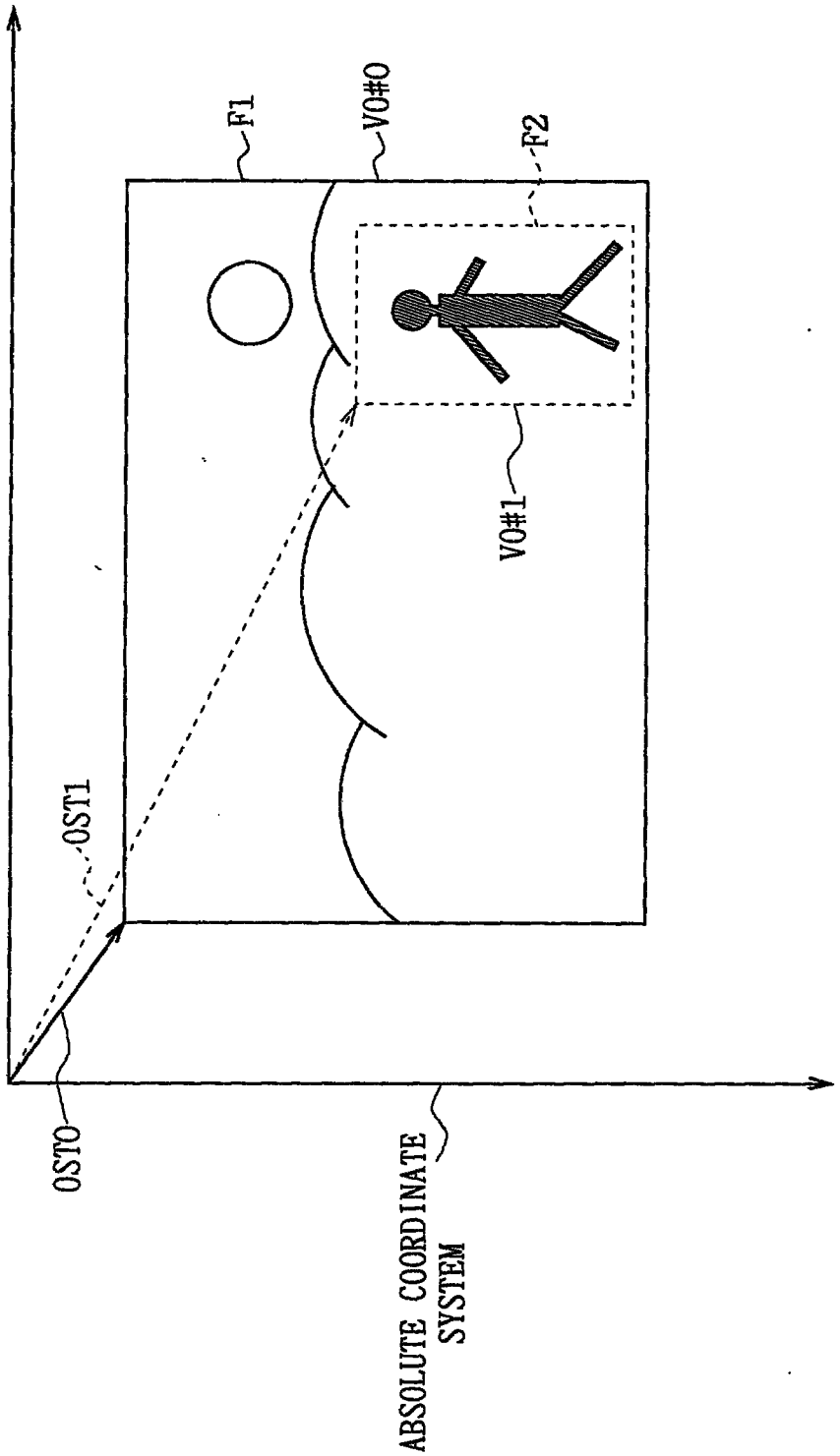
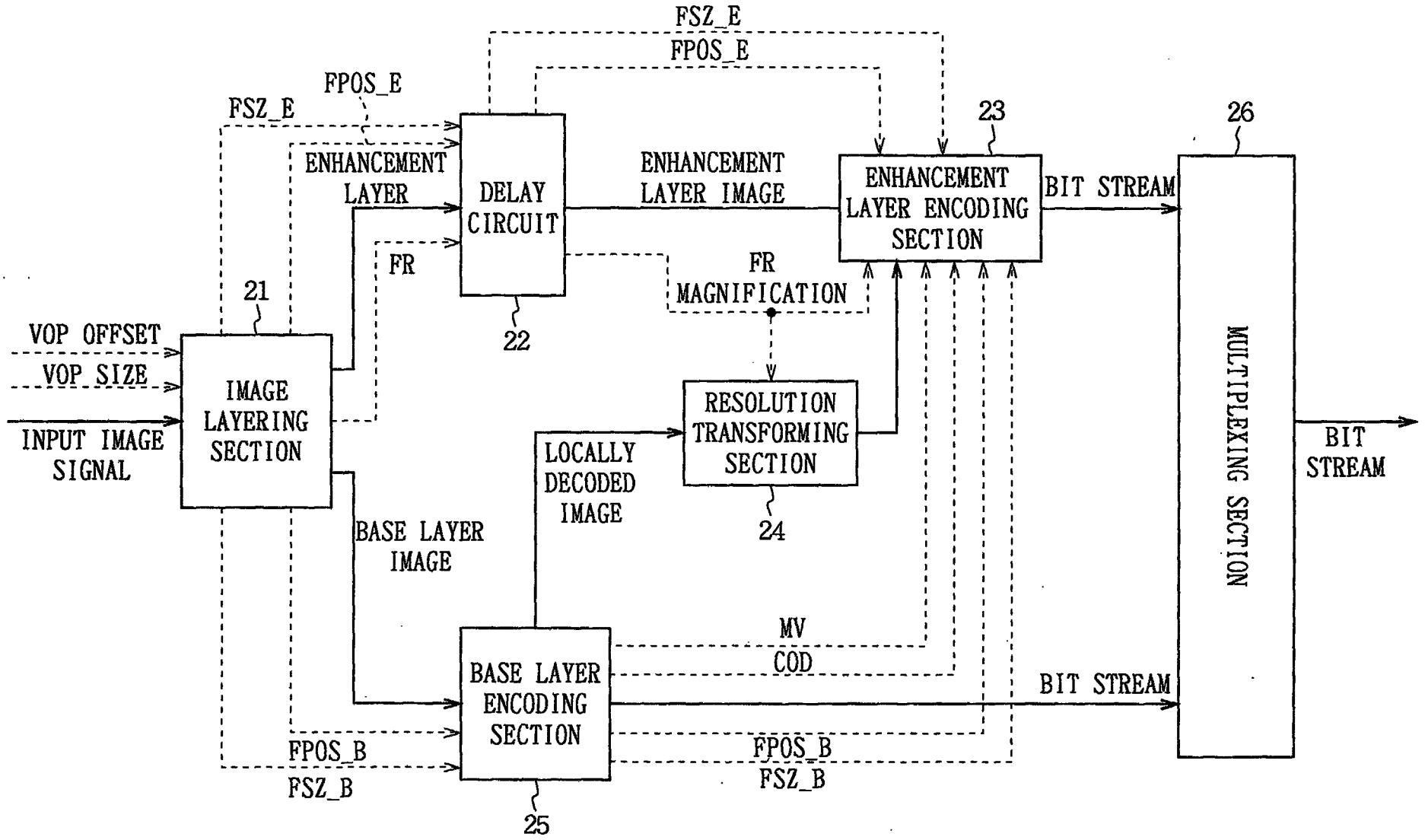


FIG. 4



VOP ENCODING SECTIONS 3n

FIG. 5

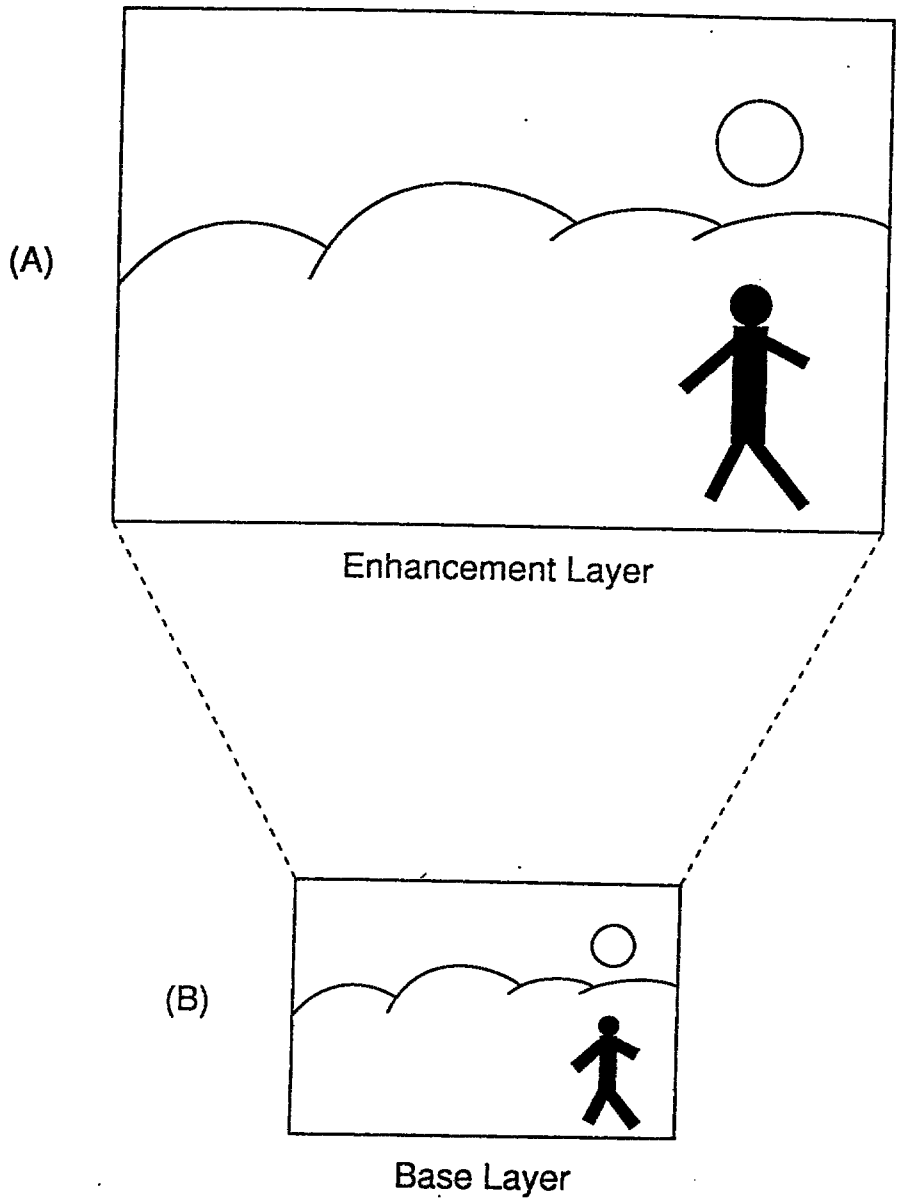


FIG. 6

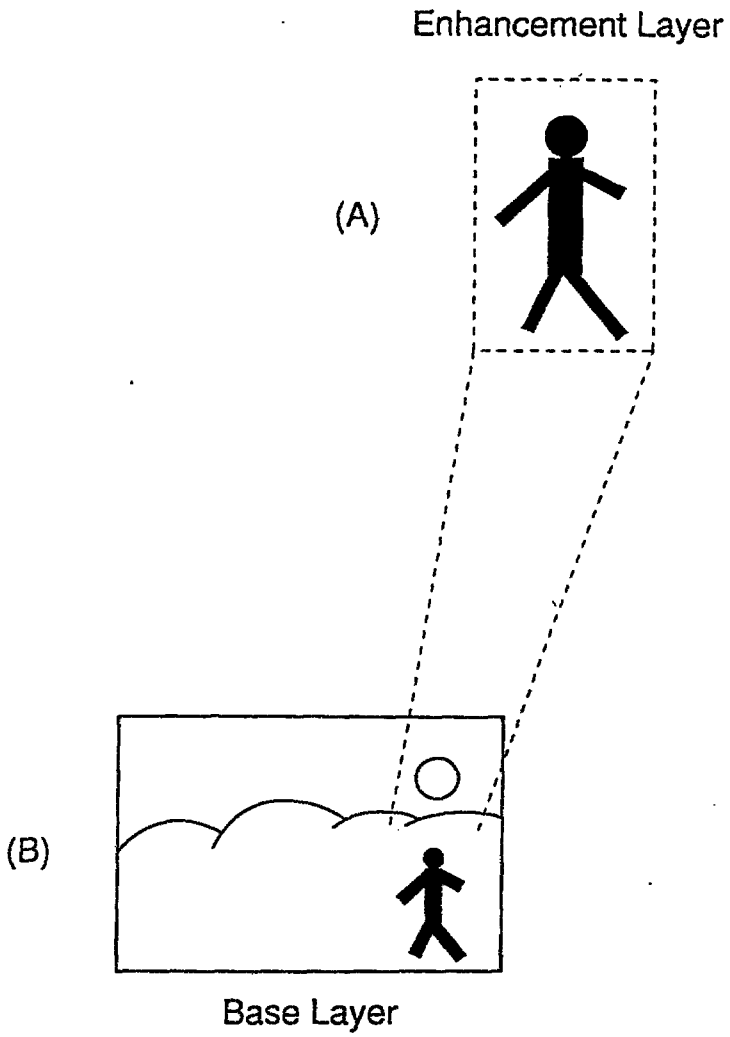


FIG. 7

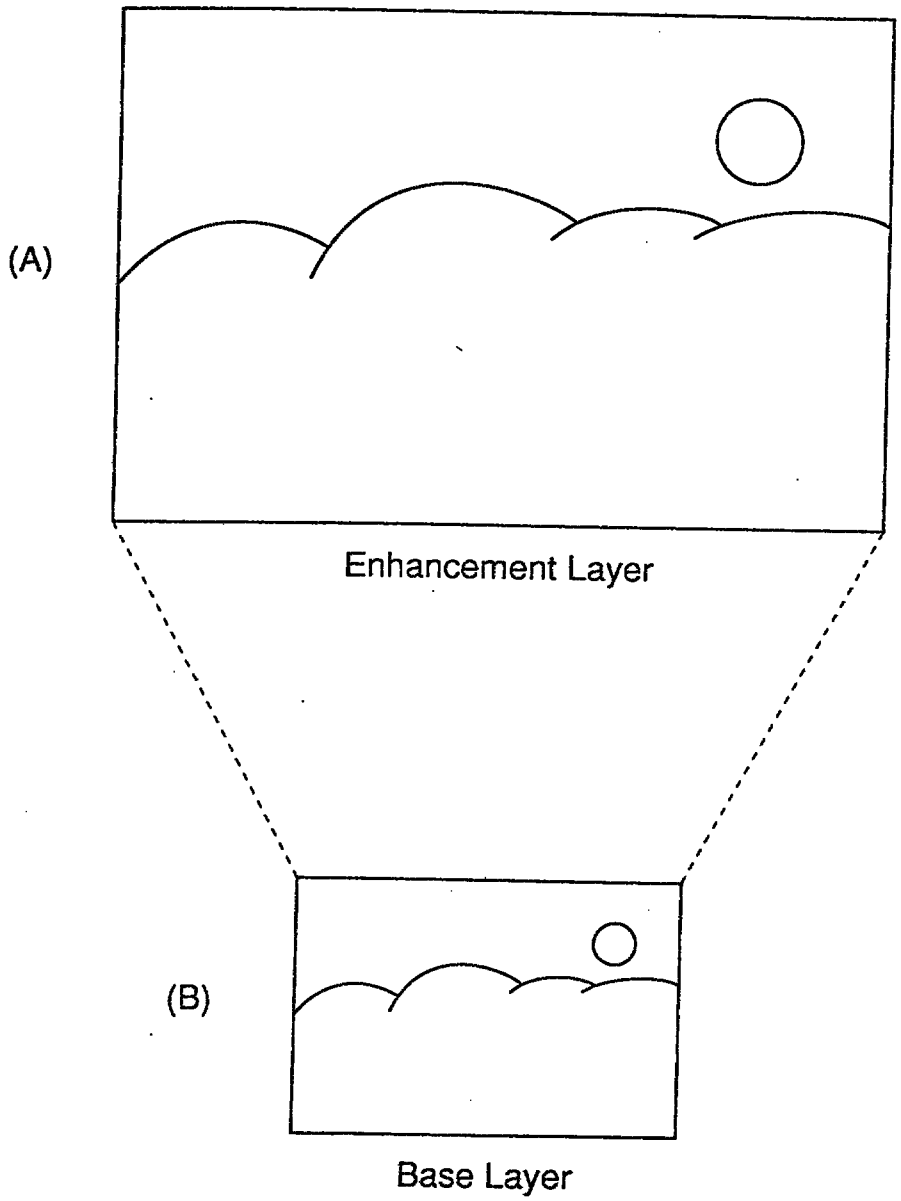


FIG. 8

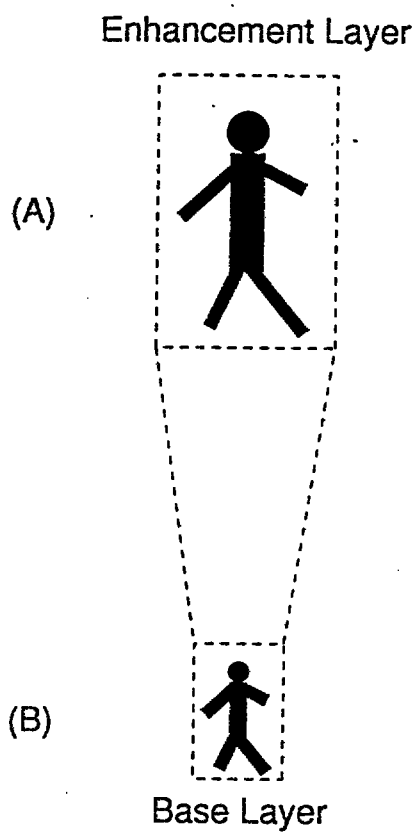
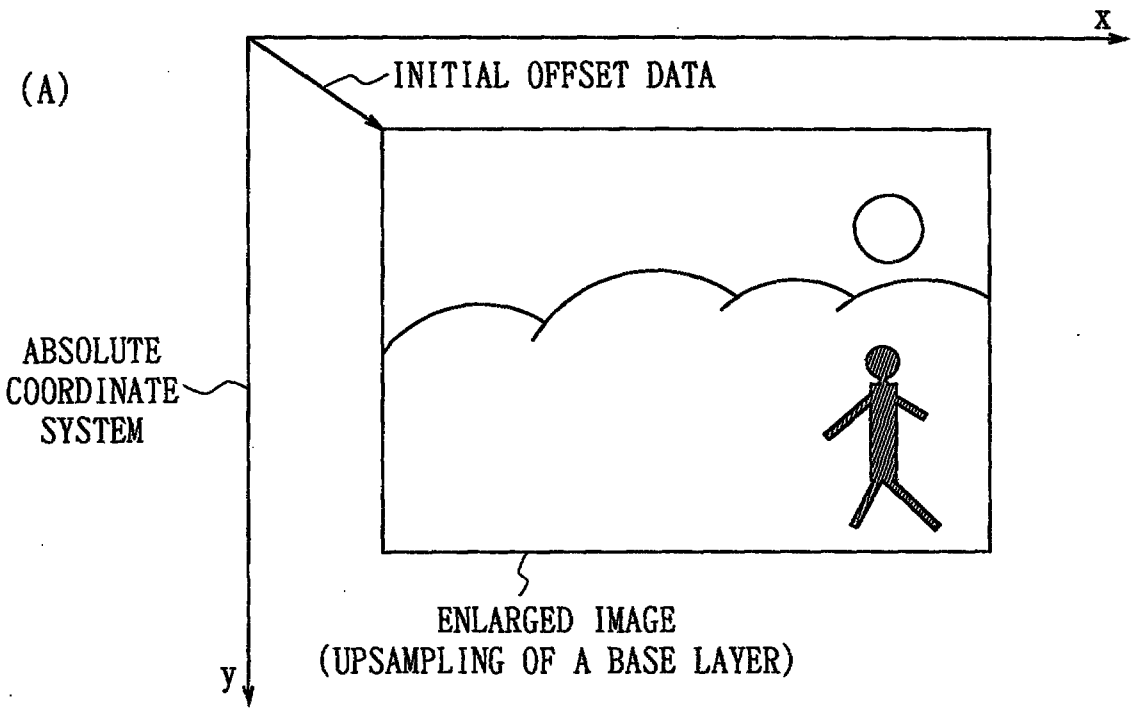


FIG. 9


(A)
x
INITIAL OFFSET DATA
ABSOLUTE
COORDINATE
SYSTEM
ENLARGED IMAGE
(UPSAMPLING OF A BASE LAYER)
y

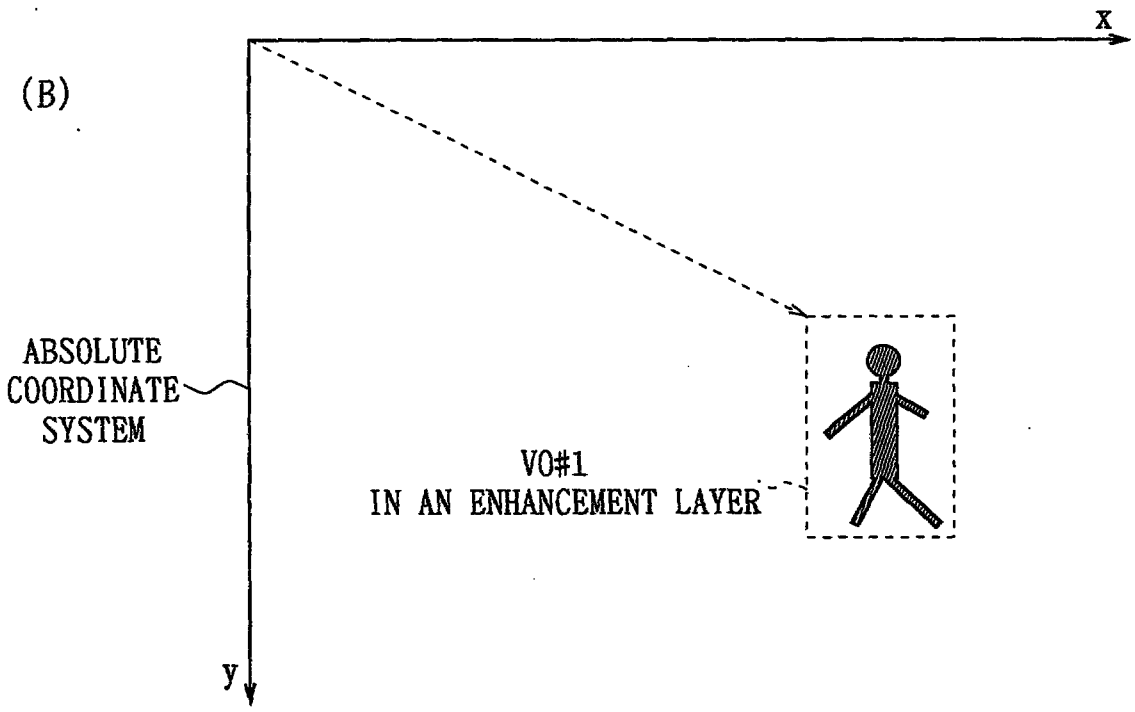

(B)
x
ABSOLUTE
COORDINATE
SYSTEM
VO#1
IN AN ENHANCEMENT LAYER
y

FIG. 10

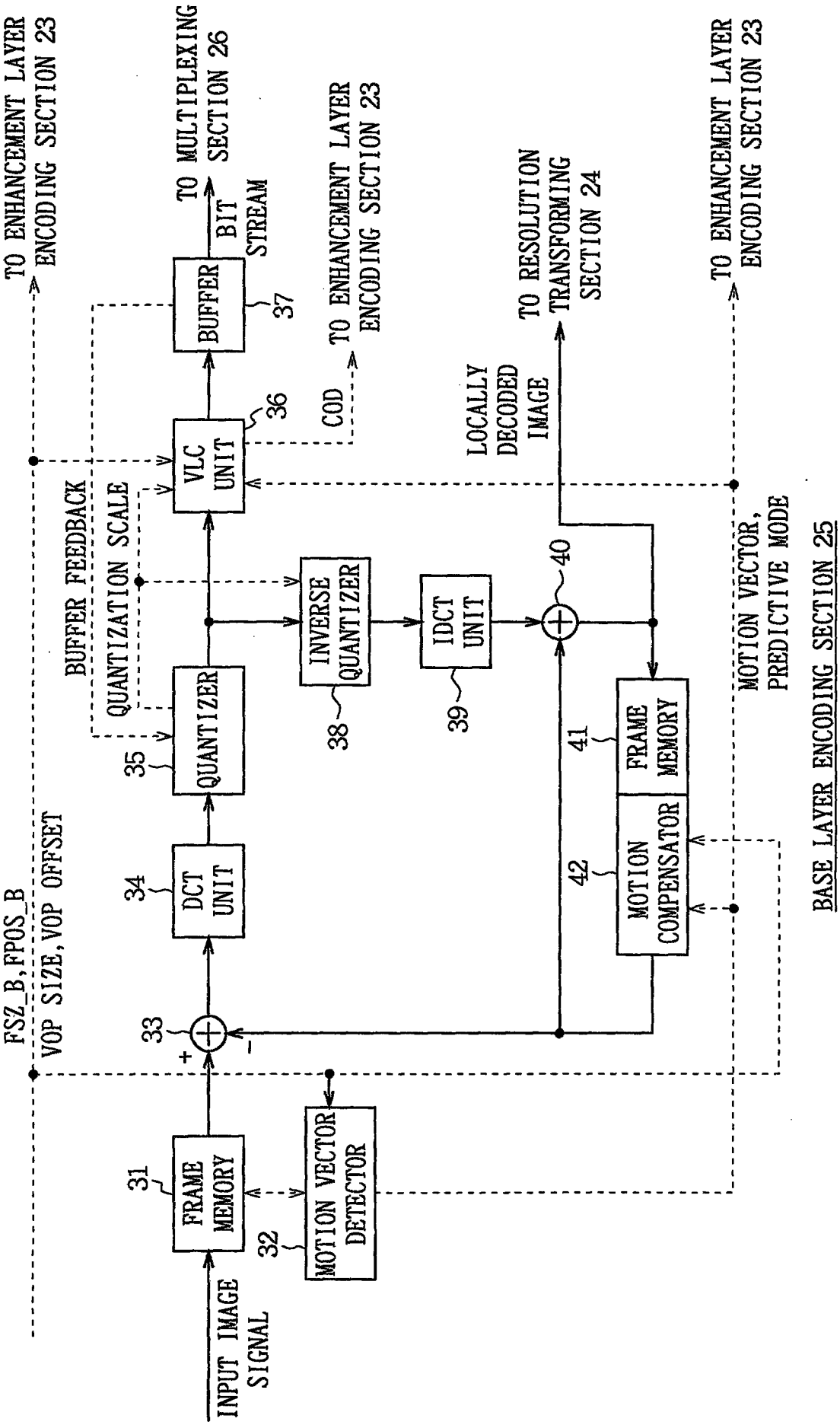

TO ENHANCEMENT LAYER ENCODING SECTION 23
TO MULTIPLEXING SECTION 26
BIT STREAM
BUFFER
37
VLC UNIT
36
COD
TO ENHANCEMENT LAYER ENCODING SECTION 23
BUFFER FEEDBACK
QUANTIZATION SCALE
LOCALLY DECODED IMAGE
TO RESOLUTION TRANSFORMING SECTION 24
QUANTIZER
35
INVERSE QUANTIZER
38
IDCT UNIT
39
40
FSZ_B,FPOS_B
VOP SIZE,VOP OFFSET
DCT UNIT
34
33
+
-
FRAME MEMORY
41
MOTION COMPENSATOR
42
TO ENHANCEMENT LAYER ENCODING SECTION 23
MOTION VECTOR, PREDICTIVE MODE
31
FRAME MEMORY
MOTION VECTOR DETECTOR
32
INPUT IMAGE SIGNAL
BASE LAYER ENCODING SECTION 25

FIG. 11

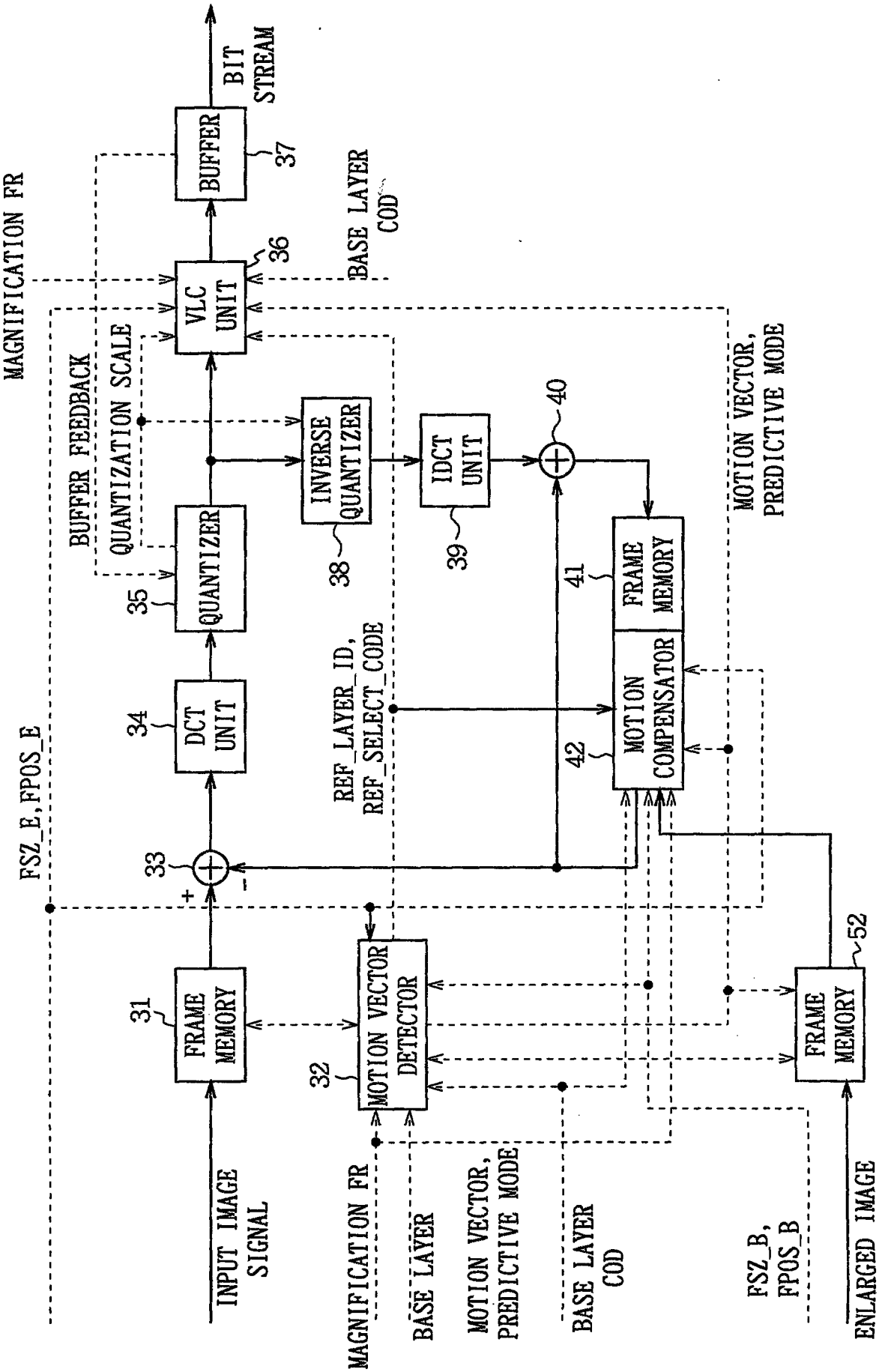

INPUT IMAGE SIGNAL
FRAME MEMORY
31
MOTION VECTOR DETECTOR
32
MAGNIFICATION FR
BASE LAYER
MOTION VECTOR, PREDICTIVE MODE
BASE LAYER COD
FSZ_B, FPOS_B
FRAME MEMORY
52
ENLARGED IMAGE
33
DCT UNIT
34
QUANTIZER
35
VLC UNIT
36
BUFFER
37
BIT STREAM
INVERSE QUANTIZER
38
IDCT UNIT
39
40
FRAME MEMORY
41
MOTION COMPENSATOR
42
REF_LAYER_ID, REF_SELECT_CODE
MOTION VECTOR, PREDICTIVE MODE
FSZ_E, FPOS_E
MAGNIFICATION FR
BUFFER FEEDBACK
QUANTIZATION SCALE
BASE LAYER COD

FIG. 12

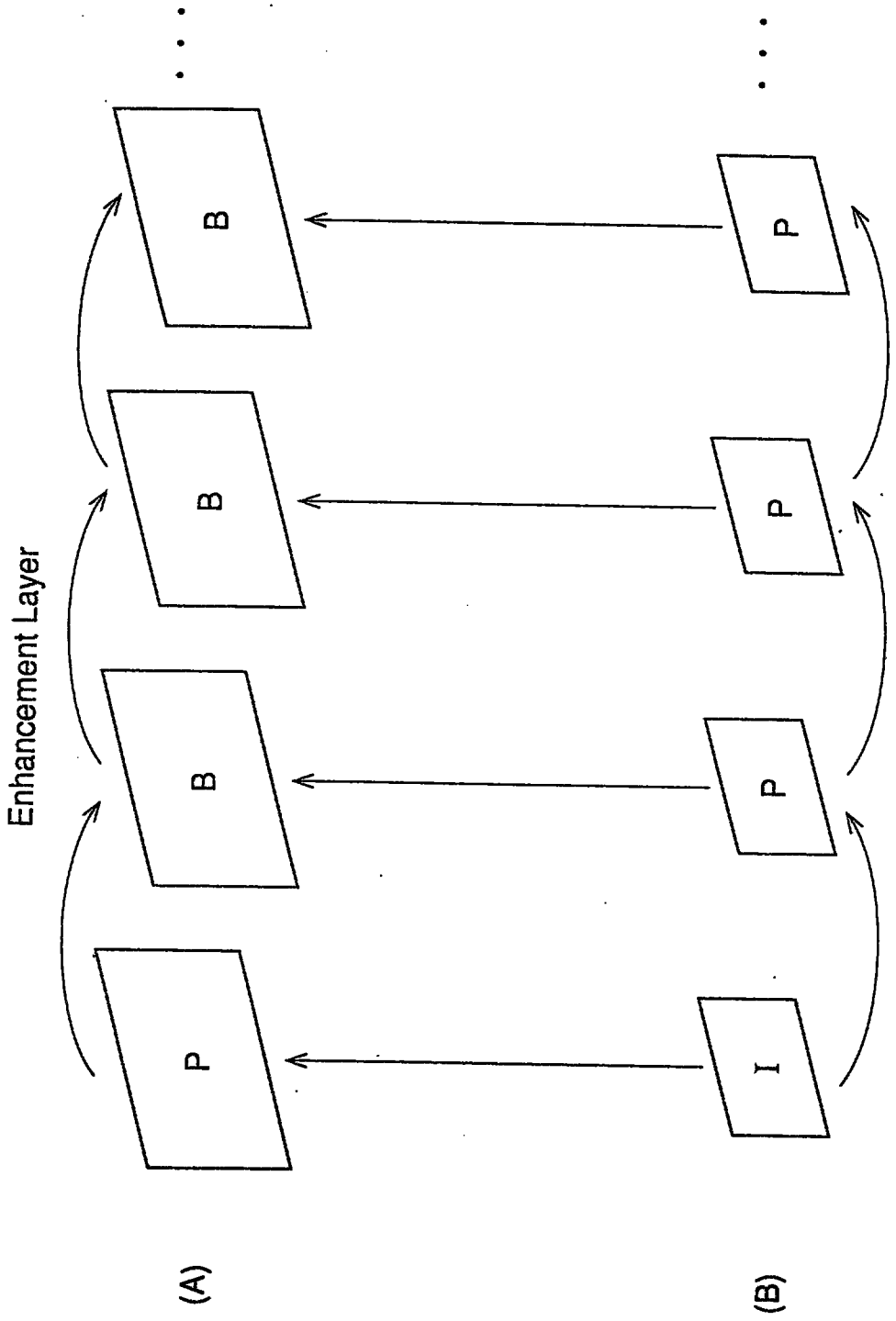


FIG. 13

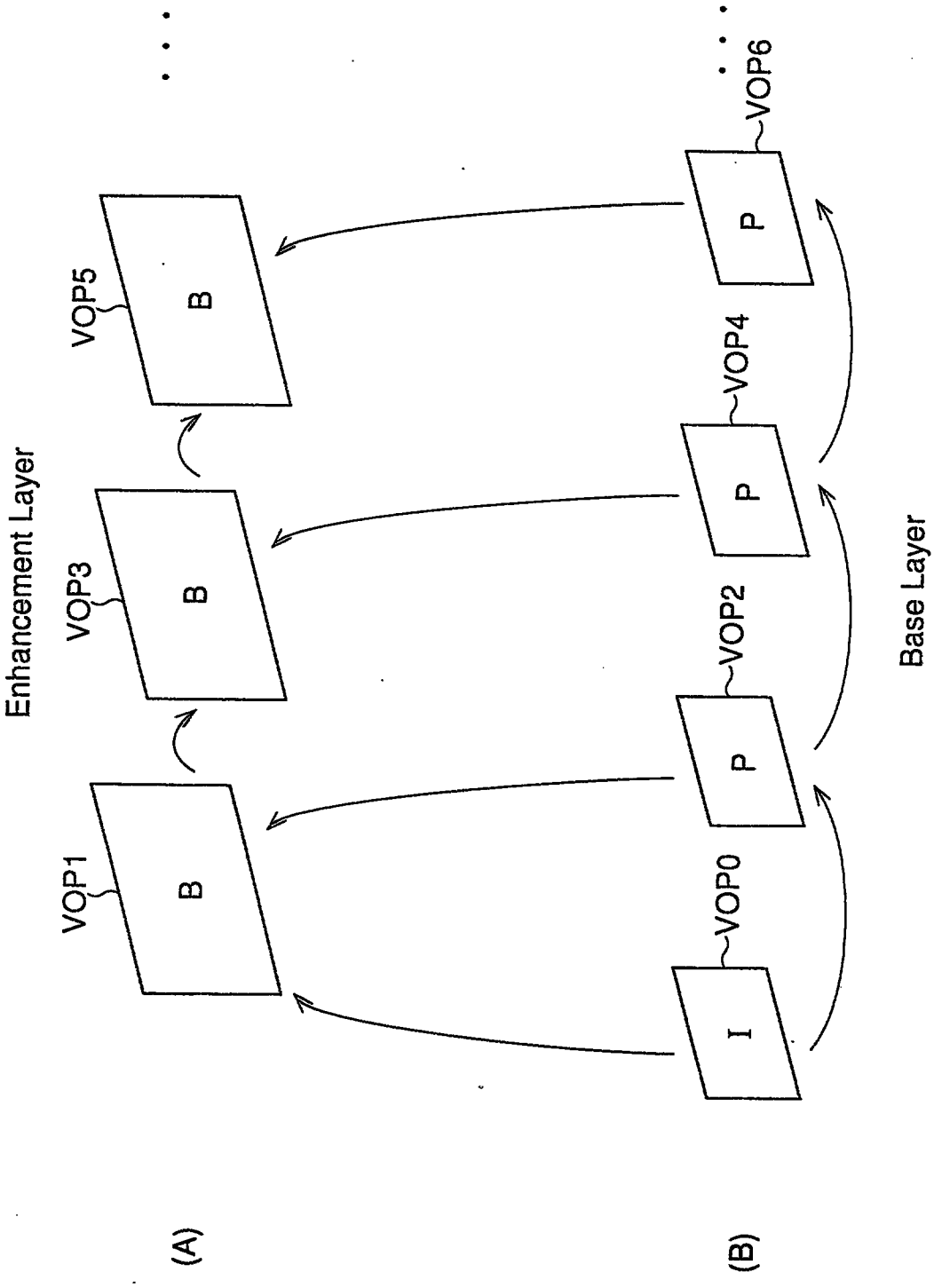

(A)
(B)
Enhancement Layer
Base Layer
VOP0
VOP1
VOP2
VOP3
VOP4
VOP5
VOP6
I
P
P
P
B
B
B
...
...

FIG. 14

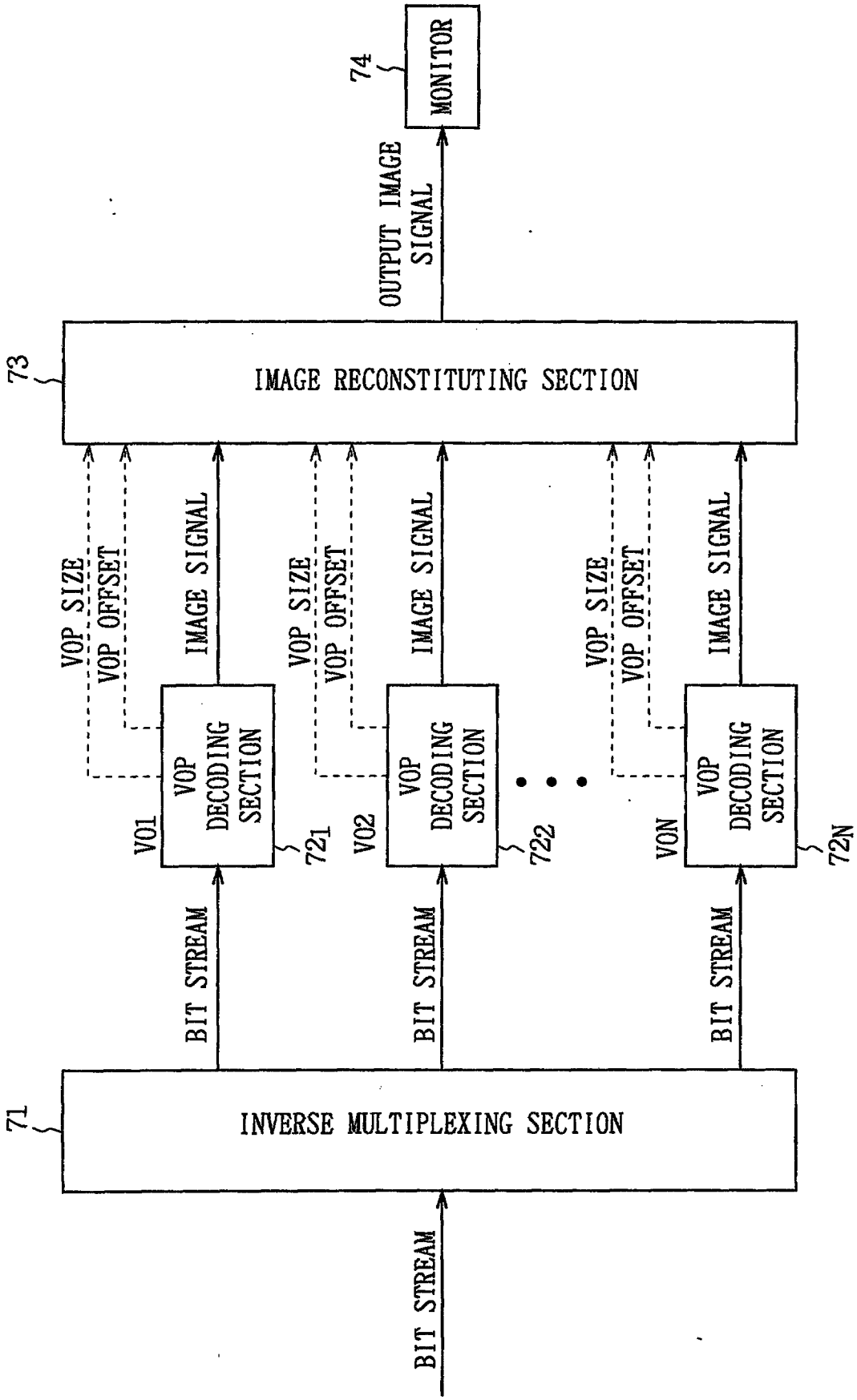
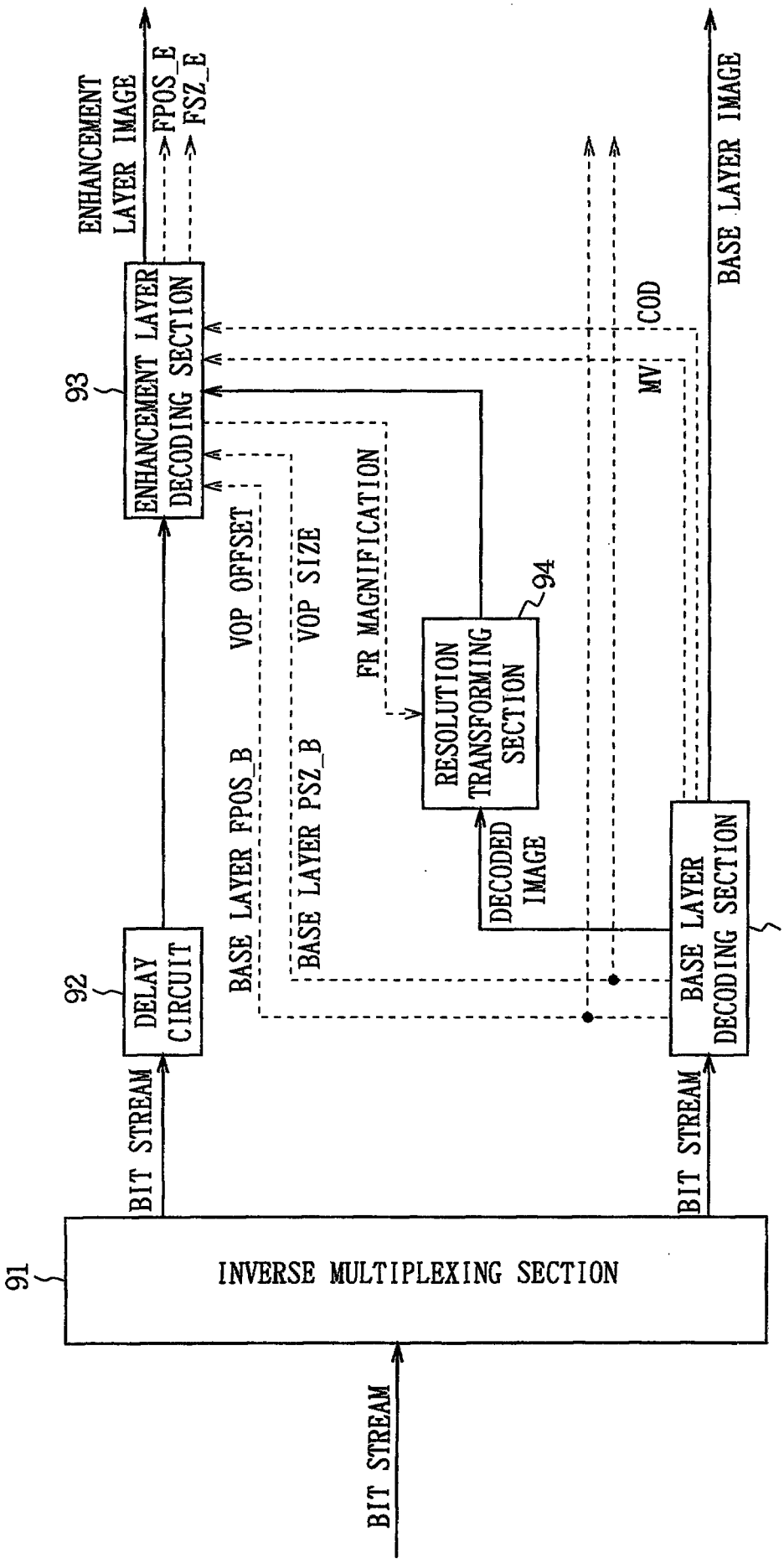
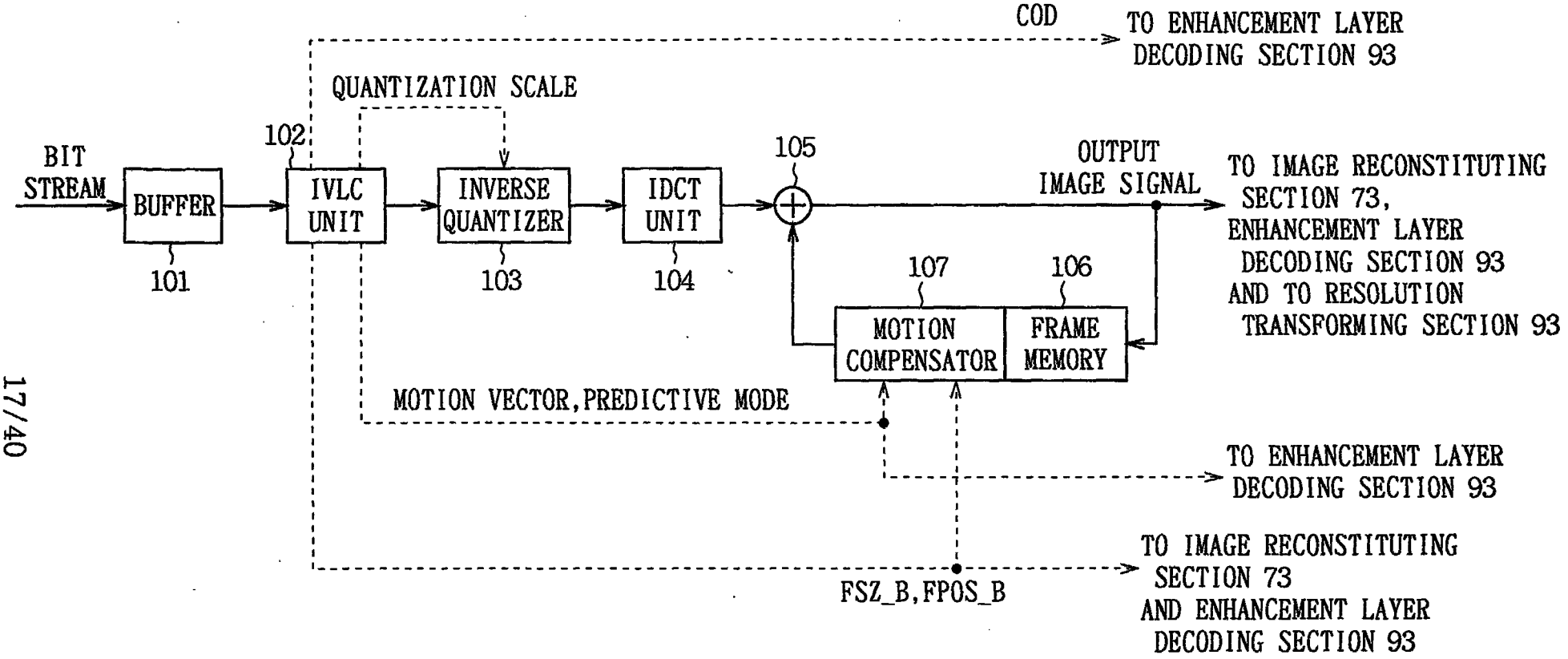

74
MONITOR
OUTPUT IMAGE SIGNAL
73
IMAGE RECONSTITUTING SECTION
VOP SIZE
VOP OFFSET
IMAGE SIGNAL
VOP SIZE
VOP OFFSET
IMAGE SIGNAL
VOP SIZE
VOP OFFSET
IMAGE SIGNAL
VO1
VOP DECODING SECTION
721
VO2
VOP DECODING SECTION
722
VON
VOP DECODING SECTION
72N
BIT STREAM
BIT STREAM
BIT STREAM
71
INVERSE MULTIPLEXING SECTION
BIT STREAM

FIG. 15



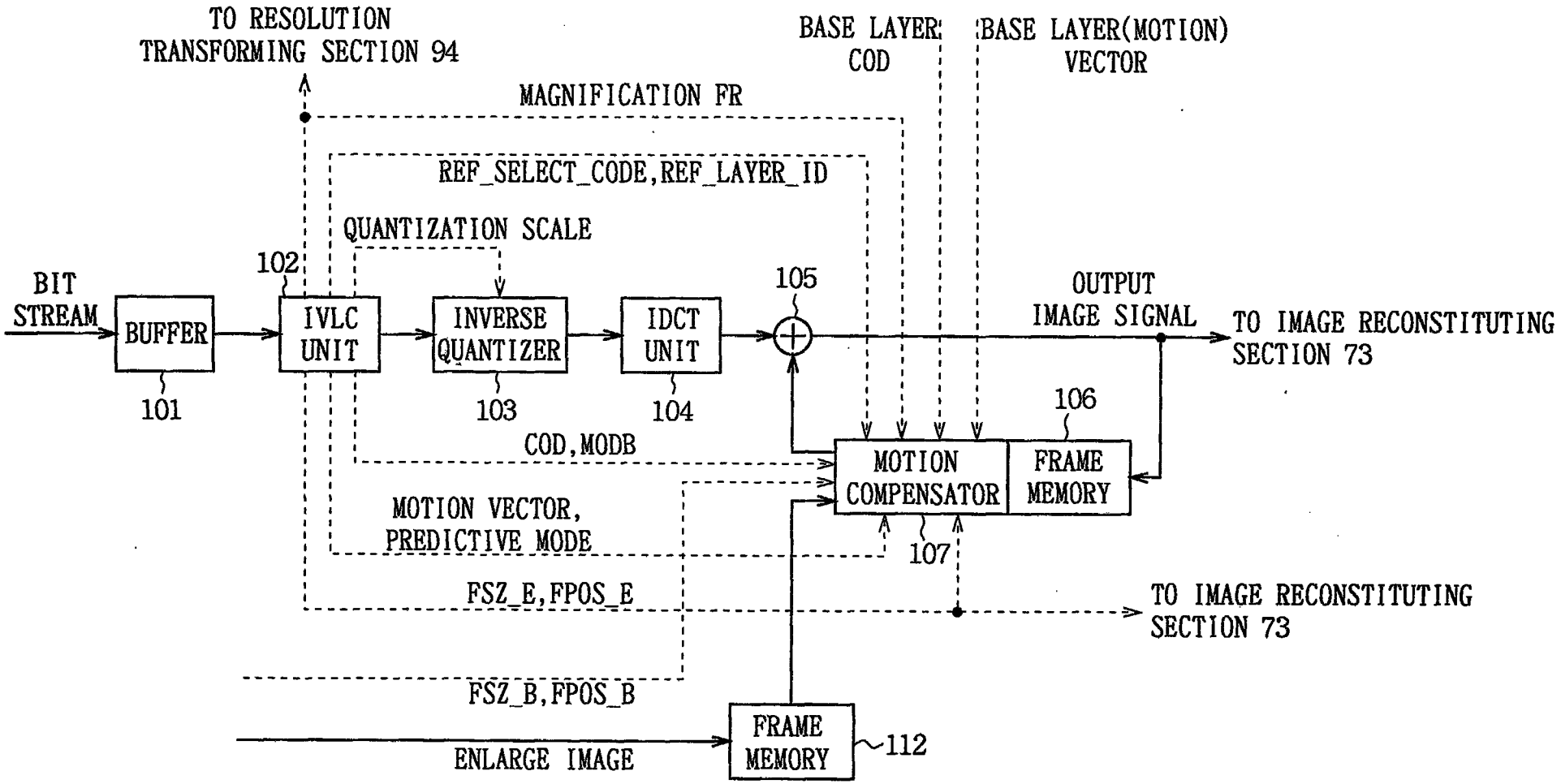
VOP DECODING SECTIONS 72n

FIG. 16



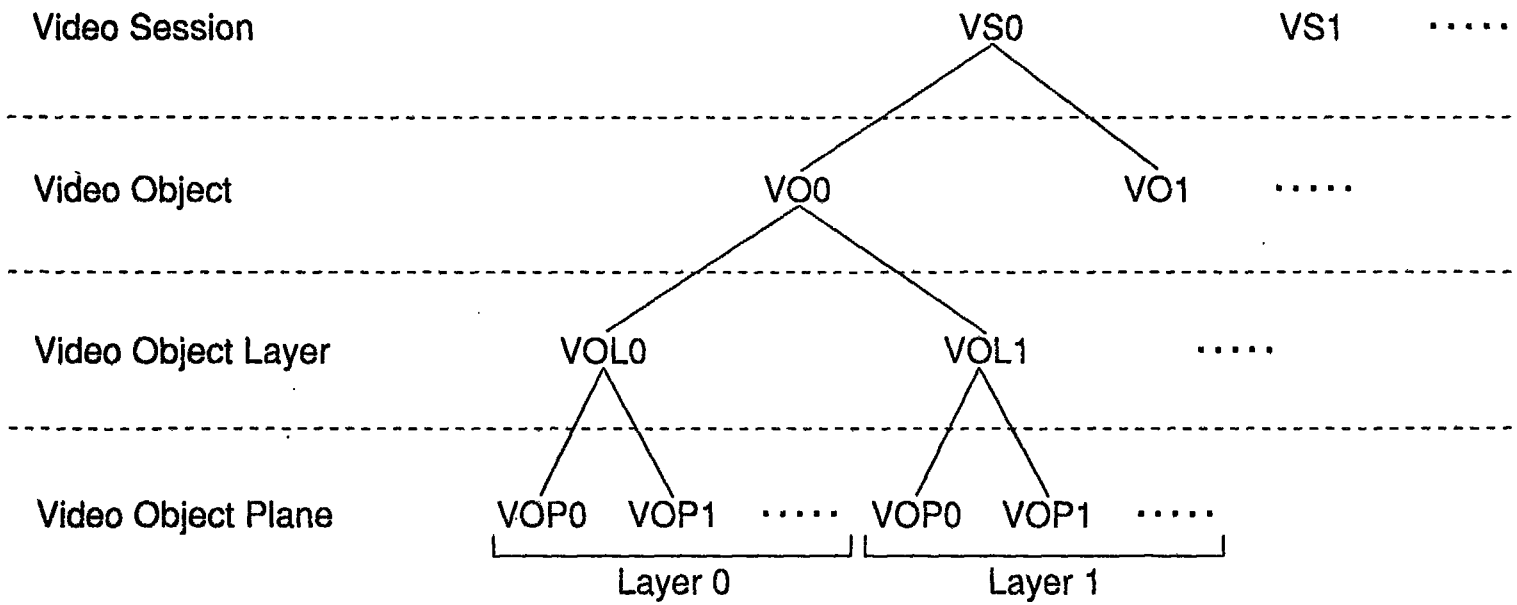
BASE LAYER DECODING SECTION 95

FIG. 17



ENHANCEMENT LAYER DECODING SECTION 93

FIG. 18



Hierarchy in the proposed video syntax

FIG. 19

Video Session Class

Syntax	No.of bits	Mnemonie
Video Seeion(){		
Video_session_start_code	sc+8=32	
do*{		
Video Object()		
}while(nextbits()==video_object_start_code)		
vedeo_session_end_code	sc+8=32	
}		

*concurrent loop solution to be provided by MSDL.

FIG. 20

Video Object

Syntax	No.of bits	Mnemonie
Video Object(){		
Video_object_start_code	sc+3=27	
Video_object_id	5	
do{		
Video Object Layer()		
}while(nextbits()==		
video_object_layer_start_code)		
}		

FIG. 21

Video Object Layer

Syntax	No.of bits	Mnemonie
Video Object Layer(){		
video_object_layer_start_code	sc+4=28	
video_object_layer_id	4	
video_object_layer_shape	2	
if(video_object_layer_shape== "00"){		
video_object_layer_width	10	
video_object_layer_height	10	
}		
video_object_layer_quant_type	1	
if(video_object_layer_quant_type){		
load_intra_quant_mat	1	
If(load_intra_quant_mat)		
intra_quant_mat[64]	8*64	
load_nonintra_quant_mat	1	
if(load_nonintra_quant_mat)		
nonintra_quant_mat[64]	8*64	
}		
intra_dcpred_disable	1	
video_object_layer_fcode_forward	2	
video_object_layer_fcode_backward	2	
separate_motion_shape_texture	1	
scalability	1	
if(scalability){		
ref_layer_id	4	
ref_layer_sampling_direc	1	
hor_sampling_factor_n	5	
hor_sampling_factor_m	5	
vert_sampling_factor_n	5	
vert_sampling_factor_m	5	
enhancement_type	1	
}		
do{		
video object Plane()		
}while(nextbits()==		
video_object_plane_plane_start_code)		
}		

FIG. 22

Video Object Plane

Syntax	No.of bits	Mnemonie
Video Object Plane(){		
VOP_start_code	sc+8=32	
do{		
modulo_time_base	1	
}while(modulo_time_base_!= "0")		
VOP_time_increment	10	
VOP_prediction_type	2	
if(video_object_layer_shape!= "0") {		
VOP_width	10	
VOP_height	10	
VOP_horizontal_mc_spatial_ref	10	
marker_bit	1	
VOP_vertical_mc_spatial_ref	10	
if(scalability && enhancement_type)		
background_composition	1	
}		
if(VOP_prediction_type== "10")		
VOP_dbquant	2	
else		
VOP_quant	5	
if(!scalability){		
if(!separate_motion_shape_texture)		
combined_motion_shape_texture_coding()		
else{		
do{		
first_MMR_code		
}while(count of macroblocks! =total number of macroblocks)	1-2	
motion_coding()		
shade_coding()		
texture_coding()		
}		
}		
else{		
if(background_composition){		
load_backward_shape	1	
if(load_backward_shape){		
bacward_shape_coding()		
load_forward_shape	1	
if(load_forward_shape)		
forward_shape_coding()		
}		
}		
ref_select_code		
if(VOP_prediction_type== "01"	2	
|| VOP_prediction_type== "10") {		
forward_temporal_ref	10	
if(VOP_prediction_type_== "10"){		
marker_bit	1	
backward_temporal_ref	10	
}		
}		
combined_motion_shape_texture_coding()		
}		
}		

FIG. 23

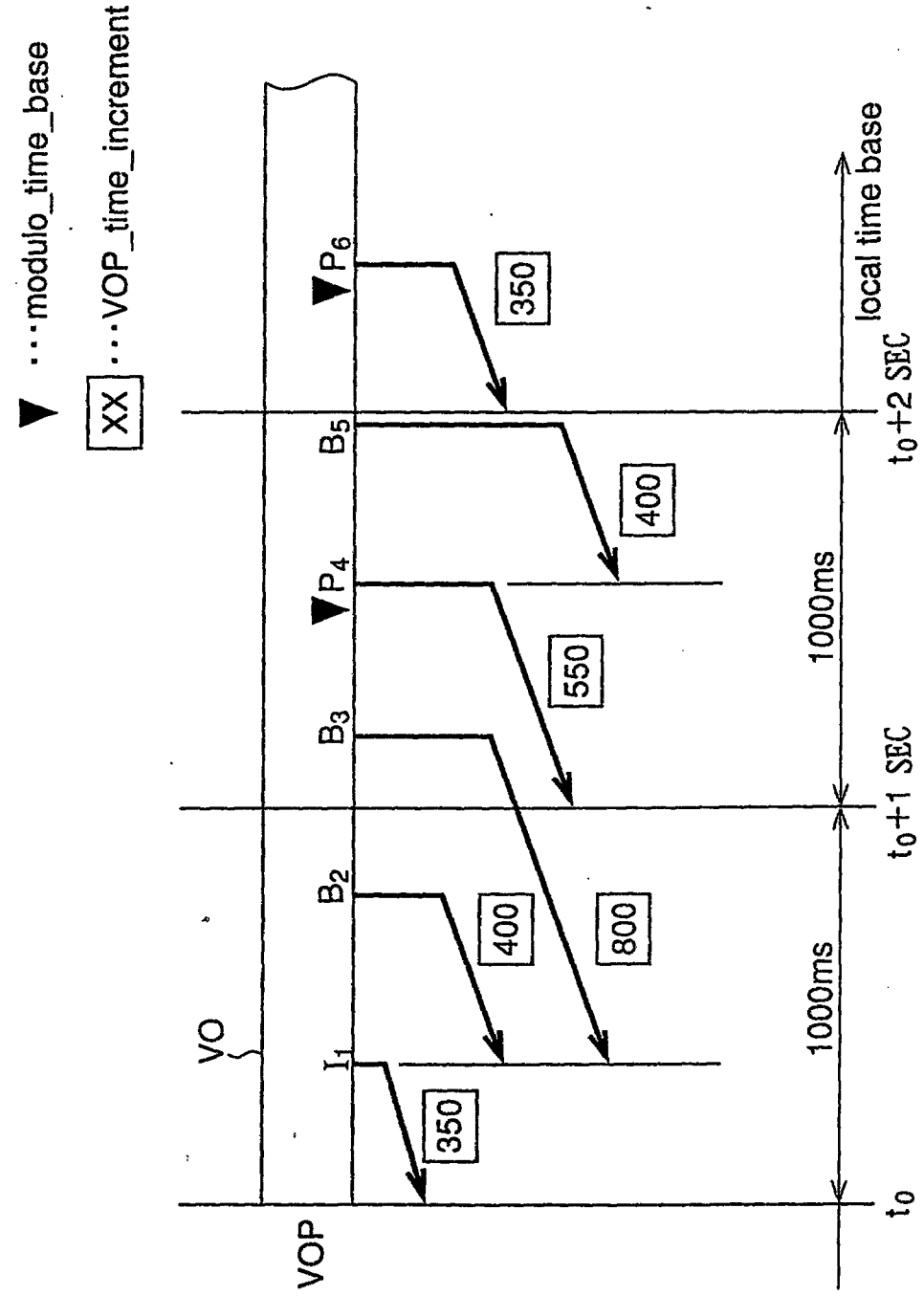

▼ ・・・modulo_time_base
XX ・・・VOP_time_increment
VO
VOP
I1
B2
B3
P4
B5
P6
350
400
800
550
400
350
1000ms
1000ms
local time base
t0
t0+1 SEC
t0+2 SEC

FIG. 24

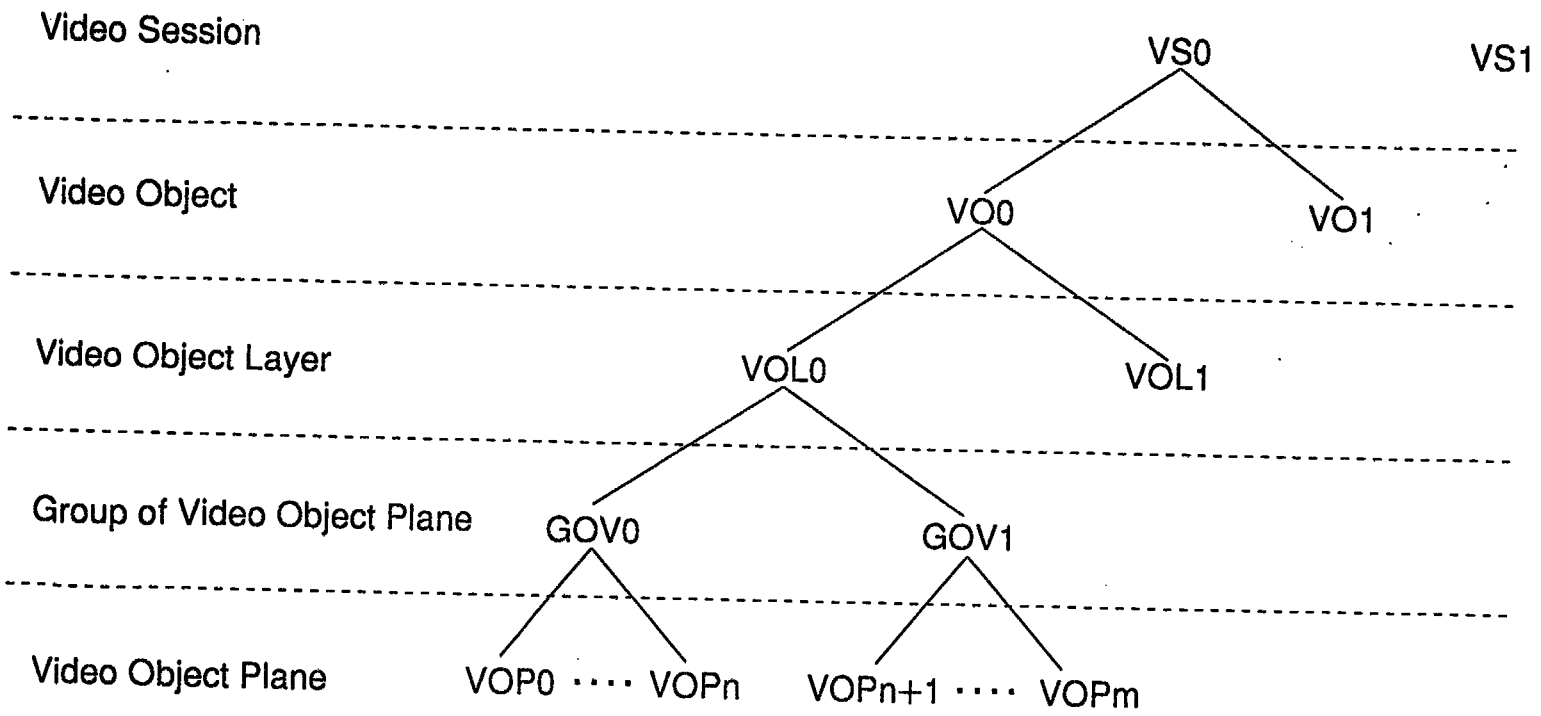

Video Session
VS0
VS1
Video Object
VO0
VO1
Video Object Layer
VOL0
VOL1
Group of Video Object Plane
GOV0
GOV1
Video Object Plane
VOP0 ···· VOPn
VOPn+1 ···· VOPm

FIG. 25

Group of Video Object Plane

Syntax	No.of bits	Mnemonie
group_of_plane(){		
group_start_code	32	bslbf
time_code	25	bslbf
closed_gop	1	uimsbf
broken_link	1	uimsbf
next_start_code()		
}		

FIG. 26

time_code

time_code	range of value	No. of bits	Mnemonic
drop_frame_flag		1	uimsbf
time_code_hours	0-23	5	uimsbf
time_code_minutes	0-59	6	uimsbf
marker_bit	1	1	bslbf
time_code_seconds	0-59	6	uimsbf
time_code_pictures	0-59	6	uimsbf

FIG. 27

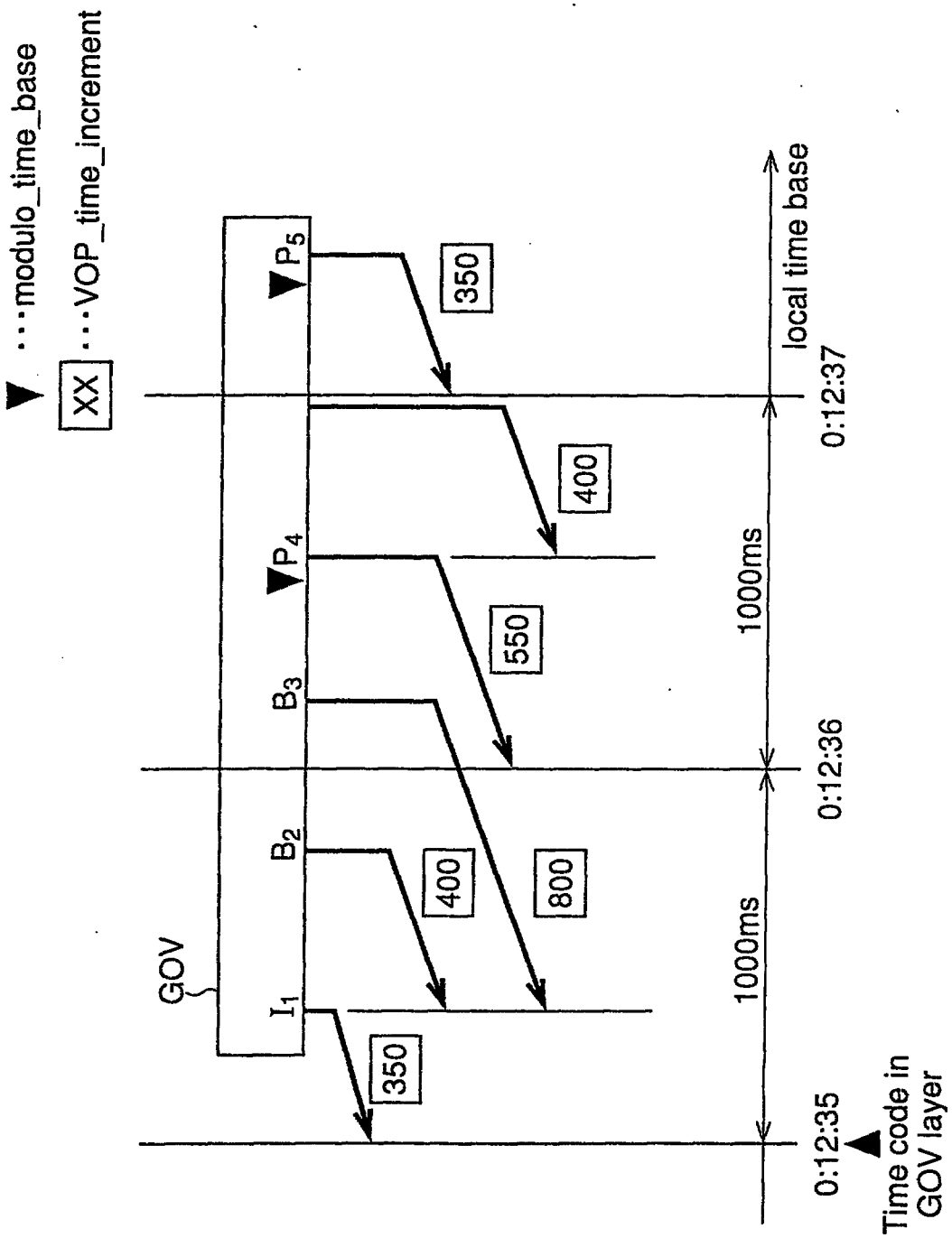

▼ ···modulo_time_base
XX ···VOP_time_increment
GOV
I1
B2
B3
P4
P5
350
400
800
550
400
350
1000ms
1000ms
local time base
0:12:37
0:12:36
0:12:35
Time code in GOV layer

FIG. 28

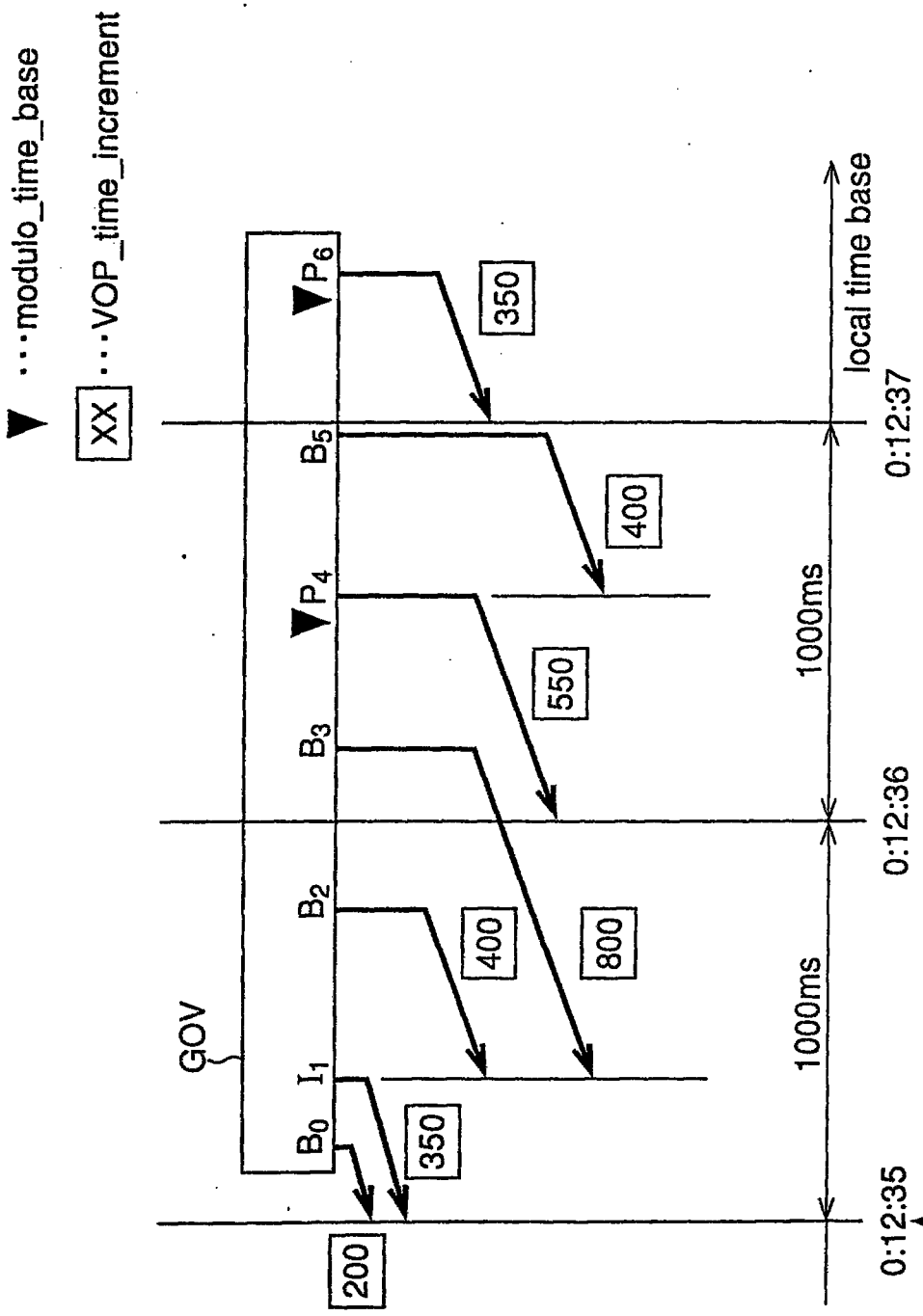

▼ ···modulo_time_base
XX ···VOP_time_increment
GOV
B_0
I_1
B_2
B_3
P_4
B_5
P_6
200
350
400
800
550
400
350
1000ms
1000ms
0:12:35
0:12:36
0:12:37
local time base
Time code in GOV layer

FIG. 29

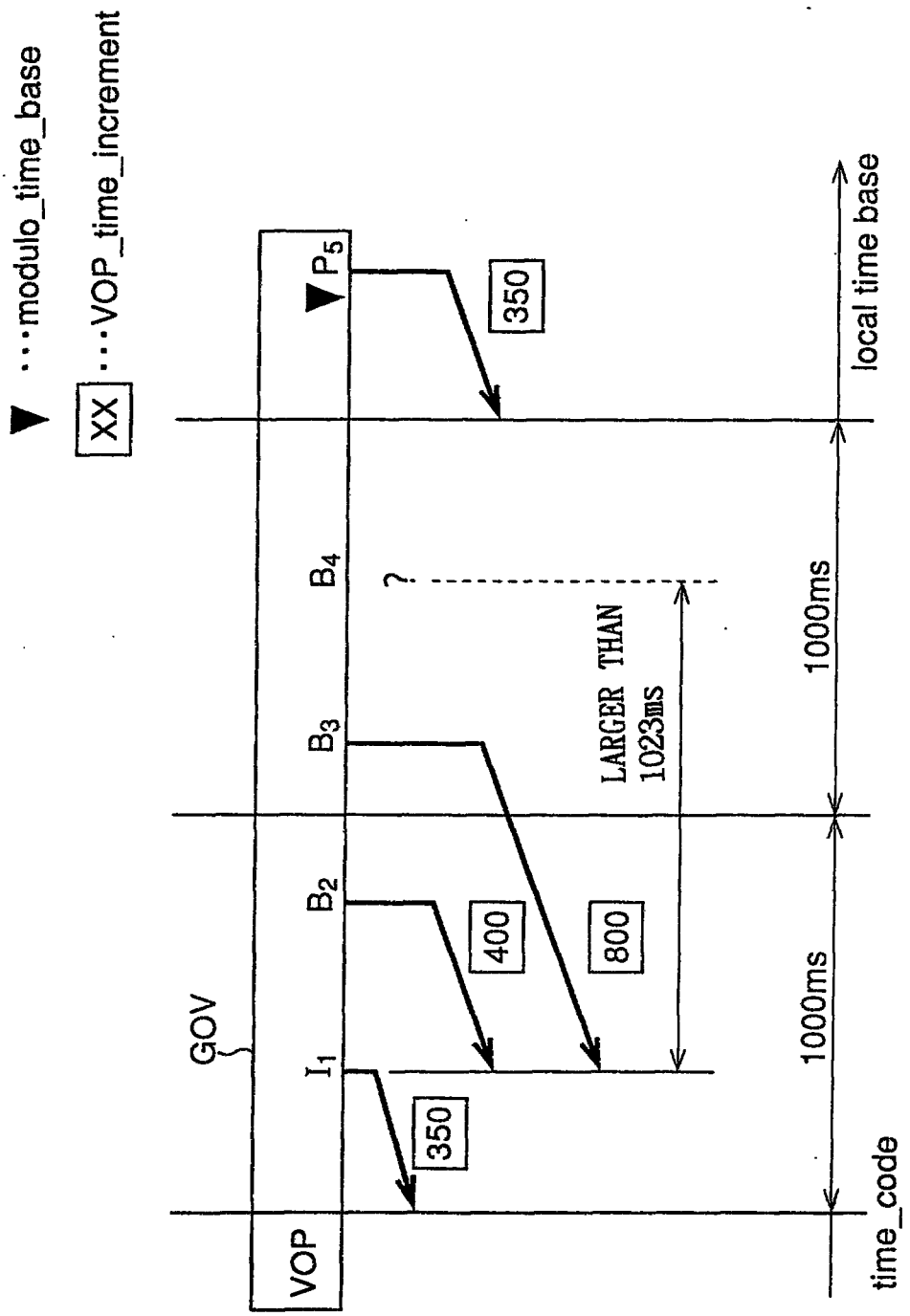

▼ ・・・modulo_time_base
XX ・・・VOP_time_increment
VOP
GOV
I_1
B_2
B_3
B_4
P_5
350
400
800
350
?
LARGER THAN
1023ms
1000ms
1000ms
local time base
time_code

FIG. 30

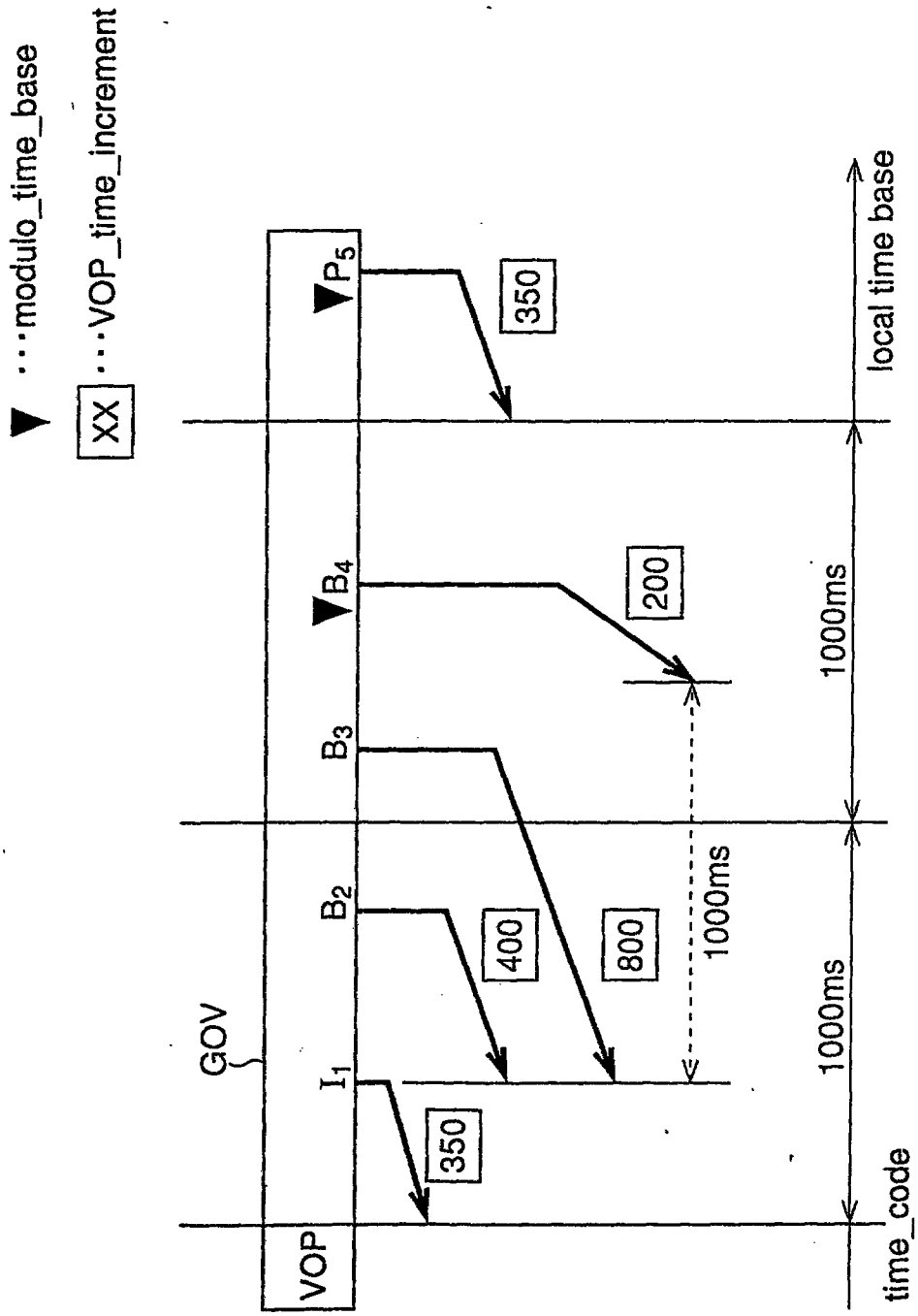

▼ ・・・modulo_time_base
XX ・・・VOP_time_increment
VOP
GOV
I1
B2
B3
B4
P5
350
400
800
200
350
1000ms
1000ms
1000ms
time_code
local time base

FIG. 31

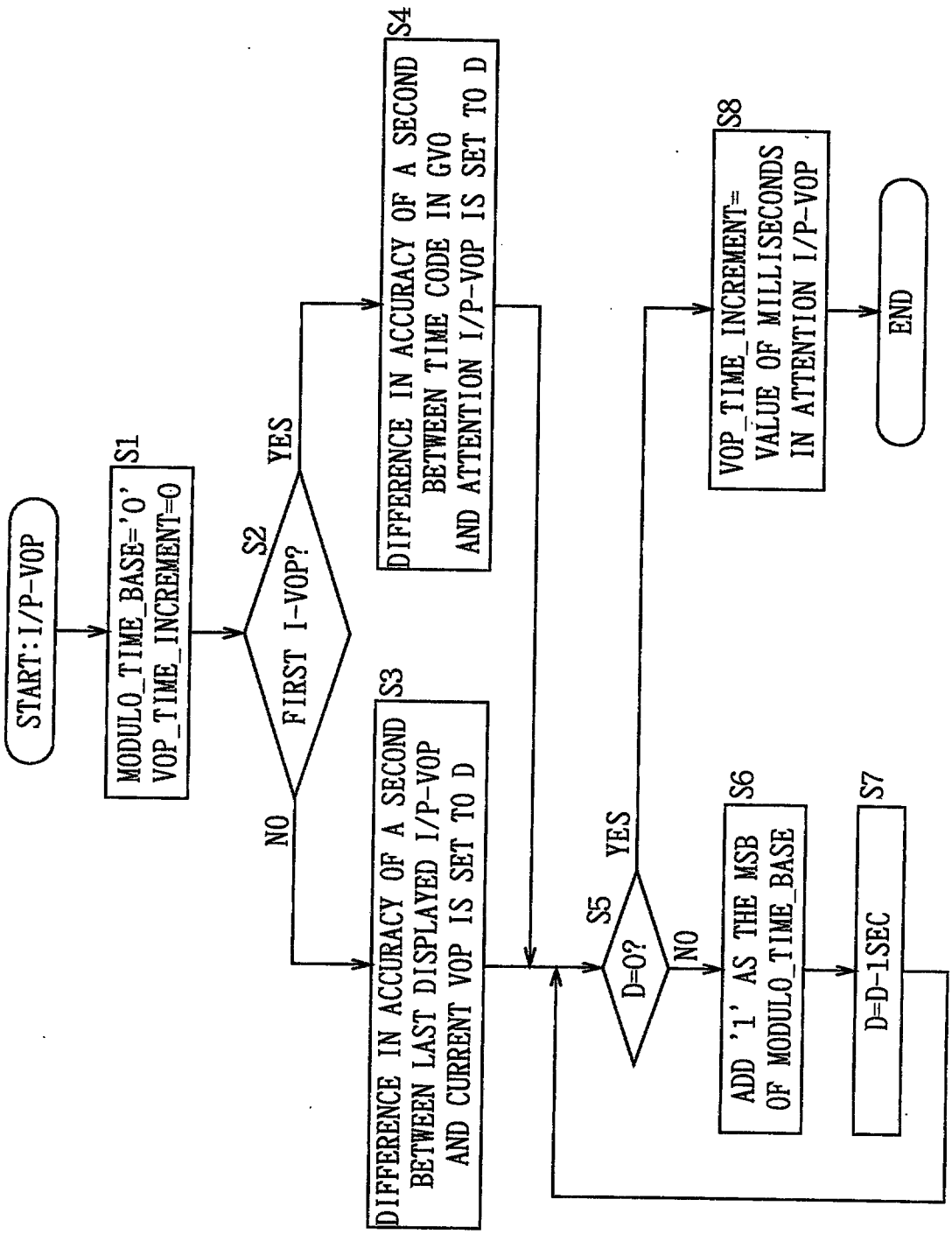

START: I/P-VOP
S1
MODULO_TIME_BASE='0'
VOP_TIME_INCREMENT=0
S2
FIRST I-VOP?
YES
NO
S4
DIFFERENCE IN ACCURACY OF A SECOND BETWEEN TIME CODE IN GVO AND ATTENTION I/P-VOP IS SET TO D
S3
DIFFERENCE IN ACCURACY OF A SECOND BETWEEN LAST DISPLAYED I/P-VOP AND CURRENT VOP IS SET TO D
S5
D=0?
YES
NO
S6
ADD '1' AS THE MSB OF MODULO_TIME_BASE
S7
D=D-1SEC
S8
VOP_TIME_INCREMENT= VALUE OF MILLISECONDS IN ATTENTION I/P-VOP
END

FIG. 32

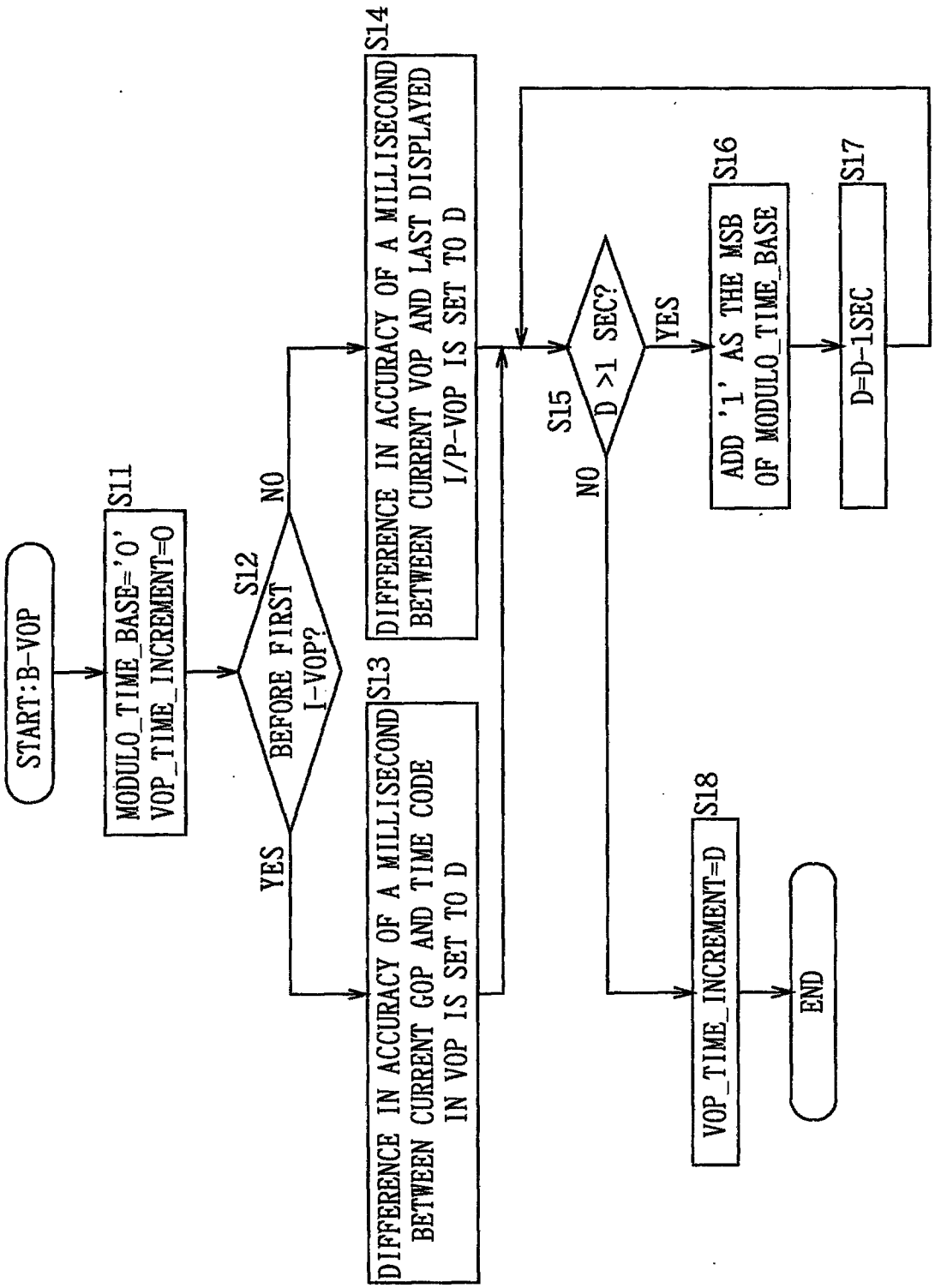

START:B-VOP
MODULO_TIME_BASE='0'
VOP_TIME_INCREMENT=0
S11
S12
BEFORE FIRST I-VOP?
YES
NO
DIFFERENCE IN ACCURACY OF A MILLISECOND BETWEEN CURRENT GOP AND TIME CODE IN VOP IS SET TO D
S13
DIFFERENCE IN ACCURACY OF A MILLISECOND BETWEEN CURRENT VOP AND LAST DISPLAYED I/P-VOP IS SET TO D
S14
S15
D >1 SEC?
YES
NO
ADD '1' AS THE MSB OF MODULO_TIME_BASE
S16
D=D-1SEC
S17
VOP_TIME_INCREMENT=D
S18
END

FIG. 33

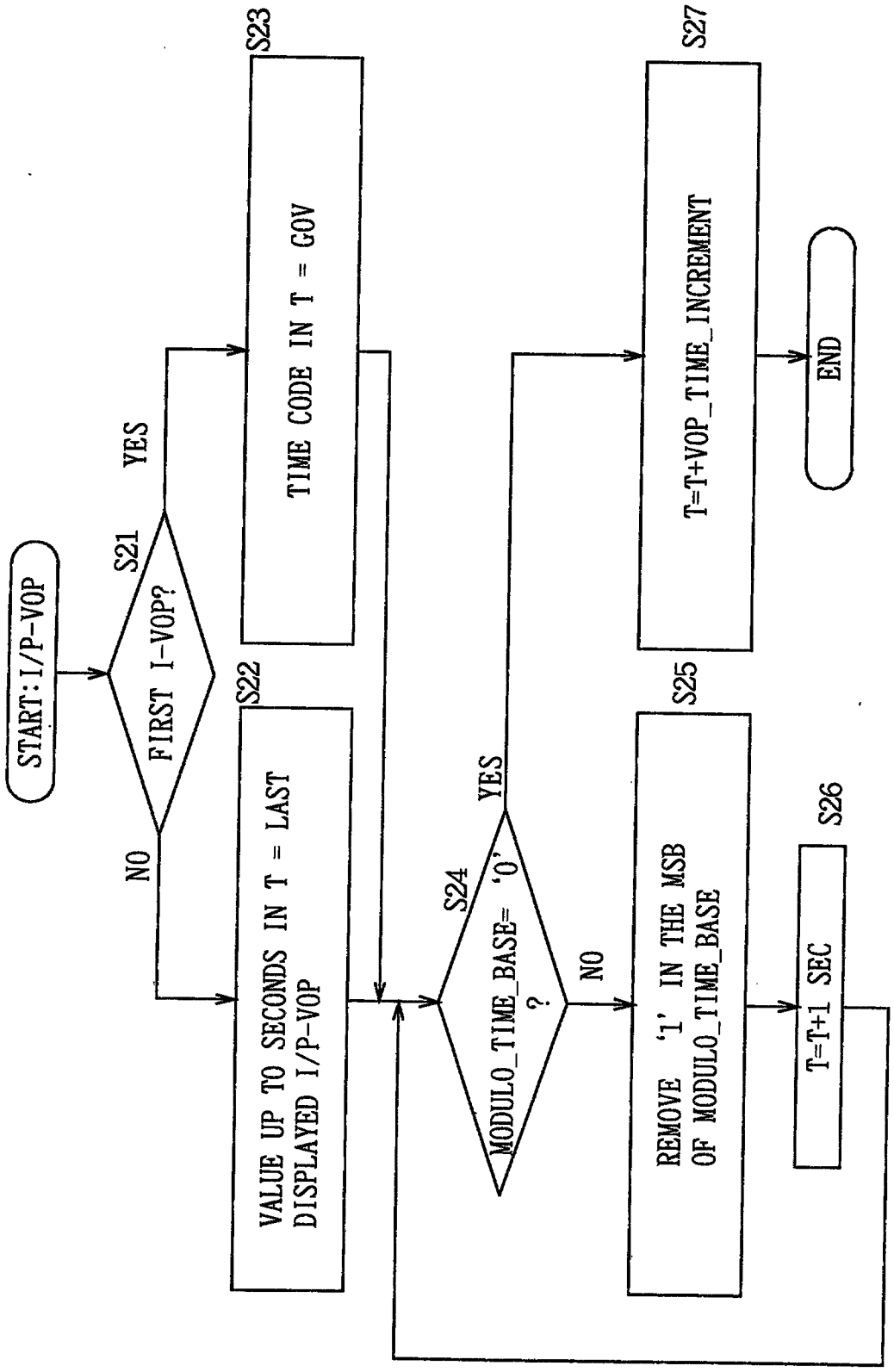

START: I/P-VOP
S21
FIRST I-VOP?
YES
NO
S23
TIME CODE IN T = GOV
S22
VALUE UP TO SECONDS IN T = LAST DISPLAYED I/P-VOP
S24
MODULO_TIME_BASE= '0' ?
YES
NO
S27
T=T+VOP_TIME_INCREMENT
END
S25
REMOVE '1' IN THE MSB OF MODULO_TIME_BASE
S26
T=T+1 SEC

FIG. 34

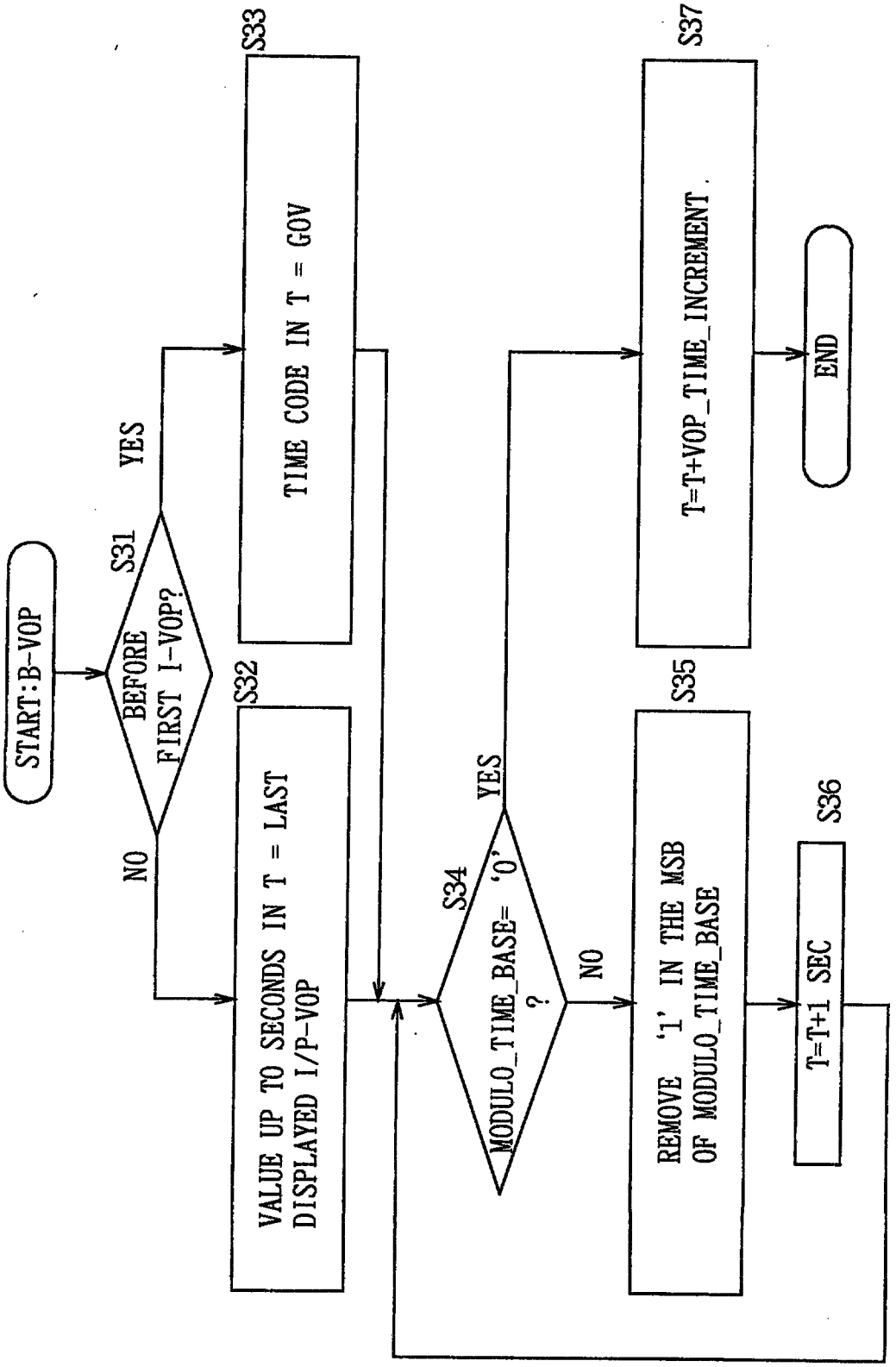

START:B-VOP
S31
BEFORE FIRST I-VOP?
YES
NO
S33
TIME CODE IN T = GOV
S32
VALUE UP TO SECONDS IN T = LAST DISPLAYED I/P-VOP
S34
MODULO_TIME_BASE= '0' ?
YES
NO
S37
T=T+VOP_TIME_INCREMENT
S35
REMOVE '1' IN THE MSB OF MODULO_TIME_BASE
S36
T=T+1 SEC
END

FIG. 35

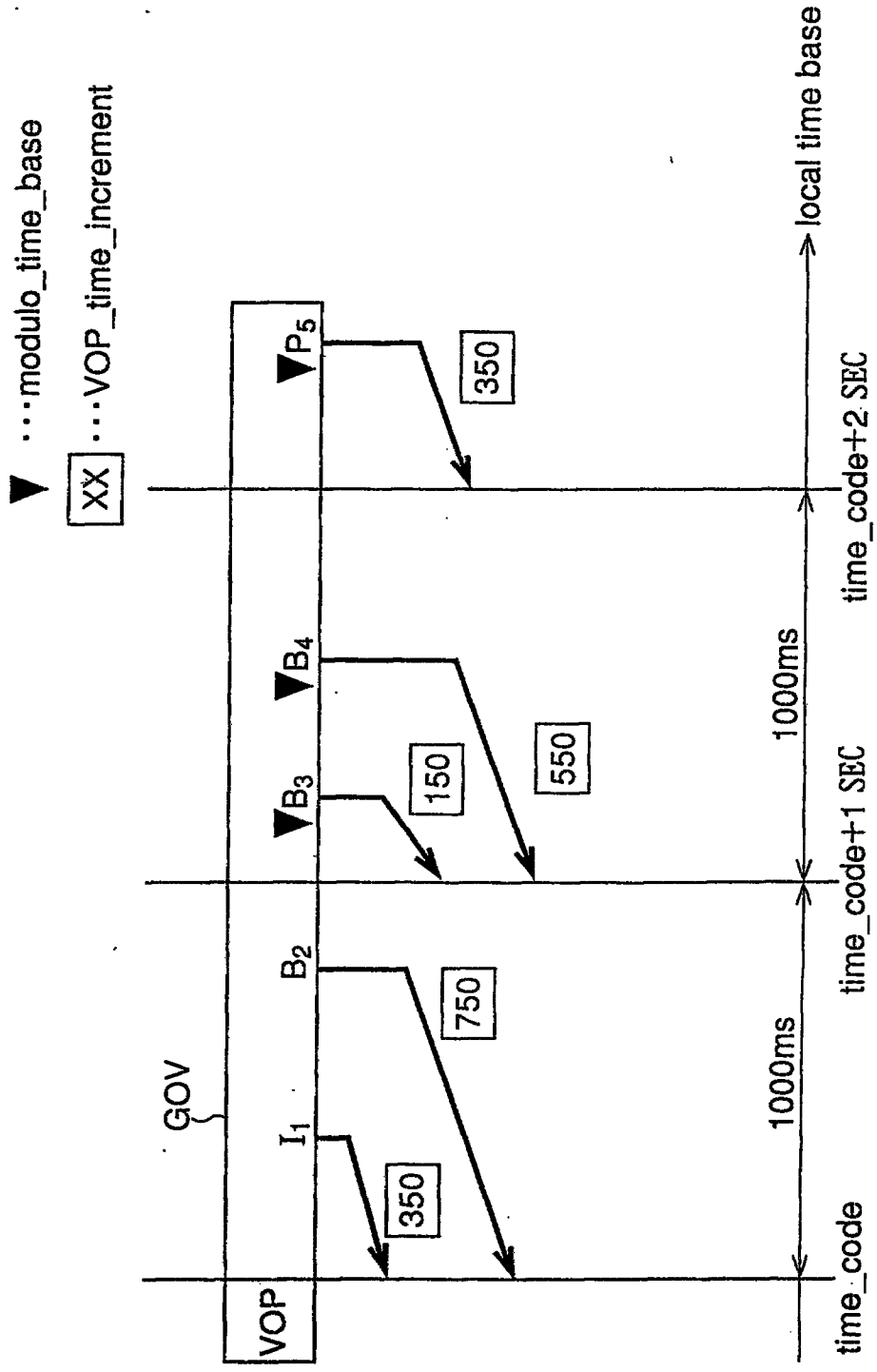

▼ ···modulo_time_base
XX ···VOP_time_increment
local time base
time_code
time_code+1 SEC
time_code+2 SEC
1000ms
1000ms
VOP
GOV
I1
B2
B3
B4
P5
350
750
150
550
350

FIG. 36

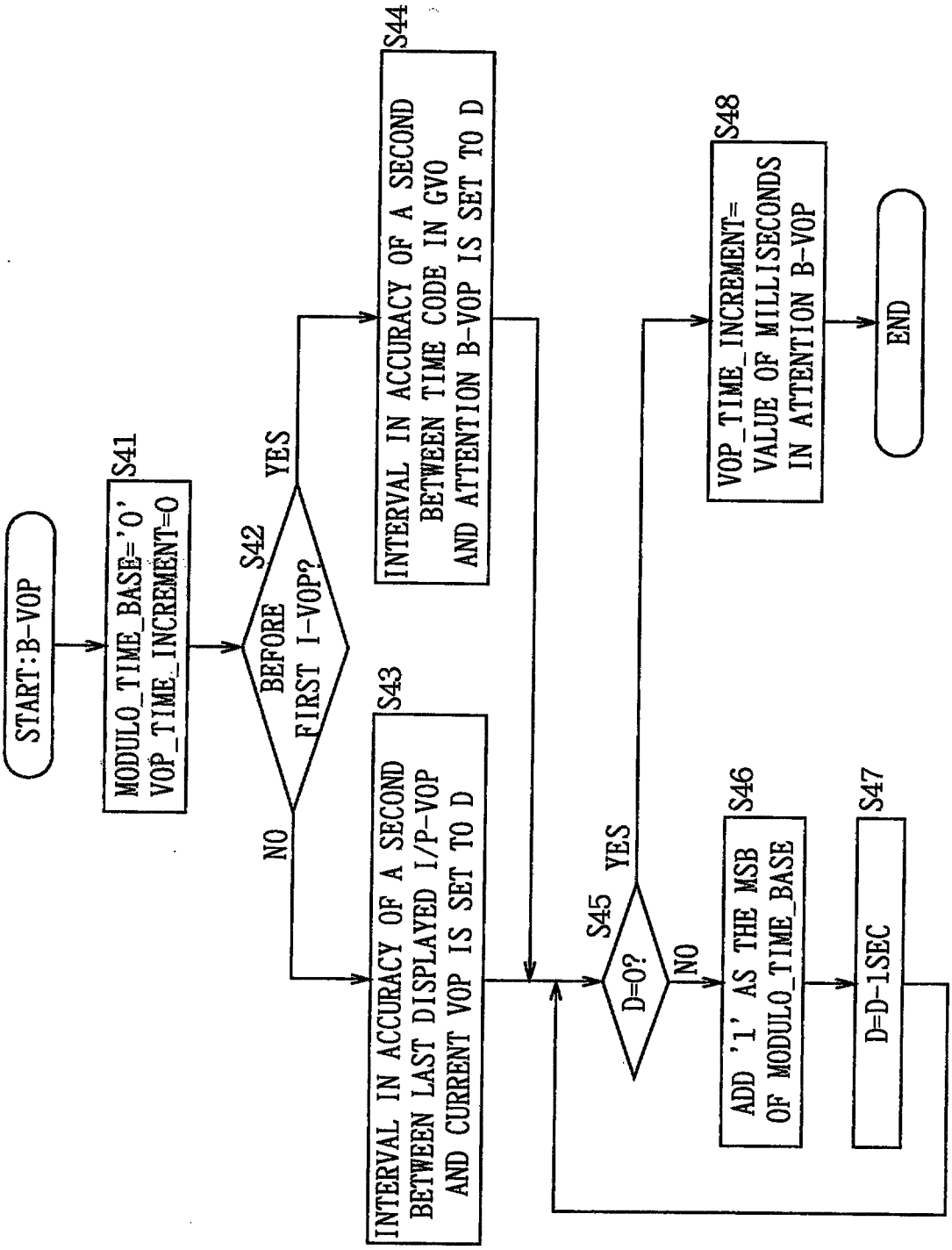

START:B-VOP
MODULO_TIME_BASE='0'
VOP_TIME_INCREMENT=0
S41
S42
BEFORE FIRST I-VOP?
YES
NO
INTERVAL IN ACCURACY OF A SECOND BETWEEN TIME CODE IN GVO AND ATTENTION B-VOP IS SET TO D
S44
INTERVAL IN ACCURACY OF A SECOND BETWEEN LAST DISPLAYED I/P-VOP AND CURRENT VOP IS SET TO D
S43
S45
D=0?
YES
NO
VOP_TIME_INCREMENT= VALUE OF MILLISECONDS IN ATTENTION B-VOP
S48
ADD '1' AS THE MSB OF MODULO_TIME_BASE
S46
D=D-1SEC
S47
END

FIG. 37

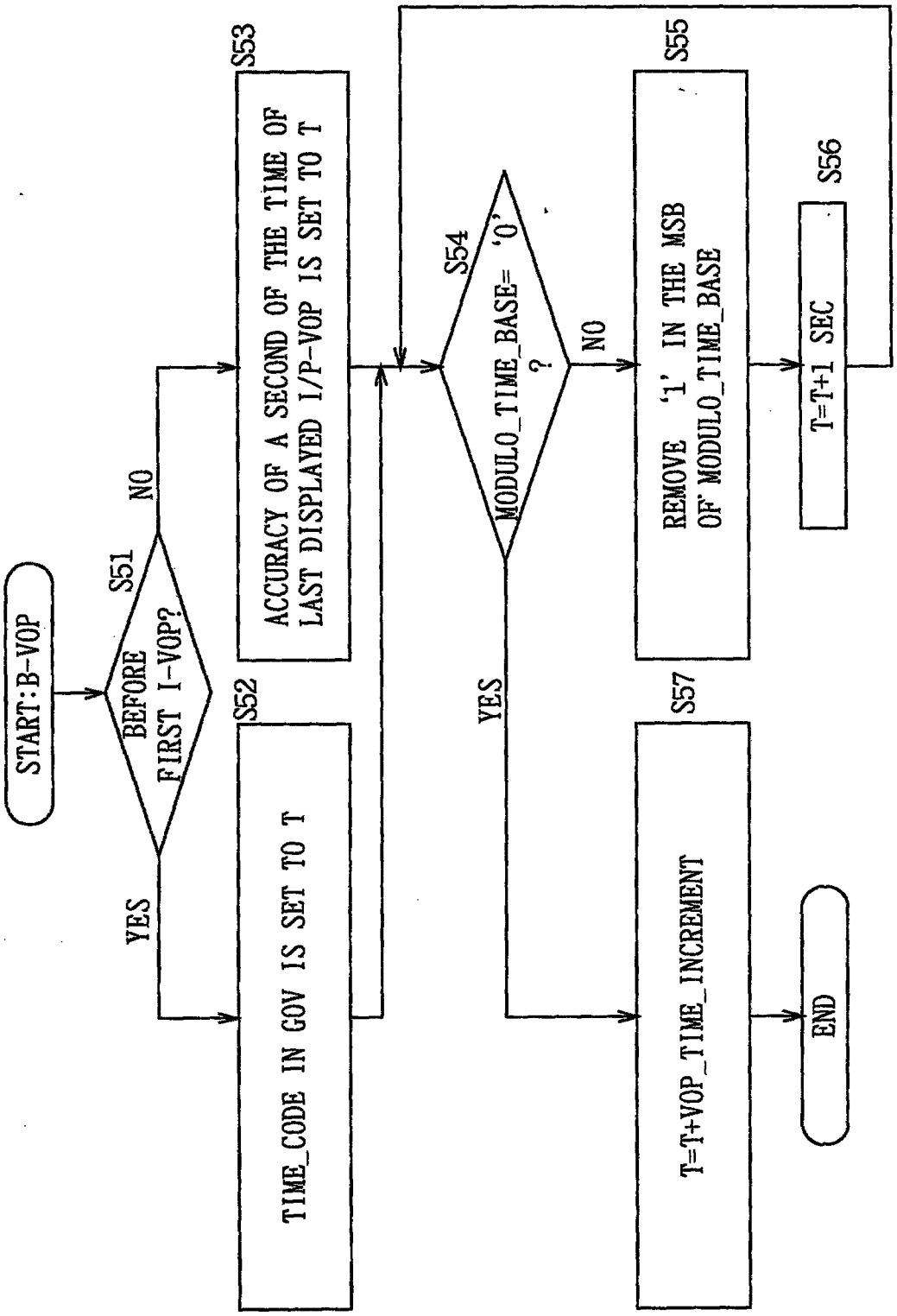

START:B-VOP
S51
BEFORE FIRST I-VOP?
NO
YES
S52
TIME_CODE IN GOV IS SET TO T
S53
ACCURACY OF A SECOND OF THE TIME OF LAST DISPLAYED I/P-VOP IS SET TO T
S54
MODULO_TIME_BASE= '0' ?
NO
YES
S55
REMOVE '1' IN THE MSB OF MODULO_TIME_BASE
S56
T=T+1 SEC
S57
T=T+VOP_TIME_INCREMENT
END

FIG. 38

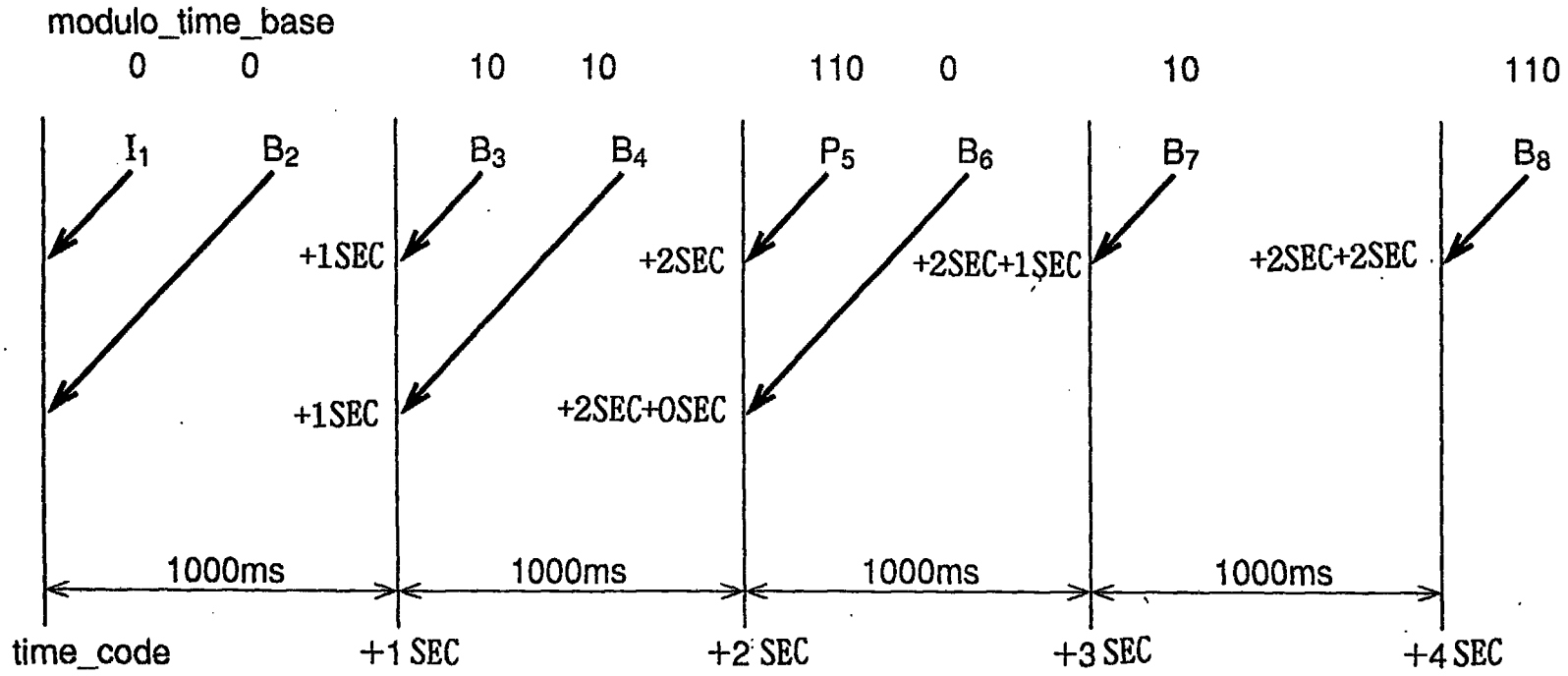

modulo_time_base
0
0
10
10
110
0
10
110
I1
B2
B3
B4
P5
B6
B7
B8
+1SEC
+1SEC
+2SEC
+2SEC+0SEC
+2SEC+1SEC
+2SEC+2SEC
1000ms
1000ms
1000ms
1000ms
time_code
+1 SEC
+2 SEC
+3 SEC
+4 SEC

FIG. 39

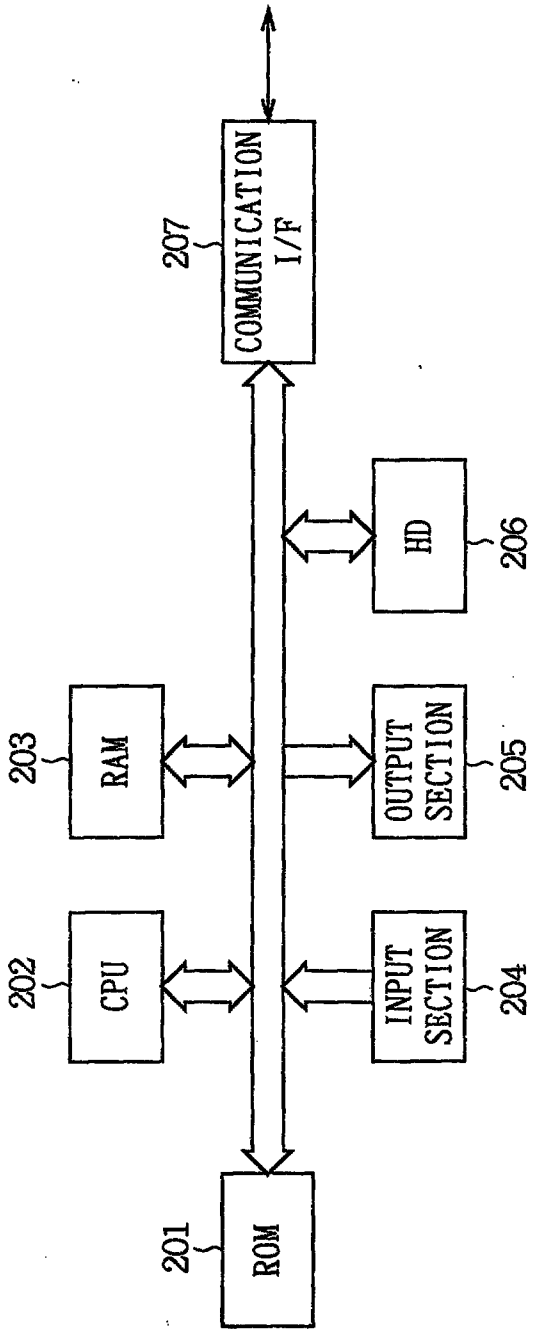

207
COMMUNICATION I/F
206
HD
203
RAM
205
OUTPUT SECTION
202
CPU
204
INPUT SECTION
201
ROM

FIG. 40

DESCRIPTION OF SYMBOLS

1 ... VO constitution section, 21 to 2N ... VOP constitution section, 31 to 3N ... VOP encoding section, 4 ... Multiplexing section, 21 ... Image layering section, 23 ... Enhancement layer encoding section, 24 ... Resolution transforming section, 25 ... Base layer encoding section, 26 ... Multiplexing section, 31 ... Frame memory, 32 ... Motion vector detector, 33 ... Arithmetic unit, 34 ... DCT unit, 35 ... Quantizer, 36 ... VLC unit, 38 ... Inverse quantizer, 39 ... IDCT unit, 40 ... Arithmetic unit, 41 ... Frame memory, 42 ... Motion compensator, 53 ... Frame memory, 71 ... Inverse multiplexing section, 72_1 to 72_N ... VOP decoding section, 73 ... Image reconstituting section, 91 ... Inverse multiplexing section, 93 ... Enhancement layer decoding section, 94 ... Resolution transforming section, 95 ... Base layer decoding section, 102 ... IVLC unit, 103 ... Inverse quantizer, 104 ... IDCT unit, 105 ... Arithmetic unit, 106 ... Frame memory, 107 ... Motion compensator, 112 ... Frame memory, 201 ... ROM, 202 ... CPU, 203 ... RAM, 204 ... Input section, 205 ... Output section, 206 ... HD, 207 ... Communication I/F.

TEXT OF THE ABSTRACT

IMAGE ENCODER, IMAGE ENCODING METHOD, IMAGE DECODER, IMAGE DECODING METHOD, AND DISTRIBUTION MEDIA

Efficient random access is made possible. A group of video plane (GOV) layer in which encoding start absolute time within accuracy of a second is arranged is provided as the layer of a coded bit stream. The GOV layer can be inserted not only at the head of the coded bit stream but at an arbitrary position in the coded bit stream. The display time of each video object plane (VOP) constituting the GOV layer is represented by modulo_time_base which represents absolute time in the unit of a second with the encoding start absolute time as standard and VOP_time_increment which represents in the unit of a millisecond the time that has elapsed since time (synchronization point) indicated by the modulo_time_base.

Fig.: 25.

This is the replacement sheet

Date: 17 JAN 2000

世界知的所有権機関
国際事務局

PCT

特許協力条約に基づいて公開された国際出願

(51) 国際特許分類6 **H04N 7/32**	**A1**	(11) 国際公開番号	**WO98/44742**
		(43) 国際公開日	1998年10月8日(08.10.98)

(21) 国際出願番号 PCT/JP98/01453

(22) 国際出願日 1998年3月31日(31.03.98)

(30) 優先権データ
特願平9/99683 1997年4月1日(01.04.97) JP

(71) 出願人（米国を除くすべての指定国について）
ソニー株式会社(SONY CORPORATION)[JP/JP]
〒141-0001 東京都品川区北品川6丁目7番35号 Tokyo, (JP)
(72) 発明者；および
(75) 発明者／出願人（米国についてのみ）
矢ケ崎陽一(YAGASAKI, Yoichi)[JP/JP]
鈴木輝彦(SUZUKI, Teruhiko)[JP/JP]
〒141-0001 東京都品川区北品川6丁目7番35号
ソニー株式会社内 Tokyo, (JP)
(74) 代理人
弁理士 田辺恵基(TANABE, Shigemoto)
〒150-0001 東京都渋谷区神宮前1丁目11番11-508号
グリーンフアンタジアビル5階 Tokyo, (JP)

(81) 指定国 AU, CA, CN, ID, IL, KR, MX, SG, US, VN, 欧州特許 (AT, BE, CH, ES, FI, FR, GB, IT, NL).

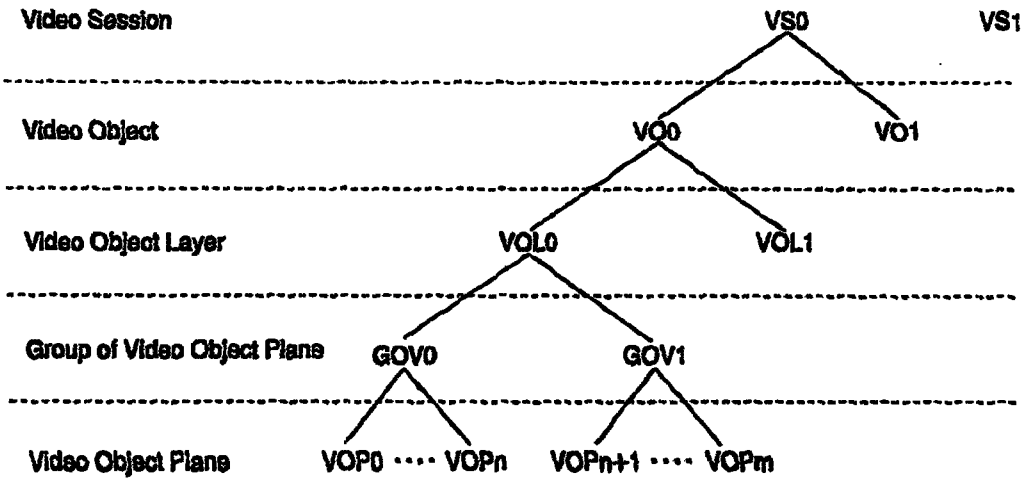
添付公開書類
国際調査報告書

980616-3

IPOS copy.

(54)Title: **PICTURE CODING DEVICE, PICTURE CODING METHOD, PICTURE DECODING DEVICE, PICTURE DECODING METHOD, AND PROVIDING MEDIUM**

(54)発明の名称 画像符号化装置および画像符号化方法、画像復号装置および画像復号方法、並びに提供媒体



(57) **Abstract**

To make an efficient random access possible, by providing, as a hierarchy of a coded bit stream, a GOV (Group of Video Object Plane) layer on which an absolute coding start time accurate to a second is arranged. The GOV layer can be inserted not only at the leading end of the coded bit stream but also at other arbitrary positions. In addition, a display time of each VOP (Video Object Plane) constituting the GOV is expressed by modulo_time_base, which expresses a relative time with reference to the absolute coding start time in seconds, and VOP_time_increment, which expresses a time elapsed from the time (synchronization point) indicated by modulo_time_base in milliseconds.

(57)要約

効率的なランダムアクセスを可能とする。符号化ビットストリームの階層として、秒精度の符号化開始絶対時刻が配置されるＧＯＶ（Group of Video Object Plane）層を設ける。ＧＯＶ層は、符号化ビットストリームの先頭だけでなく、その他の任意の位置に挿入することができるものとする。さらに、符号化開始絶対時刻を基準とする相対時刻を１秒単位で表すmodulo_time_base、およびmodulo_time_baseによって示される時刻（同期点）からの時刻を１ミリ秒単位で表すVOP_time_incrementによって、ＧＯＶを構成する各ＶＯＰ（Video Object Plane）の表示時刻を表す。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

AL アルバニア
AM アルメニア
AT オーストリア
AU オーストラリア
AZ アゼルバイジャン
BA ボズニア・ヘルツェゴビナ
BB バルバドス
BE ベルギー
BF ブルギナ・ファソ
BG ブルガリア
BJ ベナン
BR ブラジル
BY ベラルーシ
CA カナダ
CF 中央アフリカ
CG コンゴー
CH スイス
CI コートジボアール
CM カメルーン
CN 中国
CU キューバ
CY キプロス
CZ チェッコ
DE ドイツ
DK デンマーク
EE エストニア
ES スペイン
FI フィンランド
FR フランス
GA ガボン
GB 英国
GD グレナダ
GE グルジア
GH ガーナ
GM ガンビア
GN ギニア
GW ギニア・ビサオ
GR ギリシャ
HR クロアチア
HU ハンガリー
ID インドネシア
IE アイルランド
IL イスラエル
IS アイスランド
IT イタリア
JP 日本
KE ケニア
KG キルギスタン
KP 北朝鮮
KR 韓国
KZ カザフスタン
LC セントルシア
LI リヒテンシュタイン
LK スリ・ランカ
LR リベリア
LS レソト
LT リトアニア
LU ルクセンブルグ
LV ラトヴィア
MC モナコ
MD モルドヴァ
MG マダガスカル
MK マケドニア旧ユーゴスラヴィア共和国
ML マリ
MN モンゴル
MR モーリタニア
MW マラウイ
MX メキシコ
NE ニジェール
NL オランダ
NO ノールウェー
NZ ニュー・ジーランド
PL ポーランド
PT ポルトガル
RO ルーマニア
RU ロシア
SD スーダン
SE スウェーデン
SG シンガポール
SI スロヴェニア
SK スロヴァキア
SL シエラ・レオネ
SN セネガル
SZ スワジランド
TD チャード
TG トーゴー
TJ タジキスタン
TM トルクメニスタン
TR トルコ
TT トリニダッド・トバゴ
UA ウクライナ
UG ウガンダ
US 米国
UZ ウズベキスタン
VN ヴィエトナム
YU ユーゴースラビア
ZW ジンバブエ